# Engineering a self-limiting genetic control system to suppress the mosquito species *Aedes aegypti*




Xenia Karlsson

Linacre College

University of Oxford

A thesis submitted for the degree of

*Doctor of Philosophy*

Trinity 2024

*For my mom, Irena Kunovská.*

# Acknowledgements

I am grateful to several people that helped me complete this thesis. I would first like to acknowledge my supervisor Luke Alphey for this great opportunity and for his guidance throughout the research project. I want to sincerely thank my second supervisor, Michael Bonsall, for his kind mentorship and encouragement when introducing me to the field of population genetics. I am very grateful to my co-supervisor, Michelle Anderson, for sharing with me her expertise in molecular biology, as well as to Leonela Carabajal Paladino for her caring support at Pirbright.

Thank you to Ray Wilson, who taught me how to make transgenic *Ae. aegypti* and whose passion for entomology was a source of inspiration. Thank you to Joshua Ang for very helpful discussions and reading my thesis, which I am also grateful to Luke and Michelle for. Thank you to the Phoenix Lab for holding the fort at Pirbright after the lab move and making my last year more manageable than I had hoped. Mireia Larrosa Godall and Emily Levitt, thank you for sharing my daily struggles and for your care, I am so grateful to have worked with you.

Embarking on a PhD would not have been possible without the generous mentorship of Tony Nolan. Starting with a lab practical in my undergraduate degree, I was introduced to the world of gene drives and got to meet many great scientists in Andrea Crisanti's lab who helped build my scientific foundation. Thank you especially to Nace, Chryssa, Alex, Chris, Ignacio, Matteo, and Ioanna, I could not have imagined more fun people to have started this journey with.

I am endlessly grateful to be surrounded by my beautiful friends and second family Le and Nomes, Donny and Thomas, Rhi, Ravs and Tush, and Meals and TT. You make my life shine.

Thank you to my parents who believed in me more than anyone and whose love I feel every day. *Tack för allt ni har gjort för oss, ILY.* Thank you, too, for making sure I would never spend a day alone by giving me my twinnie and best friend, Al. I could not have done this without you, *thank you*. Coparenting our charismatic lab seniors/being silently therapized by our certified Emotional Support Animals, Dingo and Nala, was a dream; I am so grateful to them, too. Lastly, thank you for making us all do that thanksgiving tradition every year because I was so prepared for this section.

Finally, to Jake, thank you for your patience and for being there through it all. You kept me going and made everything better.

# Declaration

I declare that the work presented in this thesis is my own, except where explicitly acknowledged in the text. The work has not been submitted, either partially or in full, for any other qualification.

# Abstract

Mosquito-borne diseases continue to pose a great challenge to global public health despite longstanding interventions. Novel technologies based on genetic engineering of the mosquito vectors could offer promising solutions to augment the waning efficacy of traditional control methods. This thesis aimed to engineer a self-limiting genetic control system to suppress populations of *Aedes aegypti*, the principal vector of dengue and other important arboviral diseases. The system has previously been modelled to be non-invasive, while being dramatically more efficient than other self-limiting systems already trialled in the field. Named the Y-linked editor (YLE), or Sustainable, persistent SIT (SuperSIT) here, it involves linking a transgenic construct to the Y-chromosome—or male-determining locus (M locus) in *Ae. aegypti*—and imposing a fitness penalty on all daughters of transgenic males. Here, a CRISPR-Cas9 construct was designed to express in the male germline and disrupt a female-specific gene required for survival in adulthood. Depending on the haplo-sufficiency of the target gene, the nuclease could be engineered to mutagenize the pre-meiotic sperm genome, such that all sperm were affected, or it could be paternally deposited to the embryo to disrupt both parental copies of a recessive gene. Regulatory elements were thus assessed for their ability to express Cas9 in the male germline and direct efficient mutagenesis of a target locus through either mechanism, while two genes required for flight in females, namely *myo-fem* and *Act-4*, were mutated at multiple sites to assess their haplo-sufficiency. The results demonstrated that a promising route to SuperSIT development was through paternal inheritance of dominant *myo-fem* mutations. As such, transgenic lines expressing a CRISPR-Cas9 construct targeting *myo-fem* were independently established by knock-in integration at multiple loci within the M locus. Flight ability was disrupted in up to 50% of female offspring in one transgenic line, demonstrating proof-of-concept of this system in *Ae. aegypti.*

# Contents

**ACKNOWLEDGEMENTS** V

**DECLARATION** VII

**ABSTRACT** IX

**CONTENTS** XI

**CHAPTER 1 : INTRODUCTION** 1

1.1 The burden of mosquito-borne diseases 1

1.2 Field trials of novel technologies to control *Ae. aegypti* 3

*1.2.1 Sterile insect methods* 4

*1.2.2 Wolbachia-mediated population modification* 6

1.3 New designs to improve the efficacy of genetic control 8

*1.3.1 Homing endonuclease gene drives* 9

*1.3.2 SuperSIT system for non-invasive sustainable mosquito control* 10

1.4 Thesis outline 15

**CHAPTER 2 : ASSESSMENT OF GERMLINE REGULATORY ELEMENTS TO EXPRESS CAS9 AND ACHIEVE EFFICIENT MUTAGENESIS OF A TARGET LOCUS** 17

2.1 Abstract 17

2.2 Introduction 18

*2.2.1 Shutdown regulatory element identified from Ae. aegypti homing studies* 19

*2.2.2 Putative regulatory elements of protamine and β2-tubulin identified from the literature* 21

2.3 Methods 26

*2.3.1 Design and cloning of plasmids for piggyBac transgenesis* 26

*2.3.2 Mosquito rearing* 28

*2.3.3 Microinjection* 29

*2.3.4 Establishing transgenic mosquito lines expressing Cas9* 29

*2.3.5 Split-drive assessments of Cas9 activity* 31

2.4 Results and discussion 32

*2.4.1 Establishing transgenic lines* 32

*2.4.2 Assessing construct activity* 34

2.5 Conclusion and future directions 41

**CHAPTER 3 : CRISPR-CAS9 KNOCK-OUT OF FEMALE SPECIFIC FLIGHT GENES, *ACT-4* AND *MYO-FEM*** **47**

3.1 Abstract 47

3.2 Introduction 48

*3.2.1 Evolution of asynchronous indirect flight muscles* 48
*3.2.2 Harmonic convergence in mosquitos* 50
*3.2.3 IFM-specific actin and myosin genes* 51
*3.2.4 Aims* 56
3.3 METHODS 58
*3.3.1 sgRNA design* 58
*3.3.2 Testing cleavage efficiency of sgRNAs in vivo* 60
*3.3.3 piggyBac transgenesis of Ae. aegypti to express actin and myosin sgRNAs from genome* 63
*3.3.4 Experimental crosses to test activity of sgRNA-expressing transgenic lines* 65
*3.3.5 Design and construction of a purpose-built flight test device* 66
*3.3.6 Sequence verification of flight-tested females to determine sgRNA activity and dominance of mutations* 68
3.4 RESULTS AND DISCUSSION 73
*3.4.1 In vivo assessment of sgRNAs by microinjection* 73
*3.4.2 piggyBac integrations* 87
*3.4.3 Flight test device* 94
3.5 CONCLUSION AND FUTURE DIRECTIONS 95
**CHAPTER 4 : KNOCK-IN OF THE M LOCUS AND ASSESSMENT OF FLIGHT IN FEMALE DESCENDANTS OF SUPERSIT MALES 99**
4.1 ABSTRACT 99
4.2 INTRODUCTION 100
*4.2.1 Structure of the sex-determining M locus in Ae. aegypti* 100
*4.2.2 Challenges of integrating transgenes in heterogametic chromosomes* 101
*4.2.3 Considerations for safe and efficient expression of transgenes integrated in the M locus* 103
*4.2.4 Aims* 104
4.3 METHODS 105
*4.3.1 In silico analysis of the M locus for sites to integrate SuperSIT transgenes* 105
*4.3.2 Testing cleavage efficiency of sgRNAs in vivo* 107
*4.3.3 Transgenesis of male mosquitoes with a SuperSIT construct at selected sgRNA target sites* 114
*4.3.4 Assessment of transgenic SuperSIT lines* 119
4.4 RESULTS AND DISCUSSION 121
*4.4.1 In silico analysis of M locus and selection of sgRNAs for cleavage efficiency test* 121
*4.4.2 Transgenesis of male mosquitoes with a SuperSIT construct at selected sgRNA target sites* 128
*4.4.3 Future Directions* 144
**CHAPTER 5 : GENERAL DISCUSSION 153**
5.1 SUMMARY OF CHAPTERS 153
5.2 TECHNICAL CHALLENGES 154
*5.2.1 Considerations for the current design* 154

*5.2.2 Limiting position effects* ____ *155*
*5.2.3 Resistance* ____ *158*
*5.2.4 Alternative mechanisms for engineering a dominant fitness effect in females* ____ *159*
5.3 CONCLUSION ____ 162
**APPENDIX ____ 164**
APPENDIX A: SUPPLEMENTARY MATERIAL FOR CHAPTER 2 ____ 164
APPENDIX B: SUPPLEMENTARY MATERIAL FOR CHAPTER 4 ____ 166
**REFERENCES ____ 182**

# Chapter 1: Introduction

## 1.1 The burden of mosquito-borne diseases

Mosquito-borne diseases pose an acute challenge to global public health. Mosquitoes from the *Anopheles* and *Aedes* genera are responsible for over half a million deaths annually, spreading the parasites that cause malaria and arboviruses responsible for dengue, zika, chikungunya, and yellow fever, respectively. The disease burden falls predominantly on communities living in the world's poorest regions, where disproportionately higher morbidity and mortality rates also exact an immense toll on economic development (World Health Organization, 2023b; World Health Organization, 2017; Gallup and Sachs, 2000; Sachs and Malaney, 2002; Degarege *et al.*, 2019; Oviedo-Pastrana *et al.*, 2017). Despite longstanding interventions, however, there is growing concern over the increasing global incidence of these diseases (World Health Organization, 2023b; Kraemer *et al.*, 2019; World Health Organization, 2017). Many of the diseases are emerging in new areas or re-emerging in areas where they had previously subsided or been eradicated from (World Health Organization, 2023b; Stanaway *et al.*, 2016; Paixão, Teixeira and Rodrigues, 2018; World Health Organization, 2017; Weaver *et al.*, 2018). It is therefore imperative to maintain sustained funding and implement adaptive and innovative approaches to effectively control the disease burden.

The resurgence of disease is linked to many societal and environmental factors, including rapid urbanization, land use change (e.g. deforestation), global migration, travel, and trade, and climate change—although the exact effect exerted by climate change is difficult to predict (Franklinos *et al.*, 2019; Kraemer *et al.*, 2019; Nsoesie *et al.*, 2016; Sinka *et al.*, 2020). For malaria, the principal factor in the increase of mortality rates since 2015 has been attributed to the rapid evolution of insecticide resistance. This is a major problem for controlling the urban malaria vector, *An. stephensi*, that is now emerging in Eastern Africa with significant consequences (Kleinschmidt *et al.*, 2018; Kleinschmidt and Rowland, 2021; World Health Organization, 2018; Sinka *et al.*, 2020). The COVID-19 pandemic aggravated an already urgent situation with a significant disruption to vaccination programs and essential health service provisions (Sherrard-Smith *et al.*, 2020; Heuschen *et al.*, 2021). Importantly, however, the worst-case scenarios for malaria

that had been projected at the start of the pandemic were averted (World Health Organization, 2023b).

Arboviral disease outbreaks are also expected to increase in magnitude and frequency worldwide (Kraemer *et al.*, 2019). The global incidence of dengue has grown rapidly in the last fifty years, increasing 10-fold in the last two decades to 5 million reported cases, though the true symptomatic burden is likely twenty times that (Bhatt *et al.*, 2013; World Health Organization, 2023a). The largest epidemic of dengue that was ever reported in Africa occurred in Burkina Faso last year, in 2023 (Manigart *et al.*, 2024). Dengue is also now found at higher altitudes and latitudes than before, with autochthonous transmission documented in several southern European countries and in the United States (Rijal *et al.*, 2021; Cochet *et al.*, 2022; World Health Organization, 2023a). In 2013, the Zika virus emerged in South America with an unexpected surge in neurological disorders and microcephaly cases in new-born infants, causing the World Health Organization to declare an international public health emergency in 2016 (Zhang *et al.*, 2017; Faria *et al.*, 2017). Outbreaks of yellow fever are still occurring despite the availability of a highly effective vaccine (Wilder-Smith and Monath, 2017; Mayer, Tesh and Vasilakis, 2017; Grobbelaar *et al.*, 2016).

Arboviral outbreaks are inextricably linked to the global distribution of two competent vectors: *Ae. aegypti* and *Ae. albopictus* (Nsoesie *et al.*, 2016; Messina *et al.*, 2016; Kraemer *et al.*, 2019). Local outbreaks have typically followed within 5-15 years of these mosquito species invading an area, underscoring the importance of vector range expansion for arbovirus transmission (Messina *et al.*, 2014; Kraemer *et al.*, 2019). Remarkably, global dispersal of *Aedes* has occurred despite short flight ranges due to the evolution of desiccation-resistant eggs and adaptation to urban environments, where it is aided by high population densities and access to trade routes (Powell and Tabachnick, 2013; Kraemer *et al.*, 2019; Hawley *et al.*, 1989; Benedict *et al.*, 2007; Guerra *et al.*, 2014).

For most of these arboviral diseases, there are no licensed drugs or useable vaccines, with the notable exception of the vaccine against yellow fever, as mentioned above. A vaccine for dengue became available in 2016, but it is restricted to people who have previously been exposed due to the risk of severe dengue in those who have not been (World Health Organization, 2019). Vaccines against chikungunya and zika viruses are in various stages of clinical development, but it is not known when they will become available (Schrauf *et*

*al.*, 2020). Until vaccines become safe, effective, and widely available, the most effective tool to prevent disease is by controlling the vectors that spread them. This will require intersectoral collaboration, novel technologies, and community engagement (Takken *et al.*, 2021; World Health Organization, 2017).

## 1.2 Field trials of novel technologies to control *Ae. aegypti*

Conventional vector control has involved management of breeding sites and use of insecticides but these have been hampered by the development of resistance and environmental harm (Ranson and Lissenden, 2016; Hemingway *et al.*, 2016; Kleinschmidt *et al.*, 2018; World Health Organization, 2022; Gill, Ramos-Rodriguez and Raine, 2012). Insecticide-based tools like indoor residual spraying and insecticide-treated bed nets continue to be relied on for control of the *Anopheles* vectors that spread malaria but are less effective at controlling the *Aedes* species because they feed in the day and have evolved to preferentially do so outdoors in addition to resistance to insecticides (Takken *et al.*, 2021; Sene *et al.*, 2021; Toé *et al.*, 2022; Jangir and Prasad, 2022; Konkon *et al.*, 2023).

Novel technologies based on genetic engineering of vectors have emerged over the last two decades as promising solutions to supplement or replace existing integrated pest management strategies. These genetic control interventions aim to reduce disease transmission by suppressing vector populations via introduction of lethal or sterile traits or modifying them to be less able to spread disease. The technologies are attractive because of their species-specific nature—preventing off-target losses in beneficial insects like bees and enabling exploitation of natural mate-seeking behaviours to reach populations that are unknown or difficult to access—and the potential to benefit entire communities regardless of considerable infrastructure or social and economic inequities (Alphey, 2014). Characteristics of *Ae. aegypti* that power its global dissemination, like adaptation to urban environments, limited flight ranges, and desiccation resistant eggs, also make it a good candidate for genetic control, both to study and deploy (Matthews and Vosshall, 2020).

Several of these genetic control strategies are in use or being trialled in the field for control of *Ae. aegypti*, with varying success, including the sterile insect technique (SIT), *Wolbachia*-based population suppression and population modification, and various self-

limiting (or non-invasive) strains from the biotechnology company Oxitec (https://www.oxitec.com/) (Weng, Masri and Akbari, 2023). Engagement and collaboration with governments and local communities is crucial prior to and throughout such interventions, and several national and international frameworks and non-profit organizations have been established to ensure safe and effective implementation (e.g. the WHO's Global Vector Control Response program and Vector Control Advisory Group advisory body, and organizations including World Mosquito Program (http://www.eliminatedengue.com) and Target Malaria (https://targetmalaria.org/)) (World Health Organization, 2021a; World Health Organization, 2017; World Health Organization, 2021b).

### 1.2.1 Sterile insect methods

Sterile insect methods, both classical irradiation-based SIT and variants of it that use transgenic DNA or *Wolbachia*-induced incompatibility, involve the mass release of sterile males into target populations (Dyck, Hendrichs and Robinson, 2021; Alphey *et al.*, 2010; Zheng *et al.*, 2019). Sterile males seek out and induce partial or full sterility in females in the wild population, thereby reducing the number of successful matings and leading to fewer insects in subsequent generations. Classical SIT has successfully been used to eliminate or suppress pests of great agricultural significance but despite various trials, has not been widely adopted for use against mosquitos (Dyck, Hendrichs and Robinson, 2021; Bouyer *et al.*, 2020b). This is mainly because the radiation doses that render insects sterile can reduce their ability to survive and mate competitively, making determination of the ideal dose challenging and unpractical for several mosquito species (Helinski, Parker and Knols, 2009; Lees *et al.*, 2015; Bouyer *et al.*, 2020a; Bond *et al.*, 2019; Yamada *et al.*, 2019). Together with other stresses from mass-rearing and handling, the lifespan and mating performance of male *Ae. aegypti* mosquitos can be significantly reduced. Consequently, SIT programs require large and uninterrupted weekly releases, often 10-100 times that of the wild male population, to ensure that sufficient females mate with the released males (Bond *et al.*, 2019; Dyck, Hendrichs and Robinson, 2021). Where mosquito populations are especially dense or highly connected (such that reduced populations would be rapidly replenished), this technology would be prohibitively expensive or simply impracticable to pursue.

With innovations in biotechnology, numerous genetic tools have been developed that allow precise engineering of mosquito genomes for a desired effect. Engineered male sterility, referred to as bi-sex-Release of Insects carrying a Dominant Lethal (bi-RIDL), can produce similar population dynamics to classic SIT, but without the associated negative impacts of irradiation (Figure 1) (Thomas *et al.*, 2000; Phuc *et al.*, 2007). In addition, deliberate engineering of late-acting lethality enhances the efficiency of bi-RIDL compared to SIT because of the preserved suppressive effect conferred by density-dependent pressures that naturally control larval survival (Thomas *et al.*, 2000; Bouyer, 2023). A transgenic bi-sex lethal self-limiting strain from Oxitec, called OX513A, demonstrated good male competitiveness and significant suppression of *Ae. aegypti* populations in field releases in local suburban areas of Panama City, Panama, Juazeiro, Brazil, and Grand Cayman (Carvalho *et al.*, 2015; Gorman *et al.*, 2016; Harris *et al.*, 2012; Phuc *et al.*, 2007).

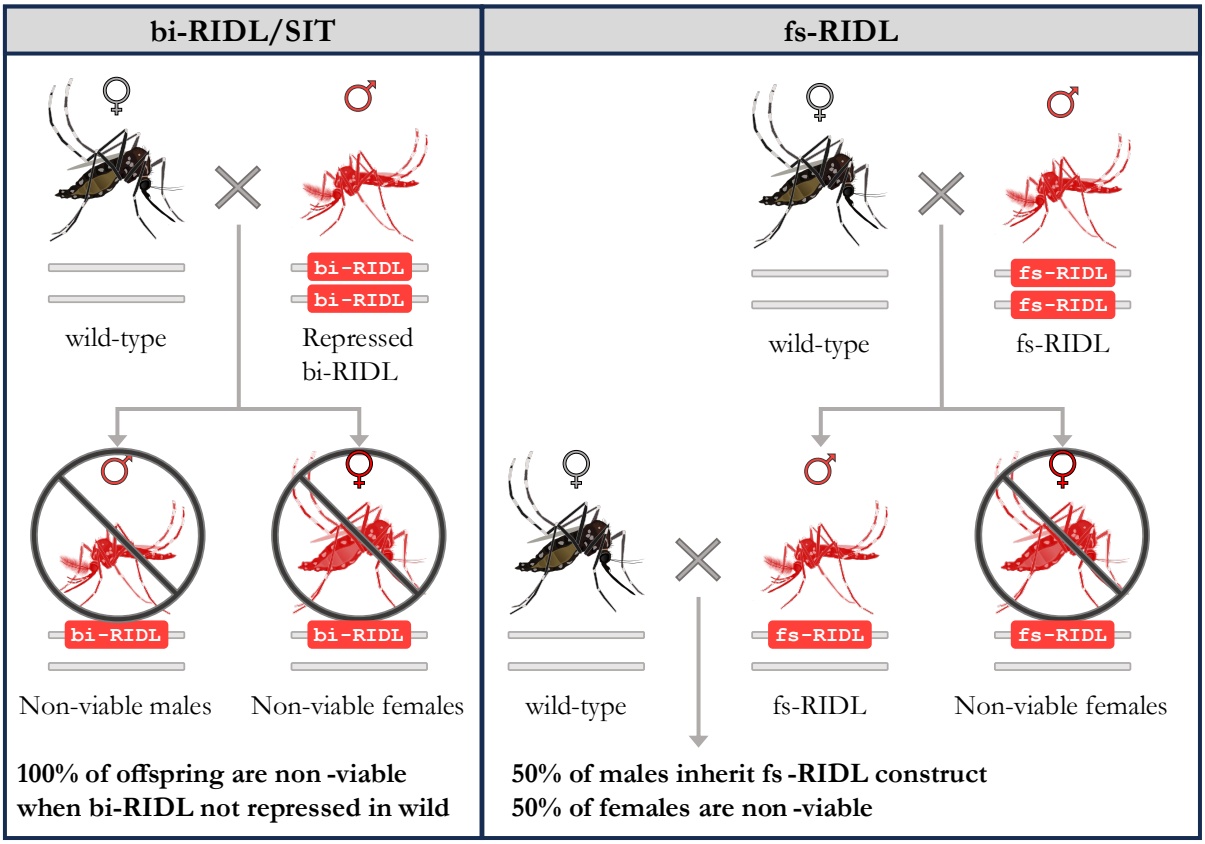



*Figure 1 Idealized inheritance patterns for the bi-RIDL (like SIT) (left) and fs-RIDL (right) genetic control systems.* Left: *In bi-RIDL and conventional SIT release programs, males that are homozygous for dominant lethal mutations seek out and induce full sterility in females in the wild population. In the subsequent generation, there are no surviving offspring, so the modification is not passed on. Transgenic bi-RIDL is possible through engineering of a repressible system: in the lab, the lethal system is repressed under specific conditions allowing bi-RIDL strains to develop normally and be bred to*

*homozygosity (Phuc et al., 2007). When the condition is removed in the wild, RIDL is no longer repressed and all offspring that inherit it are non-viable. SIT works similarly, in that radiation generates dominant lethal mutations in the gametes of otherwise healthy males (theoretically). When these gametes are passed on through mating of wild females, all offspring die.* Right*: In fs-RIDL, males homozygous for a female-specific lethal element that mate with wild females cause all daughters to die in the next generation and pass on a single copy of the fs-RIDL element to all males. These males then mate with wild females and subsequently cause half of their offspring to die. This reduces the proportion of females that represent the reproductive potential of the population, providing more efficient suppression than bi-RIDL. Anatomically accurate male and female* Ae. aegypti *images were sourced from a publicly available site (Ramírez, 2019a; Ramírez, 2019b). SIT = Sterile Insect Technique; RIDL = Release of Insects carrying a Dominant Lethal; bi = bi-sex, meaning both sexes are non-viable; fs = female-specific, meaning only females are non-viable.*

The bi-RIDL strain was later improved upon with the development of a self-limiting strain that selectively killed female offspring, referred to as female-specific RIDL (fs-RIDL) (Thomas *et al.*, 2000; Fu *et al.*, 2010). The fs-RIDL design is more effective than bi-RIDL because male offspring of released homozygous fs-RIDL males inherit a copy of the female-killing gene and thereby cause half their daughters to die in the next generation (Figure 1). This progressively reduces the proportion of females, which represents the reproductive and disease transmitting potential of the population, providing more efficient suppression than bi-RIDL (Alphey, 2014). In terms of practical advantages, this male-selecting strain allowed automatic sex sorting and dissemination of the egg stage, reducing labour intensity, costs, and operational complexity (Morrison, 2022). Deployment of Oxitec's fs-RIDL strain, OX5034 or "Friendly" mosquitos, in neighbourhoods of the densely populated Brazilian city of Indaiatuba caused a ~96% suppression of the natural *Ae. aegypti* population across two seasons and paved the way towards regulatory approval for strain deployment nationwide and in regulated trials in Florida and Texas in the United States (Spinner *et al.*, 2022; Waltz, 2022; Waltz, 2021; Burki, 2020).

### 1.2.2 *Wolbachia*-mediated population modification

Finally, *Wolbachia*-mediated population modification of *Ae. aegypti* has shown promising results in reducing dengue incidence in multiple cities worldwide (Utarini *et al.*, 2021; Ant *et al.*, 2023). *Wolbachia* is an intracellular endosymbiotic bacterium that is found in half of all insect species but does not naturally occur in *Aedes* mosquitos (Zug and Hammerstein, 2012). Injecting *Wolbachia* into *Aedes* showed that females can stably transmit it to offspring and, fortuitously, that they have a remarkably reduced ability to carry dengue, chikungunya, and Zika (Walker *et al.*, 2011). Introgression of *Wolbachia*

into a mosquito population occurs through a process called cytoplasmic incompatibility, whereby uninfected females become effectively sterile if mated with infected males, but infected females remain fertile, regardless of whether the males were infected. This manipulates reproductive outcomes in favour of *Wolbachia* transmission between mosquitos (Hoffmann *et al.*, 2011; Werren, Baldo and Clark, 2008). Incidentally, this also means that *Wolbachia* can be used as a sterile insect method for suppression if only males are released (though strict sex-sorting must be implemented).

A randomized-control trial carried out by the World Mosquito Program in Yogyakarta, Indonesia showed that *Aedes* infected with the *w*Mel strain of *Wolbachia* could within four years successfully replace the local *Aedes* population and reduce incidence of symptomatic dengue cases against the four main serotypes of dengue by 77% and reduce hospitalization by 88%, which is comparable to vaccination (Utarini *et al.*, 2021). Modelling has shown that this is already a cost-saving intervention in Indonesia (World Health Organization, 2021b; O'Reilly *et al.*, 2019). Further, introgression of *Wolbachia* in mosquito populations in north-eastern Australia led to the near elimination of dengue and its dismissal as a public health concern (Ryan *et al.*, 2019). Weaker positive outcomes were seen in Niteroi, Brazil, where release of *Wolbachia* was associated with a 69% reduction in dengue cases, 56% in chikungunya, and 37% reduction in Zika, three years after the first release (Pinto *et al.*, 2021).

Despite these observed successes, *Wolbachia*-based population modification has its limitations. Interventions over a 29-month period across parts of Rio de Janeiro, Brazil, only reduced dengue incidence by 38% and chikungunya by 10% (Dos Santos *et al.*, 2022). The exact reasons for this modest outcome are unknown, however, but the effectiveness of the tool can be impaired by strain related fitness costs of infected mosquitos, high temperatures that cause fluctuations in maternal transmission of *Wolbachia*, and the genetic background of the host that can reduce mosquito refractoriness to the arboviruses (Ross *et al.*, 2020; Liang *et al.*, 2022; Hien *et al.*, 2021; Gunasekaran *et al.*, 2022). Furthermore, any population replacement strategy that renders vectors unable to transmit disease must mitigate the risk of the pathogen evolving resistance against it (Edenborough *et al.*, 2021). This risk is heightened when an intervention limits, but does not completely inhibit, pathogen survival in the modified host. The *Wolbachia* strain wMel has been shown under experimental conditions to drive

positive selection of dengue variants with potentially enhanced ability to attach to mosquito cells, underscoring the need for regular and long-term surveillance for resistant viruses as part of *Wolbachia* control programs (Thi Hue Kien *et al.*, 2023; Edenborough *et al.*, 2021). Together, these limitations could reduce the long-term sustainability of this approach (Weng, Masri and Akbari, 2023; Edenborough *et al.*, 2021).

## 1.3 New designs to improve the efficacy of genetic control

The *Ae. aegypti* population suppression methods trialled in the field have all been strongly self-limiting; the inherent fitness penalty associated with the modified trait means it will be selected out of the target population if not maintained by iterative releases of additional modified insects (Spinner *et al.*, 2022). Despite the advances documented above, suppression of *Ae. aegypti* has been assessed on a neighbourhood-wide rather than city-wide scale. Further, for the most effective self-limiting design, OX5034, researchers employed large inundative releases of a minimum of ~25,000 males (but up to 250,000) spread over three days a week over 11 months at each site to achieve the reported suppression rates, the effect of which was not observed in the control site, which was less than 2 km away from the closest treated site. While a wealth of expertise has been gained from conducting these field trials, from best-practice in engagement of local communities and governments, to development of mass-rearing, quality-control, and release technologies, it is doubtful these genetic control designs could feasibly be scaled. More efficient designs for genetic control will be required to make a sizeable impact on areas with dense populations of *Aedes* species.

Numerous approaches have been proposed that could considerably reduce the required release ratios and increase the cost-effectiveness of genetic vector control strategies (Sinkins and Gould, 2006; Alphey, 2014; Deredec, Burt and Godfray, 2008). Gene drives have garnered much attention for their ability to bias the inheritance of a genetic element over normal Mendelian inheritance rates, even despite fitness costs associated with the engineered trait (Alphey *et al.*, 2020). This results in an eventual increase in the gene's prevalence in the population over subsequent generations and means it could be used to spread desirable traits or effects that have a suppressive effect on the population size along with it.

Gene drives differ widely in their degree of persistence and invasive potential once introduced (Alphey, 2014). Self-limiting gene drive systems have limited persistence and would gradually be removed from the natural population after release, while self-sustaining strategies spread autonomously and are intended to persist indefinitely. Self-sustaining strategies can be further graded along the size of the introduction threshold required to take hold within a target (and nontarget) population—a measure of its invasiveness and a fundamental consideration for safe release in the field. As these properties exist on a spectrum and can often be affected by subtle changes in the configuration or molecular design of the different genetic components, computer simulations have become critical for understanding the population genetics of these control systems (Davis, Bax and Grewe, 2001; Magori and Gould, 2006; Burt, 2003; Marshall and Hay, 2012b; Deredec, Burt and Godfray, 2008; Deredec, Godfray and Burt, 2011; Beaghton, Beaghton and Burt, 2016; Beaghton *et al.*, 2017; Nash *et al.*, 2019; North, Burt and Godfray, 2020).

### 1.3.1 Homing endonuclease gene drives

The discovery of Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) systems and associated proteins, of which the most widely used is the CRISPR associated protein 9 (Cas9), has transformed the field of genetic vector control (Jinek *et al.*, 2012; Cong *et al.*, 2013; Mali *et al.*, 2013). CRISPR-Cas9 has been harnessed for genetic engineering and as central components of gene drive systems (Gantz *et al.*, 2015; Hammond *et al.*, 2016). The most powerful gene drive tool that has been developed depends on the action of CRISPR-Cas9 in homing endonuclease gene (HEG) drives. HEGs are selfish genetic elements that increase their allele frequency in a host by cleaving the homologous chromosome at a unique DNA target site directly opposite the HEG and then hijacking the cell's homology-directed repair (HDR) machinery to copy itself into the cleaved allele (Burt, 2003). If this "homing" process that duplicates the HEG allele occurs efficiently in the cells that become gametes, an individual that is heterozygous for the HEG element will propagate it to the next generation at frequencies disproportionately higher than what would be expected for normal Mendelian inheritance. This effectively converts a heterozygote into a homozygote exclusively in the germline. Because of this bias that repeats itself every generation, the HEG element can rapidly invade a population from very low frequencies and drive an engineered trait, such as fitness loads sufficient to

suppress populations or antipathogen effectors, along with it (Burt, 2003; Deredec, Burt and Godfray, 2008).

CRISPR-Cas9 homing gene drives engineered against the principal malaria-transmitting mosquito, *An. gambiae*, have been shown to effectively suppress entire populations in small laboratory cages and large indoor cages designed to mimic ecologically-relevant conditions (Kyrou *et al.*, 2018; Simoni *et al.*, 2020; Hammond *et al.*, 2021b). Given that these gene drives have immense potential to spread autonomously between all connected populations, altering them within unprecedented time frames, successful deployment will require extensive stakeholder engagement (e.g. Target Malaria) and informed consent from local communities, governments, and local and international regulators alike (James *et al.*, 2018; Long *et al.*, 2020; Annas *et al.*, 2021; Thizy *et al.*, 2021; Thizy *et al.*, 2022; Hammond *et al.*, 2021b; World Health Organization, 2021a). Debates around the possibility of future field trials are actively ongoing. Notably, homing gene drives engineered to target *Ae. aegypti* have generally proved to be far less efficient due to unintended repair outcomes and poor or untimely HEG expression, so further optimization will be required before this tool could possibly be applied to *Ae. aegypti* (Anderson *et al.*, 2022; Verkuijl *et al.*, 2022a).

### 1.3.2 SuperSIT system for non-invasive sustainable mosquito control

Despite the high potential of gene drives, there is a clear desire among the scientific and stakeholder community to develop mosquito strains for genetic control that remain localized to a specific area upon release, but which are still more effective than current designs (Marshall and Hay, 2012a; Dhole *et al.*, 2018). This would be useful, for example, to control a vector population where it is most problematic: for example, within the boundaries of a city or town or specifically where outbreaks are occurring; to restrict a population from invading an area outside of its native range; or to respect the right of geo-political entities to consent to the use of these genetic technologies within their border. A set of designs have been put forward for non-invasive genetic population suppression that offer dramatically improved cost-effectiveness and sustainability of control than current self-limiting designs (Burt and Deredec, 2018). Named one of two things depending on the approach to implementation, the Y-linked editor (YLE) (here, Sustainable, persistent SIT (SuperSIT)) or fs-RIDL-with-drive have been modelled to be 20-100 times more efficient than all configurations of inundative self-limiting

technologies like the sterile insect methods discussed above, while remaining largely localized to the target population (Burt and Deredec, 2018).

The SuperSIT (or YLE) design involves linking a genetic construct to the Y-chromosome and producing a fitness effect in the modified male's (non-inheriting) female progeny (Figure 2). The genetic construct would persist through generations but not "drive" since inheritance of the genetic construct is not biased: only 50% of the offspring inherit it (and survive), but these are all transgenic males. Unlike most self-limiting strategies, the Y-linked construct would not be selected against since the construct is absent from the daughters it harms and so its inheritance is not affected by the suppressive effect it creates (Burt and Deredec, 2018). This effectively means that the population would not immediately recover if releases of transgenic males were stopped. Further, as females are generally the vectors of disease, the SuperSIT system would be more attractive than other forms of genetic control as no viable females would need to be released and the system could be deployed using eggs or other immature and mixed-sex stages.

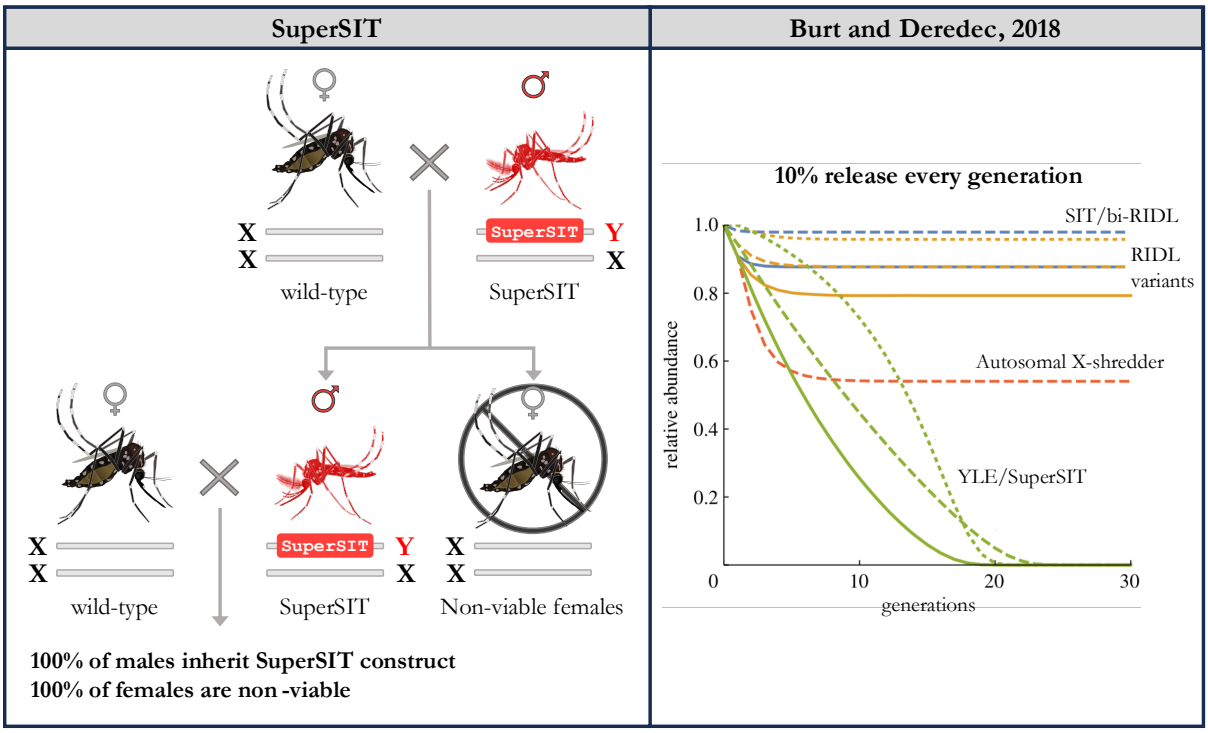



*Figure 2 Left: inheritance patterns for the SuperSIT/YLE system. The design links a transgenic construct to a Y-chromosome that is inherited by all sons and induces a fitness effect in non-inheriting daughters. Male fitness is not affected, so the transgene is not selected against, and the process can repeat itself every generation. Right: deterministic model adapted from Burt and Deredec (2018) that shows population size in terms of the number of females over time with recurrent releases of different self-limiting constructs (Burt and Deredec, 2018). A 10% release ratio of modified males to wild males occurs at every generation into an increasing target population (intrinsic rate of increase, $R_m = 6$) (SIT typically requires*

*release sizes that are 1000% that of the wild population). Deredec and Burt concluded that the YLE/SuperSIT design is 20-100 times more efficient than conventional SIT across a range of assumptions. Blue: bi-RIDL/SIT; orange: fs-RIDL; red: Autosomal X-shredder (distorts the sex-ratio towards 100% male (Galizi et al., 2014; Galizi et al., 2016)); green: YLE/SuperSIT systems. All self-limiting systems are idealized, meaning the effect of the knock-out is complete and fully dominant. Alternative modes of action of the fitness effect are denoted with solid and broken lines: solid lines indicate death after density dependent mortality, dashed lines indicate death before, and dotted lines indicate female sterility rather than death. Male and female* Ae. aegypti *images are publicly available (Ramírez, 2019a; Ramírez, 2019b). SIT = Sterile Insect Technique; RIDL = Release of Insects carrying a Dominant Lethal; bi = bi-sex and means both sexes in offspring are unviable; fs = female-specific and means only female offspring are unviable; YLE = Y-linked editor; SuperSIT = Sustainable and persistent SIT.*

This method of genetic control has immense potential for application in the field, but due to technical challenges, has only recently been built as a proof-of-concept in the *D. melanogaster* model, the agricultural pest *Ceratitis capitata*, and malaria vector *An. gambiae* (Fasulo *et al.*, 2020; Meccariello *et al.*, 2021; Tolosana *et al.*, 2024). First, genetically engineering a Y-linked insertion that efficiently expresses in the male germline is challenging (Bernardini *et al.*, 2014; Elgin and Reuter, 2013; Alcalay *et al.*, 2021). In most heterogametic species, the Y chromosome is largely heterochromatic and silenced early during male meiosis in a process called Meiotic Sex Chromosome Inactivation (MSCI). This makes it difficult to both recover Y-linked insertions and to express transgenes from the region (Turner, 2007; Namekawa and Lee, 2009; Taxiarchi *et al.*, 2019). Characteristics of the sex-determining region in the *Ae. aegypti* genome could make it possible to bypass this problem, however.

In *Aedes* and *Culex* mosquitoes, sex is determined by a male-determining locus (M locus) found on a chromosome that is mostly euchromatic and homomorphic between the sexes (Newton, Wood and Southern, 1978; Hall *et al.*, 2014; Hall *et al.*, 2015; Fontaine *et al.*, 2017; Matthews *et al.*, 2018; Aryan *et al.*, 2020). The M locus is absent from females, meaning males are M/m and females are m/m. Further, the M locus does not undergo MSCI in the germline and supports expression of several genes, including the principal gene in the sex determination pathway (*Nix,* AAEL022912), a gene required for male flight (*myo-sex,* AAEL021838), and several long-noncoding RNA that show enriched expression in the male reproductive tract (Aryan *et al.*, 2020; Matthews *et al.*, 2018; Mysore *et al.*, 2021; Turner *et al.*, 2018). Transgene insertions tightly linked to the M locus have also successfully been expressed (Krzywinska *et al.*, 2016b; Adelman *et al.*, 2008; Carpenetti *et al.*, 2012; Li *et al.*, 2020; Lutrat *et al.*, 2023). Lastly, recombination rates around the M locus are low: approximately 100 Mb surrounding the M locus is genetically differentiated

between the sexes, so expression of a male-linked transgene could be achieved in a euchromatic area while still being relatively far away from the M locus (Fontaine *et al.*, 2017).

The second criterion of the YLE system is that it must indirectly disrupt the fitness of non-inheriting daughters, which poses its own challenges. Fitness penalties could be imposed in several ways (Burt and Deredec, 2018). One of these employs a programmable nuclease to mutate genes required for female survival or fertility. This could be done by: 1) manipulating the pre-meiotic sperm genome (before the haploid gametes are formed), such that all sperm are affected; or 2) manipulating the post-zygotic embryo by way of an engineered paternal effect, whereby modifying nucleases are delivered (or "deposited") to the embryos via the sperm (Burt and Deredec, 2018). The genetic targets for the two outlined approaches differ: in the former case, a haplo-insufficient female-specific gene, essential for female viability or fertility, may be disrupted in the pre-meiotic sperm, such that all progeny inherit a damaged copy and the mutant phenotype is dominant in the daughters. In the latter case, a haplo-sufficient (recessive) female-specific gene may be disrupted due to activity of the genetic element in the embryo disrupting both paternal and maternal copies (Burt and Deredec, 2018). A challenge of the first method is that few haplo-insufficient female-specific genes exist compared to the many haplo-sufficient genes that are required specifically in females for fertility, while a challenge of the second method is that paternal deposition may be harder to engineer (for example, due to the very small sperm: egg volume ratio), though it has been done previously in *An. gambiae* using the homing endonuclease I-*Ppo*I (Windbichler, Papathanos and Crisanti, 2008; Galizi *et al.*, 2014).

Flight has been proposed as a target for genetic control as it is essential for mosquito survival and reproduction in the field; flight is required to escape water after emerging from the pupal form, avoid predation, find a blood meal required for egg laying (and disease transmission), and to produce the specific wingbeat frequency needed for courtship and mating (Cator *et al.*, 2009; Fu *et al.*, 2010; Li *et al.*, 2021a). Moreover, the actin and myosin genes that make up the indirect flight muscles (IFMs) in *Ae. aegypti* have been found to be sex-specific (in females, these are *Act-4* and *myo-fem*), making them potential candidates to target with a CRISPR-Cas9-based SuperSIT system (Muñoz *et al.*, 2004; Vyazunova and Lan, 2004; Fu *et al.*, 2010; Akbari *et al.*, 2013a; Navarro-Payá *et al.*,

2020; Aryan *et al.*, 2020; O'Leary and Adelman, 2020; Li *et al.*, 2021a). Lastly, disruption of flight would affect the later stages of female development, such that the density-dependence that controls larval survival (for example, through availability of larval nutrition), and which efficiently regulates population size, would not be disturbed (in Figure 2, solid compared to dashed lines) (Phuc *et al.*, 2007). Such late acting lethality has been modelled to be considerably more effective than lethality that is induced at earlier stages of mosquito development (Dye, 1984; Phuc *et al.*, 2007; Burt and Deredec, 2018; Evans *et al.*, 2022; Bouyer, 2023).

## 1.4 Thesis outline

This thesis aims to develop a proof-of-principle SuperSIT population suppression system in the arboviral mosquito *Ae. aegypti* for non-invasive, cost-effective, sustainable, and environmentally friendly vector control. A SuperSIT system against *Ae. aegypti* would have immense potential for application in the field and for use as a model to target other arboviral mosquitos like *Ae. albopictus* and species of the *Culex* genus (vector of Japanese encephalitis and lymphatic filariasis) in future. This thesis will be divided into three objectives, each with a dedicated chapter outlined below. Broadly, the first and second objectives explore means of disrupting the fitness of non-inheriting SuperSIT daughters by targeting females' ability to fly. The third objective attempts to assemble and express a CRISPR-Cas9 based SuperSIT construct from within the M locus and thereby ensure male-only inheritance and disruption of flight in female descendants.

In Chapter 2, the activity of three germline promoters designed to control expression of Cas9 and achieve efficient cutting of a target locus is assessed. Depending on the final target of the SuperSIT nuclease, Cas9 activity may be directed to the pre-zygotic sperm genome or delivered via a paternal effect to the post-zygotic embryo. An early germline promoter that regulates expression during the pre-meiotic stages of germline development is assessed for its ability to drive expression of Cas9 from an M-linked site. Two promoters hypothesized to be active during the later stages of spermatogenesis and thus more likely to induce paternal deposition of effectors into embryos post-mating are then assessed when driving expression of a reporter gene and Cas9. A previously validated sgRNA expressing line targeting the scorable eye phenotype gene, *kynurenine 3-monooxygenase* (*kmo*), is used to guide and assess Cas9 cleavage.

In Chapter 3, the suitability of two gene targets required for flight in females, namely *myo-fem* and *Act-4,* are investigated for use in a SuperSIT system. Cas9 guides targeting both genes are designed and tested in microinjection assays and later incorporated into multiplexed guide expressing constructs used to make transgenic lines, which are then assessed with transgenic lines expressing Cas9 created in Chapter 2. The dominance of the knock-out effect induced by each guide is explored. A flight test device is devised and constructed using inexpensive materials and simple tools to assess flight more precisely. Together, Chapter 2 and 3 inform on the design of a SuperSIT construct for integration within the M locus, which is discussed in Chapter 4.

In Chapter 4, work towards assembling and integrating a SuperSIT system within the M locus is described. The M locus is first examined for potential integration sites that would allow efficient expression of an integrated transgene without impacting male fitness. sgRNAs are designed for HDR-based knock-in at precise loci within the M locus and tested to determine cleavage efficiencies *in vivo*. Knock-in of a SuperSIT construct is attempted at multiple sites within the M locus, and resulting transgenics are tested for male-biased inheritance and the ability to impact flight in female descendants. Future directions to improve transgene expression from the M locus are discussed.

Finally, in Chapter 5, broader implications of the SuperSIT system are examined and general future directions proposed.

# Chapter 2: Assessment of germline regulatory elements to express Cas9 and achieve efficient mutagenesis of a target locus

## 2.1 Abstract

Developing a male expressed SuperSIT system hinges on precise spatiotemporal control of the effector transgene. For a system that employs CRISPR-Cas9 to disrupt a female-specific gene, mutagenic nuclease activity can be directed to the diploid gametes or paternally deposited to the zygote depending on the haplo-sufficiency of the target gene. This chapter evaluated the regulatory elements of three germline-specific genes (*shu*, *proA*, and *β2t*) to explore their potential to impose a fitness penalty through either mechanism. Transgenic *Ae. aegypti* lines expressing Cas9 regulated by these elements were generated and their activity assessed with an sgRNA-expressing element targeting the eye pigment gene, *kmo.* A fertile transgenic line expressing *shu*-Cas9 from an M-linked site was isolated from spontaneous recombination of an existing m-linked isoline. Preliminary results suggest the potential for paternal deposition of complexed Cas9 and sgRNA with *proA*, although the levels observed were minimal (<2%). The results also showed that high cleavage of the pre-meiotic sperm genome was possible with *shu* from an M-linked site, which is a key objective of a nuclease-based SuperSIT system targeting a dominant female-specific gene. Finally, challenges towards achieving the desired spatiotemporal control for paternal delivery of transgenes is discussed and possible future directions suggested.

## 2.2 Introduction

The genetic vector control approach pursued here is a highly efficient but self-limiting (and non-driving) design for population suppression of the *Aedes aegypti* mosquito vector (Burt and Deredec, 2018). The general design of the SuperSIT system involves linking a genetic construct to the Y-chromosome (or the sex determining M-locus in *Ae. aegypti*) to ensure transmission to all male progeny and inducing a fitness effect in the non-inheriting female progeny of modified males. Several designs have been proposed that would achieve similar population suppression dynamics and all of these rely on precise regulation of the effector transgenes (Burt and Deredec, 2018; Thomas *et al.*, 2000).

Here, a CRISPR-Cas9 construct was designed to be expressed in males and to indirectly disrupt female fitness through mutagenesis of a gene required for female survival or fertility. Depending on the haplo-sufficiency of the genetic target, mutagenic nuclease activity could be directed to the diploid gametes or via paternal deposition to the zygote (Burt and Deredec, 2018). In the former scenario, mutagenesis of the pre-meiotic sperm genome at a locus that was haplo-insufficient in females would affect all sperm and ensure a dominant mutant phenotype in daughters. In the latter scenario, paternal carry-over of the nuclease to the zygote would mean that both paternal and maternal copies are disrupted, allowing targeting of a haplo-sufficient female-specific gene.

Both pathways pose challenges to development, however. A challenge of the first method is that comparatively few haplo-insufficient female-specific genes exist, while a challenge of the second method is that paternal deposition may be harder to engineer. Thus, this chapter will explore pathways possible for achieving both types of Cas9 activity in the male germline, while Chapter 3 will assess candidate female-specific genes for targeting. Three putative regulatory elements were identified by the research group for potential use with the SuperSIT system; one was chosen for pre-meiotic activity based on previous homing studies, while another two were identified from the literature for their potential to induce a paternal effect. The rationale behind these choices is described below with an initial introduction into how these regulatory elements have previously been studied with homing assays.

### 2.2.1 *Shutdown* regulatory element identified from *Ae. aegypti* homing studies

Much of the progress in identifying the tools and potential pitfalls associated with achieving spatiotemporal transgene expression in the mosquito germline has been driven by the development of homing endonuclease gene (HEG) drives. These drives increase their allele frequency within a host by cleaving the homologous chromosome at a unique DNA target site directly opposite a HEG allele and then hijacking the cell's homology-directed repair (HDR) machinery to copy itself into the cleaved allele (Burt, 2003). If this "homing" process occurs efficiently in the cells that become gametes, an individual that is heterozygous for the HEG element can convert itself to a homozygote exclusively in the germline. This effectively doubles its frequency in a population every generation such that the HEG element can rapidly invade a population from very low frequencies and drive an engineered trait along with it (Burt, 2003; Deredec, Burt and Godfray, 2008).

Homing is presumed to be most efficient when the HEG nuclease is expressed in diploid germline cells during or immediately before meiosis (Champer *et al.*, 2018; Hammond *et al.*, 2021a). When the nuclease is expressed prior to this critical period in mitotically active germline cells, conversion of cleaved alleles is more likely to fail. This could be because double-stranded breaks are more prone to being repaired via the error-prone non-homologous end joining pathway (NHEJ) rather than HDR during this time, leading to mutations (Terradas *et al.*, 2022a; Pfitzner *et al.*, 2020). In some cases, a DNA break can cause destruction of the cleaved chromosome, which would lead to an increase in the HEG allele due to the loss of the non-HEG allele, termed "meiotic drive" (Verkuijl *et al.*, 2022b). The likelihood of mutagenic repair is also increased in alleles that are cleaved after segregation of gametes as these cannot be repaired using the template homologue. Finally, the persistence of transcribed or translated HEG components into fertilized embryos via gametes, termed "deposition", can have similar negative cleavage outcomes as above (though with some important differences including embryonic homing of HEG alleles and NHEJ mutations). Consequently, the optimization of gene drives has largely focused on identifying regulatory elements that restrict and enhance nuclease expression specifically within the relatively narrow pre-meiotic window (Hammond *et al.*, 2021a; Anderson *et al.*, 2023a; Anderson *et al.*, 2024; Verkuijl *et al.*, 2022a; Verkuijl *et al.*, 2022b).

While the nuclease based SuperSIT system does not rely on homing, studies of germline regulatory elements driving Cas9 expression have been informative for SuperSIT development in *Ae. aegypti* because they often assess nuclease activity in the soma and deposition into embryos, in addition to homing. The "split-drive" experimental design has been used to investigate a range of germline regulatory elements in a more high-throughput way (Anderson *et al.*, 2023a; Anderson *et al.*, 2024; Verkuijl *et al.*, 2022b; Li *et al.*, 2020). This is when a homing drive utilizing the targetable endonuclease CRISPR-Cas9 is "split", such that the Cas9 is integrated at an independent locus and only the single-guide RNA/s (sgRNA/s) are included in the drive allele and homed (this arrangement is also the basis of localized Daisy-chain gene drives) (Noble *et al.*, 2019). One drawback to this approach is that because transgenic lines expressing the Cas9 transgene are most efficiently generated by *piggyBac*-mediated transgenesis that produces semi-random integrations in the genome, the results can sometimes vary considerably due to "position effect". Position effect describes the frequently observed phenomenon that the chromosomal context in which a transgene is inserted can significantly affect expression levels (Wilson, Bellen and Gehring, 1990; Schotta *et al.*, 2003). As such, multiple "isolines" (i.e. transgenic mosquito strains that carry the same transgene but at distinct loci) must be generated to assess the reproducibility of the expression pattern in different regulatory contexts (Anderson *et al.*, 2023a).

The split-drive homing assays performed in *Ae. aegypti* have targeted two genes associated with eye colour determination as disruption of these genes results in an easily scorable and viable phenotype (Anderson *et al.*, 2024; Li *et al.*, 2017b; Li *et al.*, 2020). One of these is a recessive gene required for the synthesis of ommochrome pigments in the eyes of mosquitoes, the *kynurenine monooxygenase* (*kmo*) gene (AAEL008879). *kmo* is an attractive target gene for homing drive studies because the recessive phenotype facilitates tracking of transgene integrations and other knock-out mutations resulting from NHEJ (Anderson *et al.*, 2024; Han *et al.*, 2003). Patchy pigmentation in the eye ("mosaicism") can also be an easily observable indicator of CRISPR-Cas9 activity occurring in somatic cells or deposition into the embryo, both of which could result in only partial genotype conversion of cells or cell lineages (Figure 9).

One split-drive study carried out by the research group identified the putative regulatory regions of the *Ae. aegypti* homolog of *shutdown* (*shu*) (Anderson *et al.*, 2023a). The *shu*

gene is essential to the normal functioning of germline stem cells but is also required at later stages of oogenesis (Munn and Steward, 2000; Preall *et al.*, 2012). In *Ae. aegypti,* the *shu* regulatory regions were found to express Cas9 in the somatic and germline tissues of both males and females, inducing modest homing and significant NHEJ mutations in both the soma and germline (Anderson *et al.*, 2023a). Notably, paternal deposition of Cas9, and possibly Cas9-sgRNA complex, was observed in two of four isolines tested. Partial eye mosaicism due to paternal deposition of Cas9 was observed in up to 77% of offspring, though this result was inconsistent with similar crosses that showed mosaicism rates between 0-16%. It is important to note that the scoring of eye mosaicism can be subjective and therefore vary depending on the person screening. This regulatory promoter will therefore be studied further to assess for any paternal deposition again and more precisely quantify the germline cleavage activity. Fortuitously, one of the isolines generated in the previous study was integrated at a site linked to the m allele and exhibited low but nonzero recombination rates to the genomic regions surrounding the male-determining M locus. This line therefore offered an opportunity to study the expression of this germline promoter from an M-linked locus, which is an aim of the SuperSIT system. The following section will discuss the identification of two more promoters from the literature for use in a paternal effect system.

### 2.2.2 Putative regulatory elements of *protamine* and *β2-tubulin* identified from the literature

Parental effects involve the transfer of non-genetic factors from parent to offspring that influence the offspring's phenotype independently of their genotype. The importance of maternal effects were recognized early on due to the large volume of the female gamete that facilitated the availability of factors that could control and stimulate the earliest stages of embryogenesis (Mousseau and Fox, 1998; Marshall and Uller, 2007). In nearly all animals, the initial stages of development occur in the absence of transcription before zygotic genome activation, underscoring the role maternal effects play in initiating cell fate determination, particularly axial patterning and primary germ layer formation (Tadros and Lipshitz, 2009; Vastenhouw, Cao and Lipshitz, 2019; Mei *et al.*, 2013; Zhang *et al.*, 1998).

While maternal effects have been extensively studied, paternal effects are much less characterized. The highly reduced cytoplasmic volume of the sperm compared to the

oocyte reduces the relative concentration of biomolecules that could potentially be transmitted to offspring. Nevertheless, recent studies have uncovered evidence for a diverse population of RNA and other sperm-derived factors that are delivered directly to the fusing embryo and are essential for successful fertilization and zygote development in animal models (Fischer *et al.*, 2012; Ostermeier *et al.*, 2004; Hosken and Hodgson, 2014; Boerke, Dieleman and Gadella, 2007; Bourc'his and Voinnet, 2010; Liu *et al.*, 2012; Gur and Breitbart, 2006; Dadoune, 2009; Stoeckius, Grün and Rajewsky, 2014). Given the small sperm to oocyte volume ratio, delivery of mRNA that can be translated in the embryo has been hypothesized to be a more efficient means of contributing to embryogenesis than transporting a limited amount of proteins (Stoeckius, Grün and Rajewsky, 2014). Some of these paternal provisions have been implicated in processes known to be triggered by sperm in oocytes, including centriole-centrosome formation necessary for mitotic division and delivery of calcium oscillations required for oocyte activation (Sette *et al.*, 1997; Sette *et al.*, 2002; Wu *et al.*, 2007; Kashir, Nomikos and Lai, 2018; Dadoune, 2009; Ostermeier *et al.*, 2004). Others have speculated that paternally contributed mRNAs code for enzymes that trigger metabolic activity in the zygote after fertilization or genes involved in stress response that transmit environmental information (Stoeckius, Grün and Rajewsky, 2014).

Insights into the possible mechanisms underlining—and challenging—paternal deposition are offered by spermiogenesis, the process by which haploid spermatids mature into motile spermatozoa. Spermiogenesis involves a series of morphological changes and nuclear events that lead to a high degree of DNA compaction within the sperm head (Dadoune, 2009; Fabian and Brill, 2012). These profound changes in chromatin structure and coincidental loss of cytoplasm lead to reduced transcription from the haploid genome. Therefore, spermatocytes must provision the maturing spermatids with adequate levels of transcripts to support spermiogenesis and potentially embryogenesis as well (Fischer *et al.*, 2012; Hosken and Hodgson, 2014; Gur and Breitbart, 2006; Stoeckius, Grün and Rajewsky, 2014). As the transcripts must be stored in the developing spermatid for extended periods (up to several days in *Drosophila*), translational repression via posttranscriptional regulation is critical (Hecht, 1996; Raja and Renkawitz-Pohl, 2005; Barckmann *et al.*, 2013). This is exemplified by a process of spermatid differentiation that is common to both mammals and *Drosophila*—the compaction of chromatin into sperm nuclei.

Compaction of sperm DNA is achieved by the global replacement of somatic histones with protamines and other sperm-specific nuclear basic proteins (SNBPs) (Wilson *et al.*, 2006; Rathke *et al.*, 2007). In *Drosophila*, sperm DNA is packaged with two protamine-like proteins, the paralogous *Mst35Ba* and *Mst35Bb* (*ProtA* and *ProtB*, respectively), and the linker histone-like protein *Mst77F*. These genes are transcribed in diploid primary spermatocytes but only translated several days later during the final stages of spermiogenesis in elongating spermatids (Raja and Renkawitz-Pohl, 2005; Rathke *et al.*, 2007; Braun *et al.*, 1989; Zhong *et al.*, 2001; Kleene, 1989; Morgan *et al.*, 2021; Barckmann *et al.*, 2013; Awe and Renkawitz-Pohl, 2010). Further, fluorescent tagging of *ProtB* and *Mst77F* tracked mature spermatozoa in the female reproductive tract of *D. melanogaster* and demonstrated that protamine protein was deposited into embryos of fertilized *D. simulans* and hamster oocytes (Raja and Renkawitz-Pohl, 2005; Manier *et al.*, 2010; Barckmann *et al.*, 2013; Loppin, Dubruille and Horard, 2015; Rathke *et al.*, 2010; Zelinger *et al.*, 2024; Landmann *et al.*, 2009; Ostermeier *et al.*, 2004). Notably, evidence for protamine function in *Ae. aegypti* has recently emerged from investigations into the causes of cytoplasmic incompatibility in *Wolbachia* infected males. Researchers have shown that disruptions in the histone-protamine transition leads to the male sterility (Kaur *et al.*, 2024; Kaur *et al.*, 2022).

The SNBP genes thus represent a set of candidates whose regulatory regions could be utilized to drive Cas9 expression and induce a paternal effect in SuperSIT embryos in *Ae. aegypti*. The research group identified an orthologue of *ProtA* in *Ae. aegypti* (hereon referred to as *proA*, AAEL003177) with male biased and testes-specific expression (Akbari *et al.*, 2013a; Matthews *et al.*, 2018; Gramates *et al.*, 2022). The putative regulatory regions upstream of *proA* were then used to drive expression of a fluorescent protein gene (DsRed) in reporter constructs that were microinjected into *Ae. aegypti* embryos for *piggyBac*-mediated transgenesis. Consistent DsRed expression was observed in the base of the testes across four isolines, prompting the assessment of the same regulatory regions when driving Cas9 expression, which will be described in this chapter.

Finally, a longstanding candidate for testes specific regulation of gene expression comes from the *β2-tubulin* (*β2t*) gene (Kemphues *et al.*, 1982; Santel *et al.*, 2000). Studies in *D. melanogaster* determined that *β2t* is expressed at high levels throughout spermatogenesis and is important for sperm structure, specifically for the formation of

axoneme microtubules found at the sperm head (Kemphues *et al.*, 1982; Santel *et al.*, 2000; Rudolph *et al.*, 1987). The regulatory regions of *β2t* orthologs have since been shown to drive expression of reporter genes in the testes of *An. stephensi* and *Ae. aegypti,* as well as other mosquito and insect pest species (Catteruccia, Benton and Crisanti, 2005; Smith *et al.*, 2007; Yamamoto, Sumitani and Matsuoka, 2016; Zimowska, Nirmala and Handler, 2009; Xu *et al.*, 2015; Chen *et al.*, 2021; Scolari *et al.*, 2008). In *Ae. aegypti, β2t*-driven reporter proteins were also shown to persist in the spermathecae of wildtype females mated to transgenic males (up to two weeks post mating), although sperm-specific DsRed presence was not demonstrated (Agudelo *et al.*, 2022; Smith *et al.*, 2007). Finally, proof of principle that a *β2t* promoter could induce a paternal effect was established, albeit inadvertently, from experiments with the X-shredding endonuclease I-PpoI when it was used to develop a synthetic sex-distorter system in *An. gambiae* (Windbichler, Papathanos and Crisanti, 2008; Galizi *et al.*, 2014). *β2t*-driven expression of I-PpoI in males caused a bias towards Y-bearing sperm but also sterility due to carry-over of stable endonuclease to zygotes, which led to unintentional cleavage of the maternally inherited X chromosome and dominant embryonic lethality.

Considering these results, the *β2t* gene also emerged as a candidate gene with an expression pattern that could be desirable for SuperSIT effector gene expression. As such, the research group made attempts to reproduce the *β2t* expression pattern observed in *Ae. aegypti* using the promoter region (<1 kb) specified by Smith *et al.* (2007) but was unsuccessful. This prompted the engineering of reporter lines with a longer putative promoter segment (>4 kb), which ultimately showed strong testis specific expression in male pupae and adults but showed no persistence in spermatheca or sperm. This promoter was chosen for further assessment when driving expression of Cas9.

In summary, this chapter aimed to assess the activity of three putative germline promoters derived from the *Ae. aegypti* homologs of *shu*, *ProtA*, and *β2t* when driving the expression of Cas9 *in vivo.* The regulatory elements driving Cas9 expression will be assessed with homing assays utilizing the previously validated sgRNA element targeting *kmo*. In addition to identifying promoters that can achieve spatiotemporal control of Cas9, a key objective of SuperSIT development is ensuring canonical expression of the promoter-Cas9 construct within the chromatin context of the M locus. Therefore, genomic integrations of *proA*-Cas9 and *β2t*-Cas9 will be attempted via phiC31-recombinase-

mediated cassette exchange at an M-linked docking site, as well as produced via *piggyBac*-mediated transgenesis, which integrates transgenes semi-randomly in the genome and can often produce M or m-linked integrations by chance. As an m-linked isoline expressing *shu*-Cas9 was available from previous transgenesis experiments performed by the research group, fertile male recombinants expressing the construct from an M-linked site will be isolated from mass screens of $F_1$ offspring from transgenic males. Finally, challenges to achieving the desired spatiotemporal control of transgenes in *Ae. aegypti* is discussed and possible future directions are suggested.

## 2.3 Methods

### 2.3.1 Design and cloning of plasmids for *piggyBac* transgenesis

Two integration vectors were designed by Leonela Carabajal Paladino and Michelle Anderson to test expression of Cas9 under the control of the putative regulatory regions of *β2t* (AAEL019894, plasmid AGG-2141) and *proA* (AAEL003177, plasmid AGG-2142) (Figure 3). The Cas9 was designed to be fused at its N-terminus to a DsRed fluorescent protein gene using a ubiquitin followed by a serine residue (UbS) to increase the stability of Cas9 (Chassin *et al.*, 2019; Varshavsky, 2011). The backbone vector was designed to facilitate genome transformation via *piggyBac* transposase or phiC31-recombinase mediated cassette exchange (RCME) and CRE-lox recombination and thus included *piggyBac* and attB and loxP sites, respectively (Figure 4) (Haghighat-Khah *et al.*, 2015).

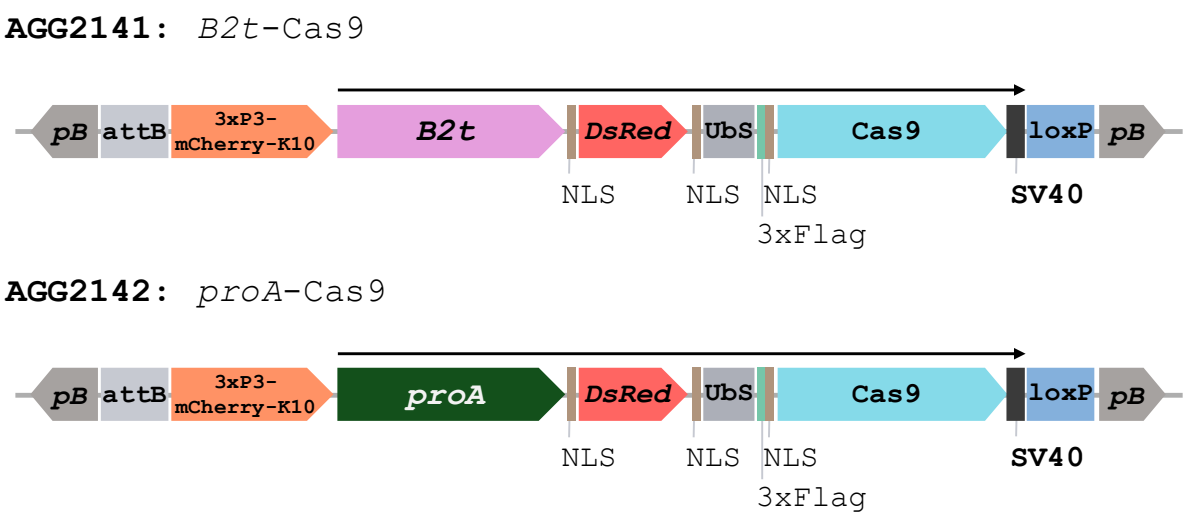



*Figure 3 Schematic of plasmids AGG-2141 and AGG-2142 for piggyBac-mediated transgenesis and phiC31-recombinase mediated cassette exchange (RCME). The plasmids are made up of two expression cassettes: the 3xP3-mCherry-K10 transformation marker and a male germline promoter (β2-tubulin (β2t) or protamine A (proA)) driving expression of a DsRed-Cas9 fusion (black line indicates entire transcript). NLS refers to nuclear localization signal.*

Precursor plasmids AGG-2083 and AGG-2084 were cloned from existing plasmids cloned by the research group using four-fragment NEBuilder HiFi DNA Assembly (NEB). Primer sequences were designed to create a minimum overlap of 25 bp between adjacent fragments (Appendix A Table 1). The *β2t*-DsRed fragment was amplified from AGG-1918 and the *proA*-DsRed fragment from AGG-1920. Both promoters include exactly 4.5 kb of genomic sequence upstream of the respective gene's start codon, including the 5' UTR. The regulatory regions were originally amplified from genomic DNA isolated from wild-

type *Ae. aegypti* of the Liverpool strain (LVP). The *SpCas9* coding sequence was codon optimized for *Ae. aegypti* (Anderson et al., 2023a).

The UbS-Cas9 fragment was amplified from AGG-2139, and the SV40 terminator was amplified from AGG-1918. The backbone of both plasmids was derived from a parent insect transformation plasmid, AGG-2071, by restriction enzyme digest using FseI and AscI (NEB). The AGG-2083 and AGG-2084 precursor plasmids were adapted by replacing the HR5/IE1-AmCyan marker cassette with 3xP3-*Ae*mCherry to minimize basal expression of the Cas9 due to the enhancer elements in HR5/IE1 and to keep the fluorescent marker uniform for all Cas9 lines (Anderson *et al.*, 2019). This was done by first excising the HR5/IE1-AmCyan fragment from AGG-2083 and AGG-2084 using restriction enzymes PacI and AflIII, and PacI and Bsu36I (NEB), respectively. The 3xP3-mCherry-K10 cassettes were amplified from AGG-1927 using primers in Appendix A Table 1 and cloned into the linearized vectors using NEBuilder HiFi DNA Assembly. The entire sequence of the final assembled plasmids was verified by Sanger sequencing (Appendix A Table 1). Plasmids with the expected sequence were midi-prepped using NucleoBond Xtra Midi kit for endotoxin-free (EF) plasmid DNA (Macherey-Nagel, UK) and eluted in EF water. Midi-prepped plasmids were then sequenced again.

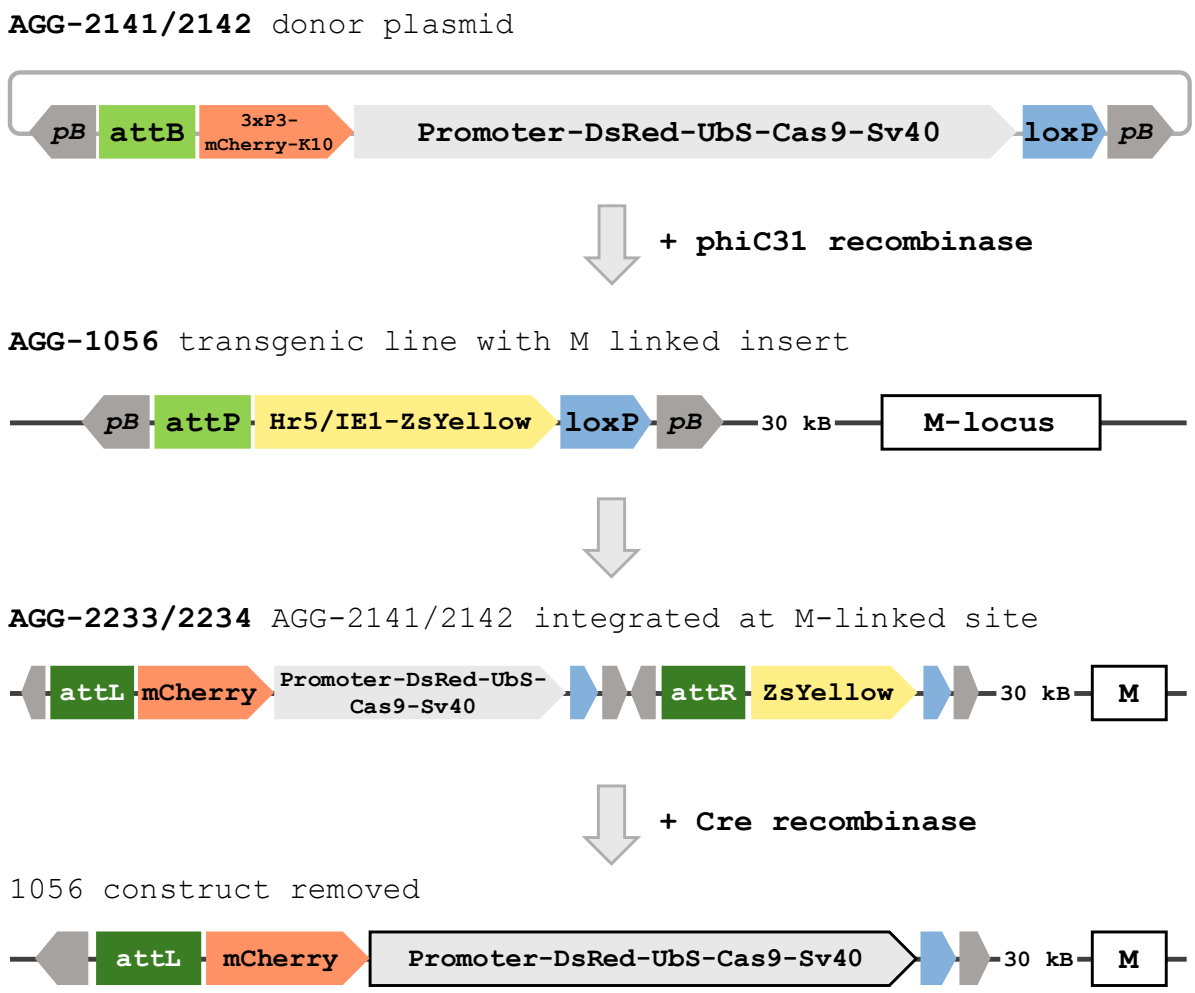



*Figure 4 Schematic of phiC31 recombinase mediated cassette exchange and CRE-Lox in M-linked docking line AGG-1056. Following phiC31-mediated recombination at M-linked site, canonical integration of donor plasmids AGG-2141 and AGG-2142 would result in expression of both transgenic markers: Hr5/IE1-ZsYellow from the original docking line and 3xP3-mCherry from the integrated donor plasmids. In a separate micro-injection experiment, Cre-recombinase is used to recognize the two identical and similarly oriented loxP sites and excise the intervening DNA, yielding only the donor plasmid at the original M-linked site.*

### 2.3.2 Mosquito rearing

*Ae. aegypti* of the Liverpool strain (LVP) were reared under standard conditions at 26-28°C, 65-75% relative humidity, and a 12:12 h light/dark (1 h dawn/dusk) cycle (Kuno, 2014). Larvae were reared in purified water and fed fish food (Tetramin Tropical Flakes, Tetra GmbH, Germany). Males and females were separated as pupae using a Larval Pupal separator (John Hock Company https://www.johnwhock.com/products/laboratory-equipment/larval-pupal-separator) and verified under a light microscope to ensure virginity prior to experimental crosses (Focks, 1981). Adults were maintained in cages with continuous access to a 10% sucrose solution. Mated adult females were fed

defibrinated horse blood (TCS Biosciences Ltd., UK) through a Parafilm and hog gut membrane using a Hemotek system (Hemotek Ltd, Blackburn, UK) for 30 minutes. Eggs were collected on dampened unbleached filter paper (Edesia Espress Ltd., UK) three days post bloodmeal and hatched a minimum of five days after in a vacuum chamber.

### 2.3.3 Microinjection

Microinjection of *Ae. aegypti* embryos was carried out following published protocols (Lobo *et al.*, 2006; Morris, Eggleston and Crampton, 1989). Injection mixes for *piggyBac*-mediated transgenesis consisted of 500 ng/µl *piggyBac* donor plasmid (eluted in EF water), 300 ng/µl AePUb-hyperactive-*piggyBac* transposase (AGG-1245), and injection buffer (Coates *et al.*, 1998). Injection mixes for phiC31 recombinase mediated cassette exchange (phiC31-RCME) consisted of 500 ng/µl donor plasmids, 300 ng/µl AePub-phiC31-SV40 (AGG-1733), and injection buffer (Carabajal Paladino *et al.*, 2023). For *piggyBac* transgenesis, constructs AGG-2141 and AGG-2142 were injected into LVP. For phiC31-RCME, constructs AGG-2141 and AGG-2142 were injected into eggs from transgenic line AGG-1056 that carried attP sites at an M-linked site (Figure 4). Injected embryos were hatched after five days.

### 2.3.4 Establishing transgenic mosquito lines expressing Cas9

#### 2.3.4.i From microinjection

For *piggyBac* transgenesis, male and female $G_0$ survivors were reared separately and crossed in pools of 3-5 $G_0$s to excess wildtype (Figure 5). For phiC31-RCME, only male $G_0$ survivors were kept and crossed in a single pool to wild-type females. $G_1$ larvae from each pool were then screened under fluorescence for presence of the transformation markers using a Leica MZ10 microscope (Leica Biosystems, Germany). Transgenic lines were established from single $G_1$ positives crossed to wild type. $G_2$ offspring were then screened for transgene inheritance and sex bias to estimate the number of insertions and whether any were sex-linked. $G_2$ individuals suspected of having multiple insertions were outcrossed again and allowed to lay individually to isolate transgenic isolines that produced a 50% transgene inheritance rate, indicating a single insertion.

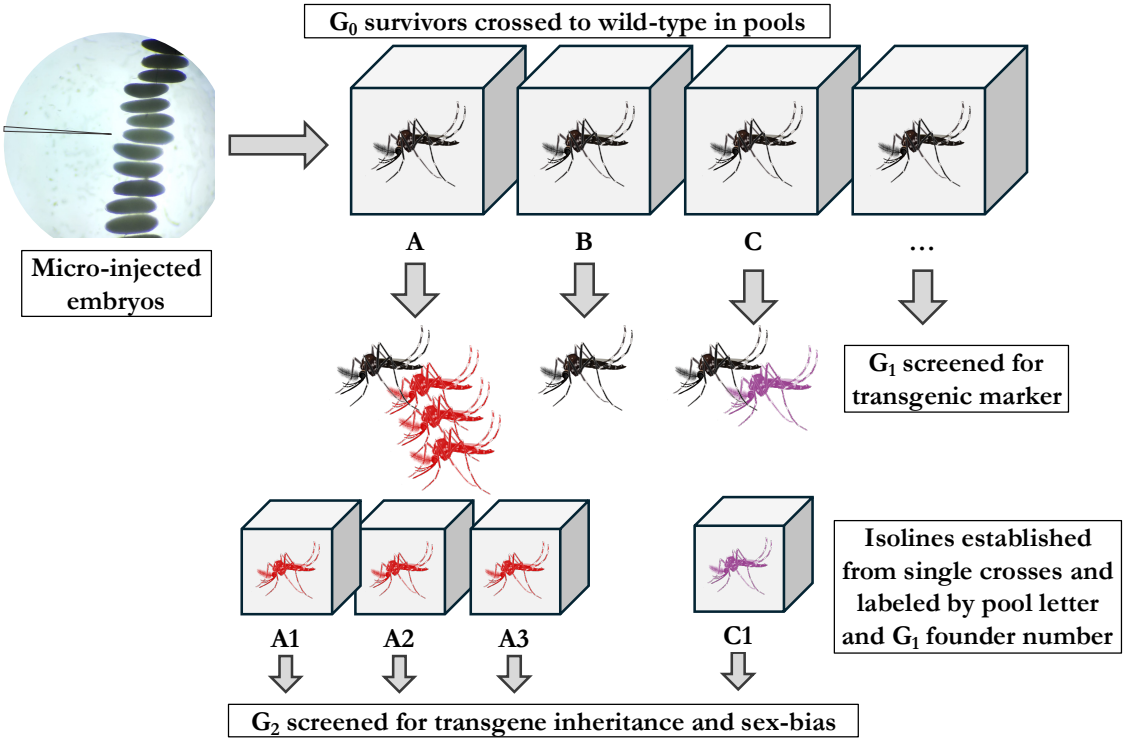



*Figure 5 Establishing transgenic lines from* piggyBac*-mediated transgenesis. Surviving micro-injected embryos ($G_0$) are reared to adults, separated by sex, and crossed to wildtype. Each $G_0$ cage is designated with the construct name and a pool letter. $G_1$ offspring are screened for transformation events. Transgenic $G_1$s from within a pool (e.g.* ***A****) could be the result of one or more insertion events, but will all be distinct from those in pool* ***C****. Thus, isolines are established from single $G_1$ positives crossed to wild type. $G_2$ offspring are then screened for transgene inheritance and sex bias to estimate the number of insertions and whether any were sex-linked. Isolines should carry only a single transgene insertion.*

### 2.3.4.ii From spontaneous recombination of an existing line

To generate an M-linked *shu*-Cas9 expressing line, an existing m-linked *shu*-Cas9 line (AGG-1551B1) was mass-reared, and the offspring screened for rare males with a paternally inherited transgene (Figure 6). Such an event would occur by spontaneous recombination of the transgene from the m-locus to the M during meiosis, allowing a transgenic male to pass the transgene on to its male (M-inheriting) progeny. Approximately 200 transgenic male AGG-1551B1 mosquitoes were crossed to 300 wildtype females in duplicate cages. The entire $F_1$ progeny were then screened as $L_1$ larvae five hours post hatching for the presence of the transgene marker (*Ae*PUb-mCherry) using a BioSorter large particle flow cytometer (Union Biometrica). Positive larvae were reared to the pupal stage and mass-sorted for males using the larval pupal glass plate separator. Positive males were crossed separately to an excess of wildtype females and their $F_2$ progeny screened for male-biased inheritance of the transgene.

Progeny from three M-linked lines were pooled to ensure fitness of the line and reared by crossing to LVP females. The M-linked *shu*-Cas9 line was named AGG-2235.

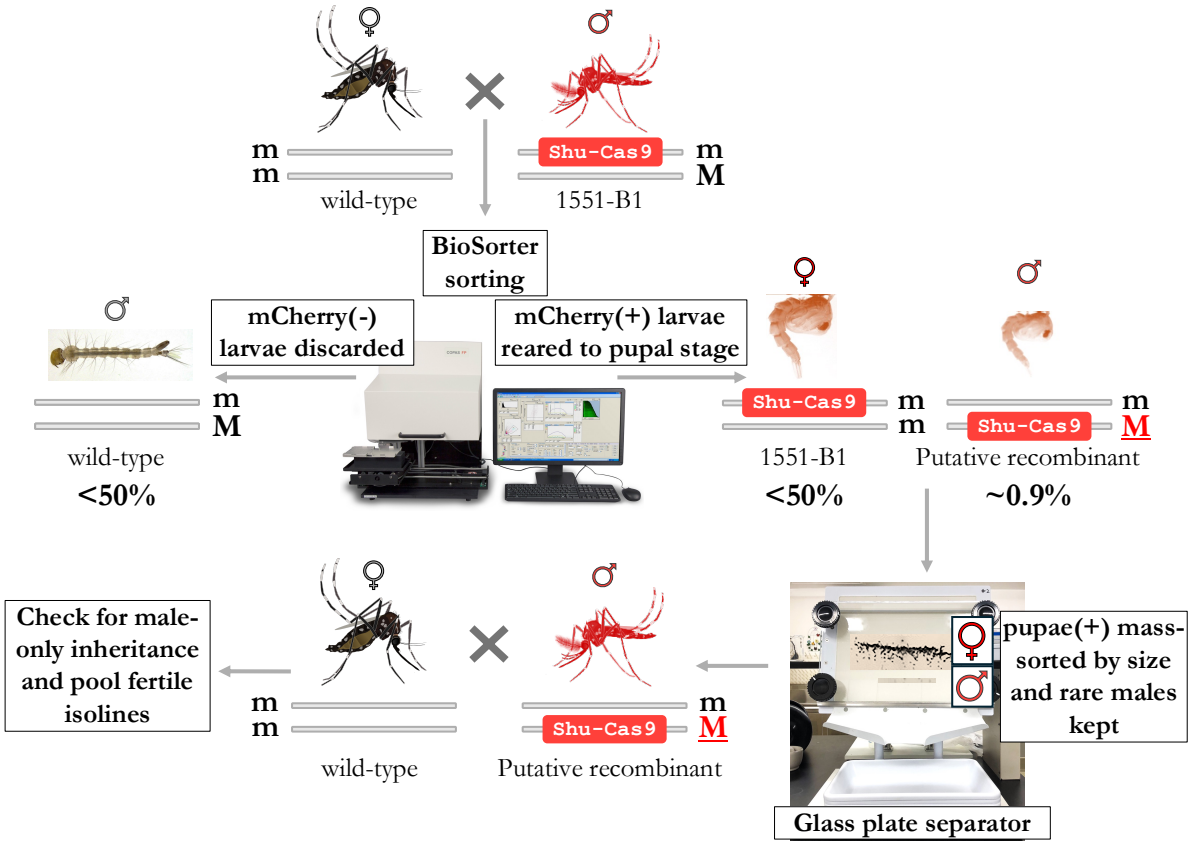



*Figure 6 Method for isolation of M-linked recombinant shu-Cas9 transgenic line. M-linked recombinant line from pooled isolines was named AGG-2235. Recombination of the shu-Cas9 construct from the m- to the M-carrying chromosome was estimated at 0.9-1%.*

### 2.3.5 Split-drive assessments of Cas9 activity

Transgenic males (>20) were crossed in groups to females (>30) from transgenic line AGG-1095N1, which expressed sgRNAs targeting *kmo* from a knock-in integration in *kmo* (Anderson *et al.*, 2024). $F_1$ progeny were then screened for transgene inheritance and eye mosaicism to test for somatic expression and paternal deposition of Cas9. $F_1$ trans-heterozygous males carrying the Cas9 construct and sgRNA element were crossed to females that were homozygous for a knock-out (KO) mutation of *kmo* (*kmo*-KO) that could not be re-cleaved by Cas9. $F_2$ progeny from both crosses were then scored again for transgene inheritance and eye mosaicism (Figure 7).

## 2.4 Results and discussion

Controlling the spatiotemporal expression of Cas9 is critical for developing a CRISPR/Cas9-based SuperSIT system that can disrupt the fitness of daughters descending from male carriers. Germline promoters that regulate expression during the early stages of sperm development could be suitable for targeting a female-specific gene that leads to dominant mutant phenotypes in daughters, while promoters active during later stages might be more likely to induce paternal deposition of effectors into embryos post-mating and disrupt both copies of a recessive female-specific gene. Cas9 activity driven by the germline-specific regulatory regions of *shu*, *β2t,* and *proA* was assessed here with an sgRNA-expressing element targeting the recessive eye pigment gene, *kmo* (Anderson *et al.*, 2024). The activity of each Cas9 line was assessed by recording the rates of inheritance bias of the sgRNA element (indicating homing or meiotic drive), non-homologous end joining in the germline (NHEJ), paternal deposition, and somatic expression (Figure 7).

### 2.4.1 Establishing transgenic lines

A transgenic line expressing *shu*-Cas9 from a site linked to the sex-determining M-locus was isolated from spontaneous recombination of an existing m-linked line (AGG-1551B1) and named AGG-2235 (Anderson *et al.*, 2023a) (Figure 6). The regulatory regions of *shu* consisted of the entire 5'UTR and any additional putative promoter elements within approximately 2kb of genomic DNA upstream of the transcription start site, as well as the 3'UTR and putative poly-adenylation signal to terminate Cas9 expression. Recombination of the *shu*-Cas9 construct to the M-locus was estimated at 0.9% (Table 1).

*Table 1 Linkage of sex-linked (M or m allele) transgene integrations. Linkage is defined as the proportion of males that were positive and of females that were negative, due to recombination from the M to the m in males and off the M in females, as a percentage. Parentheses indicate the parental cross; the left side of the colon specifies the genotype of the father, and the right side specifies the genotype of the mother. *AGG-1349 is an sgRNA expressing line, but it is assumed to be neutral here and should not affect transgene inheritance of the M-linked shu-Cas9.*

| Line | Cross | Tot screened (males) | Recombination from m to M | Tot screened (females) | Recombination from M to m |
|---|---|---|---|---|---|
| 2235 | [(2235: LVP): (1349*: LVP)] | 767 | 99.1% | 762 | 99.0% |
| 2141C3 | ((LVP: 2141C2 (G4)): LVP): LVP | 437 | 95.4% | 583 | 96.7% |

Two constructs were then cloned for semi-random *piggyBac*- and site-specific phiC31-mediated transgenesis of *Ae. aegypti* embryos. These plasmids incorporated the putative promoter regions of *β2t* and *proA* to initiate expression of a DsRed fluorescent protein

gene fused with a ubiquitin (UbS) to the N-terminal end of a Cas9 gene. The selected promoter regions included 4.5 kb of genomic DNA upstream of each gene's respective start codon, while the SV40 terminator sequence was incorporated downstream of the Cas9, which had previously been shown to be adequate for testes specific expression (Smith *et al.*, 2007). Both regulatory pairs had been shown by the lab group to drive expression of a fluorescent reporter protein in the testes.

The DsRed-UbS-Cas9 fusion was included for three reasons: one, to track Cas9 expression by spatiotemporal expression of DsRed, two, to efficiently identify active *piggyBac* integrations that showed a high degree of Cas9 expression, assuming DsRed fluorescence correlated with Cas9 activity, and three, to increase the half-life of the translated Cas9 protein, and thus the possibility of paternal deposition, by engineering a stabilizing N-terminal residue (in this case serine (S)) that becomes exposed upon ubiquitin cleavage (Varshavsky, 1996; Varshavsky, 2011; Chassin *et al.*, 2019). To test *β2t*-Cas9 without the intermediary fluorescent protein fusion, a third construct was cloned by replacing the promoter region in the transformation plasmid used to produce the *shu*-Cas9 isolines (1551) with that of the 4.5 kb putative *β2t* promoter (Anderson *et al.*, 2023a). A p10 terminator sequence could not be assembled into this plasmid following multiple attempts, so the final construct included the putative *β2t* promoter upstream of the Cas9 and the 3'UTR *shu*-terminator sequence downstream of Cas9.

Transgenesis of *Ae. aegypti* mosquito embryos was performed with hyperactive *piggyBac* transposase for random genomic integration and phiC31 recombinase for site-specific integration at an existing M-linked docking site (*β2t* and *proA* only), but only *piggyBac*-mediated transgenesis was successful. Integration of each construct was verified by screening $G_1$ larvae for mCherry fluorescence after which positive males from each pool were used to establish transgenic isolines. Due to the potential for construct expression to vary by *piggyBac* integration site ("position effects"), multiple independent transgenic lines were generated per construct. Isolines represent independent single integrations of each Cas9 construct and were named with a letter that denoted the pool they originated from and a number indicating the $G_1$ founder male (Table 2). The number of isolines generated for each construct is shown in Table 2 and isolines carrying insertions linked to the sex-determining regions, i.e. the M or m locus, are highlighted.

*Table 2 piggyBac mediated transgenesis of* Ae. aegypti *embryos with Cas9 expressing constructs driving by germline regulatory regions* B2t *and* proA*. Survival rates of microinjected embryos, number of $G_0$ founder pools set up, and number of isolines generated for each micro-injection experiment are shown. *As the survival rates from the initial microinjection experiment of 2142 were low, i.e. <10%, this plasmid was midiprepped and micro-injected again. Transgenic isolines were established from single $G_1$ positives crossed to wildtype females. Isolines carrying single insertions of transgenes were isolated by screening for normal 50/50 Mendelian inheritance rates that would be expected for a single dominant allele. Independent isolines are represented with a letter denoting the pool and $G_1$ founder male they originated from, respectively. Isolines carrying sex-linked insertions are underlined; an * indicates the isoline is m-linked (female-biased inheritance from males), while a ^ indicates the isoline is M-linked (male-biased inheritance). nr = not recorded.*

| AGG: promoter | Embryos injected | $G_0$ Survival (%) | No. $G_0$ founder pools | Isolines generated |
|---|---|---|---|---|
| 2141: *B2t* | 205 | 28% | 13 | 8 (C1, C2*, C3^, D1, H1, K1, M1, M2^) |
| 2142: *proA* | 295 | 4% | 6 | 1 (E1) |
| 2142: *proA** | 400 | 19% | 12 (only 5 screened) | 2 (D1, E2^) |
| 2349: $B2t_{shu}$ | nr | nr | 7 | 2 (B1, G2) |
| 2233: *B2t* in 1056 | 400 | 17% | 4 | 0 |

## 2.4.2 Assessing construct activity

### 2.4.2.i DsRed expression

Males expressing the *β2t*- and *proA*-driven DsRed-UbS-Cas9 fusions were screened for testes-specific DsRed expression. Preliminary results from minimal reporter constructs driving fluorescent protein expression with identical regulatory regions (promoter-DsRed) had shown strong testes expression that was clearly visible through the cuticle of undissected pupae in two *β2t*-DsRed isolines, while dimmer expression was observed in four protamine isolines, only two of which showed expression in undissected pupae. Contrary to results reported in the literature for a similar *β2t*-DsRed reporter construct (but a shorter promoter fragment), no DsRed activity was observed in the spermathecae of mated females (Smith and Wakimoto, 2007; Agudelo *et al.*, 2022).

Expression of DsRed from the DsRed-UbS-Cas9 construct was low in six *β2t* isolines, while no DsRed expression was observed in three *proA* isolines. For the *β2t* isolines, expression was lower than the minimal *proA*-DsRed designs discussed above, meaning all had to be dissected to observe any fluorescence. Multiple variables could have contributed to this lower expression, making it unclear what the exact cause was. A recent finding by the lab group showed that the Hr5/IE1 enhancer/promoter element, which was used to drive expression of the genomic marker cassette in the minimal reporter constructs, can lead to allele overexpression in *An. stephensi*. This promoter was replaced by a 3xP3 promoter in the fusion constructs, so it is possible that this reduced overall expression from the entire transgene insertion. It is also possible that position effects around the *β2t*-DsRed

construct in the assessed isoline contributed to higher expression, although the expression of the reporter gene in the *β2t* and *proA* fusion constructs was consistent across the four isolines assessed. Lastly, the ubiquitin fusion could be functioning unexpectedly and reducing the expression of both the DsRed and Cas9 transgenes. This hypothesis gains some support from the observation that a *β2t*-Cas9-*shu* expressing isoline, which lacked the fusion, showed comparatively higher rates of cleavage to the fusion constructs (see below), although a compounding variable is the inclusion of the *shu* terminator sequence. Unfortunately, an unfused Cas9 construct was not performed as a control to determine the significance of the ubiquitin and putatively stabilizing S-terminal amino acid. Further investigation into the mechanisms underlying the activity of the ubiquitin fusion would be warranted to fully understand its effects on gene expression in this context.

### 2.4.2.ii Cas9 expression

Isolines expressing *β2t*-DsRed-UbS-Cas9, *proA*-DsRed-UbS-Cas9, and *β2t*-Cas9-*shu*, and the one M-linked *shu*-Cas9 isoline were tested for Cas9 activity with a transgenic line ubiquitously expressing sgRNAs targeting the eye pigment gene, *kmo* (Anderson *et al.*, 2024). The *kmo* sgRNA element is inserted within the *kmo* gene within the exact region that the sgRNAs target and includes an AmCyan expression cassette for genotype scoring. In these experiments, the effects of paternal inheritance of the Cas9 nuclease were assessed across two generations ($F_{1\text{-}2}$). In the first set of crosses ($F_0$ cross), $F_1$ male trans-heterozygotes (i.e. those that carry both a copy of the Cas9 transgene and the sgRNA element) and $sgRNA^+$-only heterozygotes (i.e. those that carry only the sgRNA element and no Cas9 transgene) were generated (Figure 7). The trans-heterozygous and heterozygous males were then crossed separately to females homozygous for a knock-out mutation in *kmo* (*kmo*-KO) to generate $F_2$ progeny that carry one maternally inherited disrupted copy of *kmo*. This allows the attribution of the eye phenotype to the paternal kmo allele that did not carry the sgRNA expressing transgene. Larval eye phenotype was scored in the $F_1$ and $F_2$ generations as homozygous loss of *kmo* results in lack of pigmentation in the eye. This can occur across the entirety of the eye ("white-eyed") or in patches ("mosaic-eyed") depending on the spatiotemporal activity of the nuclease.

First, paternal deposition of Cas9 from the $F_0$ father can induce mutations in the $F_1$ progeny inheriting the sgRNA element ($sg^+/Cas9^-$, $sg^+/Cas9^+$, and $sg^+$/ns in Figure 7) and

could manifest in both mosaic- and white-eyed phenotypes (Figure 9). Nuclease activity in embryonic somatic cells of trans-heterozygous individuals can also induce mosaic/white eyes and by comparing their phenotype to that of the sgRNA$^+$ heterozygotes, the rate of somatic activity can be deduced. Germline activity in trans-heterozygotes that results in NHEJ mutations in the wildtype *kmo* allele is revealed by the frequency of white-eyed individuals in those not inheriting the sgRNA element (sg$^-$/Cas9$^-$, sg$^-$/Cas9$^+$, and sg$^-$/ns in Figure 8A and C). This frequency could theoretically be attributed to paternal deposition into the $F_2$ as well, but the likelihood of paternal deposition is assessed by screening the $F_1$ pool for eye pigment abnormalities (assuming ubiquitous expression of sgRNAs but see below). Further, paternal deposition of Cas9-sgRNA complex that results in a mosaic-eyed phenotype in the $F_2$ is assessed in those that do not inherit the sgRNA element. Finally, crossing sgRNA$^+$-only $F_1$ heterozygotes to the *kmo*-KO line reveals whether $F_0$ paternal deposition induced NHEJ in the $F_1$ germline that was inherited in the $F_2$ offspring.

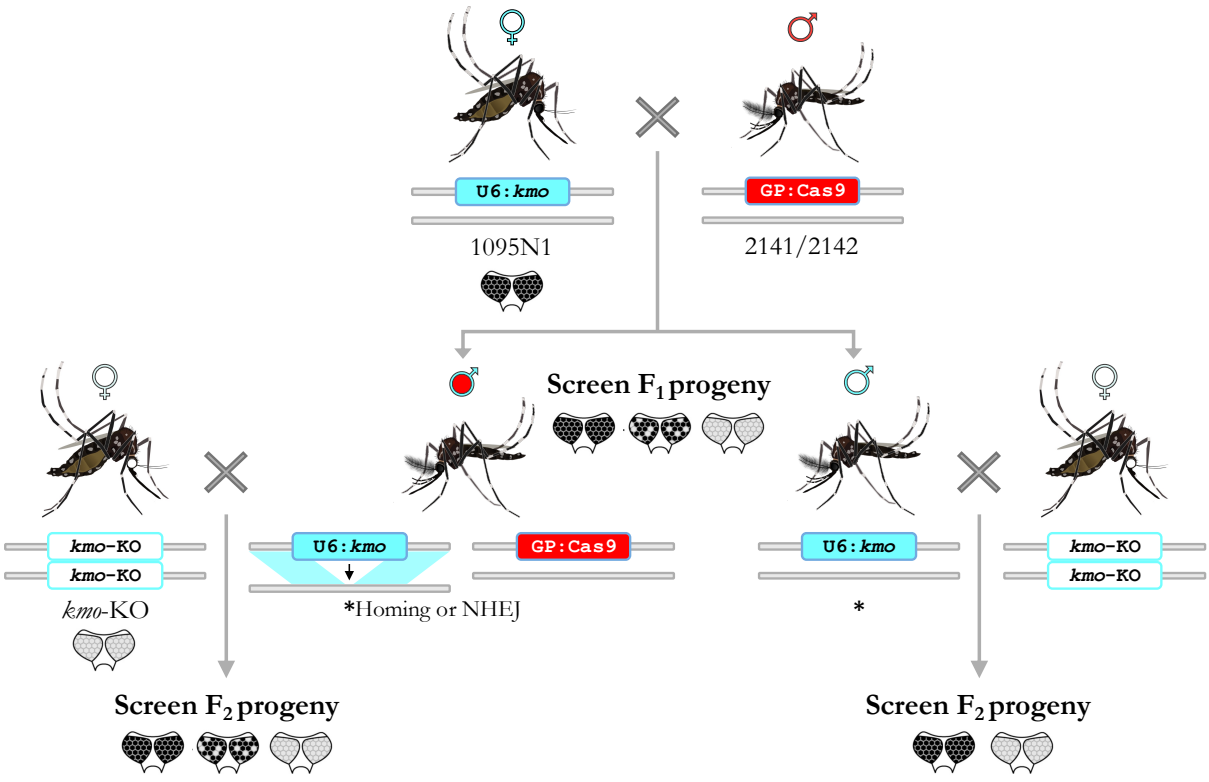



*Figure 7 A) Diagram of the genetic cross used to assess the spatiotemporal activity of germline-specific (GP) promoters shu, β2t, and proA when expressing Cas9 (GP:Cas9). Paternal deposition of Cas9 from the $F_0$ father can induce mutations in sgRNA+ only (U6:kmo) heterozygotes (teal) and transheterozygotes (U6:kmo + GP:Cas9) (red and teal). Somatic expression in transheterozygous individuals can also induce cutting and by comparing the phenotype of the sgRNA+/Cas9- and sgRNA+/Cas9+ individuals, the rate of somatic expression can be deduced. Crossing sgRNA+ only individuals to a* kmo *knock-out (KO) line reveals how much of deposition-induced cutting in the $F_1$ occurs in the germline and thus inherited in*

*the $F_2$ offspring. When transheterozygous individuals are crossed to kmo-KO, mutant phenotypes in Cas9+ only heterozygotes (sgRNA-/Cas9+, red) and wildtype (sgRNA-/Cas9-, grey) can arise from inherited deposition-induced mutations as in the previous cross, as well as germline cutting from expression in the transheterozygous parent (which includes somatic expression), and lastly from paternal deposition of complexed Cas9/gRNA.*

Results from the $F_0$ and $F_1$ crosses are shown in detail in Figure 7 and Figure 8, respectively, and summarized in Table 3. The primary finding from these sets of experiments was that a weak signal of paternal deposition was observed in the $F_2$ of a single proA-Cas9 isoline (Figure 9). As this group was unfortunately not scored for the inheritance of Cas9, deposition of complexed Cas9/sgRNA cannot be conclusively differentiated from deposition of only sgRNA/s and somatic expression of Cas9, although the latter scenario is unlikely. First paternal deposition of sgRNAs has not normally been observed, second it is unlikely that the sgRNAs would be stable on their own until Cas9 was expressed from the activated zygotic genome, and third no mutant phenotypes were observed in the $F_1$ transheterozygotes, indicating that the promoter is not normally active in somatic tissues and therefore would not have been likely to pair with deposited sgRNAs. The other two *proA* isolines showed no mosaicism in the $F_2$ but did show higher rates of white-eyed individuals, the highest of which was incidentally in an M-linked isoline (2142 E2). These two lines were not screened at the $F_1$ stage to ascertain paternal deposition of Cas9, so their results should be interpreted with caution as paternal deposition can also result in complete loss of eye pigmentation, though in this case it is unlikely. No white-eyed individuals were observed in the $sg^-$ progeny of the $sgRNA^+$ heterozygote crosses, which indicates no paternal deposition occurred from the $F_0$ cross that caused mutations in the $F_1$ germline cell lineages. Otherwise, there was no paternal deposition observed for *β2t*-Cas9 nor *shu*-Cas9, for the latter of which paternal deposition of Cas9 (but not complexed Cas9-sgRNAs) had previously been observed (although not in the 1551B1 isoline from which the M-linked *shu*-Cas9 line (2235) was derived).

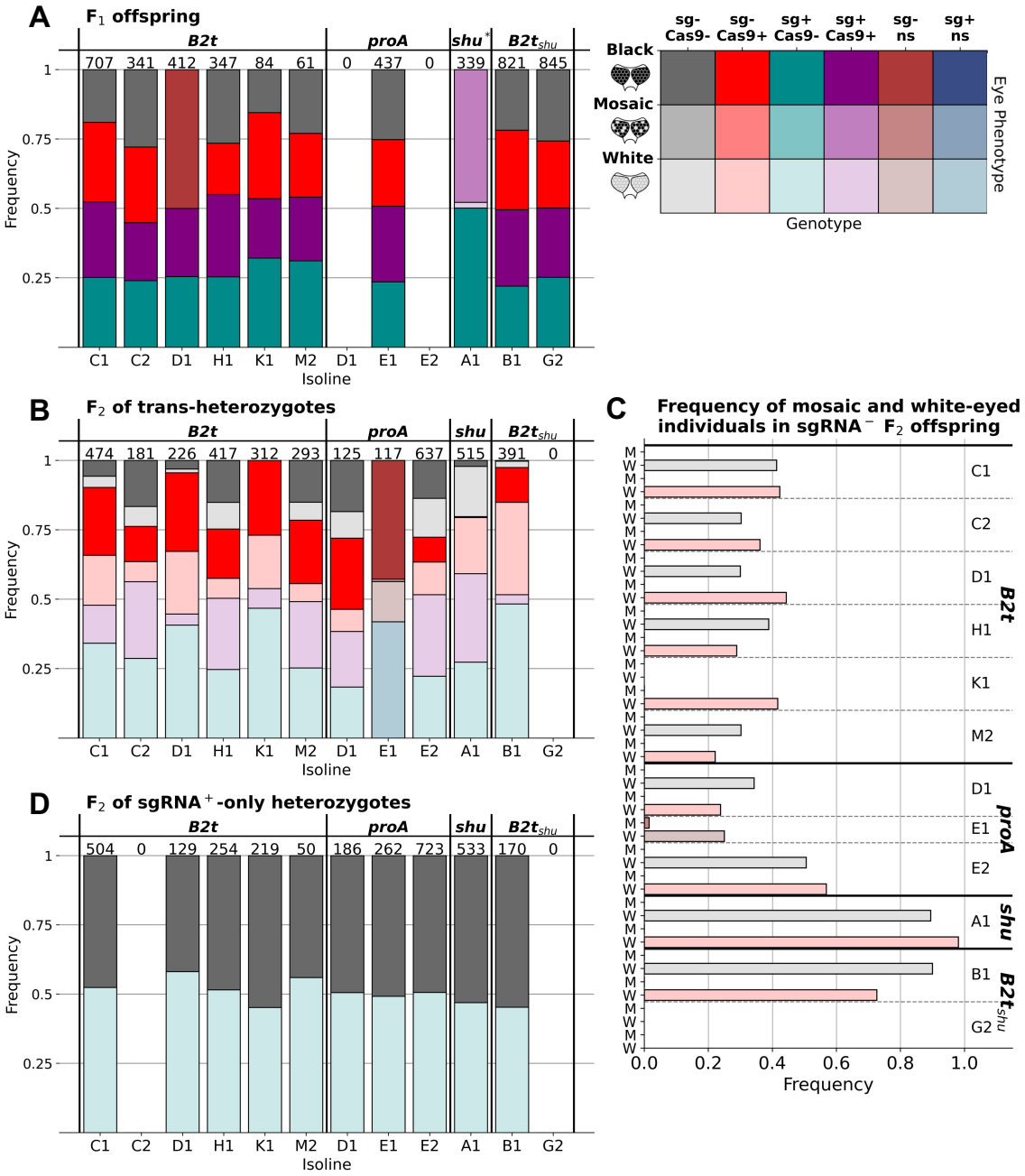



*Figure 8 A) Inheritance bias of sgRNA-expressing element targeting kmo and frequency of black-, mosaic-, and white-eyed phenotypes (designated by lighter hues of respective genotype colours). Shu* indicates that the F0 cross was performed with females that were homozygous for the kmo sgRNAs and thus all offspring inherited the sgRNAs. Larval eye phenotype and sgRNA inheritance rates in F₂ offspring of B) trans-heterozygous (sgRNA⁺/Cas9⁺) and D) single heterozygous (sgRNA⁺/Cas9⁻) males. C) Frequency of mosaic- (M) and white (W)-eyed larvae/pupae in individuals not-inheriting the sgRNA element. Homozygous loss of* kmo *results in lack of pigmentation in the eye (see key). Black-eyed is the wild-type phenotype, while total loss of kmo results in white eyes and partial loss of kmo across somatic tissue results in patches of white eyes, called "mosaic" phenotype.*

The results from the first assessment of paternal deposition in the $F_1$ generation should be interpreted with caution, though, as they may not be directly comparable to deposition in the $F_2$ generation. This is because sgRNAs may not be equally expressed in both the

developing embryo ($F_1$ sgRNA$^+$ heterozygotes) and the testes ($F_1$ transheterozygotes). In the $F_1$ sgRNA$^+$ heterozygotes, due to the delay in zygotic genome activation, the maternally inherited sgRNA element may not express in time to form a complex with the paternally deposited Cas9 before it is possibly degraded. While maternally deposited Cas9 under the control of other germline promoters is stable enough in the embryo to induce high rates of NHEJ, it may be that paternal factors are recognized and degraded earlier. In contrast, deposition into the $F_2$ generation from $F_1$ transheterozygotes could still cause NHEJ in the two hours when the zygotic genome is not yet activated. Paternal deposition could therefore have been more explicitly assessed by crossing $F_2$ transheterozygous males carrying a knock-out *kmo* mutation to wildtype females since germline sgRNA targeting and NHEJ would be blocked and thus any signal in the $F_3$ would be attributable only to paternal deposition. As a preliminary study this would be useful, but it will ultimately be necessary to demonstrate paternal deposition leads to knock-out of two wildtype copies of a gene rather than just one as has been shown here.

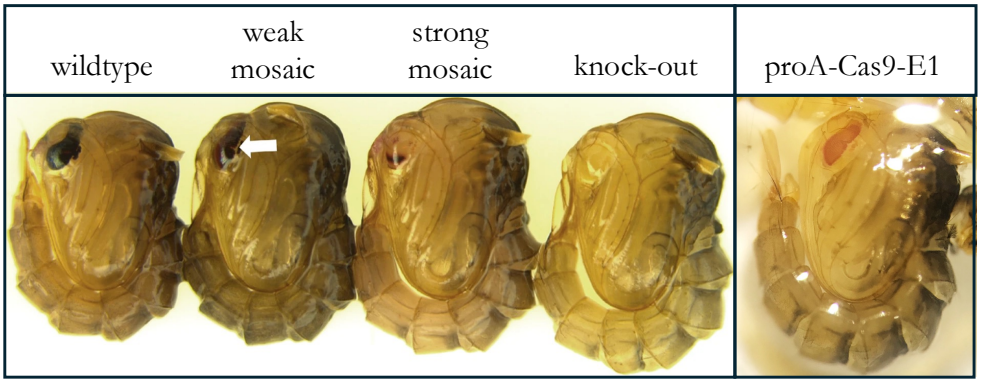



*Figure 9 Left) Pupae displaying eye phenotypes ranging from wildtype (dark) to mosaic to white eye (far right); white arrow in the weak mosaic indicated site of mosaicism (image reproduced from Anderson* et al.*, 2024). Right) Image of pupa isolated from cross of $F_1$ transheterozygous males (carrying the* proA*-DsRed-UbS-Cas9 fusion construct and sgRNA element targeting* kmo*) to* kmo*-KO (homozygous) females.*

The second main finding was that *shu*-Cas9 induced high rates of knock-out mutations—ranging from 89-98%—in the wild-type *kmo* allele inherited from trans-heterozygous males. Such high cleavage activity occurring in the male germline is desirable for development of a SuperSIT system targeting a dominant female-specific gene, such that all offspring inherit a mutated copy. The *β2t*-Cas9-*shu* line also achieved high germline disruption of the wild-type allele, between 73-90%; no somatic activity was observed in this line, which could limit off-target mutations in males. The germline cutting rate was around double that observed for the six isolines carrying the *β2t*-DsRed-UbS-Cas9

construct. As discussed above, it is possible that the UbS fusion reduced activity (and thus germline cleavage) in both the *β2t* and *proA* lines. It is also possible, however, that the *shu* 3'UTR influenced the activity of the Cas9 in some manner, through post-transcriptional modifications that increased the stability of the Cas9. Finally, the third main finding was that position effects were not strongly observed for either the *β2t* or *proA* fusion lines for which multiple isolines were tested, suggesting that the transgene insertion site did not significantly influence the observed outcomes. This could have been because the longer promoter fragments used in both cases (4 kb) helped to insulate the construct from significant position effects.

*Table 3 Summary of results from genetic crosses shown in detail in Figure 8. *Somatic activity judged by eye mosaicism in transheterozygotes.*

| Regulatory element | Isolines | Paternal deposition Cas9 only ($F_1$) | Paternal deposition Cas9-sgRNA complex ($F_2$) | Germline NHEJ (%) | sgRNA inheritance | Somatic activity in males ($F_1$)* |
|---|---|---|---|---|---|---|
| B2t | 6 | No | No | 28-44 | 45-54 | No |
| proA | 3 | No | Possible (1.5%) | 24-57 | 38-52 | No |
| shu | 1 | No | No | 89-98 | 59 | Yes (100%) |
| B2t-(shu) | 1 | No | No | 73-90 | 52 | No |

While these sets of experiments were useful in determining preliminary Cas9 deposition and putative germline cleavage, it will ultimately be necessary to demonstrate that paternal deposition leads to knock-out of two wildtype copies of a gene in the subsequent generation rather than just one as has been putatively shown here. This could be assessed by crossing $F_1$ transheterozygous males to wildtype.

## 2.5 Conclusion and future directions

This chapter aimed to evaluate the *in vivo* activity of three putative germline promoters derived from the *Ae. aegypti* homologs of *shu, proA,* and *β2t* in driving expression of Cas9. First, paternal deposition leading to eye mosaicism was shown with *proA*-Cas9, although the levels observed were minimal (<2%). Second, the results showed that high cleavage of the pre-meiotic sperm genome was possible with *shu*-Cas9 from an M-linked site, which is a key objective of a nuclease-based SuperSIT system targeting a dominant female-specific gene. This section discusses limitations of the study, the challenges of achieving the desired spatiotemporal control for paternal delivery of transgenes, and possible future directions.

To engineer a paternal effect with a transgene expressed from a hemizygous locus (i.e. the M locus in a SuperSIT system), transcription of the effector transgene must occur before spermatid individualisation, such that either the transcript or protein is provided to all spermatids and eventually deposited to the fertilized zygote. While most genes required for spermiogenesis in *Drosophila* are transcribed in diploid spermatocytes and translationally repressed until required—meaning pre-meiotic transgene transcription is possible and probably even necessary—challenges remain in 1) ensuring the stability of Cas9 transcript or protein throughout meiotic divisions, spermiogenesis, and fertilization, and 2) generating enough Cas9 product to impose a substantial effect in the embryo (Raja and Renkawitz-Pohl, 2005; Awe and Renkawitz-Pohl, 2010; Barckmann *et al.*, 2013; Vibranovski *et al.*, 2010). Thus, recapitulating the endogenous post-transcriptional and post-translational regulatory mechanisms that lead to efficient accumulation of gene products in size-restricted spermatids is of most importance.

The regulatory regions of two *Ae. aegypti* homologs of the *Drosophila* genes *β2t* and *ProtA* were investigated in this chapter based on robust expression throughout spermatogenesis and effective post-transcriptional regulation in spermatocytes, respectively. A potential limitation of the study was the use of the viral SV40 sequence for transcription termination in the *proA*-Cas9 and *β2t* -Cas9 constructs rather than the respective endogenous 3'UTRs. The 3'UTR, however, is important for post-transcriptional regulation, playing a crucial role in the localization, stability, and translational repression and efficiency of transcripts (Pfeiffer, Truman and Rubin, 2012; Mayr, 2019). While the function of the 3'UTR was acknowledged at the time of construct synthesis, several

reasons led to its exclusion from both final constructs. First, earlier attempts were made by the group to include the 3'UTR in the *β2t* reporter construct but these were unsuccessful due to plasmid instability. Moreover, a previous study of the putative *β2t* promoter in *Ae. aegypti* achieved reporter expression in the testes and spermatheca with an SV40 terminator but not with the endogenous 3'UTR (Smith *et al.*, 2007). For *proA*, a reporter construct that included the 3'UTR was made and transgenic isolines established, however no DsRed expression was visible in the testes of two isolines. In contrast, consistent, albeit low, DsRed expression was observed in testes across four *proA* isolines whose reporter construct included the SV40 terminator.

Nevertheless, the absence of reporter gene expression in specific tissues and cell types such as the seminal vesicles, spermatheca, and sperm heads, despite using the SV40 terminator suggests that the desired spatiotemporal expression pattern was not achieved prior to the incorporation of the promoter/SV40 regulatory elements into the Cas9 constructs. So, while exogenous terminators such as the SV40, or rather the p10, may be necessary for achieving sufficient levels of transgene expression—and would be recommended amongst other translation enhancing modifications (e.g. the synthetic 5'UTR Syn21 and specific translation initiation sequences) for constructs used to target haplo-insufficient genes—they may be inadequate for achieving post-transcriptional control (Pfeiffer, Truman and Rubin, 2012; Leftwich *et al.*, 2023). This may be especially important for accumulation of Cas9 in late-stage spermatids and inducing a paternal effect in embryos. On the other hand, lack of reporter protein in the desired spermatid tissues could be indicative of post-translational protein instability due to the lack of relevant post-translational signals rather than the timing of translation. While DsRed is thought to be a highly stable protein, with an estimated half-life of 4.6 days *in vivo* in *Drosophila*, it may be that active degradation of proteins during specific stages of spermiogenesis could clear all but certain specifically marked proteins (Verkhusha *et al.*, 2003).

Improving Cas9 retention in spermatids was attempted via a ubiquitin fusion designed to attach a stabilizing residue at Cas9's N-terminal end, but the impact of this was inconclusive as a non-fused construct was not tested. However, minimal paternal deposition could indicate that Cas9 activity was not stabilized or not to sufficient levels, though other factors like Cas9 abundance could have influenced these results, too. It was,

however, hypothesized that the ubiquitin destabilized DsRed expression, as no or only very dim DsRed protein was observed in both the *proA* and *β2t* lines, which contrasted with evidence of modest CRISPR-Cas9 activity (30-50% mutation rate) in the same lines and brighter reporter gene expression in other reporter-only lines. Therefore, correlating Cas9 expression with that of DsRed, as was initially an aim of this study to better understand the significance of DsRed absence in sperm cells, was unfortunately not possible here. A quantitative PCR or visualisation of RNA levels through spermatogenesis via FISH could be more informative and performed in future.

Low levels of paternally deposited complexed Cas9/sgRNA were possibly demonstrated with *proA*-Cas9, despite the challenges discussed above. Modifications to the construct should therefore be considered to potentially enhance the paternal effect, though the following suggestions would be relevant to other genes that are post-transcriptionally regulated in maturing spermatids. First, sequences in the open reading frame (ORF) and intron of two SNBP genes, *ProtB* and *Mst77F*, have been shown to enhance expression in reporter constructs in *Drosophila* (Barckmann *et al.*, 2013). Nuclear localization signals were also putatively identified in the C-terminal coding sequence of *Mst77F.* The importance of introns in enhancing transcript abundance has been recognized in a broad range of organisms and reaffirmed in *Ae. aegypti* with evidence that the internal intron of the polyubiquitin gene boosted expression of a transgene when included in the proximal promoter (Anderson *et al.*, 2023b; Shaul, 2017). While the putative *proA* gene does not have introns, the ORF could still be important in delimiting transcript survival. Therefore, a modified *proA*-Cas9 construct could be expressed as a fusion to the entire *proA* gene with a ribosomal skipping T2A peptide to help localize the transcript spatiotemporally without interfering with protein function (i.e. promoter-*proA* 5'UTR-*proA* ORF-T2A-Cas9-NLS-*proA* 3'UTR) (Matthews, Younger and Vosshall, 2019; Li *et al.*, 2017b; Diao and White, 2012; Daniels *et al.*, 2014). Perhaps the 5'UTR and ORF would be sufficient to localize the transcript as it was in Barckmann *et al.,* (2013), in which case a translational enhancer like the p10 sequence (almost eight times more efficient than the SV40 sequence) could be used in place of the endogenous 3'UTR to increase polyadenylation and protein production (Pfeiffer, Truman and Rubin, 2012). In the case of *β2t*, which is expressed throughout spermatogenesis and where post-translational modifications possibly limit protein turnover, direct fusions could be considered.

Further, assuming that *proA* transcript survival is strictly regulated, then perhaps earlier transcription as long as it enhances transcript abundance could be possible. While *proA* and *β2t* are among the most abundant transcripts in the testes, the selected promoter sequences could lack the necessary regulatory sequences to reproduce the high levels of transcription. It has been observed that regulatory regions are challenging to identify in *Ae. aegypti* due to its relatively large genome (Matthews and Vosshall, 2020; Nolan and Hammond, 2022; Anderson *et al.*, 2023a). Therefore, perhaps another promoter sequence, like *shu*, which is known to drive high Cas9 cleavage rates with a shorter (and thus possibly more compact) promoter fragment and endogenous 3' UTR, and from which paternal deposition has even been observed, could be used (though chromatin state at the time of Cas9 expression can also lead to high cleavage rates). Alternatively, a Q-binary expression system could be used to increase transcript abundance together with the *shu* or *β2t* promoter (i.e. *shu/β2t-QF2* | *QUAS-proA 5'UTR+ORF*-T2A-Cas9-NLS-*p10*), though there is a risk that with *β2t* it would express too late and transcription of the responder construct would not be possible before transcriptional silencing occurs (Matthews, Younger and Vosshall, 2019; Afify *et al.*, 2019). Together, these modifications could ensure high levels of transcripts are delivered to early spermatids, while the endogenous UTRs and ORF (and introns if another gene with a similar expression profile was identified) would improve transcript survival and translation in spermatids, increasing the chances of protein deposition into the embryo. With these designs, expression of U6-driven sgRNAs may also need to be optimized to ensure colocalization with Cas9 due to uncertainties in the level of U6 expression in spermatids as well as the stability of sgRNAs after expression in spermatocytes (see 4.4.3.i.d).

A simpler approach to enhancing transgene expression within a target tissue could involve identifying a highly expressed endogenous gene using available transcriptomics data. While many transcriptomic studies exist that can be mined for genes that express exclusively in testes, the majority are bulk RNA-sequencing (RNA-seq) studies that could obscure important differences between related cell types, especially in tissues with ongoing cellular differentiation, such as the testes (Sun, Shao and Iqbal, 2023). Over the past decade, transcriptomic analysis has transitioned from RNA-seq to single-cell RNA-seq (scRNA-seq), enabling a much finer resolution of gene expression patterns at the level of individual cells and cell types. Notably, the recently published Fly Cell Atlas is a powerful minable resource that presents an annotated transcriptome of more than 250

distinct cell types in adult *D. melanogaster* and includes mapped trajectories of differentiating sex-specific cells (Li *et al.*, 2022). Additionally, two scRNA-seq studies on the testes of the malaria mosquito species *An. gambiae* have also been made publicly available (Taxiarchi *et al.*, 2019; Page *et al.*, 2023). These studies along with existing testes specific RNA-seq studies will be important in characterizing novel promoters for effector transgene expression with enhanced cell-type specificity (Akbari *et al.*, 2013a; Cassone *et al.*, 2017; Izquierdo *et al.*, 2019; Raz *et al.*, 2023).

Future mining of single-cell transcriptomic datasets could pursue gene transcripts that are found in most/all cell types of spermatogenesis, from diploid spermatocytes (late primary could be sufficient) to late elongation stage spermatids. These transcripts could then be compared to those found in young, recently fertilized embryos and oocytes to identify any RNAs with high abundance in sperm, low abundance in the oocyte, and decreasing levels in the fertilized zygote as zygotic genome activation is initiated (Pérez-Mojica *et al.*, 2023; Kwasnieski, Orr-Weaver and Bartel, 2019; Castillo, Jodar and Oliva, 2018). Coupled with proteomics studies on similar tissues, transcripts that were delivered by the sperm and translated in the fertilized embryo could then be identified (Garlovsky, Sandler and Karr, 2022; McDonough-Goldstein, Pitnick and Dorus, 2021; McCullough *et al.*, 2022). Other studies have made use of differences in single nucleotide polymorphisms found in a set of highly inbred lines of *D. melanogaster* to identify paternally and maternally derived transcripts in embryos (Cridland and Begun, 2023).

While transcriptomic studies greatly facilitate the identification of endogenous genes with desired expression profiles, characterizing the regulatory sequences that recapitulate these patterns—let alone at sufficient levels—can be challenging. Generally, regulatory sequences for transgene expression comprise a few kilobases upstream and downstream of the start and stop codons, respectively (Anderson *et al.*, 2023a; Nolan and Hammond, 2022). This template for regulatory sequence design has been useful for *Anopheles* mosquitoes but may be less appropriate for *Ae. aegypti*, whose genome is much larger and characterized by long intronic sequences (often tens of kilobases) that can make 5'UTRs and proximal promoters difficult to identify, as well as an enrichment of transposable elements whose regulatory function is unknown (Nolan and Hammond, 2022; Melo and Wallau, 2020; Sharakhov and Sharakhova, 2015; Matthews *et al.*, 2018). Performing a single-cell assay for transposase-accessible chromatin sequencing (scATAC-

seq) on *Ae. aegypti* testes could reveal regions of open chromatin that signify gene regulatory activity and thus be highly informative for promoter design. Separately, it has been suggested that promoters from closely related species with more compact genomes could be used to drive transgene expression with a desired profile in *Ae. aegypti* instead (Nolan and Hammond, 2022).

Finally, although challenging, future split-drive studies in *Ae. aegypti* should identify neutral sites for transgene integration that could limit position effects and allow better comparison between different regulatory elements or synthetic constructs (Matthews and Vosshall, 2020). While phiC31 site-specific cassette exchange was tested here at an M-linked site, the method did not prove to be as efficient as reported, perhaps due to the docking site used (Carabajal Paladino *et al.*, 2023; Anderson *et al.*, 2023b). For the development of a SuperSIT system, it will be necessary to test regulatory elements from within the M locus or at an M-linked site, as the uniquely heterochromatic context may distort gene expression significantly; this will be the focus of Chapter 4. Finally, if paternal deposition can be demonstrated, it greatly expands the mechanisms that SuperSIT (or YLE) could be designed to operate through, for example, by using molecules (RNA or protein) that disrupt the sexual differentiation pathway in females but do not harm males or via deposition of toxins that are rescued by an antidote expressed zygotically, though this is more like the Medea design (O'Leary and Adelman, 2020; Adelman and Tu, 2016; Burt and Deredec, 2018; Ward *et al.*, 2011).

# Chapter 3: CRISPR-Cas9 knock-out of female specific flight genes, *Act-4* and *myo-fem*

## 3.1 Abstract

Flight has been proposed as a target for genetic control as it is essential for mosquito survival and reproduction in the field. In *Ae. aegypti*, the actin and myosin genes that make up the indirect flight muscles have been found to be sex-specific, making them candidates for a CRISPR-Cas9-based SuperSIT system. This chapter investigated the knock-out effect of the female-specific versions of these genes, namely *myo-fem* and *Act-4,* focusing on estimating their dominance. The *myo-fem* gene was mutagenized with CRISPR-Cas9 at two N-terminal exons following recent research that indicated the gene could be haplo-sufficient at exon three. The *Act-4* gene was targeted at the C-terminal end because it was hypothesized to produce higher rates of dominant antimorphic mutations akin to those observed in the *D. melanogaster* homolog, *Act88F.* To assess the effect of gene knock-out on flight more precisely, a flight test device was devised and constructed using inexpensive materials. The results indicated that high rates of dominant mutations could be produced at the selected sites. Dominance was found to be close to complete (0.9) for *myo-fem* targeted at exon two and incomplete in *Act-4,* varying around 0.5 at four sgRNA sites*.* The dominance rate for *myo-fem* exon 2 mutants reflected a null phenotype in females, so would correspond to a substantial reduction in fitness. These results hold promise for developing a nuclease based SuperSIT system in *Ae. aegypti*.

## 3.2 Introduction

Insects are the most species rich and abundant group of animals on earth, representing over 60% of all known taxa (Zhang, 2013a). With a million insect species described and tentative estimates ranging between 3 and 13 million, they have evolved to occupy a diverse set of niches (Stork, 2018). Of all 29 extant insect orders, five—Diptera (mosquitoes, midges, and flies), Hymenoptera (ants, bees, and wasps), Coleoptera (beetles), Lepidoptera (moths and butterflies), and sometimes Hemiptera (true bugs)—have been classified as hyper-diverse because they make up more than 80% of all insect species (Zhang, 2013b; Deora, Gundiah and Sane, 2017). This evolutionary success is unique among animal taxa and has been attributed to two major factors: the evolution of small body size and flight (Deora, Gundiah and Sane, 2017; Grimaldi and Engel, 2005).

### 3.2.1 Evolution of asynchronous indirect flight muscles

A small body size, however, imposes physical constraints on flight ability (Deora, Gundiah and Sane, 2017). Smaller insects have a larger surface area to body mass ratio, which increases drag relative to wing inertia and reduces overall aerodynamic forces for flight (Chin and Lentink, 2016; Deora, Gundiah and Sane, 2017). This means miniature insects must generate much greater forces by flapping their wings harder and at higher frequencies and do so despite relatively small muscle mass. In response to this high energetic cost of sustaining flight, most winged insects (except e.g. locusts and dragonflies) have evolved asynchronous indirect flight muscles (IFMs) that increase the efficiency of muscle activity (Dickinson and Tu, 1997; Dickinson, 2006; Conley and Lindstedt, 2002). This evolutionary adaptation has incidentally attracted mechanical engineers in the field of biomimicry, who seek inspiration for how to build miniaturized flying robots (Abas *et al.*, 2016).

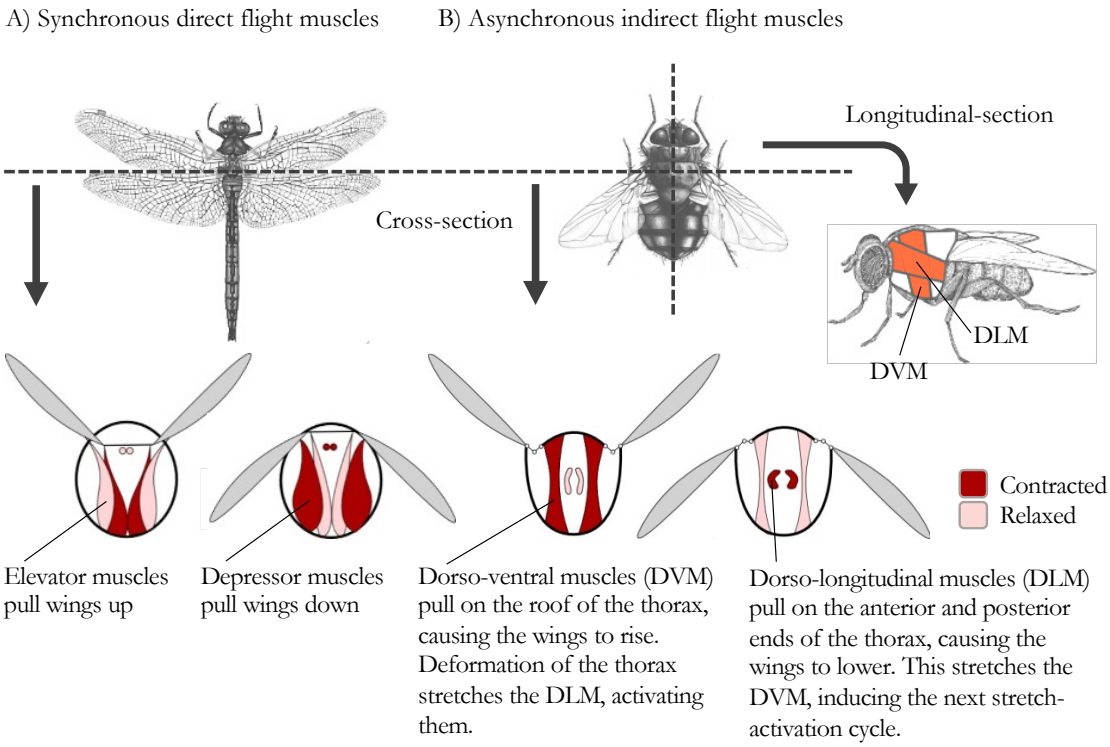



*Figure 10 Direct and indirect insect flight muscles in the dragonfly and fly, respectively. Left) Synchronous direct flight muscles drive wing movement by directly pulling on wings. When elevator muscles contract, wings are pulled up, and when depressor muscles contract, wings are pulled down. Right) Asynchronous indirect flight muscles drive wing movement by deforming the thorax. This alternately activates the dorso-ventral (DVM) and dorso-longitudinal (DLM) muscles in a cycle of stretch-activation and produces the high oscillation frequencies necessary for flight. Illustration of the emperor dragonfly and blue bottle fly by Lizzie Harper https://lizzieharper.co.uk/. Cross-section diagrams freely available from https://www.amentsoc.org/. General figure structure adapted from (Hill, Wyse and Anderson, 2012).*

The IFMs are found in the thorax and attach to the thoracic exoskeleton rather than directly to the wing bases (Figure 10) (Conley and Lindstedt, 2002). Two sets of IFMs, the dorso-longitudinal (DLM) and dorso-ventral muscles (DVM), are arranged in cross orientation and work together to drive deformation of the thorax, which is what causes the wings to move. When taking off, a low frequency nerve impulse generates multiple cycles of high frequency contractions via a process called "stretch activation" (this contrasts with synchronous direct muscles that require a single nerve impulse for each contraction) (Deora, Gundiah and Sane, 2017). The process occurs due to the elastic properties of the thorax: when one set of IFMs contracts, the thorax is squeezed one way and stretched the other, which activates the other set of IFMs, repeating the cycle of contraction and stretching (Iwamoto, 2011; Hedenström, 2014). This oscillation drives resonance of the entire thorax, producing high wingbeat frequencies that often exceed 100 Hz, though typical wing beat frequencies are around 200 Hz in *Drosophila* and around

750 Hz and 500 Hz in males and females of *Ae. aegypti*, respectively (Deora, Gundiah and Sane, 2017; Kim *et al.*, 2021; Bartussek *et al.*, 2013; Andrés *et al.*, 2020).

### 3.2.2 Harmonic convergence in mosquitos

A precisely coordinated flight mechanism is essential for courtship and mating success in many dipterans, especially mosquitos (Cator *et al.*, 2009; Ewing, 1977). Mosquitos from several genera mate in dense aerial swarms that are heavily dominated by males (Downes, 1969; Cator and Zanti, 2016; Sawadogo *et al.*, 2014; Charlwood and Jones, 1980). In *Aedes*, these swarms typically occur above vertebrate hosts, so when an individual female approaches in search of a blood meal, males can detect her flight tone and pursue her mid-flight (Cator *et al.*, 2011). These acoustic cues initiate a series of courtship behaviours, termed "harmonic convergence", where males and females dynamically adjust their flight tones to harmonize with each other; this has been documented in several mosquito species (Cator *et al.*, 2009; Cator and Harrington, 2011; Cator and Zanti, 2016; Aldersley and Cator, 2019; Gibson and Russell, 2006; Gibson, Warren and Russell, 2010). The acoustic duet, mediated by wing beat frequency, has been shown to be a predictor of female acceptance or rejection of the male partner and the mating success of the converging pairs' male offspring (Cator and Harrington, 2011; Aldersley and Cator, 2019). As harmonic convergence strongly correlates with mating outcomes—and can be an indicator of male fitness and immune performance of offspring—it has been suggested to be under sexual selection and linked to assortative mating and reproductive isolation (League *et al.*, 2021; Reitmayer *et al.*, 2021; Garcia Castillo *et al.*, 2021; Gibson, Warren and Russell, 2010).

Because flight is essential for mosquito communication, mating success, and survival, it has become a focus of various vector control objectives. Flight has been used as an indicator of insect quality in sterile male programs, proposed as a phenotypic target for suppressive population control both in terms of reducing flight ability and the acoustic perception of flight, and incorporated into the development of automated tools for trapping and surveillance by taking advantage of the high species specificity of flight tones (Carpenter, Blomefield and Vreysen, 2012; Dor *et al.*, 2020; Culbert *et al.*, 2020; Fu *et al.*, 2010; O'Leary and Adelman, 2020; Navarro-Payá *et al.*, 2020; Li *et al.*, 2021a; Mukundarajan *et al.*, 2017; Rohde *et al.*, 2019; Andrés *et al.*, 2020; Kim *et al.*, 2021).

Therefore, identifying the genes that play a role in the precisely coordinated function of flight has been of great interest to the field.

### *3.2.3* IFM-specific actin and myosin genes

The indirect flight muscles introduced above are similar in structure to the striated muscles of vertebrates (Lemke and Schnorrer, 2017; Celestino-Montes *et al.*, 2021; Iwamoto, 2011). The muscle fibres (myofibrils) are made up of basic contractile units, called sarcomeres, arranged in hexagonal arrays and composed of thick and thin filaments formed by polymerization of myosin, actin, and associated proteins (e.g. troponin and tropomyosin) (Figure 11A) (Beall, Sepanski and Fyrberg, 1989; Perkins and Tanentzapf, 2014; Shadrin, Khodabukus and Bursac, 2016; Celestino-Montes *et al.*, 2021). Each sarcomere is bordered by Z-discs that anchor the thin actin filaments. These filaments extend toward the middle of the sarcomere (the M-line), where they interact and crosslink with thick myosin filaments that drive contraction. Longitudinal and transverse sections of the myofibrils of adult *Ae. aegypti* clearly show Z-discs and M-lines in each sarcomere and hexagonal arrangement of the myosin and actin filaments (Figure 11B) (Celestino-Montes *et al.*, 2021).

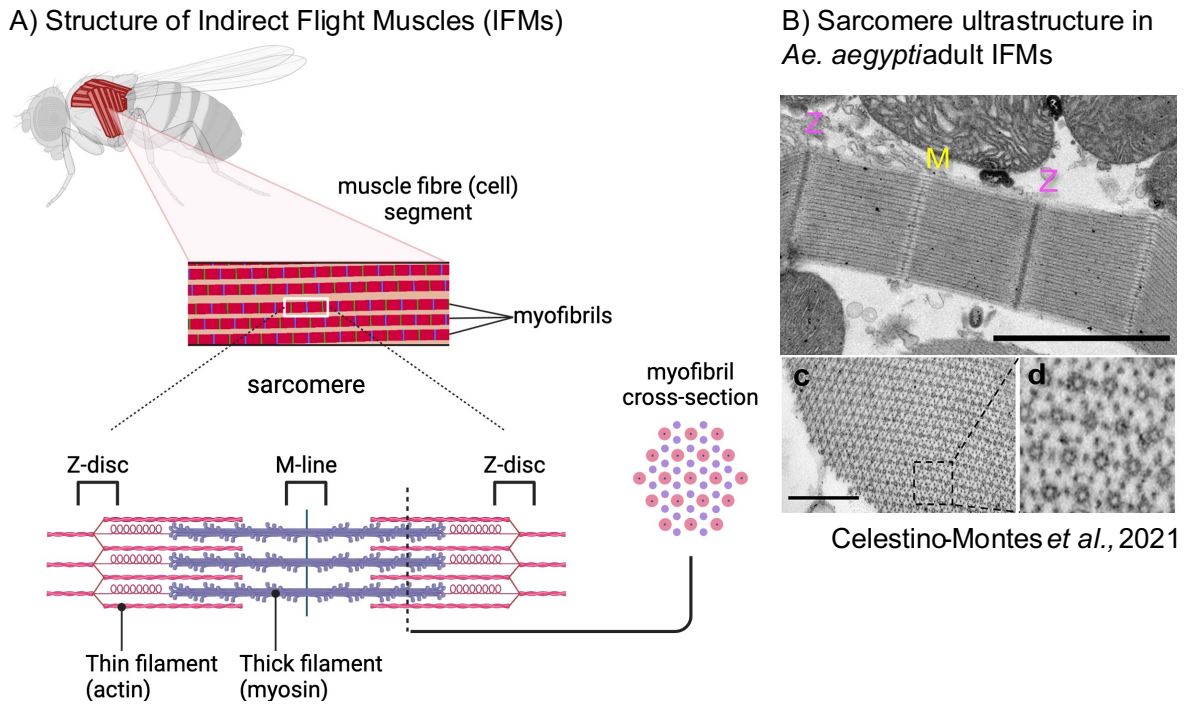



*Figure 11 A) Basic structure of a sarcomere unit within the indirect flight muscles. Created with BioRender.com. B) Sarcomere ultrastructure in* Ae. aegypti *adult IFMs analysed by transmission electron microscopy (TEM) from Celestino-Montes et al 2021. Top image shows longitudinal sections of IFMs with clearly defined Z-discs and M-lines in the sarcomere. Images c) and d) show a cross-section of a mature myofibril with hexagonal arrangement of the thick and thin filaments.*

### 3.2.3.i IFMs in Drosophila melanogaster

Studies in the *Drosophila melanogaster* model first showed that the actin, myosin, and some associated proteins that make up the filaments of myofibrils in the IFMs were specialized (Mahaffey *et al.*, 1985; Wells, Edwards and Bernstein, 1996). Numerous highly conserved actin isoforms with different functions in cytoskeletal structure and motility exist in most organisms (Perrin and Ervasti, 2010; Emerson Jr and Bernstein, 1987). *D. melanogaster* has four muscle-specific actins that are distinguished by a combination of 22 out of 377 amino acids. One of these, *Act88F*, is the only actin isoform to express in the IFMs and is itself expressed almost exclusively there, except some additional low activity observed in leg, wing, and uterine muscles (Fyrberg *et al.*, 1983; Fyrberg *et al.*, 1998; Courchesne-Smith and Tobin, 1989; Nongthomba *et al.*, 2001). Depletion of *Act88F* is not compensated by up-regulation of any other actin isoforms, though the jump muscle actin isoform, *Act79B,* is rescued by upregulation of *Act88F* alone (Okamoto *et al.*, 1986; Dohn and Cripps, 2018). Lastly, replacement of multiple residues (15/375) in *Act88F* with that of the cytoskeletal *Act42A* actin isoform permitted synthesis and accumulation of thin actin filaments but did not allow normal assembly of these into IFM myofibrils or support flight (Fyrberg *et al.*, 1998). These results suggest that flight is a highly coordinated process potentially requiring multiple specialized muscle proteins compared to other activities like walking or jumping and that other actin isoforms cannot replicate this precise function.

Remarkably, all null mutants of *Act88F* exhibit a dominant flightless phenotype, meaning that this actin isoform is a rare example of a haplo-insufficient gene in *Drosophila* (Okamoto *et al.*, 1986; An and Mogami, 1996; Marygold *et al.*, 2007; Haigh *et al.*, 2010; Cripps *et al.*, 1994). Null mutants show a reduction of thin actin filaments in the sarcomeres and a range of structural aberrations in the IFM myofibrils but are otherwise viable (Mahaffey *et al.*, 1985; Nongthomba *et al.*, 2001; Okamoto *et al.*, 1986). Some of these dominant mutants are *antimorphic*, meaning that the altered gene product acts antagonistically and causes a mutant phenotype that cannot be rescued by two copies of the wild-type allele (in contrast to *hypomorphic* or *amorphic* dominant mutations that can be rescued); notably three doses of *Act88F* had no impact on flight (Hiromi *et al.*, 1986; Okamoto *et al.*, 1986; Drummond, Hennessey and Sparrow, 1991; An and Mogami, 1996; Karlik, Saville and Fyrberg, 1987). The antimorphic mutations are primarily found in the

C-terminal region of the gene, suggesting they only impair the local structure of actin and interfere with normal polymerization or cause improper interactions with other molecules, like myosin and the associated protein, tropomyosin (Table 4 in Methods) (Okamoto *et al.*, 1986; An and Mogami, 1996; Haigh *et al.*, 2010; Noguchi *et al.*, 2010).

Myosins are complex multimeric proteins that associate with actin and play fundamental roles in contractile processes required for eukaryotic motility (Berg, Powell and Cheney, 2001; Emerson Jr and Bernstein, 1987). The myosin proteins have a conserved hexamer structure made up of three types of subunits: the myosin alkali light chain (MLC-ALK), the myosin regulatory light chain (MLC-2), and the myosin heavy chain (MHC) subunits (Emerson Jr and Bernstein, 1987). The MHC functions as the motor protein of the thick filaments and is responsible for mediating contraction of sarcomeres in myofibrils. Like actin, the MHC exists in multiple isoforms that differ in their muscle-specific function and spatiotemporal expression pattern (Emerson Jr and Bernstein, 1987; Beall, Sepanski and Fyrberg, 1989; O'Donnell *et al.*, 1989). In *D. melanogaster*, these isoforms are produced by alternative splicing of five of 19 exons in a single *MHC* gene; (Bernstein *et al.*, 1983; Rozek and Davidson, 1983; Zhang and Bernstein, 2001; Miller *et al.*, 2007; Hastings and Emerson Jr, 1991).

Specific null mutants of the *MHC* gene were first found to induce dominant flightless phenotypes in flies that were otherwise viable (O'Donnell *et al.*, 1989; Emerson Jr and Bernstein, 1987; Bernstein, O'Donnell and Cripps, 1993; Chun and Falkenthal, 1988). These were later found to disrupt an IFM-specific exon or the IFM-specific splicing mechanism and ultimately revealed the splicing pattern of the IFM-specific isoform (Hastings and Emerson Jr, 1991). Later, expression of an embryonic *MHC* isoform (that differs from the IFM isoform in all five alternatively spliced exons) in a wild-type and null genetic background produced viable flies with dominant antimorphic flightless phenotypes (Wells, Edwards and Bernstein, 1996). The authors observed that while IFM myofibril assembly was normal in transgenic flies, their stability was heavily affected, beginning to degenerate a few days after eclosion. The authors hypothesized that the slow synchronous embryonic myofibrils could not produce enough power to generate the high asynchronous oscillation frequencies needed for flying and that highly specialized interactions between the IFM-specific MHC and actin isoforms were required (Wells, Edwards and Bernstein, 1996; Singh *et al.*, 2014).

Indeed, sequence comparisons and functional studies *in vitro* and *in vivo* demonstrated that specific structural domains determine the muscle kinetics and power-generating abilities of the MHC isoforms (Bernstein and Milligan, 1997; Swank *et al.*, 2000; Swank *et al.*, 2001; Swank *et al.*, 2002; Swank *et al.*, 2004; Zhao and Swank, 2017). For example, the ATP-dependent motor domain and the converter domain are encoded by the alternatively spliced exons 3 and 11, respectively, and both have been found to independently and dramatically influence actin sliding velocity and power output (Swank *et al.*, 2004; Swank *et al.*, 2002). Interestingly, four of the five alternatively spliced exons are found in the N-terminal regions of the MHC polypeptide that form the myosin head. The myosin head interacts with the thin actin filaments and has recently been proposed to act as the stretch sensor for stretch-induced activation of the asynchronous indirect flight muscles (Swank *et al.*, 2000; Swank *et al.*, 2001; Iwamoto, 2017).

*3.2.3.i.a Evolution of haplo-insufficiency in IFMs*

Haplo-insufficiency, which arises when a single copy of a gene is not sufficient for normal development, is rare but present in all eukaryotes (Morrill, 2020). Haplo-insufficient genes predominantly encode structural components or members of protein complexes and show high rates of conservation across eukaryotic species and retention of the haplo-insufficient state (Morrill, 2020; de Clare, Pir and Oliver, 2011; Papp, Pál and Hurst, 2003). One hypothesis for the persistence of haplo-insufficiency over evolutionary time, despite a high fitness penalty, suggests that it ensures that sufficient amounts of a protein are produced to reach a threshold amount needed to maintain wild-type function (Sæbøe-Larssen *et al.*, 1998; Wilkie, 1994). A second—the "stoichiometric" or "balance" hypothesis—suggests that an imbalance in the concentrations of components of a multi-protein complex can result in deleterious effects (Marygold *et al.*, 2007; Papp, Pál and Hurst, 2003).

The "balance" hypothesis is thought to mostly explain the haplo-insufficiency of the IFM-specific actin and myosin isoforms in *Drosophila* (Beall, Sepanski and Fyrberg, 1989; Reedy, Bullard and Vigoreaux, 2000). Beall et al. demonstrated that individuals heterozygous for null mutants of either actin or myosin had myofibrillar defects and were unable to fly but individuals heterozygous for both nulls had more normal myofibrils and a restored flight ability, indicating that stoichiometric balance between actin and myosin is crucial (Beall, Sepanski and Fyrberg, 1989). These dynamics, or a combination of both

hypotheses, may be maintaining the haplo-insufficiency of other genes that play a role in IFM myofibril structure and which are required for flight. These include the IFM-specific isoforms of tropomyosin, troponin, and MLC-2, the *flightin* gene (though dominance is incomplete), and an unidentified gene with antimorphic alleles found at the *Laker* locus whose function is also unknown (Mogami and Hotta, 1981; Cripps *et al.*, 1994; Vigoreaux *et al.*, 1998; Warmke, Kreuz and Falkenthal, 1989; Reedy, Bullard and Vigoreaux, 2000; Nongthomba *et al.*, 2007; Casas-Tintó and Ferrús, 2021). Finally, both MLC-2 (by shortening the N-terminal side) and *flightin* have been implicated in courtship song production in flies (which in contrast to mosquitos does not occur aerially), suggesting they could partly be under sexual selection (Chakravorty *et al.*, 2017; Chakravorty *et al.*, 2014; Chakravorty, Wajda and Vigoreaux, 2012). In summary, strong evolutionary pressures maintain strict stoichiometric quantities of the major (i.e. most abundant) and some minor structural components of the myofibril.

### 3.2.3.ii IFMs in Ae. aegypti

IFM-specificity of actin and myosin isoforms is conserved in *Ae. aegypti* paralogs of *Act88F* and the MHC gene from *Drosophila*. Remarkably, these IFM-specific isoforms have also been found to be sex specific (Muñoz *et al.*, 2004; Akbari *et al.*, 2013a; Aryan *et al.*, 2020; O'Leary and Adelman, 2020; Li *et al.*, 2021a; Vyazunova and Lan, 2004). Of sixteen actin paralogues in *Ae. aegypti*, two show localization to the IFMs and sex-biased expression in pupae: the male-specific *Actin-3* (*Act-3*, AAEL009451) and female-specific *Actin-4* (*Act-4*, AAEL001951) (Muñoz *et al.*, 2004; Akbari *et al.*, 2013a; Vyazunova and Lan, 2004). Mutagenesis of *Act-4* at the N-terminal end of the protein indicates that loss-of-function mutations are recessive, though some short in-frame deletions have an antimorphic (dominant negative) effect on flight, and does not affect male flight or fitness (Navarro-Payá *et al.*, 2020; O'Leary and Adelman, 2020). Sex-specificity of IFM actin appears to be conserved across *Culicine* (*Aedes* and *Culex* mosquitos) and *Anopheline* mosquitos (Labbé *et al.*, 2012; Jiang *et al.*, 2015; O'Leary and Adelman, 2020; Paiz-Reyes *et al.*, 2020).

Similarly, two of sixteen MHC paralogues show sex-biased expression in *Ae. aegypti*: *myo-sex* (AAEL021838), expressed in male pupae and found in the male-determining M locus (so absent in females), and *myo-fem* (AAEL005656) expressed in female pupae (Akbari *et al.*, 2013a; Aryan *et al.*, 2020; O'Leary and Adelman, 2020). CRISPR-Cas9 mutagenesis of *myo-fem* at exons 2 and 3 renders females flightless, while the flight ability of males

remained unaffected (O'Leary and Adelman, 2020; Li *et al.*, 2021a). A null mutation in exon 3 (hypothesized to be the motor domain of *myo-fem*) was shown to be somewhat dominant—53% of heterozygous mutant females were deemed flightless—indicating the gene could be haplo-insufficient (O'Leary and Adelman, 2020). A putative orthologue of *myo-fem* (LOC109402113) and two non-M linked orthologues of *myo-sex* (LOC109430926 and LOC109412105) have been identified in *Ae. albopictus* based on strong sex-biased expression, but other homologs have not yet been found in other mosquito species (Lutrat *et al.*, 2022).

The presence of sex-specific paralogues of the actin and myosin genes suggests concurrent development of regulatory networks to ensure correct expression (and repression) of the sex-specific forms. The *doublesex (dsx)* gene, for example, functions as a molecular switch in the sex determination pathway of insects, triggering male or female differentiation with alternatively spliced isoforms (Burtis and Baker, 1989; Price, Egizi and Fonseca, 2015; Scali *et al.*, 2005). Studies in *Drosophila* have shown that *dsx* and another sex-determining transcription factor, *fruitless* (*fru*), are expressed in neuronal circuits that control male courtship songs (Dickson, 2008; Su and Kamikouchi, 2023). In *Anopheles gambiae,* homozygous mutants of the female-specific splicing isoform (*dsxF*) produce a sterile intersex phenotype (Kyrou *et al.*, 2018). Interestingly, the wing-beat frequency of intersex mosquitos is approximately halfway between the sexually dimorphic flight tones of males and females (Su *et al.*, 2020).

### 3.2.4 Aims

The SuperSIT system and other population suppression approaches to vector control can involve generating loss-of-function mutations in genes that reduce the reproductive potential of females, leading to population decline (Fu *et al.*, 2010; Hammond *et al.*, 2016; O'Leary and Adelman, 2020; Burt and Deredec, 2018; Alphey, 2014). The actin and myosin genes that modulate flight activity in *Ae. aegypti* females have been proposed for these systems, both as genetic targets and to drive female-specific lethal effects using their promoters (Fu *et al.*, 2010; Labbé *et al.*, 2012; Navarro-Payá *et al.*, 2020; O'Leary and Adelman, 2020; Li *et al.*, 2021a). Importantly, these genes would induce "late-acting" lethality, which has been modelled to increase the efficacy of vector control considerably when compared to lethality that is induced at earlier stages of mosquito development (Dye, 1984; Phuc *et al.*, 2007; Burt and Deredec, 2018; Evans *et al.*, 2022; Bouyer, 2023).

This is because the density-dependence that controls larval survival (for example, through availability of larval nutrition), and which efficiently regulates population size, would not be disturbed (Phuc *et al.*, 2007).

The design of the SuperSIT population suppression system requires modified males to impose a fitness cost on female descendants, but to do so *in trans* as females do not inherit the construct. Fitness penalties could be imposed using a programmable nuclease to mutate female-specific genes required for survival or fertility. This could be done by manipulating both paternal chromosomes in the modified male's germline cells, causing a dominant effect in female offspring that inherit a mutated copy. Alternatively, both the maternal and paternal chromosomes could be mutated in a fertilized embryo by way of an engineered paternal effect, which would allow a recessive gene to be targeted (Burt and Deredec, 2018). In the previous chapter, progress towards identifying novel germline promoters that could precisely regulate a CRISPR-Cas9 construct for a desired mutagenesis effect was described. Because paternal deposition of complexed Cas9 and sgRNA was not observed for the promoters tested (and only limited paternal deposition of Cas9), but high germline cleavage was, a haplo-insufficient gene could provide the preferred route to developing a SuperSIT system.

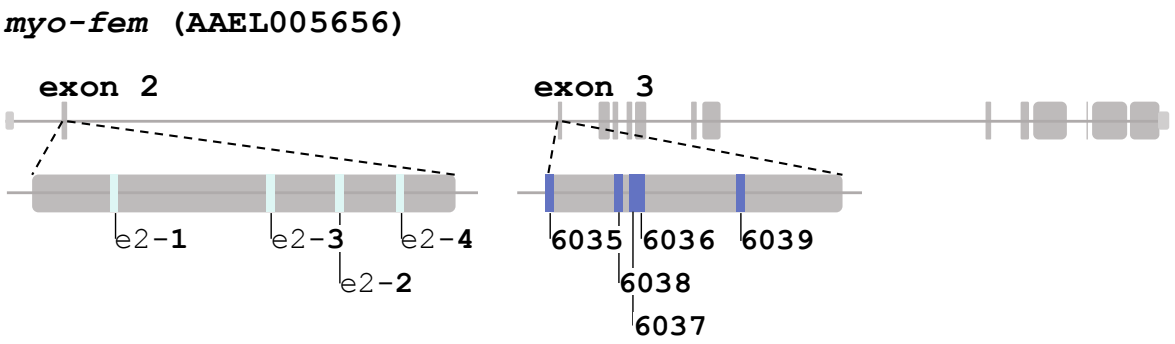
In this chapter, *Act-4* and *myo-fem* were studied further to assess the dominance of each knock-out effect and thereby evaluate their suitability as targets for a nuclease based SuperSIT system operating in the pre-meiotic germline. CRISPR-Cas9 was used to mutagenize *Act-4* at its C-terminal end, which was hypothesized to be more likely to produce truncated proteins that interfere with its ability to polymerize, creating a dominant antimorphic mutant phenotype akin to those observed in *Drosophila*. The *myo-fem* gene was instead targeted at the N-terminal end of the protein, in exons two and three, to study the extent of the haplo-insufficient phenotype that was previously reported. Finally, a simple and inexpensive flight test device was also devised and constructed using available literature to assess the impact of knock-out mutations on female flight more precisely.

## 3.3 Methods

### 3.3.1 sgRNA design

#### *3.3.1.i myo-fem*

sgRNAs were designed to target *myo-fem* (AAEL005656) in exons two and three for gene knock-out using the gene sequence obtained from VectorBase and the CHOPCHOP web tool (Labun *et al.*, 2019) (Giraldo-Calderón *et al.*, 2022). The first exon is non-coding, so was excluded from the analysis. Four of the top-scoring sgRNAs in exon two were previously published by Li et. al and sgRNA 6036 in exon three was published by O'Leary and Adelman (Li *et al.*, 2021a; O'Leary and Adelman, 2020). Four more sgRNAs targeting exon three were selected for a total of nine sgRNAs targeting *myo-fem* (Figure 12).




**exon 2**

ATGCCTAAGCCAGTGGTACAGGTTGGTGGCGACCCCGATCCCTCCGAATGGCTCTTCATTTCGGAGGAG
ATGAAACGAATTGATCAGAGCAAGCCTTACGATGCCAAAAAGGCCTGCTGGGTTCCAGATGAAGCCGAA
GGCTATCTTCTGGGTGAAATCAAAGCAACCAAGGGTGAATTGGTAACCGTTGCCCTACCCGGTGGTGAG

**exon 3**

*ccgcag*ACCAAGGATTTCAAGAAGGATCTGGTAGGCCAGGTGAATCCCCCCAAATACGAAAAATGCGAG
GATATGTCGAATTTAACCTACCTCAACGATGCCTCTGTGTTACATAACTTGAGAGAACGATACAGGGCT
AGGCTGATCTAT

*Figure 12 Nine sgRNAs targeting* myo-fem *(AAEL005656) in* Ae. aegypti *were designed within exon 2 (regions in light blue) and 3 (regions in purple). Each sgRNA is highlighted in the coding sequence of the exons; PAM sites are underlined, intron sequence in exon 3 is indicated by lower case and italicized bases.*

#### 3.3.1.ii Actin-4

To induce dominant mutations in *Actin-4* (*Act-4*, AAEL001951), sgRNAs were designed to target sites known to cause antimorphic mutations in the *D. melanogaster* homolog, *Act88F*. Antimorphic mutations for *Act88F* were obtained from FlyBase (version FB2021_04) and published material (Table 4) (Haigh *et al.*, 2010; An and Mogami, 1996;

Hiromi *et al.*, 1986; Okamoto *et al.*, 1986; Drummond, Hennessey and Sparrow, 1991; Karlik, Saville and Fyrberg, 1987; Sevdali *et al.*, 2013; Cripps *et al.*, 1994; Gramates *et al.*, 2022). Antimorphic mutants generated by a premature stop codon (i.e. W353@ or W356@ in Table 4) were of highest interest here. The proximity of both mutations to the end of the protein and to each other could suggest that termination in a short, random chain of amino acids would have a similar antimorphic effect. As CRISPR can only generate random end-joining mutations, these sites would be preferable to other sites caused by specific point mutations.

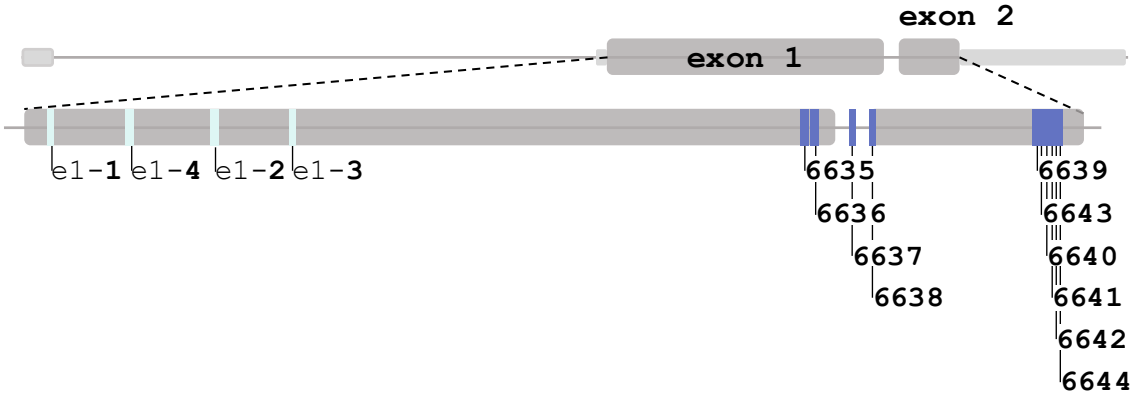
*Act-4* was aligned to the male-specific actin paralog, *Actin-3* (*Act-3,* AAEL009451), the putative jump muscle actin paralog, *Actin-1* (*Act-1*, AAEL001928), and putative orthologs of *Act-4* from *Ae. albopictus* (AALF020552), *Culex quinquefasciatus* (CPIJ012572), *Anopheles gambiae* (AGAP011515), and *D. melanogaster* (*Act88F*) using the Clustal Omega sequence alignment program (version 1.2.4) and sequences obtained from VectorBase and FlyBase (Giraldo-Calderón *et al.*, 2022; Madeira *et al.*, 2019; Matthews *et al.*, 2016). sgRNAs targeting *Act-4* were designed with the CHOPCHOP and CRISPR GuideXpress web tools using the *Ae. aegypti* AaegL5 genome assembly. Ten sgRNAs predicted to cleave at sites in the C-terminal region of *Act88F* were then selected for CRISPR-Cas9 mutagenesis (Figure 13).

*Table 4 Antimorphic mutations observed in the IFM-specific actin isoform,* Act88F, *in* D. melanogaster *sourced from available literature and FlyBase. @ indicates a stop codon.*

| Mutant | Morphism | Reference |
|---|---|---|
| G63D | Anti | An and Mogami 1996 |
| R95C | Anti | An and Mogami 1996 |
| A138V | Anti | Haigh et al 2010; Cripps et al 1994 |
| G156S | Anti | An and Mogami 1996 |
| G156D | Anti | An and Mogami 1996 |
| E226K | Anti | An and Mogami 1996 |
| R256C | Anti | Haigh et al 2010 |
| G268D | Anti | An and Mogami 1996; Haigh et al 2010 |
| D292V | Anti | Sevdali et al 2013 |
| G301D | Anti | An and Mogami 1996 |
| G302D | Anti | An and Mogami 1996 |
| E317K | Anti | Drummond et al 1991 |
| V339I | Anti | Drummond et al 1991 |
| Q353@ | Anti | An and Mogami 1996 |
| W356@ | Anti | An and Mogami 1996; Okamoto et al 1986 |
| W356X | Anti | Fyrberg et al 1988; An and Mogami 1996 |
| G366D | Anti | Drummond et al 1991; Okamoto et al 1986 |
| E57K | Hypo | An and Mogami 1996 |
| V103E | Hypo | An and Mogami 1996 |
| Q121@ | Hypo | An and Mogami 1996 |
| P307L | Hypo | An and Mogami 1996 |
| G366S | Hypo/Anti | An and Mogami 1996; Fyrberg et al 1988 |

| | | |
|---|---|---|
| R372H | Late acting dominant | Haigh et al 2010 |
| A296S | Recessive | Viswanathan et al 2017 |
| E334Q | Recessive | Nongthomba et al 2003 |

***Actin-4* (AAEL001951)**




**exon 1**

ATGTGTGACGATGATGCTGGAGCACTAGTCATTGACAACGGATCCGGCATGTGTAAGGCCGGTTTCGCT
GGTGATGATGCCCCA**CGT**GCCGTCTTCCCGTCCATTGTCGGCCGCCCTCGCCACCAGGGTGTGATGGTC
GGTATGGGTCAAAAAGATGCCTACGTCGGTGATGAAGCCCAATCCAAGCGA**GGT**ATCCTCACCCTGAAA
TATCCCATAGAGCACGGTATC**ATC**ACCAACTGGGATGATATGGAGAAGATTTGGCATCACACCTTCTAC
AACGAGTTG**CGA**GTAGCTCCTGAAGAACATCCAGTATTGCTGACTGAGGCTCCCTTGAATCCAAAGTCC
AATCGCGAGAAGATGACTCAGATCAATCACCAACTGGGATTGTTTGAAACATTCGCTTCGCCAGCTGTG
TATGTTGCCATCCAA**GCT**GTTCTGTCCCTGTACGCCTCCGGTCGTACTACTGGTATTGTTCTGGATTCC
**GGA**GATGGTGTCTCCCATACCGTCCCAATCTACGAAGGTTATGCTCTGCCACATGCCATCCTCCGTATG
GATTTGGCTGGTCGTGATCTGACCGACTACCTGATGAAGATCTTGACCGAACGTGGATACTCTTTCACC
ACCACCGCTGAACGTGAAATCGTTCGTGACATCAAGGAGAAGCTGTGCTACGTCGCTCTGGACTTCGAG
CAG**GAA**ATGCAAGCCGCTGCCGCTACGTCTTCATCCGAGAAGTCTTATGAACTTCCCGAT**GGC**CAAGTC
ATCACAATCGGCAACGAGCGTTTC**CGT**GCTCCAGAAGCCCTTTTCCAGCCATCCTTCCTG**GGA**ATGGAA
TCAACTGGCATTCATGAAACGGTCTACAACTCGATCATGCGTTGCGATGTCGACATCCGCAAG**GAT**CTC
TATGCTAACAGCGTCTTGTCT**GGTGGT**ACCACCATGTACCCAG

**intron 2/exon 2**

*gtatttcttatatttaaccactcaactctgcatcatactcaaaacctccctctattacaG*GTATTGCTG
ATCGTATGCAGAAG**GAA**ATCACTTCCCTGGCTCCATCCACCATCAAGATCAAGATCATTGCCCCACCGG
AACGTAAATACTCC**GTC**TGGATCGGTGGATCCATCCTGGCCTCGCTGTCTACCTTC**CAA**GCTATG**TGG**A
TCTCCAAGCAGGAATACGACGAAGGT**GGC**CCAGGAATTGTCCACCGCAAGTGCTTCTAA

*Figure 13 Ten sgRNAs targeting* Act-4 *(AAEL001951) in Ae. aegypti were designed within the C-terminal region of the coding sequence (regions in purple). Triplet codons associated with antimorphic mutations in* Drosophila *are highlighted in orange and in bold. sgRNAs targeting exon 1 (in blue) were previously targeted by Navarro-Payá et al. and induced recessive mutations in* Act-4 *(Navarro-Payá et al., 2020). Each sgRNA is highlighted in the coding sequence of the exons; PAM sites are underlined, intron sequence immediately before exon 2 is indicated by lower case and italicized bases.*

### *3.3.2* Testing cleavage efficiency of sgRNAs *in vivo*

Predicted sgRNA efficiencies can vary considerably from measured ones, so testing sgRNA efficiencies *in vivo* is recommended before incorporating them into plasmids used for transgenesis (Kistler, Vosshall and Matthews, 2015; Konstantakos *et al.*, 2022; Kotwica-

Rolinska *et al.*, 2019; Labuhn *et al.*, 2018). sgRNAs targeting *myo-fem* exon two were not tested *in vivo* prior to making transgenics expressing these guides because their activity had already been verified by Li et. al (Li *et al.*, 2021a).

### 3.3.2.i sgRNA synthesis by in vitro transcription

sgRNAs targeting *myo-fem* exon three and *Act-4* were synthesized by *in vitro* transcription of a PCR template generated from overlapping primers containing the sgRNA sequence and a T7 polymerase binding site (Bassett *et al.*, 2013; Purusothaman *et al.*, 2021) (Table 5). sgRNAs were *in vitro* transcribed using the MEGAscript T7 Transcription kit (ThermoFisher Scientific, USA) and purified using the MEGAclear Transcription Clean-Up kit (ThermoFisher Scientific).

*Table 5: Primer sequences used in each PCR with common reverse primer (LA137) to generate a template for* in vitro *transcription of sgRNAs. sgRNA sequences (excluding PAM) are in purple.*

| Primer no. | Oligo sequence |
|---|---|
| 6035 | GAAATTAATACGACTCACTATAGGTTCTTGAAATCCTTGGTCTGGTTTTAGAGCTAGAAA |
| 6036 | GAAATTAATACGACTCACTATAGGCCCCAAATACGAAAAATGCGGTTTTAGAGCTAGAAA |
| 6037 | GAAATTAATACGACTCACTATAGGTCGCATTTTTCGTATTTGGGGTTTTAGAGCTAGAAA |
| 6038 | GAAATTAATACGACTCACTATAGGCGTATTTGGGGGGATTCACCGTTTTAGAGCTAGAAA |
| 6039 | GAAATTAATACGACTCACTATAGGTCTCAAGTTATGTAACACAGGTTTTAGAGCTAGAAA |
| 6635 | GAAATTAATACGACTCACTATAGGTATGCTAACAGCGTCTTGTCGTTTTAGAGCTAGAAA |
| 6636 | GAAATTAATACGACTCACTATAGGGCTAACAGCGTCTTGTCTGGGTTTTAGAGCTAGAAA |
| 6637 | GAAATTAATACGACTCACTATAGGgagtatgatgcagagttgagGTTTTAGAGCTAGAAA |
| 6638 | GAAATTAATACGACTCACTATAGGGCAATACCtgtaatagagggGTTTTAGAGCTAGAAA |
| 6639 | GAAATTAATACGACTCACTATAGGAGCTTGGAAGGTAGACAGCGGTTTTAGAGCTAGAAA |
| 6640 | GAAATTAATACGACTCACTATAGGTGTCTACCTTCCAAGCTATGGTTTTAGAGCTAGAAA |
| 6641 | GAAATTAATACGACTCACTATAGGACCTTCGTCGTATTCCTGCTGTTTTAGAGCTAGAAA |
| 6642 | GAAATTAATACGACTCACTATAGGTCCAAGCAGGAATACGACGAGTTTTAGAGCTAGAAA |
| 6643 | GAAATTAATACGACTCACTATAGGGGAGATCCACATAGCTTGGAGTTTTAGAGCTAGAAA |
| 6644 | GAAATTAATACGACTCACTATAGGGAATACGACGAAGGTGGCCCGTTTTAGAGCTAGAAA |
| 137 | AAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAAC |

### 3.3.2.ii Synthetic sgRNAs

Six of the ten sgRNAs designed to target *Act-4* were later commercially sourced to improve *in vivo* cleavage efficiencies (Standard Synthetic SygRNA®, Merck, UK) (Table 6). SygRNAs were ordered "modified" with a stabilizing 2' 0 methyl & phosphothioate to increase stability toward exonucleases and increase genome editing efficiency (Hendel et al., 2015). Lyophilized SygRNAs were resuspended to a concentration of 30 µM (1500 ng/µl) in 10mM Tris containing buffer.

*Table 6 Sequence of SygRNAs targeting* Act-4 *sourced commercially from Merck, UK. Capital bases indicate sequence is found in exons, lower case and italicized bases indicate intronic sequence. In bold are sgRNAs incorporated into a piggyBac plasmid for mosquito transgenesis.*

| sgRNA | Sequence (excluding PAM) |
|---|---|
| 6636 | GCTAACAGCGTCTTGTCTGG |
| 6638 | GCAATACCtgtaatagaggg |
| 6640 | TGTCTACCTTCCAAGCTATG |
| 6643 | GGAGATCCACATAGCTTGGA |
| 6642 | TCCAAGCAGGAATACGACGA |
| 6644 | GAATACGACGAAGGTGGCCC |

### 3.3.2.iii Microinjection to generate knock-out mutants

To generate knock-out mutants, 100 ng/µl of each sgRNA, 300 ng/µl of purified NLS-SpCas9 protein (recombinant Cas9 protein from Streptococcus pyogenes expressed in Escherichia coli with an NLS sourced from PNA Bio, USA for in vitro transcribed sgRNAs or Merck, UK for synthetic sgRNAs) resuspended in water and 1X injection buffer (final concentration 5mM KCl and 0.1mM NaH2PO4 at pH 6.8) was combined for each injection mix. Three groups of sgRNAs targeting *myo-fem* in exon 3 were injected in groups of three that all included sgRNAs 6035 and 6039 and differed by inclusion of one of the following three sgRNAs: 6036, 6037, and 6038 (Table 7). sgRNAs targeting *Act-4* were injected individually with the help of Ray Wilson.

*Table 7 sgRNAs targeting* myo-fem in *exon 3 that were included in each of the three groups injected for knock-out mutant screening. Blue text denotes two sgRNAs (6035 and 6039) that were used in all three microinjection groups, unique sgRNA sequences are in bold. sgRNAs in groups 1 and 3 were incorporated into piggyBac plasmids for mosquito transgenesis.*

| gRNA group | gRNAs | sgRNA sequence (5'-3'), excluding PAM |
|---|---|---|
| Group 1 | 6035 | TTCTTGAAATCCTTTGTCTG |
| | 6039 | TCTCAAGTTATGTAACACAG |
| | 6036 | GGCTTCATCTGGAACCCAGC |
| Group 2 | 6035 | TTCTTGAAATCCTTTGTCTG |
| | 6039 | TCTCAAGTTATGTAACACAG |
| | 6037 | CGTATTTGGGGGGATTCACC |
| Group 3 | 6035 | TTCTTGAAATCCTTTGTCTG |
| | 6039 | TCTCAAGTTATGTAACACAG |
| | 6038 | CCCCAAATACGAAAAATGCG |

For *in vitro* transcribed sgRNAs, the assembled injection mix was incubated at 37°C for 20 mins. For synthetic SygRNAs (for re-targeting of *Act-4* only), sgRNAs and Cas9 were first combined and incubated for 5-15 minutes at room temperature to complex the RNPs before injection buffer was added. Injection mixes were stored at -80°C until ready to inject or immediately centrifuged at 14,000×*g* at 4 °C for at least 10 min prior to injection.

*Ae. aegypti* embryos were microinjected as before (see 2.3.3) and hatched after five days. Male and female injection survivors ($G_0$) were outcrossed in groups of 5-15 to wild type of the opposite sex (LVP strain). $G_1$ females were then assessed for their ability to fly by a simple visual assessment or by use of a purpose-built flight test device (described below) at 3-7 days post eclosion. For the simple flight test, females were visually assessed by tapping the cage and gently agitating the adult females with a paintbrush. Individuals that could not fly in response to the trigger were deemed flightless. The proportion of flightless individuals was determined by dividing the number of flightless individuals by the total number of females in the cage. Females were labelled by whether they could fly or not and stored at -20°C for downstream sequencing.

### 3.3.3 piggyBac transgenesis of *Ae. aegypti* to express actin and myosin sgRNAs from genome

#### 3.3.3.i sgRNA-expressing piggyBac donor plasmids

Four sgRNA expression plasmids were designed for piggyBac-mediated transgenesis of mosquito embryos; AGG-2199 included four sgRNAs targeting *myo-fem* in exon 2 (previously published by Li et. al), AGG-2198 and AGG-2227 included three sgRNAs each targeting *myo-fem* in exon 3 and differing by a single sgRNA (sgRNA 6036 vs. 6038), and AGG-2695 included three sgRNAs targeting *Act-4* in the C-terminal end of exon 2 (Figure 14) (Li *et al.*, 2021a; O'Leary and Adelman, 2020). Each sgRNA within a cassette was designed to be under the control of a different RNA Polymerase III promoter and sgRNA scaffold to minimize repetitive regions in the construct (Table 8) (Anderson *et al.*, 2020; Noble *et al.*, 2019). sgRNA cassettes were synthesized by Genewiz (UK) and cloned into a piggyBac transformation plasmid, AGG-1910, that expressed the fluorescent transgenic marker cassette 3xP3-AmCyan-SV40. Synthesized plasmids were midi-prepped by Tim Harvey-Samuel using NucleoBond Xtra Midi kit for endotoxin-free (EF) plasmid DNA (Macherey-Nagel, UK) and eluted in EF water.

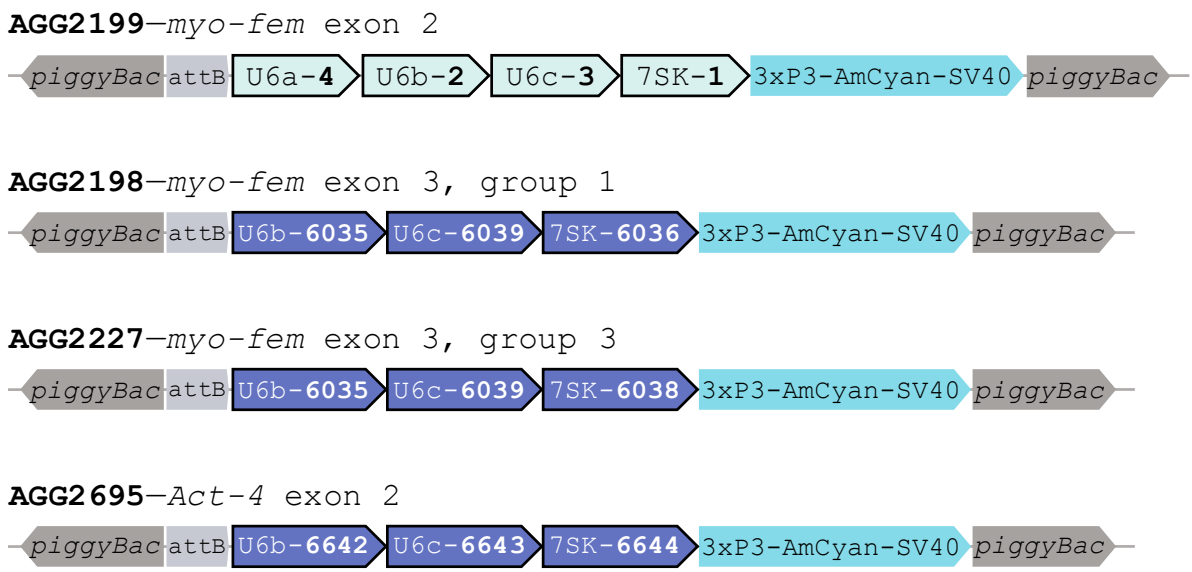



*Figure 14 Schematic of four sgRNA expressing plasmids targeting* myo-fem *and* Act-4 *for piggyBac-mediated transgenesis of mosquito embryos. All sgRNA cassettes were cloned into the same backbone plasmid (AGG-1910) that contained the genetic marker cassette 3xP3-AmCyan-SV40 and piggyBac flanks to facilitate genome integration. Expression of each sgRNA in a cassette was driven by a different Pol III promoter (U6a-c and 7SK). sgRNA labels and colours correspond to Figure 12 and Figure 13.*

*Table 8 DNA sequence of each Polymerase III sgRNA expression cassette used in Figure 14 is shown: polymerase III promoter sequences are in blue, placeholders for each sgRNA sequence used are* ***black****, different sgRNA scaffold sequences in green, and the terminator sequence in black.*

| Cassette | Sequence (5'-3') |
|---|---|
| U6a-sgRNA-scaffold WT-Pol IIII terminator | GGCCGGCCGTTTCCAGACTTTCCCTCCCGGTAAACGGAGACAAAACGACAGACGTAAGTAGGTACATATGCATA CCGCACGGACAAATCAAATTTGTCTGGCAGCTCCAATTAGAGTCGTTAAAAATTTAACGATGCGCTAAATAACTT CAAGCTATTTGTCTCGCTGGATTGGCTTCGAGTGGTAAGATCCTATCAAATGCCGAAAACAAAAAACTTTCTTCTT AATTGTTCGTTCTTCAACACCTCTCCATGGTGATAACGGATACGGTTTCATTGTCAGCATCCATCCTCCGAAAAAT ACATTACGCCTTGAAATATGCAATCGCAAACACGGATCTGTTTGGAACATTTATTTTACTATGAAGAGATGCGAT AGGTAATATTTATTTGAGCGTTTAAGATACTCATTGTTCTCTCAAAGAATGTCATTGAAAGCCAACGAGGTCAAA TCAAATATTATAATAAAAAGGTCAAAGAGGACTAACTTAAAGCTCTCTTTATGGATAGGAAAAAATATTTTCGCC CATCGCTAGAACTTTTACCGTTTCCATTGAGTATATAACTAAGATGAATGAGGCTAATTGAT-**sgRNA**-*GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGT GC*-TTTTTTT |
| U6b-sgRNA-scaffold 23-Pol IIII terminator | GGAACCCGAATTTAGTGCTATATAATTTAATTCCACTAGAGTTTGTATCCTTTGATAGATACGCGTATTTCGACCT CAACTGCAAGGCCGTCGTGTACTAGACTTGACTAATCCAGACTGGTCTTTTAGTTATGACTTCTGTCCACATCTCC ATACATTCAACGCACTGTGCGGCTGTGCTGTGCGACTCCGTCGAGTCGACCAACATAGTTGAAACAAATTGAAT ATTTAATTGATCGTTATAGGAATGGTGTTAGATGAGTCATCCTTTACAGTAAGCACATACAGTATTATAATTGAA GATCGTCGGCAGATAGGTGTGTAGGGTAGAGTATCAGCAATAAGTTGGGACGTTTGACTTTTTGTAGGTAGACA AAAACTAAACTTTTTTTCGCTTCTCTATGTGTGCCCCCCCGGGTAGCGTATCGTTCCGATTGTGGTGCGAACGAAT GAAATCGCCCATCGAGTTGATACGTCCATCCATCGCTAGAACCGCGTTCGCTGTAGAAGACTATATAAGAGCAG AGGCAAGAGTAGTGAAAT-**sgRNA**-GTTCCAGAGTCGGGAACGACAAGTTGGAATAAGGCAAGTCCGTTATCATGCCGGAAGGCAGGCACCGATTCGG TGC-TTTTTTT |
| U6c-sgRNA-scaffold 22-Pol IIII terminator | GAGTCTGGTAACCCTAGTGCACGCAAATATCTCGCGGGCATATTTGGTTGCTGAGGTATATTTATATTTGAACGC CATGAGAAAAAGCGGAAGAAATTGGCTCATGGCCGATTTTAAGGATATTTAAAAATTGTACAATGTACATATAA TAGGCCAGGAGAAGTGGATGAACTTGTCATTCATTTTTCTGTCAATTCTCATACAAAATCTACTTTTTCTTTGACA TAAATTCACTCTAGGTGAACCACTTCCCCTGGCCTATTAAACATCCGTTCCTTCAATGTGTTCTTTTTTTTAAGCGT GTGTTAAAAGTTTGCTCTGCTGGTGAATTCACGCTCTACCCGTTCAGGCAGCATTCATCGAAAAGCCCTATCTGC TCGCACACATTTACAAAATGCTGATTGCGTTGTGTGCTGAATGGGTCACTCGTCCGTCACTGCTTGCTGTGTACA CTGTACAGTTACGCAGTCTGTGCATCGCTAGAATCATATTTACGGAAGAGTATTATATATACCCGATGCGTTGCT CTTCGATT-**sgRNA**- |

| | |
|---|---|
| | GTTCGAGAGGACGAGAGTCCAAGTTCGAATAAGGCCAGTCCGTTATCACAGGGAGACCTGGGCACCGAGTCGGTGC-TTTTTTT |
| 7SK-sgRNA-scaffold 15-Pol IIII terminator | ATGCTCCATTCTCCGCCACTTGTTGATGCGGACCCTAACCACGTGGTCGCTCCTCTGCTCACCGGAGCACGTTTCATACAGCCTGACGACGACGAGCAATCAGAGGTATGGTGAGCATGCGCATGGAGAGTGGACAGCAGTGCACCCTAAAATCAATTCACACATCATGTGTCAATAGCTGTGTCAATGTTGCACAGCCTTTTCTTATTAAATTTACTCCTTTTGTGACCATTTCTCTTTCATCCACCGTTATTTTAATGAGTTTTGTGTTCCGGTGGACGAACGTTCACACAAAAAATGTGTAAATCTTAATCAACCAGAACACAAAGTATAGTGAAAAAATTAAAGTGTGTGGCTTTTATACATCCTAACTGTAAATTATTTTTAGAGTGCGTGCGATCGTTCTCTCGAACCACGCTCTCCGCTACACATTCGCAGCGAATGGCGTGAATGGATGAAAGAACAAACTAAAGTTTATTTTTAGATTCGTCTCAAAACAACTGCTGTGCATCGCTAGAACCAAGAAATACGCCACTCAGTATATATAGCACTTCCAACCCCGCTTTCCTC-**sgRNA**-GTGTTAGAGGGATAGAGATATCCCAAGTTAACATAAGGCTAGTCCGTTATCACTGCAGGAATGCAGGGCACCGAGTCGGTGC-TTTTTTT |

### 3.3.3.ii Microinjection and establishing transgenic isolines

Injection mixes for piggyBac transgenesis consisted of a sgRNA-expressing piggyBac donor plasmid at 500 ng/µl, a piggyBac helper plasmid expressing AePUb-hyperactive-piggyBac transposase (AGG-1245) at 300 ng/µl, and injection buffer as above. *Ae. aegypti* embryos were microinjected as before and hatched after five days. Male and female $G_0$ survivors were reared separately and crossed in pools of 3-5 to wild type of the opposite sex. $G_1$ offspring from each pool were then screened under fluorescence using a Leica MZ10 microscope (Leica Biosystems, Germany) for presence of the transformation marker. Transgenic lines were established from single $G_1$ positives crossed to wild type. $G_2$ offspring were screened for transgene inheritance and sex bias to estimate the number of insertions and whether any were sex-linked. $G_2$s suspected of having multiple insertions were outcrossed again and allowed to lay individually to isolate transgenic isolines that produced a 50% transgene inheritance rate.

## 3.3.4 Experimental crosses to test activity of sgRNA-expressing transgenic lines

Transgenic isolines from each sgRNA set were tested for expression of the integrated construct, which can be considerably impacted by the chromosomal context in which the transgene is inserted (termed position effects) (Anderson *et al.*, 2023a; Schotta *et al.*, 2003). sgRNA-expressing isolines were crossed to a transgenic line expressing Cas9 from the male-determining M-locus and driven by the *shutdown* promoter (*shu*-Cas9; AGG-2235; generated in Chapter 2 by recombination of the m-linked sister line AGG-1551B1). Where sgRNA-expressing isolines were M-linked, they were instead crossed to AGG-1551B1. $F_1$ offspring were then screened, and trans-heterozygous males selected to cross to wild type females. For M-linked sgRNA expressing isolines that had been crossed to AGG-1551B1, single-heterozygotes that did not inherit *shu*-Cas9 were additionally

selected and crossed to wild type to control for maternal deposition of *shu*-Cas9, which has been documented before (Anderson *et al.*, 2023a). Depending on the activity of each sgRNA-expressing cassette, *shu*-Cas9 was expected to induce germline mutations at the sgRNA target sites that would be inherited in the $F_2$ (that also inherit a non-mutated maternal copy) and produce flightless females, if dominant. F2 females were screened for their ability to fly using a flight test device (described below) at 3-4 days post eclosion. Females were labelled by whether they could clear the flight test device tubes or not and were stored at -20°C for downstream sequencing.

### 3.3.5 Design and construction of a purpose-built flight test device

A flight test device was designed with the help of Michelle Anderson to measure the impact of mutating flight muscle genes on flight ability more precisely (Figure 15). The final form of the flight test device is described below. Briefly, the simplest types of flight test devices estimate flight ability by calculating the escape rate of insects from a cylinder (Rowley and Graham, 1968; Carpenter, Blomefield and Vreysen, 2012; Bond *et al.*, 2019; Culbert *et al.*, 2018; Dor *et al.*, 2020; Balestrino *et al.*, 2017; Maïga *et al.*, 2022; O'Leary and Adelman, 2020). The sensitivity of this method is influenced by tube diameter, height, and duration of the test, as well as the interactions between each of the parameters (Carpenter, Blomefield and Vreysen, 2012). Versions of this method are regularly used because they are inexpensive, portable, and low tech (Carpenter, Blomefield and Vreysen, 2012; Benedict *et al.*, 2018; Maïga *et al.*, 2022; Dor *et al.*, 2020). As such, a similar device was designed here and built with the help of the electrical and mechanical services division at the Pirbright Institute.

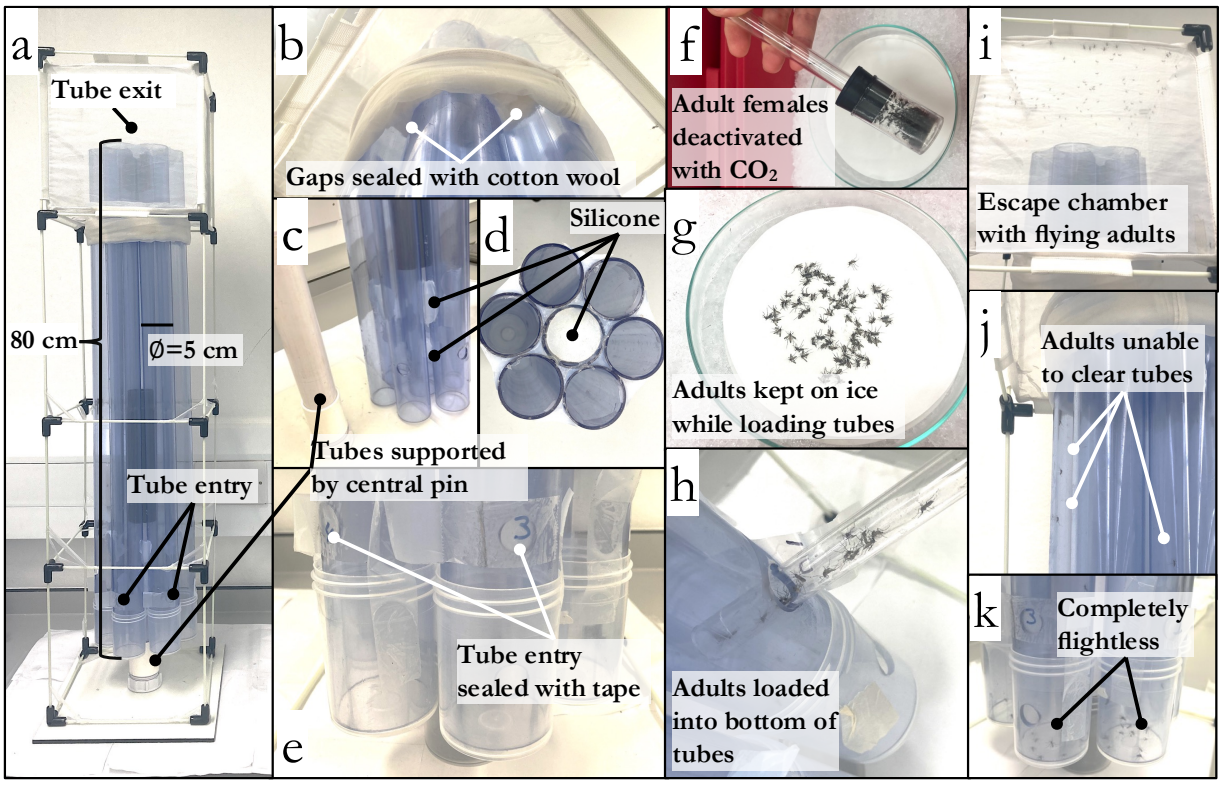



*Figure 15 Final design of the flight test device used to assess flight of CRISPR-Cas9 mutated* Ae. aegypti *females. The device is made up of six clear PVC tubes 80 cm tall and 5 cm wide. Other structural components (a-e) are highlighted and described further in the text. f) and g) show adult females deactivated with $CO_2$ and chilling on ice prior to selection of 20 individuals at random to load into the bottom of each tube (h). After 1.5 hours, females were deemed able to fly if they escaped the tubes into the top cage (i) and flightless if they did not clear the tubes (j and k). A conservative estimate of flight ability was made by counting the number of females at the bottom of the tubes and dividing that by the total number added at the start (k).*

The flight test device was made up of six vertically positioned, clear PVC tubes (https://www.plasticpipeshop.co.uk/) that were 5 cm wide and 80 cm long (Figure 15a). Silicone was used to attach adjacent tubes to each other and each tube to an inner structural tube that was used to suspend the assembled tube structure over a central pin that allowed it to rotate (Figure 15a, c, and d). A 1 cm wide hole bored into the bottom 7 cm of each tube functioned as the entry point for loading adult female mosquitoes into the flight test device using an insect aspirator (Figure 15e, f, and h). A medium cage (24.5 $cm^3$, BugDorm, USA) suspended at the top of the device with its opening facing down provided an escape chamber for mosquitoes that had flown up through each tube. The cage was supported by a lattice structure of joints and rods. To prevent mosquitoes from escaping into the room, the cage opening was secured with a rubber band, that fastened the nylon sleeve over the tube structure, and cotton wool to seal any gaps (Figure 15a and

b). The device was designed to allow around 80% of wild type females to escape to improve the discriminatory power of the device.

Adult females between 3-7 days old that had been allowed to emerge in a separate cage were aspirated and then anaesthetized with $CO_2$ for 5-10 s before chilling on a glass petri dish kept on ice (Figure 15f and g). A dry filter paper was placed between the mosquitoes and the petri dish to prevent condensation dampening the wings. A random selection of 20-30 females was then aspirated and loaded into the bottom of each tube (Figure 15f). The total number of females found in the bottom 5 cm and top 75 cm was recorded every 15 minutes for 1.5 hours. Recordings started the moment all females had woken up, which was normally within a few minutes of loading the tubes. Prior to taking a recording, tubes were agitated by lightly thumping the tube structure on the table a few times to encourage flight and minimize re-counting of mosquitoes that had already escaped and flown back into the tubes to rest.

The proportion of flightless individuals per tube was determined by dividing the number of individuals remaining in the tube at the end of the experiment by the total number of females added into it at the start. A mean flightless rate was calculated using the total number of tubes that were used for the experiment. A conservative estimate was also calculated by removing all females remaining in each tube at the end of the experiment and assessing flight ability in a cage. This was done by removing the top cage and quickly placing a piece of parafilm on top of the opening of the tubes (this is also how the flight test was ended). $CO_2$ was then used to deactivate the females in the tubes, which were then transferred to a smaller cage. Flight ability was visually assessed as before by tapping the cage and gently agitating the adult females with the head of an aspirator tube. Individuals that could not fly in response to the trigger were deemed flightless.

### 3.3.6 Sequence verification of flight-tested females to determine sgRNA activity and dominance of mutations

#### 3.3.6.i gDNA extractions using the KingFisher Flex

Genomic DNA (gDNA) from heterozygous flightless and flying females collected from the experiments described above was extracted for downstream sequencing. gDNA was extracted using the MagMAX™ CORE Nucleic Acid Purification Kit (Thermo Fisher Scientific, UK) and KingFisher™ Flex Purification System (Thermo Fisher Scientific),

following the "Digestion" workflow (manual MAN0015944) of the MagMAX kit. 70% of the recommended reaction volumes was used to save on kit reagents. Tissues were automatically homogenized using the TissueLyser II (Qiagen, UK) instrument and 5 mm stainless steel beads operated at 30 Hz for 30s. Samples were homogenized in individual tubes in "PK solution" (Thermo Fisher Scientific, UK) as per the MagMAX protocol and transferred to a 96-well plate using a wide-bore pipette tip. The plate was incubated at 55°C for 2 h and spun at 3,000 x g for 2 min to separate the tissue from the supernatant. The supernatant was then purified using the KingFisher Flex instrument, proceeding from step 3 of the MagMAX CORE Digestion workflow.

### 3.3.6.ii PCR and Sanger sequencing

Preliminary Sanger sequencing of heterozygous flightless females with suspected mutations in *myo-fem* resulted in a high proportion of sequencing traces that were wild type or only displaying the mutated trace. This suggested that the primer pairs used to PCR amplify *myo-fem* were preferentially amplifying one haplotype. Because both exons are relatively short, primer pairs were limited to binding in the surrounding intron sequences, which contained greater variation due to reduced sequence selection there. Thus, to minimize the impact of primer bias, individuals were sequenced out from the respective exons to identify new primer sites with low sequence variation that could be used to re-sequence the experimental samples. As such, several PCR primer pairs and sequencing primers were tested to ensure both alleles were amplified in a PCR; the final primer pairs showing the least bias are highlighted in Table 9. Using these primers, a selection of flightless and flying females was sequenced for each sample group.

*Table 9 Primer pairs used to amplify and sequence each targeted exon in* myo-fem *and* Act-4*. Sequencing primers are underlined.*

| Target | F primer | F no. | Sequence | R primer | R no. | Sequence | Length (bp) |
|---|---|---|---|---|---|---|---|
| myo-fem exon 2 | Myofem_E2_F6 | 8092 | GACACTAGGGTGCTGTTATCTC | Myofem_E2_R8 | 8095 | GTAGAGGATATCGTTAGGGCT | 592 |
| myo-fem exon 3 | Myofem_E3_F12 | 8087 | CGTGATCAGGTATTCCGTAT | Myofem_E3_R9 | 8090 | GTTACACAAGGGTATACACGG | 774 |
| Act-4 | Act4_Cterm_F1 | 7623 | TCCGAGAAGTCTTATGAACTTC | Act4 3'UTR_R01 | 4241 | ATAACGTGATTTGATGTTCGATGCA | 1055 |

### 3.3.6.iii Tracy variant calling

Sanger chromatogram traces of sequenced heterozygous samples required deconvolution to identify mutant alleles. This was performed in a high-throughput way using a free and

open-source command-line application called Tracy (Rausch *et al.*, 2020). Sanger sequencing files were aligned to a wildtype reference input as a chromatogram trace file or fasta sequence. Variants were output in BCF format, which contained the variant position, reference and alternate alleles, quality measures, estimated allelic fractions, and three alignments: align1: first allele against the reference; align2: second allele against the reference; and align3: two alleles aligned to each other (Rausch *et al.*, 2020). Bcftools was then used to identify indel variants for each sample (Li, 2011). Mutations were attributed to specific sgRNAs by the position of the indel relative to the reference sequence.

### 3.3.6.iv Estimating the dominance of indel variants

Flight-tested females were categorised into flightless and flying categories and sequenced to determine whether they carried a mutation at each sgRNA cleavage site. Where only a subset of samples from both categories had been sequenced, a weighted estimate of the total number of mutated females was calculated (Table 10 and Table 11). The dominance of the knock-out effect was then estimated by dividing the weighted number of flightless mutants over the weighted total that were expected to be mutated.

*Table 10 Total number and proportion of total flight tested $G_1$s carrying a putative mutant allele at* Act-4 *that were Sanger sequenced. NF1-6 represent the non-flying groups in tubes 1-6 of the flight test device.*

| sgRNA | Group | Total | Clear variant | Sequenced % |
|---|---|---|---|---|
| 6636 | Flying | 27 | 26 | 96% |
| | NF.1 | 14 | 13 | 93% |
| | NF.2 | 12 | 12 | 100% |
| | NF.3 | 10 | 10 | 100% |
| | NF.Tot | 36 | 35 | 97% |
| 6638 | Flying | 96 | 27 | 28% |
| | NF.1 | 4 | 3 | 75% |
| | NF.2 | 4 | 1 | 25% |
| | NF.3 | 3 | 2 | 67% |
| | NF.4 | 2 | 0 | 0% |
| | NF.5 | 6 | 0 | 0% |
| | NF.6 | 3 | 0 | 0% |
| | NF.Tot | 22 | 6 | 27% |
| 6640 | Flying | 57 | 24 | 42% |
| | NF.1 | 9 | 9 | 100% |
| | NF.2 | 6 | 6 | 100% |
| | NF.3 | 10 | 10 | 100% |
| | NF.4 | 14 | 0 | 0% |
| | NF.5 | 11 | 0 | 0% |
| | NF.6 | 13 | 0 | 0% |
| | NF.Tot | 63 | 25 | 40% |
| 6642 | Flying | 78 | 25 | 32% |
| | NF.1 | 8 | 8 | 100% |
| | NF.2 | 8 | 8 | 100% |
| | NF.3 | 6 | 6 | 100% |

| | | | | |
|---|---|---|---|---|
| | NF.4 | 6 | 0 | 0% |
| | NF.5 | 6 | 0 | 0% |
| | NF.6 | 6 | 0 | 0% |
| | NF.Tot | 40 | 21 | 53% |
| 6643 | Flying | 59 | 24 | 41% |
| | NF.1 | 7 | 7 | 100% |
| | NF.2 | 11 | 11 | 100% |
| | NF.3 | 14 | 14 | 100% |
| | NF.4 | 6 | 0 | 0% |
| | NF.5 | 11 | 0 | 0% |
| | NF.6 | 12 | 0 | 0% |
| | NF.Tot | 61 | 32 | 52% |
| 6644 | Flying | 89 | 23 | 26% |
| | NF.1 | 6 | 6 | 100% |
| | NF.2 | 2 | 0 | 0% |
| | NF.3 | 4 | 4 | 100% |
| | NF.4 | 8 | 0 | 0% |
| | NF.5 | 6 | 5 | 83% |
| | NF.6 | 5 | 0 | 0% |
| | NF.Tot | 31 | 15 | 48% |

*Table 11 Total number and proportion of total flight tested $F_2$s carrying a putative mutant allele at* myo-fem *that were Sanger sequenced. NF1-6 represent the non-flying groups in tubes 1-6 of the flight test device.*

| Line | Flight test | Group | Total | Clear variant | Sequenced % |
|---|---|---|---|---|---|
| 2199C1 (sgRNA+) | Simple | Flying | 21 | 20 | 95% |
| | | Non-flying | 56 | 56 | 100% |
| 2199C1 (wt) | FTD | Flying | 10 | 10 | 100% |
| | | NF.1 | 13 | 11 | 85% |
| | | NF.2 | 12 | 10 | 83% |
| | | NF.3 | 12 | 9 | 75% |
| | | NF.4 | 12 | 0 | 0% |
| | | NF.5 | 12 | 0 | 0% |
| | | NF.6 | 12 | 0 | 0% |
| | | Tot NF | 73 | 30 | 41% |
| 2199C1 (sgRNA+) | FTD | Flying | 15 | 9 | 60% |
| | | NF.1 | 17 | 15 | 88% |
| | | NF.2 | 17 | 14 | 82% |
| | | NF.3 | 20 | 17 | 85% |
| | | NF.4 | 17 | 0 | 0% |
| | | NF.5 | 17 | 0 | 0% |
| | | NF.6 | 20 | 0 | 0% |
| | | Tot NF | 108 | 46 | 43% |
| 2227E1 (wt) | Simple | Flying | 64 | 50 | 78% |
| | | Non-flying | 33 | 32 | 97% |
| 2227E1 (wt) | FTD | Flying | 60 | 30 | 50% |
| | | NF.1 | 7 | 4 | 57% |
| | | NF.2 | 7 | 4 | 57% |
| | | NF.3 | 6 | 6 | 100% |
| | | NF.4 | 11 | 0 | 0% |
| | | NF.5 | 7 | 0 | 0% |
| | | NF.6 | 4 | 0 | 0% |
| | | Tot NF | 42 | 14 | 33% |
| 2227E1 (sgRNA+) | FTD | Flying | 77 | 28 | 36% |
| | | NF.1 | 4 | 4 | 100% |
| | | NF.2 | 8 | 7 | 88% |
| | | NF.3 | 8 | 5 | 63% |
| | | NF.4 | 6 | 0 | 0% |
| | | NF.5 | 7 | 0 | 0% |

| | | | | | |
|---|---|---|---|---|---|
| | | NF.6 | 11 | 0 | 0% |
| | | Tot NF | 44 | 16 | 36% |
| 2198B1 (sgRNA+) | Simple | Flying | 70 | 56 | 80% |
| | | Non-flying | 53 | 27 | 51% |

## 3.4 Results and Discussion

The actin and myosin genes (*Act-4* and *myo-fem*) that modulate flight activity in *Ae. aegypti* females have been proposed as genetic targets for genetic population suppression systems (Navarro-Payá *et al.*, 2020; O'Leary and Adelman, 2020; Li *et al.*, 2021a). In this chapter, CRISPR-Cas9 mutagenesis was used to investigate the dominance of each gene and their overall suitability as targets for a nuclease-based SuperSIT system. The *myo-fem* gene was targeted at the N-terminal end to study the extent of the haplo-insufficient phenotype observed in a previous study (Li *et al.*, 2021a; O'Leary and Adelman, 2020). *Act-4* was mutagenized at the C-terminal end, which was hypothesized to be more likely to interfere with actin's ability to polymerize and thereby induce dominant negative mutations analogous to the ones observed in *Drosophila*. Heterozygous mutants were assessed using a flight test device constructed to measure female flight more precisely.

### 3.4.1 *In vivo* assessment of sgRNAs by microinjection

*In vivo* activity of the designed sgRNAs was initially assessed by microinjection of sgRNAs and Cas9 protein into mosquito embryos and assessing the effect on female flight in subsequent generations (Figure 16). Surviving $G_0$ individuals were outcrossed to wildtype to identify dominant non-flying female mutants in the heterozygous $G_1$ cohort. In this experiment, *myo-fem* was targeted in the third exon only (as the sgRNAs targeting exon 2 had previously been shown to be active in Li et at. 2020) with three groups of sgRNAs that differed by the central guide, while *Act-4* was targeted with ten sgRNAs that were injected individually. The *myo-fem* experiment was performed with *in vitro* transcribed sgRNAs, while the *Act-4* experiment was first tested with *in vitro* transcribed sgRNAs and later repeated with synthetic sgRNAs sourced commercially. A simple flight test was used for the initial experiments targeting *myo-fem* and *Act-4*, while a flight test device was used to assess females from the repeated *Act-4* experiment.

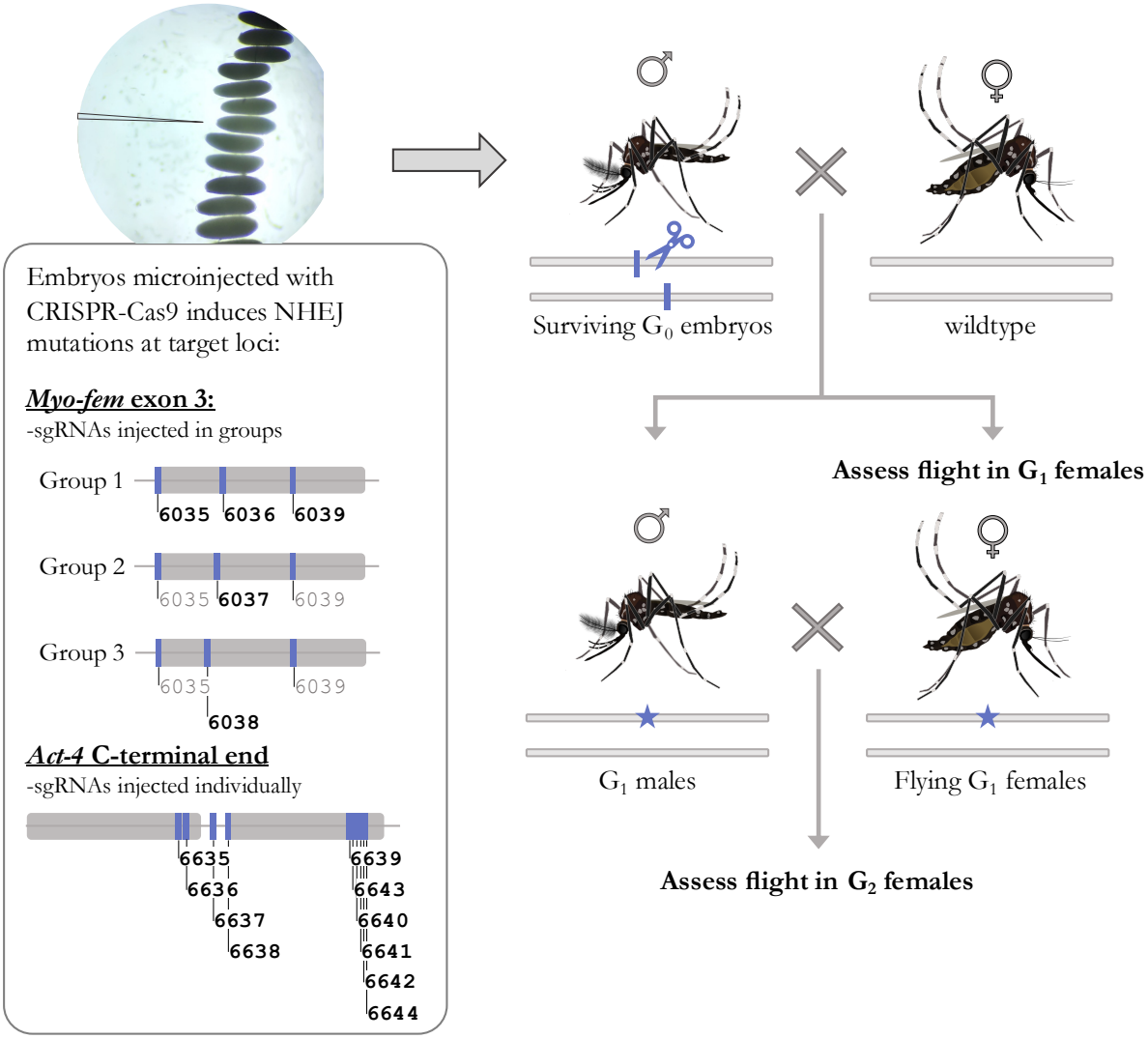



*Figure 16 Activity of sgRNAs targeting* myo-fem *in exon 3 and* Act-4 *at the C-terminal end was tested in separate experiments by microinjection of mosquito embryos. Top left image shows microinjection of* Ae. aegypti *embryos through the lens of a light microscope with the needle containing the CRISPR-Cas9 injection mixture outlined for clarity. Surviving $G_0$ individuals were outcrossed to wildtype and the $G_1$ assessed for female flight to identify any dominant mutants. To determine a dominance rate for the regions targeted, flying $G_1$ females were crossed to $G_1$ male siblings to generate recessive mutations in the $G_2$. $G_2$ females were then assessed for flight in the same way as the $G_1$ group.*

### *3.4.1.i Myo-fem*

Dominant completely flightless mutants were observed at low frequencies in the $G_1$ of embryos targeted with two of the three *myo-fem* guide groups (Table 12). The second guide group did not produce any dominant $G_1$ mutants, but data was sparse due to a low survival rate from those microinjections. To determine whether the low frequency of dominant mutants in the $G_1$ females was due to poor sgRNA efficiency or low dominance of the region targeted without having to sequence the samples, $G_1$s were crossed to each other (with flightless $G_1$ females removed) to generate recessive mutations in the $G_2$

(Table 12). Because mutated males do not display a mutant phenotype, $G_1$ males were still a potential source of dominant mutations in the $G_2$ population. As such, the frequency of the dominant mutations was expected to halve in the $G_2$ and any additional proportion of flightless females in the $G_2$ was expected to be caused by recessive mutations. There was no noteworthy extra proportion of recessive mutations arising in the $G_2$, however, which suggested some degree of dominance for the mutations generated. Nevertheless, a reliable dominance rate could not be estimated from this data due to the low rates of flightless females in all the groups as well as the high variation between the pools of each guide group. As such, the sgRNAs from the two groups that produced dominant mutations were incorporated into sgRNA expression plasmids for piggyBac transgenesis for further testing.

*Table 12* in vivo *assessment of* in vitro *transcribed sgRNAs targeting* myo-fem *in exon 3. sgRNA groups correspond to those in Figure 16. Rate of non-flying females is shown for the first generation ($G_1$) of microinjected embryos ($G_0$) outbred to wildtype of the opposite sex and for the $G_2$ offspring of a $G_1$ sibling cross with flightless females removed. Females were assessed using a simple flight test. For the simple flight test, females were visually assessed by tapping the cage and gently agitating the adult females with a paintbrush. Individuals that could not fly in response to the trigger were deemed flightless. The proportion of flightless individuals was determined by dividing the number of flightless individuals by the total number of females in the cage.*

| sgRNA group | $G_0$ sex | Pools | No. $G_1$ females (mean of pools) | Flightless $G_1$ females (mean of pools ± SE) | No. $G_2$ females (mean of pools) | Flightless $G_2$ females (mean of pools ± SE) |
|---|---|---|---|---|---|---|
| 1 | M | 4 | 392 | 6.31 ± 1.3 | 338 | 2.89 ± 0.9 |
| | F | 4 | 89 | 3.66 ± 1.2 | 135 | 1.47 ± 1 |
| 2 | M | 1 | 113 | 0 | 400 | 0 |
| | F | 1 | 41 | 0 | 400 | 0.5 |
| 3 | M | 4 | 239 | 1.52 ± 0.5 | 400 | 1 ± 0.5 |
| | F | 6 | 72 | 0.44 ± 0.3 | 352 | 1.17 ± 0.4 |

## *3.4.1.ii Act-4*

### *3.4.1.ii.a Results from initial experiment using in vitro transcribed sgRNAs*

sgRNAs targeting *Act-4* were first assessed using *in vitro* transcribed sgRNAs, as was done for *myo-fem*. One of the ten sgRNAs assessed, sgRNA 6640, produced a notable rate of dominant flightless mutants in the $G_1$, up to 10% from female $G_0$s (Table 13). Individuals from this sgRNA group were sequenced to determine the rate of dominance of the knock-out effect (i.e. the proportion of mutated samples producing a non-flying phenotype)); these results are discussed below. High variability in the $G_2$ data, however, prevented making an estimate of dominance at this site: when the $G_2$ groups of the 6640-injected individuals were assessed for flight, a ~460% increase in the proportion of flightless

individuals was observed in those originating from the male $G_0$ pool, while a 70% decrease in flightless individuals was observed in those originating from female $G_0$s. The experiment was therefore repeated with commercially sourced synthetic sgRNAs that were expected to be more efficient and reduce variability between experiment; synthetic sgRNAs carried engineered stabilizing modifications and required minimal RNA handling compared to *in vitro* transcription (Hendel *et al.*, 2015).

*Table 13* in vivo *assessment of* in vitro *transcribed sgRNAs targeting* Act-4 *at the C-terminal end. sgRNA numbers correspond to those in Figure 16. Rate of non-flying females is shown for the first generation ($G_1$) of microinjected embryos ($G_0$) outbred to wildtype of the opposite sex and the $G_2$ offspring of a $G_1$ sibling cross with flightless females removed. Females were assessed with a simple flight test. For the simple flight test, females were visually assessed by tapping the cage and gently agitating the adult females with a paintbrush. Individuals that could not fly in response to the trigger were deemed flightless. The proportion of flightless individuals was determined by dividing the number of flightless individuals by the total number of females in the cage. *indicates individuals were sequenced.*

| sgRNA | $G_0$ sex | Pools | No. $G_1$ females (mean of pools) | Flightless $G_1$ females (%, mean of pools ± SE) | No. $G_2$ females (mean of pools) | Flightless $G_2$ females (%, mean of pools ± SE) |
|---|---|---|---|---|---|---|
| 6635 | M | 1 | 100 | 0 | | |
| | F | 1 | 127 | 0 | | |
| 6636 | M | 2 | 122 | 0.52 ± 0.5 | | |
| | F | 3 | 85 | 0 | | |
| 6637 | M | 1 | 76 | 0 | | |
| | F | 1 | 39 | 0 | | |
| 6638 | M | 2 | 69 | 0 | | |
| | F | 2 | 66 | 0 | | |
| 6639 | M | 4 | 118 | 0.72 ± 0.4 | | |
| | F | 3 | 62 | 0.52 ± 0.5 | | |
| 6640 | M | 4 | 79 | 3.05 ± 3.0 | 203 | 16.7 (rep = 1) |
| | F* | 3 | 47 | 10.0 ± 5.1 | 261 | 3 ± 2.4 (rep = 2) |
| 6641 | M | 1 | 246 | 0 | | |
| | F | 1 | 100 | 0 | | |
| 6642 | M | 1 | 40 | 0 | | |
| | F | 1 | 376 | 0 | | |
| 6643 | M | 3 | 149 | 0.83 ± 0.4 | | |
| | F | 1 | 40 | 0 | | |
| 6644 | M | 3 | 100 | 0 | | |
| | F | 1 | 130 | 0 | | |

### *3.4.1.ii.b Results from repeated experiment using synthetic sgRNAs*

Four of the six sgRNAs (6636, 6640, 6642, and 6643) that were retested using the synthetic synthesis method caused non-flying rates in $G_1$ heterozygotes that were significantly different to the wildtype control: up to a mean non-flying rate of 57.1% (mean ± SE 5.5%, n=6) was observed compared to 19.0 % (mean ± SE 2.4 %, n=5) in the wildtype control (Figure 17). Importantly, in this experiment, flight was assessed using the flight test device rather than the simple binary method employed previously. Therefore, the results cannot be directly compared with the previous experiment, though

a notably higher rate of non-flying females was observed in the $G_0$ using the commercially sourced synthetic sgRNAs, indicating that they were likely much more efficient.

The "non-flying" groups in Figure 17 can include mutated females that are completely flightless ("null"), mutated females that can fly at some rate comparable to wildtype (either reduced flight ability ("semi-flightless") or equal to wildtype), and finally some non-mutated wildtype individuals. A conservative estimate of flight was unfortunately not recorded in this experiment to be able to define the non-flying proportions more precisely (i.e. completely flightless/"null" vs reduced flight ability/"semi-flightless"). It is important to note, however, that while null mutants were observed in the $G_1$ of groups 6640-6644, the non-flying signal arose mostly from females that were semi-flightless. Nevertheless, the results suggested that a partial dominant effect could be produced with four (or possibly five) of the sgRNAs targeting the C-terminal region of *Act-4.* To determine the rate of dominance of the knock-out effect (i.e. the proportion of mutated samples producing a non-flying phenotype), samples from all groups were collected and sequenced.

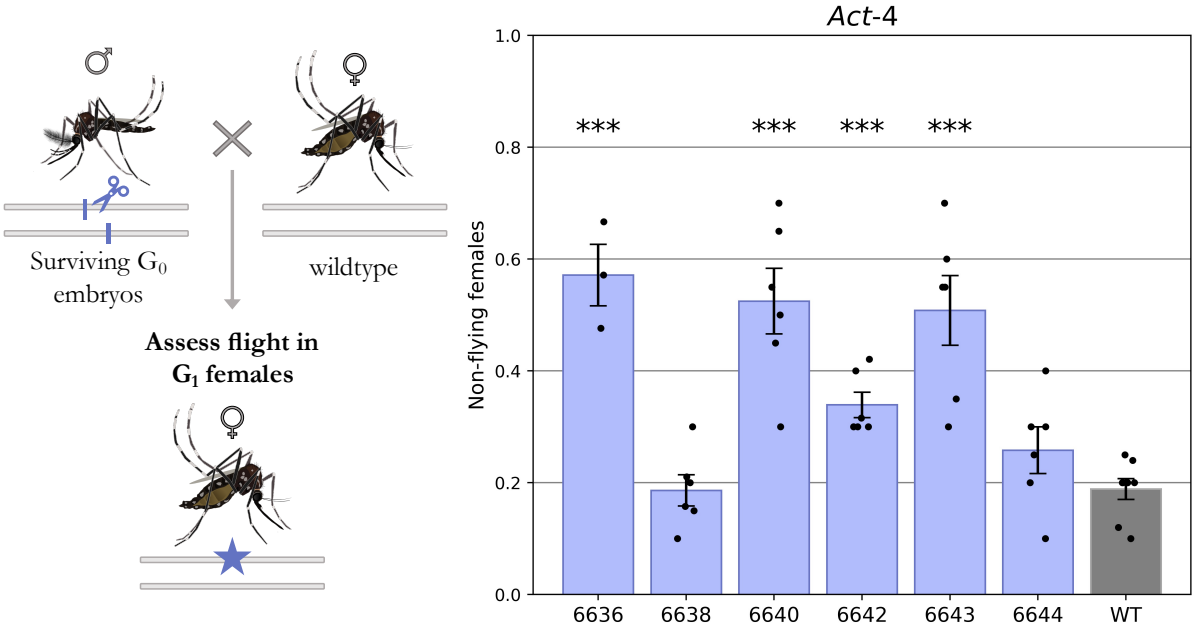



*Figure 17* In vivo *assessment of synthetic sgRNAs targeting* Act-4 *at the C-terminal end. Crossing scheme shows microinjected surviving $G_0$ individuals were outcrossed to wildtype and the $G_1$ assessed for female flight to identify any dominant mutants. Rate of non-flying females is shown for the first generation ($G_1$) of microinjected male embryos ($G_0$) outbred to wildtype females. Only male $G_0$ survivors were outbred in this experiment to prevent negative selection of mutants from the gene pool due to the possibility of flight affected $G_0$ females with a reduced ability to reproduce. Females were assessed with a flight test device and collected for sequencing. WT = wildtype. The statistical significance of the*

*difference between each group's mean and the wildtype control was tested using a Student's t-test. *** indicates a p-value <0.001.*

### 3.4.1.iii Dominance of Act-4 at the C-terminus

Flight-tested $G_1$ female offspring carrying putative mutant alleles of the *Act-4* gene were collected from the flight test device for individual gDNA extraction and Sanger sequencing of the *Act-4* C-terminal region. Due to the design of the flight test device, all tubes opened into the same escape cage, so "flying" individuals were collected as one cohort and a subset of these were sequenced. Non-flying individuals remaining in each tube at the end of the experiment were collected separately and a subset of individuals from the first three tubes were sequenced (Table 10). To identify the mutant alleles of sequenced heterozygous samples, Sanger chromatogram traces were systematically deconvoluted using the command-line application Tracy (Rausch *et al.*, 2020).

Figure 18A shows four categories of flight-tested and sequenced females—non-flying mutated, non-flying wildtype, flying mutated, and flying wildtype—as proportions of the total female population. Overall, all sgRNAs except 6638 cleaved their target efficiently according to the rate of non-homologous end joining (NHEJ) mutations produced at each site: the highest rate of mutations was observed for sgRNA 6643 (79.5%), followed by 6640 (61.6%), 6642 (55.7%), 6636 (37.6%), and 6644 (23.8%), while no mutations were observed for sgRNA 6638.

The dominance of the *Act-4* mutations can be approximately determined with the formula used to calculate the fitness of heterozygotes (Aa):

$$\omega_{Aa} = 1 - hs$$

where 1 is the full fitness of a dominant homozygote, *s* represents the selection coefficient or the proportion of fitness lost due to a genotype, and $h$ is the dominance coefficient. Assuming the selection coefficient of *Act-4* is 1, which was determined by generating recessive mutations at the N-terminal (published by Navarro-Payá *et al*., (2020)) and C-terminal (due to somatic cleavage of target sites 6642-6644 resulting in completely flightless females discussed in the transgenic line section below) ends of the gene, then the measure of the dominance, $h$, of the mutant alleles can be determined by:

$$h = 1 - \omega_{Aa}$$

where $\omega_{Aa}$ is the flight ability of heterozygotes carrying a mutant *Act-4* allele. While the flight test device was modelled on similar devices used to test fitness of males intended for release in sterile insect technique (SIT) programs, it is not known what level of flight is required for full fitness in females and is a limitation of the study, as is discussed later. Completely flightless individuals (nulls) are expected to have zero fitness, but these were unfortunately not recorded in this experiment. Dominant null mutants were observed at some rate for 6640-6644, but not for 6636 or 6638. Nevertheless, a dominance calculation of *Act-4* was made by taking the number of mutated females that could not clear the flight test tubes (non-flying) as a proportion of all mutated samples (Figure 18B and Table 14). Because ~19% of wildtype females were not expected to clear the tubes (as observed with the wildtype control), a lower range estimate of dominance was calculated by subtracting this proportion relative to the flying mutated group from the non-flying mutated group.

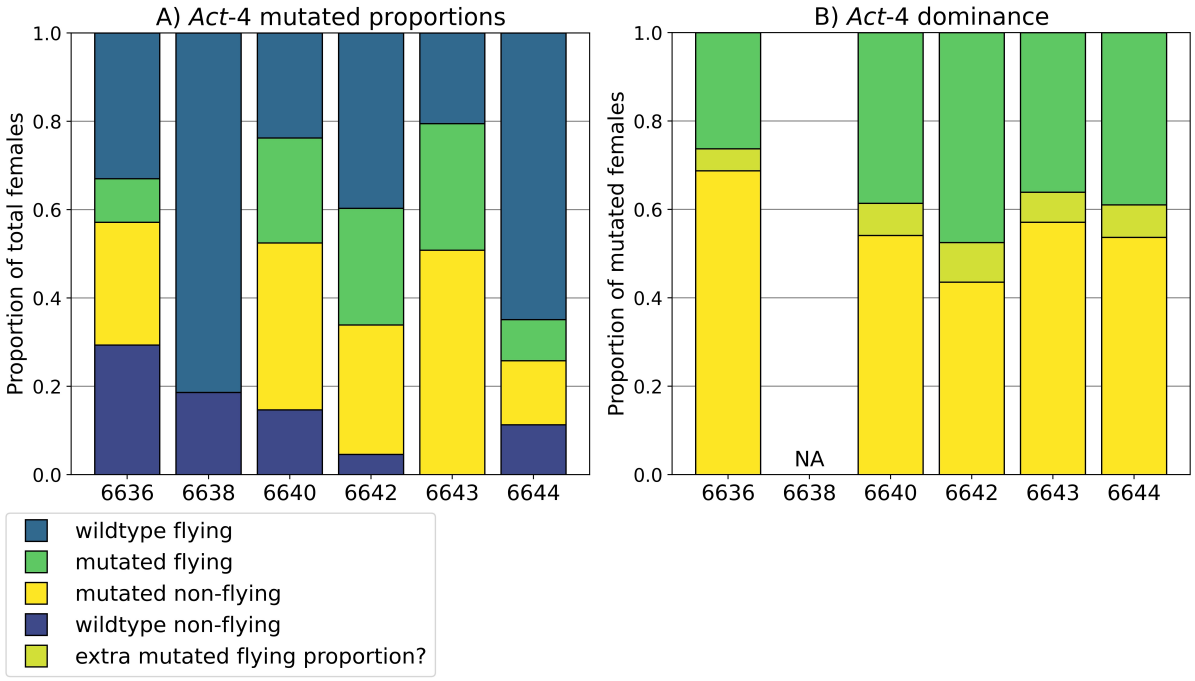



*Figure 18 Results from sequencing $G_1$ female offspring carrying putative mutations generated in the* Act-4 *gene. A) Four categories of $G_1$ females classed by flight ability (flying or non-flying) and sequence verified allele (wildtype or mutated) are shown as proportions of the total female population; the four categories are wildtype females classed as flying (dark teal bars) and non-flying (dark purple bars) and mutated females classed as flying (green bars) and non-flying (yellow bars). The sum of the mutated proportions (green and yellow bars) indicates the sgRNA efficiency. B) Mutated proportions normalized to 1 are shown. The lime green bars represent the proportion of non-flying mutated females that would be expected to fly given the proportion of females that were non-flying in the wildtype control. The yellow bars thus represent the lower range estimate of dominance of each target site. 6638 did not produce any mutations so a dominance calculation could not be performed (NA).*

The lower range dominance rates induced by NHEJ mutations caused by five of the sgRNAs ranged from 44-69% (Figure 18B and Table 14). Both the uncorrected and corrected 'lower' range rates were significantly different to two flight test device assessments of wildtype and two transgenic lines each carrying a copy of a knock-in integration at a haplo-sufficient site of *Act-4* (Table 14, Table 15, Table 16). Similar rates of non-flying wildtype females (mean=22.5% ± SE 4.2%, n=6) to those reported here (mean=19.0% ± SE 2.4%, n=5) were observed by colleagues at York University using an independently built replica of the flight test device. Similar rates were also reported for heterozygous transgenic females carrying a transgene at the start of exon 1 in Act-4 (AGG-1070 carrying a fluorescent genetic marker transgene integrated at sgRNA site e1-1 in Figure 2 of methods), which had previously been shown to be recessive using a simple flight test and examination of myofibrils using transmission electron microscopy (mean=21.7% ± SE 2.8%, n=6) (Navarro-Payá et al., 2020). A comparable haplo-sufficient transgenic knock-in line of Act-4 (AGG-1349 carrying a marker and sgRNA cassette integrated within a resected region between sgRNA sites e1-1-4 in Figure 2 of methods) was also tested here with similar results (mean= 22.7% ± SE 6.7%, n=3). All sgRNAs were significantly different to the full knock-out of Act-4 (1349/1070), showing that the dominant effect was incomplete.

*Table 14 Summary of non-flying rates recorded for female progeny of nine parental crosses: wildtype control (experiment 1), cross of two transgenic lines (AGG-1349 and AGG-1070) each heterozygous for a transgene knock-in at a haplo-sufficient site of* Act-4 (exp 2)*, cross of one transgenic line to wildtype (AGG-1070) (exp 3), and six CRISPR-Cas9-injected male G0 survivors crossed to wildtype (exp 4-10). LVP is the wildtype strain of* Ae. aegypti *reared in the insectary. The male genotype of the parental cross is shown on the left of the expression and the female genotype is on the right. *Represents CRISPR-Cas9-injected individuals with a heterogenous mix of cleaved (NHEJ mutant) and non-cleaved (wildtype)* Act-4 *alleles. Non-flying rates represent the mean percentage of females that had not escaped the flight test tubes in the flight test device within 60 mins of the start of the flight test assay. Experiment 3 was performed by colleagues at York University. Standard errors of the means (SE) and number of tubes used for the flight test experiment (n) are shown; genotypes with the same experiment group number are siblings from the same parental cross. Mean non-flying rates of* Act-4*-mutagenized $G_1$ progeny were calculated from sequenced samples of three tubes verified to carry a mutant* Act-4 *allele. Values in parentheses represent the lower range dominance values that were calculated by subtracting the proportion of wildtype females not expected to clear the tubes relative to the flying mutated group (19%) from the non-flying mutated group.*

| Genotype | wt copies of *Act-4* | Experiment (exp) | Parental cross | Reps | Mean non-flying at 60 min (%) | SE (%) |
|---|---|---|---|---|---|---|
| LVP | 2 | 1 | LVP | 5 | 19.0 | 2.4 |
| wt-1 | 2 | 2 | (1349: LVP): (1070: LVP) | 3 | 18.7 | 3.5 |
| 1349 | 1 | 2 | (1349: LVP): (1070: LVP) | 3 | 22.7 | 6.7 |
| 1349/1070 | 0 | 3 | (1349: LVP): (1070: LVP) | 3 | 100.0 | 0.0 |
| wt-2 | 2 | 3 | (1070: LVP): LVP | 6 | 22.5 | 4.2 |
| 1070 | 1 | 3 | (1070: LVP): LVP | 6 | 21.7 | 2.8 |
| 6636 | 1 | 4 | (6636*: LVP) | 3 | 73.5 (68.5) | 1.5 (1.8) |

| | | | | | | |
|---|---|---|---|---|---|---|
| 6638 | 1 | 5 | (6638*: LVP) | 3 | 18.9 | 1.6 |
| 6640 | 1 | 6 | (6640*: LVP) | 3 | 53.7 (44.9) | 7.5 (9.0) |
| 6642 | 1 | 7 | (6642*: LVP) | 3 | 54.6 (46.0) | 2.8 (3.3) |
| 6643 | 1 | 8 | (6643*: LVP) | 3 | 63.9 (57.1) | 4.7 (5.6) |
| 6644 | 1 | 9 | (6644*: LVP) | 3 | 60.7 (53.3) | 4.8 (5.7) |
| 6640 (simple) | 1 | 10 | (6640*: LVP) | 2 | 39.2 | 8.3 |

*Table 15 Summary of significance levels between differences in non-flying rates for each experiment, calculated using a pairwise t-test. Statistical significance is denoted as follows: *** p < 0.001, ** p < 0.01, * p < 0.05.*

| wt copies of *Act-4* | | 2 | 2 | 1 | 0 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | genotype | LVP | WT-1 | 1349 | 1349/1070 | WT-2 | 1070 | 6636 | 6638 | 6640 | 6642 | 6643 | 6644 |
| 2 | LVP | | | | *** | | | *** | | ** | *** | *** | *** |
| 2 | WT-1 | | | | *** | | | *** | | * | ** | ** | ** |
| 1 | 1349 | | | | *** | | | ** | | * | * | ** | ** |
| 0 | 1349/1070 | | | | | *** | *** | *** | *** | ** | *** | ** | ** |
| 2 | WT-2 | | | | | | | *** | | ** | ** | *** | *** |
| 1 | 1070 | | | | | | | *** | | ** | *** | *** | *** |
| 1 | 6636 | | | | | | | | *** | | ** | | |
| 1 | 6638 | | | | | | | | | * | *** | *** | ** |
| 1 | 6640 | | | | | | | | | | | | |
| 1 | 6642 | | | | | | | | | | | | |
| 1 | 6643 | | | | | | | | | | | | |
| 1 | 6644 | | | | | | | | | | | | |

*Table 16 Summary of significance levels between differences in non-flying rates for each experiment using the lower range estimates for* Act-4 *$G_1$ progeny (values in parentheses in Table 14), calculated using a pairwise t-test. Statistical significance is denoted as follows: *** p < 0.001, ** p < 0.01, * p < 0.05.*

| wt copies of *Act-4* | | 2 | 2 | 1 | 0 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | genotype | LVP | WT-1 | 1349 | 1349/1070 | WT-2 | 1070 | 6636 | 6638 | 6640 | 6642 | 6643 | 6644 |
| 2 | LVP | | | | *** | | | *** | | * | *** | *** | *** |
| 2 | WT-1 | | | | *** | | | *** | | | ** | ** | ** |
| 1 | 1349 | | | | *** | | | ** | | | * | * | * |
| 0 | 1349/1070 | | | | | *** | *** | *** | *** | ** | *** | ** | ** |
| 2 | WT-2 | | | | | | | *** | | * | ** | ** | ** |
| 1 | 1070 | | | | | | | *** | | * | ** | *** | *** |
| 1 | 6636 | | | | | | | | *** | | ** | | |
| 1 | 6638 | | | | | | | | | * | ** | ** | ** |
| 1 | 6640 | | | | | | | | | | | | |
| 1 | 6642 | | | | | | | | | | | | |
| 1 | 6643 | | | | | | | | | | | | |
| 1 | 6644 | | | | | | | | | | | | |

### *3.4.1.iii.a Variants*

Variants generated by each sgRNA and their distribution among non-flying and flying groups are illustrated in Figure 19. Variants generated at site 6640 are shown for two pools of $G_1$ individuals tested with a simple flight test (results from microinjections with *in vitro* transcribed sgRNAs) and one pool tested with the flight test device (synthetic sgRNAs, comparable to the other four sgRNA groups shown). Individuals classed as non-flying in a simple flight assessment displayed completely null phenotypes, while flying

individuals exhibited varying flight abilities. Notably, variant 415-(2) (indicating an indel variant at position 415 bp relative to the amplicon start site, a deletion of 2 bp, indicated by brackets), was exclusively detected in non-flying individuals across all three pools, suggesting a potential dominant negative mutation (Figure 19). Similarly, variants 413-(3) and 417-(2) were also exclusively detected in null mutants, again indicating possible dominant negative mutations, although the sample sizes were small; variant 413-(9), another in-frame mutation at position 413, observed in group 6643, was also only found in non-flying females.

Other variants at the same positions, however, did not induce a non-flying phenotype in the more sensitive flight assay (e.g. 415-(1), 415-(7), and 413-(1) and 413-1 in group 6643), indicating that while both in-frame and out-of-frame mutations could induce null phenotypes, the null phenotype may be caused by specific variants or that variants can produce a range of phenotypic effects, possibly depending on the age of the individual. As the sample sizes were small for each sequenced variant (Table 11), it is difficult to draw specific conclusions. The latter hypothesis, however, is supported by the detection of variant 416-(17) in females exhibiting a null phenotype as well as in the flying groups of those assessed with the flight test device, although most individuals carrying this mutation could not clear the flight test tubes. It is possible that these mutated females could be delayed in gaining the ability to fly, such that the females could be at different stages of flight development when they were flight tested. This has been observed for *myo-fem* mutants, which regained some ability to fly two days after eclosion (O'Leary and Adelman, 2020). In this experiment, females were flight tested earliest at three days post eclosion but sometimes up to seven days, suggesting this could be a possibility.

The dominance attributed to site 6636, which was the highest among all groups (69%, Table 14), was curiously caused by only one out-of-frame variant. The presence of this mutation in the flying groups suggests that this out-of-frame variant negatively impacted flight, but the dominance was incomplete and varied between individuals. Indeed, a previous CRISPR-Cas9 study of *Act-4* produced a 10 bp deletion 5 bp downstream of the cleavage site of sgRNA 6636 that caused a null phenotype in 3% of heterozygous females in a relatively conservative assessment of flight (O'Leary and Adelman, 2020). Both in-frame and out-of-frame variants were found in non-flying individuals from groups 6642-6644, though conclusions about the dominance of these variants cannot be drawn on a

per-variant nor in-frame/out-of-frame basis due to small sample sizes and non-normal distribution. It is still possible that in-frame deletions would be more likely to cause antimorphic mutations, but more data would be needed to support this. Nonetheless, sgRNAs 6640-6644, which targeted sites within a 40-bp region, displayed similar dominance rates, suggesting a diverse set of mutations contribute to incomplete dominance effects and that these can vary between individuals. The sgRNAs with the highest dominance rates (6642-6644) were thus taken forward to be incorporated into one multiplexed sgRNA expressing construct for piggyBac transgenesis; 6640 could have been included but was not.

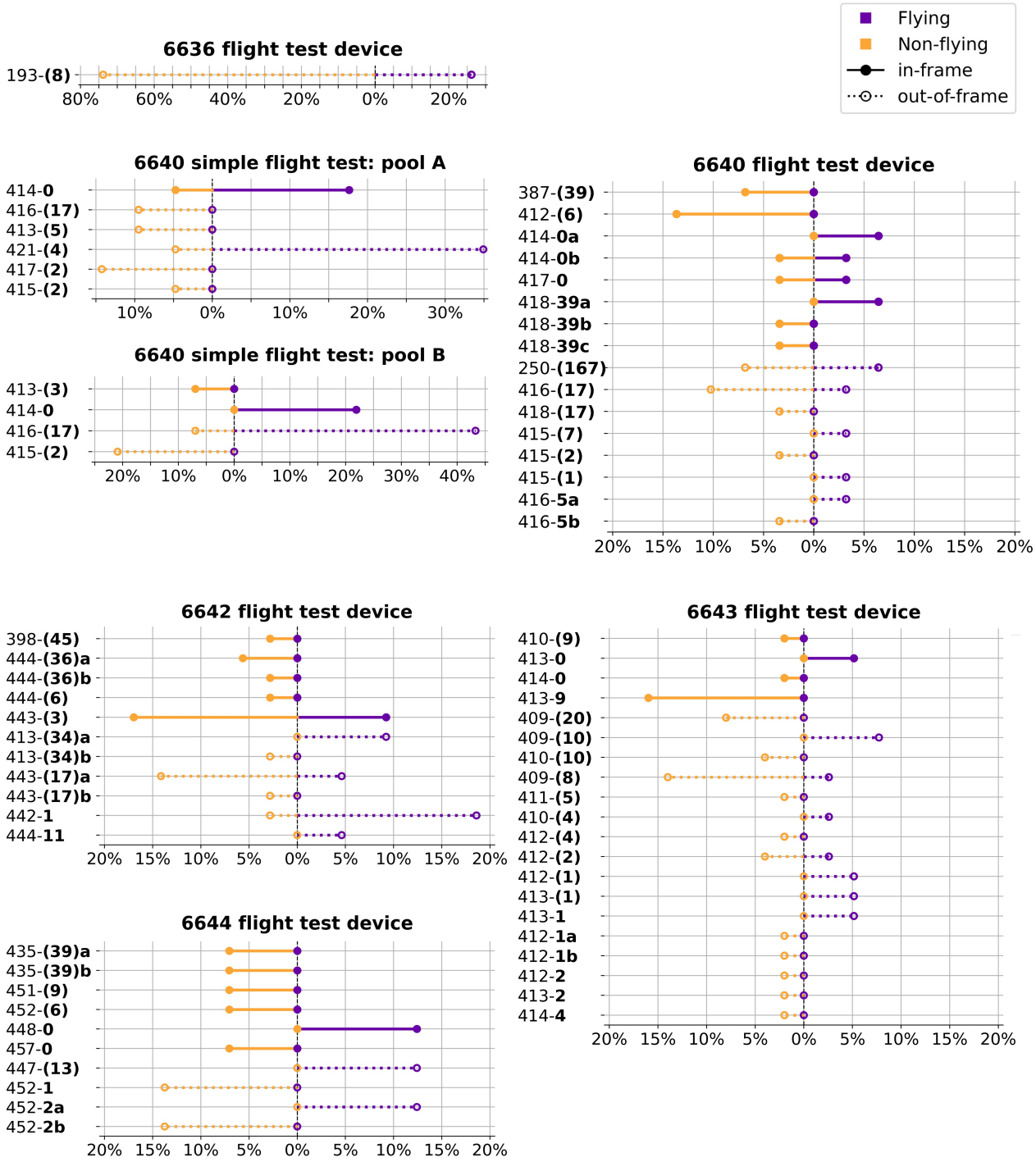



*Figure 19 Variants detected by Sanger sequencing of the* Act-4 *target sites 6636, 6640, and 6642-6644. Plots show the distribution of genetic variants among flying (purple) and non-flying (orange) $G_1$ females as a proportion of the total mutated population. Variants are labelled by the position of the indel or SNP mutation relative to the reference sequence followed by the size of each indel in bold; deletions are indicated by parentheses and letters are added as unique identifiers where more than one variant was found at the same site and of the same size. In-frame mutations are represented by solid lines, while out-of-frame mutations are represented by dashed lines.*

*3.4.1.iii.b Limitations of the Act-4 mutagenesis experiment*

Results from the proportions of wildtype females present in the non-flying groups of the heterozygous $G_1$ offspring indicated some issues with this experiment, however. The percentage of non-flying wildtype individuals taken as a proportion of the total wildtype female population was high for two of the sgRNAs tested (6636 at 47.1% and 6640 at 38.2%), where the expected non-flying rate for wildtype individuals was around ~19% (Figure 17 and Figure 18A). An overrepresentation of wildtype in the non-flying population would reduce the calculated dominance value because it reduces the number of recorded mutations resulting in a non-flying phenotype. It would be inappropriate to treat the wildtype proportion as an internal control for the proportion of mutated samples found in the non-flying group that should be assumed to fly because there could be compounding factors linked to the calling of mutant variants, as is discussed below. A lower-than-expected proportion of wildtype is not as concerning because the proportions observed in 6642 and 6644 (10.4% and 14.8%, respectively) were not statistically different to the control and 6643 (0%) had a low sample number of wildtype flying females (n=25); also, for this experiment (but not the wildtype control) the tubes were agitated by thumping the device up and down prior to each recording to encourage females to fly and prevent escaped females from flying back into the tubes, which could have led to the lower proportions (this also makes the high wildtype proportions in 6636 and 6640 more concerning).

Possible reasons for the overrepresentation of wildtype alleles in the non-flying groups include issues with the calling of mutant variants, flight test device, off target mutations, and primer bias. sgRNAs were assessed for off-targets during the design process, however, and primer binding sites were chosen within conserved sequence. Issues with the flight test design are also unlikely, given the similarity between the controls, including the two wildtype assessments, the two transgenic lines deemed to be haplo-sufficient, and group 6638 that did not appear to be mutated (and could thus function as a negative control). Second, females were generally believed to fly independently of each other, at least at the ranges of females added into the tubes here; there is a risk that enclosing too many females in a tube could lead to injuries through increased contact and thereby increase the rate of non-flying wildtype females, but with only 20 females added to the 5cm diameter tubes, the density was very low and no overcrowding was observed (see

Figure 15h). So, while some variation around the control non-flying rates was expected, the high levels of wildtype genotype found in groups targeted with sgRNA 6636 and 6640 suggest another source of variation.

The most probable reason for the overrepresentation of wildtype alleles in the non-flying groups could be due to issues with calling of mutant variants. Mutant alleles with large deletions could have missing primer binding sites that prevented amplification or sequencing or alignment by the Tracy tool (the default trim size on each end is 50 bp). This would have been a greater risk for sgRNA 6636 (non-flying wildtype proportion=47.1%) because the position of the sgRNA cleavage site was 194 bp into the 623 bp amplicon, of which the initial ~50-100 bp of a sequencing read can be of poor quality. The cleavage site of sgRNA 6640 (non-flying wildtype proportion=38.2%), in contrast, was 417 bp into the amplicon, so relatively distant from the sequencing primer and nearby sgRNA sites 6642-6644, which had relatively low non-flying wildtype rates (10%, 0, and 15%, respectively), however deletion of the reverse primer binding site or difficulties in alignment could still have occurred. NHEJ repairs are strongly influenced by sequence context, so mutational outcomes, including length of indels, could be different for each sgRNA site (Chen *et al.*, 2019; Aubrey *et al.*, 2015; van Overbeek *et al.*, 2016). Indeed, deletions up to 167 bp were observed for sgRNA 6640, whereas other sgRNAs had much shorter indel ranges (Figure 19). The implications of this reasoning are complicated, however, because it assumes that these long indel variants (i.e. >~200 bp) would have had some impact on flight, but this would be unexpected because the *Act-4* protein is relatively short (327 residues) and haplo-sufficient at the N-terminal end.

It would have been more informative to assess flight conservatively in addition to the assessment using the flight test device and collect and sequence three groups of females: those that were completely flightless (null), those that could not clear the flight test tubes (semi-flightless), and those that could (flying). Samples were collected into three flight categories when transgenic lines expressing a multiplexed sgRNA cassette targeting *Act-4* at sgRNA sites 6642, 6643, and 6644 were tested (discussed below), but these samples have not yet been sequenced. A longer amplicon and sequencing from both ends might have produced a more accurate result, too; Tracy can discover indels up to 1000 bp.

### 3.4.2 *piggyBac* integrations

Transgenic lines expressing multiplexed sgRNA cassettes targeting *myo-fem* and *Act-4* were generated by piggyBac-mediated transgenesis of mosquito embryos. Three Pol III-driven sgRNA constructs targeting *myo-fem* in exon 2 (AGG-2199) and exon 3 (AGG-2198 and AGG-2227), and one construct targeting *Act-4* at sgRNA sites 6642, 6643, and 6644 (AGG-2695), were individually microinjected into embryos with a helper plasmid expressing hyperactive piggyBac transposase (Figure 14). Integration of each construct was verified by screening $G_1$ larvae for AmCyan fluorescence after which positive males from each pool were used to establish transgenic isolines. Isolines represent independent single integrations of each sgRNA cassette and were named with a letter that denoted the pool they originated from and a number indicating the $G_1$ founder male (Table 17). The total number of isolines generated for each construct is shown in Table 17 and isolines carrying insertions linked to the sex-determining regions, i.e. the M or m locus, are highlighted.

*Table 17 piggyBac mediated transgenesis of* Ae. aegypti *embryos with sgRNA expressing constructs targeting* myo-fem *in exon 2 and 3 and* Act-4 *at the C-terminus. Survival rates of microinjected embryos, number of $G_0$ founder pools set up, and number of isolines generated for each micro-injection experiment are shown. Transgenic isolines were established from single $G_1$ positives crossed to wildtype females. Isolines carrying single insertions of transgenes were isolated by screening for normal 50/50 Mendelian inheritance rates that would be expected for a single dominant allele. Independent isolines are represented with a letter denoting the pool and $G_1$ founder male they originated from, respectively. Isolines carrying sex-linked insertions are underlined; an * indicates the isoline is m-linked (female-biased inheritance from males), while a ^ indicates the isoline is M-linked (male-biased inheritance). nr = not recorded.*

| AGG: gene target | Embryos injected | $G_0$ survival (%) | No. founder pools | No. isolines |
|---|---|---|---|---|
| **2198**: *myo-fem* exon 3 | 298 | 6.7 | 6 | 3 (A1, B1*, D2) |
| **2199**: *myo-fem* exon 2 | 300 | 7.7 | 5 | 4 (A2, C1, D1, E1) |
| **2227**: *myo-fem* exon 3 | 337 | 14.2 | 11 | 5 (A2, B2, D1, E1, F1^) |
| **2695**: *Act-4* C-terminus | nr | nr | 3 | 5 (A2, B1^, B2, C1, C2*) |

Two lines from each *myo-fem* set and three from *Act-4* were tested for construct activity with a transgenic line expressing Cas9 driven by germline promoter *shutdown* from an M-linked locus (AGG-2235) (Figure 20A). One transgenic line expressing the sgRNA construct targeting *Act-4* (AGG-2695B1) was itself M-linked, so was instead tested with an m-linked *shu*-Cas9 line (AGG-1551B1). Male $F_1$ transheterozygotes, i.e. those that carry both a copy of the *shu*-Cas9 transgene and the transgenic sgRNA cassette, were produced and crossed to wildtype females. Depending on the activity of the sgRNA cassette, *shu-Cas9* was expected to induce germline mutations that would be inherited in the $F_2$ generation and produce flightless females if dominant.

### *3.4.2.i myo-fem*

Expression of sgRNAs targeting *myo-fem* from integrated transgenes produced high rates of dominant completely flightless mutants in the $F_2$ when assessed with a simple flight assay (Figure 20B). A null phenotype was observed for up to 73% of $F_2$ offspring from transgenic line AGG-2199-C1 that targeted *myo-fem* in exon 2. There was a wide range of activity observed between transgenic lines of the same construct, indicating position effects strongly influenced Pol III promoter-driven sgRNA expression. For example, a lower rate of flightless females (~26%) was observed for AGG-2199-A2 compared to the C1 isoline. No difference was observed in the flightless rates of sisters that had inherited the sgRNA construct compared to those that had not, indicating no paternal deposition of *shu*-Cas9.

### *3.4.2.ii Act-4*

Lower rates of non-flying females were observed in the $F_2$ of transheterozygotes compared to direct injection of single guides that was discussed above (Figure 20C). This could have been due to low expression of sgRNAs from the genome resulting in lower rates of end-joining mutations or due to lower dominance rates when larger regions were resected due to targeting multiple sites with the multiplexed sgRNA cassette. Sequencing of flying and non-flying mutants will be necessary to determine the dominance rate here (i.e. the proportion of mutated individuals that exhibit a phenotype), but this has yet to be performed. Thus, at the moment, no conclusions can be drawn about the lower rates of non-flying females observed in these crosses. Due to the general rate of ~50% dominance across the second exon of *Act-4* and the discovery of long deletions (up to 167 bp) that could have reduced flight, it would be surprising if multiplexing reduced this dominance rate. Null phenotypes were indeed observed in the $F_2$ of each isoline, and though the sample sizes were small (3-24 total non-flying of each group), they ranged from 25-67% of the females in the non-flying groups.

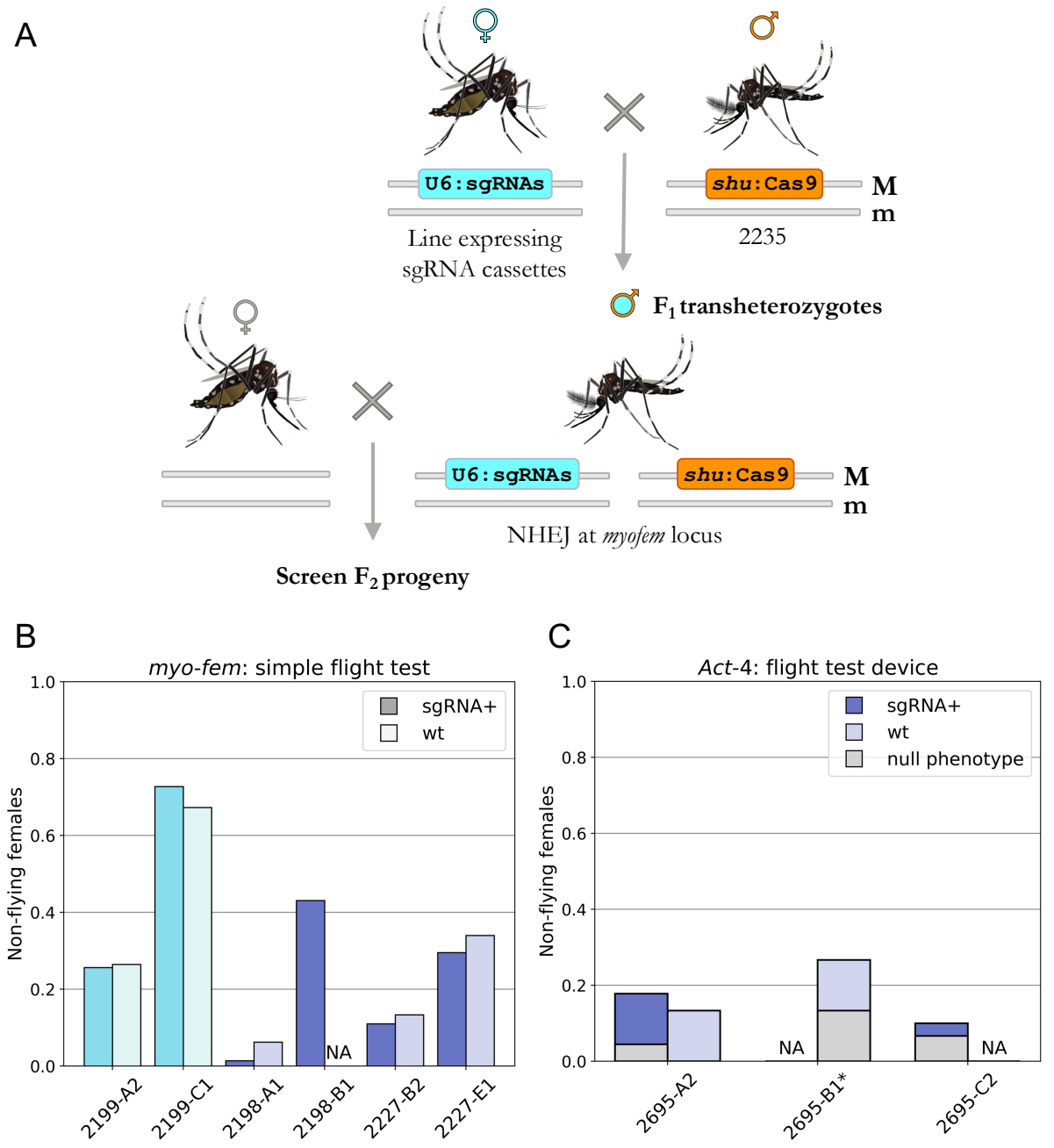



*Figure 20 A) Crossing scheme to determine activity of piggyBac integrated sgRNA cassettes with a source of Cas9 driven by the shu promoter and expressed from an M-linked locus (AGG-2235). *2695-B1 was M-linked so was instead tested with an m-linked shu-Cas9 line (AGG-1551B1); in the schematic the sexes would therefore be reversed. $F_1$ transheterozygous males carrying both the sgRNA cassette and shu-Cas9 were isolated and crossed to wild-type females. The $F_2$ progeny were then screened for inheritance of the sgRNA cassette and then tested for flight (B and C).* myo-fem *sgRNA cassette groups were tested with a simple flight test so flightless rates represent completely flightless null phenotypes, while* Act-4 *sgRNA cassette groups were tested with a flight test device so could represent a range of flight abilities. Flightless rates are shown as a proportion of total females tested. The proportion of females exhibiting a null phenotype are additionally shown in grey in C).*

### 3.4.2.iii Dominance of *myo-fem* in exon 2 and 3

To determine the dominance of mutagenizing exons two and three of *myo-fem* (i.e. the proportion of mutated samples producing a non-flying phenotype), samples from the isolines with the highest activity (2199-C1, 2198-B1, and 2227-E1) that had been assessed by the simple flight assay were sequenced (Figure 21 and Table 18). Isolines 2199-C1 and 2227-E1 targeting exons two and three, respectively, were then retested with the flight test device to gain a more precise estimate of flight, and samples from both genotypes (sgRNA+ and wildtype) were collected for Sanger sequencing of the targeted regions.

*Table 18 Summary of dominance rates calculated from mutagenesis of* myo-fem *exon 2 and exon 3. The male genotype of the parental cross is shown on the left of the expression and the female genotype on the right. Dominance values represent the percentage of females carrying a single copy of a mutant allele that had not escaped the flight test tubes in the flight test device by 60 mins. Values in parentheses represent the lower range dominance values that were calculated by subtracting the proportion of wildtype females not expected to clear the tubes relative to the flying mutated group (19%) from the non-flying mutated group. Standard errors of the means (SE) and number of tubes used for the flight test experiment (reps) are shown; genotypes in bold are siblings from the same parental cross.*

| Transgenic line | $F_2$ genotype | Flight test | Parental cross | Reps | Dominance | SE |
|---|---|---|---|---|---|---|
| 2199-C1 | wildtype | FTD | (2235: 2199-C1): LVP | 3 | 0.95 (0.94) | 0.01 (0.01) |
| 2199-C1 | sgRNA+ | FTD | (2235: 2199-C1): LVP | 3 | 0.91 (0.89) | 0.01 (0.01) |
| 2199-C1 | sgRNA+ | simple | (2235: 2199-C1): LVP | 1 | 0.93 | - |
| 2227-E1 | wildtype | FTD | (2235: 2227-E1): LVP | 3 | 0.91 (0.89) | 0.03 (0.04) |
| 2227-E1 | sgRNA+ | FTD | (2235: 2227-E1): LVP | 3 | 0.84 (0.81) | 0.04 (0.04) |
| 2227-E1 | wildtype | simple | (2235: 2227-E1): LVP | 1 | 0.72 | - |
| 2198-B1 | sgRNA+ | simple | (2235: 2198-B1): LVP | 1 | 0.68 | - |

High rates of dominance were observed after mutagenesis of each exon. Dominance was higher for exon two compared to exon three when assessed with a simple flight assay: 0.93 compared to 0.70 (Figure 21B and Table 18). When mutant females were assessed with the flight test device, however, dominance rates were not statistically different for the two exons: 0.93 compared to 0.87. This suggests that a large proportion of mutations in exon three induced an intermediate phenotype that reduced flight ability but did not cause a full null phenotype.

Lastly, the calculated dominances likely vary around the reported values due to issues with variant calling, which lead to a potential underestimation of mutant alleles in all groups, but particularly in the females exhibiting a null phenotype (non-flying group, conservative assessment). The relative allelic fractions reported by the Tracy tool that was used to deconvolute the heterozygous Sanger reads were sometimes exact 50/50

splits, which suggested only one allele had been analysed. These samples were included in the analysis, however, because otherwise the proportion of truly wildtype samples would have been underestimated (as these were more likely to carry two identical haplotypes that would also have appeared as only one allele). This could suggest that the dominance rates were underestimated, particularly for the conservative assessments and mainly exon two. There may be mutated samples in the flying groups that would not have been called either, however, but one could potentially assume that these would also have been excluded from the non-flying groups. Longer amplicons and sequencing from both directions would have produced more accurate results, though primer design to minimize amplification bias was difficult in the highly variable intronic sequences that flanked the relatively short exons.

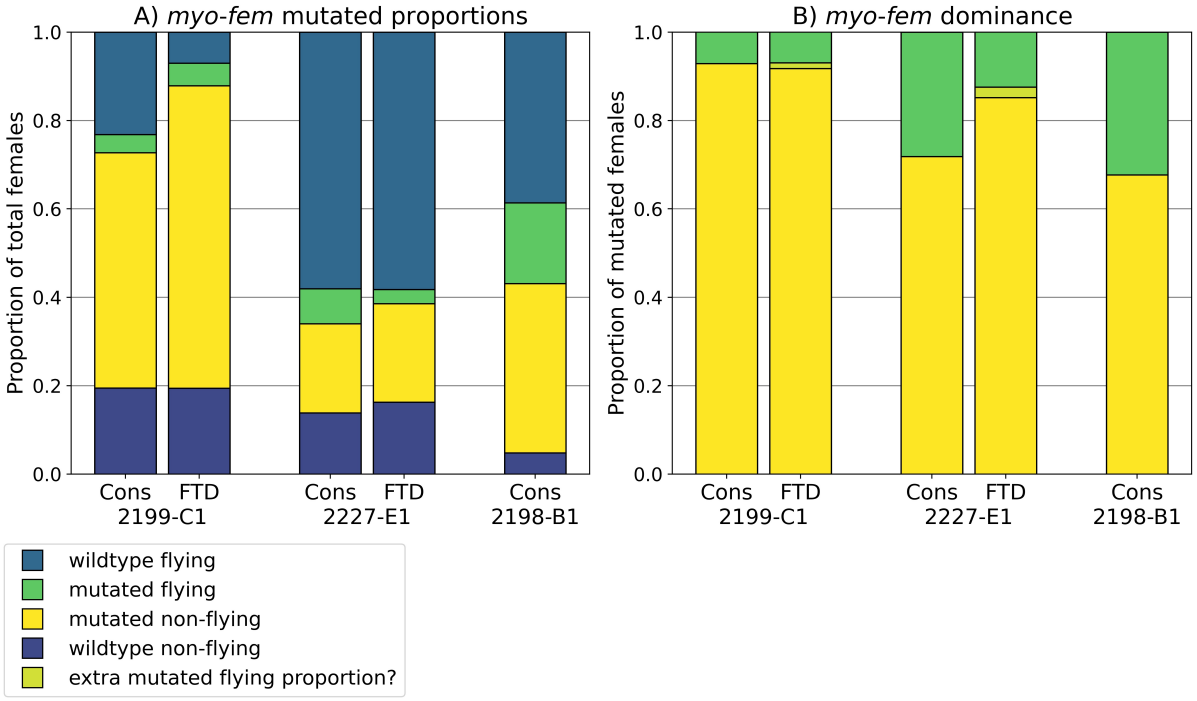



*Figure 21 Results from sequencing of $F_2$ female offspring carrying putative mutations generated in the* myo-fem *gene. A) Four categories of $G_1$ females classed by flight ability (flying or non-flying) and sequence verified allele (wildtype or mutated) are shown as proportions of the total female population (wildtype females classed as flying (dark teal bars) and non-flying (dark purple bars) and mutated females classed as flying (green bars) and non-flying (yellow bars)). The sum of the mutated proportions (green and yellow bars) indicates the cleavage efficiency. B) Mutated proportions normalized to 1 are shown. The lime green bars represent the proportion of non-flying mutated females that would be expected to fly given the mean proportion of females that were non-flying in the wildtype control. The yellow bars thus represent the lower range estimate of dominance of each target site. Flight test device (FTD) results shown are means of sgRNA+ and wt genotypes.*

*3.4.2.iii.a sgRNA cleavage efficiencies*

All sgRNAs in the 2199 cassette that targeted exon two of *myo-fem* generally cleaved at equal rates and thus were more efficient at generating large deletions (Figure 22A and B). sgRNAs from both the 2198 and 2227 cassette that targeted exon three seemed to vary in their cleavage efficiencies. No variants were observed around the cleavage position of sgRNA 6035 likely because of a SNP in the seed region that likely prevented Cas9 binding at its target site. Finally, sgRNAs 6038 and 6039 seemed to cleave more efficiently than 6036, as inferred from the respective distributions in Figure 22B. These results are a rough estimate, however, since as sequencing was only performed from one end, large mutations that span both sgRNAs would be more likely to be attributed to the sgRNA that comes first. Nevertheless, sgRNA 6039 comes after 6036, so was likely more efficient. Together these results also lend confidence to the use of the four different Pol-III promoters in multiplexed constructs since cleavage did not seem to vary with promoter.

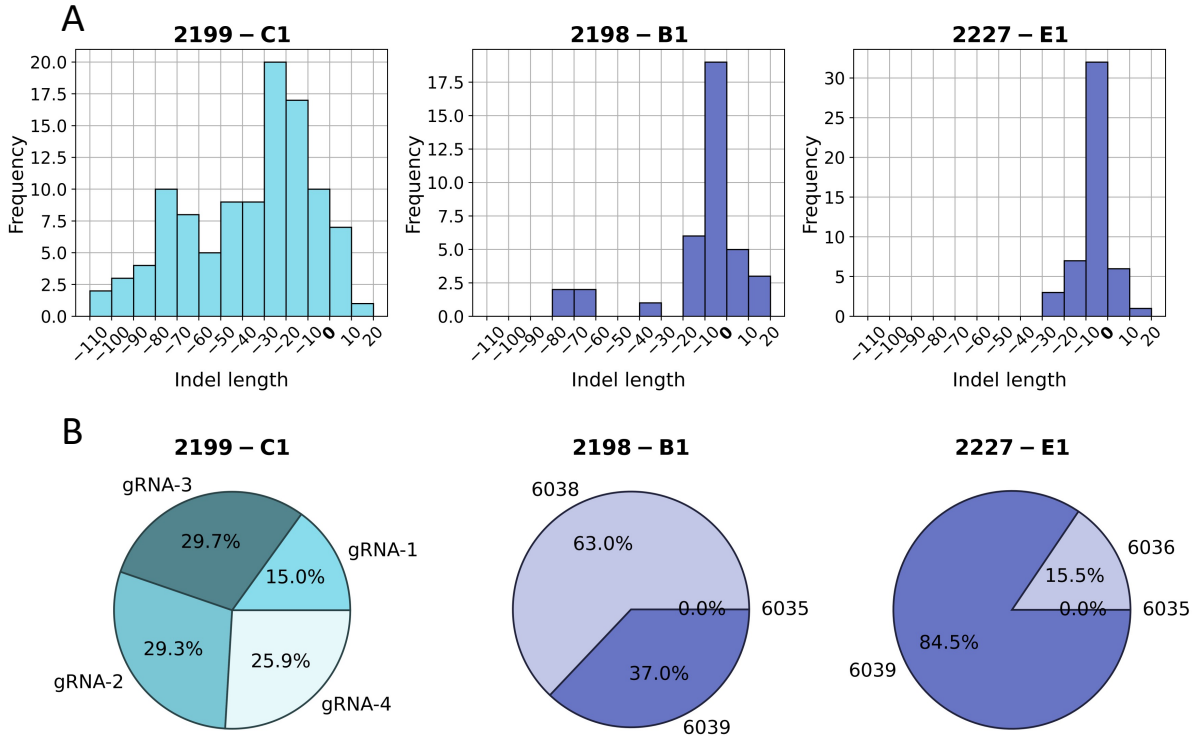
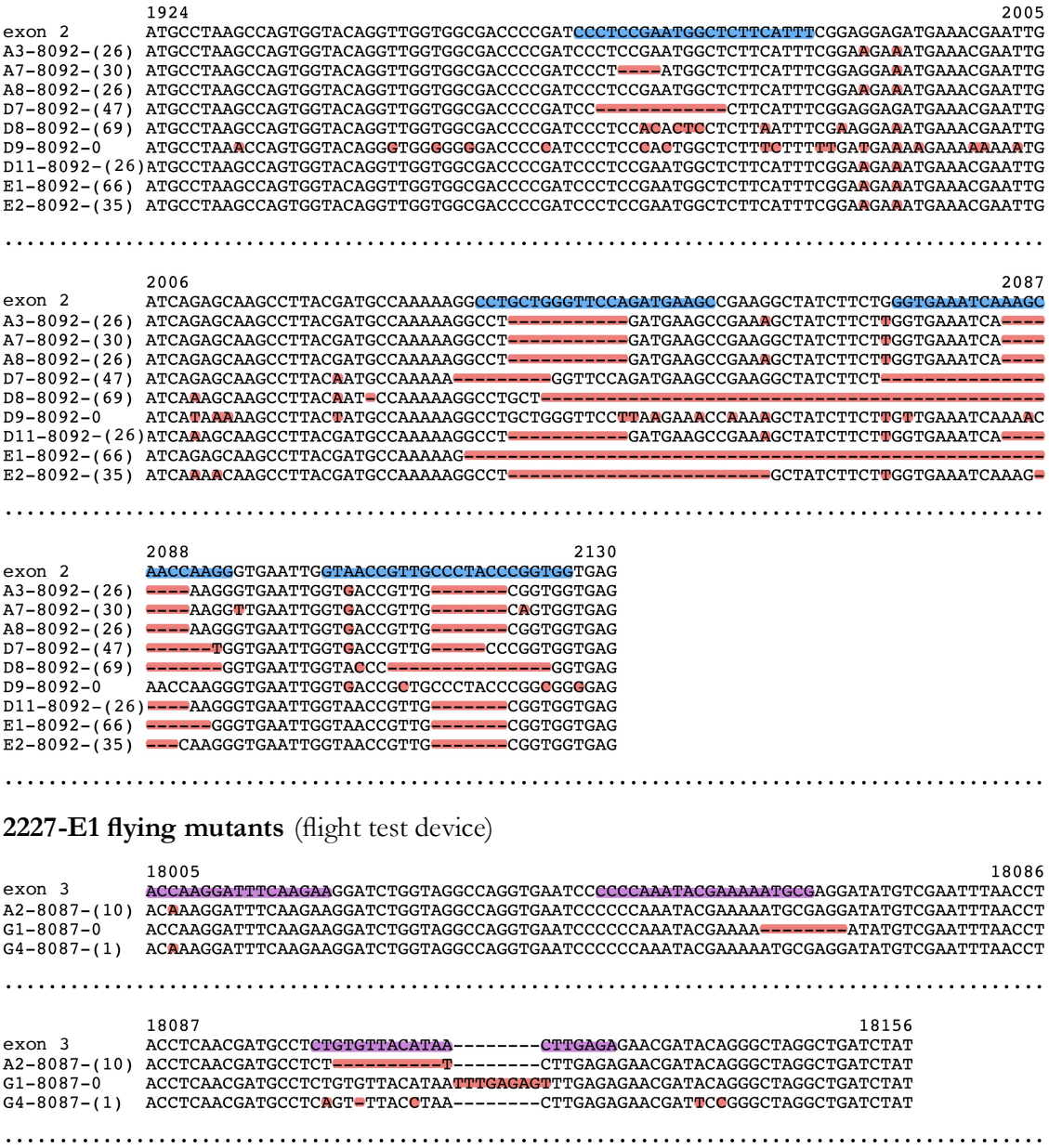



*Figure 22 A) Distribution of indel sizes in the offspring of $F_1$ transheterozygotes carrying one of three sgRNA cassettes targeting* myo-fem *in exon 2 (2199-C1) and exon 3 (2198-B1 and 2227-E1). B) Pie charts showing proportion of mutations attributed to each sgRNA included in each sgRNA cassette. 2199-C1 and 2227-E1 include variants from groups tested with both flight tests.*

**2199-C1 flying mutants** (flight test device)




**2227-E1 flying mutants** (flight test device)




*Figure 23 Alignment of mutant Sanger sequencing traces identified within the 'flying' groups of 2199-C1 and 2227-E1 to the reference sequence of each respective* myo-fem *exon. sgRNAs targeted by 2199-C1 are highlighted in blue, while sgRNAs targeted by 2227-E1 are highlighted in purple. Sample IDs include the sequencing plate well number and the number of the sequencing primer used and are followed by the total size of the indel. Numbers in parentheses indicate a deletion, while absence of parentheses indicates an insertion. Formatted sequence alignment was exported from Benchling.com and edited to highlight sgRNA target sites.*

### *3.4.2.iii.b Variants*

All variants identified in the "flying" mutants, as assessed by the flight test device, are shown in Figure 23 for each targeted exon. Notably, large in-frame and out-of-frame mutations in exon 2 permitted flight through the flight test device. Despite the diverse

range of variants represented in the $F_2$, one variant was observed three times in the flying group and was not discovered in the non-flying group, suggesting a potentially functional mutant allele. Nonetheless, the "flying" variants were not that dissimilar from other mutations identified in null and non-flying mutants, with similar overall deletion sizes and sgRNA sites cleaved. Together these results could imply that these specific variants permit flight, or alternatively, that there could be some variability in the dominance of the knock-out effect between individuals, akin to that observed in *Act-4*.

### 3.4.3 Flight test device

The flight test device developed in this study was valuable for achieving efficient and sensitive assessments on flight using relatively inexpensive materials. It offers a means of conducting an initial assessment of flight with greater precision over conservative flight assessments, enabling researchers to determine whether further investigation into a mutant phenotype is warranted. Further, it has been possible to adapt this device to assess flight in another mosquito species, *An. stephensi*, with some modifications to the methodology to reflect different behaviour (e.g. assessment at dawn/dusk only). To assess males, a smaller tube diameter of 2 cm compared to the 5 cm used for females is recommended. Although timepoints were not analysed in this study, a time series analysis could be incorporated to further investigate flight dynamics, including any delays to gaining flight ability. More advanced techniques for evaluating flight ability, like determining wingbeat frequencies or the use of flight mills, would offer additional insights and could help better interpret the findings from the device. Additionally, incorporating fitness assessments, survival assays, and observing the time taken to develop flight post eclosion could provide more comprehensive insights.

## 3.5 Conclusion and future directions

This chapter explored the dominance of gene knock-out in two genes required for female flight in *Ae. aegypti*: *myo-fem* and *Act-4*. A key motivation for analysing the haploinsufficiency of the two genes was to assess their relevance for genetic vector control, specifically a nuclease based SuperSIT system. CRISPR-Cas9 was used to mutagenize each gene at several target sites after which females carrying a single copy of a mutant allele were assessed for flight. The *myo-fem* gene was targeted at two N-terminal exons (exons two and three, the latter of which encodes the putative motor domain of the myosin gene), following recent research that indicated the gene could be haplo-sufficient at exon three (O'Leary and Adelman, 2020). The *Act-4* gene was instead targeted at the C-terminal end because it was hypothesized to produce truncated or altered proteins rather than nulls that would be more likely to induce dominant antimorphic mutations akin to those observed in the *D. melanogaster* homolog, *Act88F,* and thereby interfere with actin's ability to polymerize*.*

The results indicated that high rates of dominant mutations could be produced at the selected target sites, but that they varied by site in both *myo-fem* and *Act-4.* Dominance was close to complete when *myo-fem* was mutagenized in exons two (0.93) and three (0.87) and females were assessed with a flight test device. When using a conservative assessment of flight that considered only null phenotypes, dominance was similar in exon two (0.93), but lower in exon three (0.68), suggesting a higher tolerance of indels at exon three. As most females carrying a mutant allele at exon two showed a null phenotype, it is not known whether the large out-of-frame alleles carried by flying individuals represented functional variants that permitted flight or whether dominance can sometimes vary significantly between individuals. Mutant females could also have developed the ability to fly but at a delayed rate post eclosion, as has been reported when comparing flight of *myo-fem* mutants at one day post eclosion, when wildtype females can already fly, versus three (O'Leary and Adelman, 2020). The significance of such a delay in terms of female fitness is not known, but it could reduce the females' chances of survival. As flight ability of the mutants generated in this chapter was assessed at three (and up to seven) days post eclosion of all pupae, it is possible that some females could have gained the ability to fly in that time; thus, the dominance could have been somewhat underestimated. Assessing the flight ability of these flying mutants' offspring at discrete

time points would show whether this flight ability was heritable and could improve with age and thus inform on the risk of resistance developing at this site.

For *Act-4,* dominance of knock-out was greatly increased when the C-terminal end was mutated compared to the previously targeted N-terminal end, where disruption has been shown to be recessive (Navarro-Payá *et al.*, 2020). Dominance of the flightless phenotype at the five CRISPR targeted sites, as measured by a flight test device, was incomplete, calculated as 0.69 at the end of exon one and slightly varying around 0.50 at four sites in exon two. While the dominance rates of the first exon relied on a single variant, an independent study of an overlapping out-of-frame mutation reported null phenotypes in a small proportion of heterozygous females when flight was assessed conservatively, suggesting a nonzero impact on flight perhaps similar to the effect observed here (O'Leary and Adelman, 2020). Multiple variants (both in-frame and out-of-frame) were discovered in exon two with cleavage occurring within a 40 bp region, indicating that the incomplete dominance rate was general rather than variant specific in this region. Dominant mutations in *Act-4* have been reported in the literature, with some suggestion that these are biased to in-frame mutations, though this was not observed here (Navarro-Payá *et al.*, 2020). Putative dominant antimorphic mutations were highlighted above.

The impact of intermediate flight ability on female fitness remains unexplored in this chapter and in the literature, although females with no flight ability have been shown to have zero fitness (Li *et al.*, 2021a). The relationship between flight ability and fitness has been better studied in males, especially in the context of sterile insect (SIT) release programs. Ionizing radiation is commonly used to sterilize insects but can negatively impact insect quality and thus field competitiveness. Therefore, various quality control assessments have been developed to help determine the optimal dose that balances male sterility with quality; among these, a key method involves measuring flight ability (Culbert *et al.*, 2018; Dor *et al.*, 2020; Carpenter, Blomefield and Vreysen, 2012; Balestrino *et al.*, 2017). One of these flight test devices, on which the device used in this chapter was modelled, was able to predict the first significant decrease in insemination rates of females resulting from incrementally increasing radiation doses (and other physiological stresses), which occurred at relatively high flight rates (>70% escape rate through the device, i.e. 30% non-flying rate) (Culbert *et al.*, 2018).

This underscores the importance of investigating the relationship between varying degrees of flight ability in females and their survival rates and mating competitiveness. It could be that minor reductions in female flight ability are not as deleterious as they would be in males, as females are normally heavily outnumbered by males in mating swarms, resulting in intense competition between males for selection by females. On the other hand, harmonic convergence is performed by both sexes at a shared frequency that is triple that of females' normal output, emphasising the role females play in the courtship ritual (Cator *et al.*, 2009). Further, because flight tones are highly conserved within mosquito species, and particularly among those belonging to the human-biting genera (*Aedes, Anopheles,* and *Culex*), females that exhibit an atypical flight tone may go undetected by mate-seeking males, thereby preventing mating initiation (Kim *et al.*, 2021). A recent study that tested male *An. gambiae* responses to auditory stimuli of females showed that males showed a strong preference to the wildtype flight tone even when compared to a flight tone produced by a ~13% increase in the wing beat frequency (Su *et al.*, 2020). Such preferences may be further amplified in large experimental cages or the field where there is greater competition between wildtypes. Lastly, females must be capable of powering flight at much higher end ranges when carrying heavier loads such as a blood meal or developing eggs, the former of which can double their weight (Jové *et al.*, 2020). It is thus plausible that small degrees of flight changes could reduce mating competitiveness or survival enough such that females are functionally sterile in the wild.

In conclusion, this chapter highlights a significant finding that high rates of dominance exceeding 0.9 were achievable with a multiplexed sgRNA construct cleaving exon two in *myo-fem*. This dominance rate reflected a null phenotype in females, so would correspond to a substantial reduction in fitness. This holds promising implications for developing a nuclease based SuperSIT system, which requires modified males to impose a fitness cost on female descendants indirectly as females are not expected to inherit the construct. Even incomplete dominance, as long as the dominant effect is above 0.8, has been modelled to achieve sufficient population suppression within a reasonable time frame using a low release ratio (Burt and Deredec, 2018). Furthermore, the partially dominant sites identified in *Act-4* could potentially be included in a SuperSIT construct targeting *myo-fem* with the aim of increasing the impact on flight. However, studies into the stoichiometric balance between *Act-4* and *myo-fem* would be required to ensure that

function is not restored with two nulls, as reported for *D. melanogaster.* Finally, while the primary focus of this research was the development of the SuperSIT system in *Ae. aegypti*, gaining a deeper understanding of the genes that are involved in reproduction or survival has implications for the success of future genetic vector control strategies.

# Chapter 4: Knock-in of the M locus and assessment of flight in female descendants of SuperSIT males

## 4.1 Abstract

The SuperSIT system involves a transgenic construct that is only transmitted through males and yet disrupts the fitness of female descendants. In previous chapters, work towards achieving the second criterion—a reduction in the fitness of female descendants—showed that CRISPR targeting of the female-specific flight gene *myo-fem* could induce strong dominant fitness costs in female offspring. This chapter discusses work towards achieving the first criteria—male-only inheritance—by assembling a *β2t*-driven CRISPR-Cas9 construct targeting *myo-fem* within the male-determining locus, the M locus, of *Ae. aegypti*. To do this, the M locus was first analysed *in silico* for potential integration sites that would allow efficient expression of the SuperSIT construct and not impact fitness. sgRNAs for knock-in of the M locus were then designed and tested *in vivo* in day-old mosquito embryos, which showed that the M locus is accessible to sgRNA cleavage and that a high cleavage activity could be achieved (up to 85%). Thirteen of these sgRNAs targeting a wide range of the M locus were chosen for knock-in integration. Transgenic lines with male-only inheritance were attained for five of these sites, including two in the *Nix* intron and one in the *myo-sex* intron. Two of the five sites supported expression of the full integrated CRISPR-Cas9 construct, causing disruption of flight in female descendants. Together, these results showed proof of principle of a SuperSIT system in *Ae. aegypti,* the first built to date, and offered insight into directions for future research, which were discussed.

## 4.2 Introduction

The SuperSIT system involves a transgenic construct that is only transmitted through males and disrupts the fitness of female descendants. Linking a transgene to a male-determining chromosome or locus—i.e. the Y chromosome in heterogametic species—ensures that modified males pass it on through all their surviving sons. This is essential to the SuperSIT design (and what determines its efficiency) because the construct would be absent from the daughters it harms and so its inheritance would not be affected by the suppressive effect it creates (Burt and Deredec 2018). The previous two research chapters explored mechanisms of disrupting the fitness of non-inheriting SuperSIT daughters using CRISPR-Cas9. Chapter 2 showed that the regulatory elements of Ae. aegypti homologs of the Drosophila genes shut-down (shu, AAEL011632), *β2*-tubulin (*β2t*, AAEL019894), and ProtA (*proA*, AAEL003177) could be used to express Cas9 in the male germline and generate mutations in the pre-meiotic sperm genome. Chapter 3 then showed that high rates of dominance exceeding 0.9 were achievable with a multiplexed sgRNA construct cleaving exon two of myo-fem. This dominance rate reflected a null phenotype, so would correspond to a substantial reduction in female fitness. Together, the first two research chapters showed that a promising route for SuperSIT development was through paternal inheritance of dominant myo-fem mutations. This chapter will thus discuss the assembly of a SuperSIT CRISPR-Cas9 construct targeting *myo-fem* within the M locus to first achieve male-only inheritance, and second, to manipulate flight in non-inheriting female progeny.

### *4.2.1* Structure of the sex-determining M locus in *Ae. aegypti*

In *Aedes* and *Culex* mosquitoes, sex is not determined by a Y chromosome, but rather by a male-determining locus—the M locus—found in a 1.5 Mb region on an autosome that is otherwise mostly homomorphic between the sexes (Newton, Wood and Southern, 1978; Hall *et al.*, 2014; Hall *et al.*, 2015; Fontaine *et al.*, 2017; Matthews *et al.*, 2018; Aryan *et al.*, 2020). The M locus is absent from females, meaning males are essentially heterogametic (M/m) and females are homogametic (m/m) (Matthews *et al.*, 2018). The M locus supports expression of two protein-coding genes, including the principal gene in the sex determination pathway, the male-determining factor *Nix* (AAEL022912), and a gene required for male flight, *myo-sex* (AAEL021838), as well as several long-noncoding RNA

(lncRNA) that show enriched expression in the male reproductive tract (Aryan *et al.*, 2020; Matthews *et al.*, 2018; Mysore *et al.*, 2021; Turner *et al.*, 2018).

The M locus is found in a heterochromatic region of suppressed recombination at the pericentromeric region of chromosome 1 (AegL5_1:151.68..152.95 Mb) (Juneja *et al.*, 2014; Matthews *et al.*, 2018; Timoshevskiy *et al.*, 2014; Sharakhov and Sharakhova, 2015). The low recombination frequency is further maintained by sexually antagonistic factors near the male-determining factor that reduce the number of successful crossover events (Krzywinska *et al.*, 2016b; Campbell *et al.*, 2017). Due to the lack of recombination, this region has accumulated retrotransposons and other repetitive elements, such that more than 73.7% of the M locus is repetitive DNA (Matthews *et al.*, 2018). The intron of *Nix* alone is 100 kb, almost 75% of which is repetitive elements, which is exceptional for a zygotically expressed gene that normally benefits from having short introns to enable rapid transcription and in turn rapid cell cycles (Biedler *et al.*, 2012; Artieri and Fraser, 2014; Turner *et al.*, 2018). This sequence degeneration is characteristic of Y chromosomes in various species, including *Drosophila melanogaster* and *Anopheles gambiae* (Muller, 1964; Charlesworth, 1991; Kaiser and Bachtrog, 2010; Turner *et al.*, 2018; Bachtrog, 2013; Carvalho *et al.*, 2001; Carvalho, 2002).

### 4.2.2 Challenges of integrating transgenes in heterogametic chromosomes

The highly repetitive and heterochromatic nature of the M locus and heteromorphic chromosomes makes transgene integration and robust transgene expression challenging (Bernardini *et al.*, 2014; Elgin and Reuter, 2013). Moreover, because repetitive DNA hampers sequencing, assembly, and annotation of genomes, data for Y chromosomes is often incomplete, unassigned, or absent from genome assemblies, which prevents precise genome editing (Tørresen *et al.*, 2019; Hall *et al.*, 2016; Tomaszkiewicz, Medvedev and Makova, 2017). Even if a representative sequence assembly is made available, knock-in by homology-directed repair (HDR) can be inefficient in low-complexity and repetitive regions (Tan *et al.*, 2020).

Successful transgene integrations in insect Y chromosomes have thus mainly been achieved by rare and random events from transposon-mediated transgenesis or chromosomal recombination (Condon *et al.*, 2007; Adelman *et al.*, 2008; Carpenetti *et al.*,

2012; Bernardini *et al.*, 2014; Bernardini *et al.*, 2017; Hall *et al.*, 2014; Hall *et al.*, 2015; Krzywinska *et al.*, 2016b; Chen *et al.*, 2022). In *Ae. aegypti*, male-linkage of transgenes can be attained while still being relatively distant from the M locus due to the lack of recombination; strong male-female genetic differentiation extends to 63-123 Mb around the M Locus (representing 20%-40% of the entire chromosome length) (Fontaine *et al.*, 2017). In our lab, multiple M-linked insertions have been produced through random *piggyBac* mediated transgenesis or through spontaneous recombination of relatively common m-linked insertions to M-linked regions, which show comparable expression to autosomal sites. Together, this suggests that the M-locus, or at least proximal autosomal sequence, may be less refractory to integration and expression of transgenes than the Y chromosome is in other insect species.

Recent advances in sequencing and scaffolding technologies greatly improved the contiguity and chromosome anchoring of genome assemblies, including those of various mosquito species (Matthews *et al.*, 2018; Waterhouse *et al.*, 2020; Zamyatin *et al.*, 2021; Kingan *et al.*, 2019). As a result, the newest genome assembly for *Ae. aegypti* provided a more complete understanding of the genomic context of the M locus and made possible site-specific transgenesis around it (Matthews *et al.*, 2018; Li *et al.*, 2020; Lutrat *et al.*, 2023). Similarly, site-specific knock-in of the *Drosophila* Y-chromosome was also recently achieved (Buchman and Akbari, 2019). An *Ae. aegypti* transgenic genetic sexing strain (GSS) was developed by integrating a fluorescent marker 9 Mb from *Nix* using CRISPR-Cas9 knock-in into the *mucin-3A* gene (AAEL019619) (Lutrat *et al.*, 2023). The rate of recombination for this transgene was 0.01%, which is an improvement over the 0.29-0.9% rates achieved using random transgenesis methods and 0.2% using CRISPR-Cas9 knock-in into the *white* gene (AAEL016999), which is 45 Mb from *Nix* (Krzywinska *et al.*, 2016b; Li *et al.*, 2020).

Nevertheless, this nonzero recombination rate would mean that the GSS, developed for automated sex sorting of insects being mass-reared for field release, would require frequent manual checking of colonies to avoid accumulation of recombinant females (Lutrat *et al.*, 2023). This would be relevant for upscaling production of an effective SuperSIT line were it to be used in control programmes in the field. Male-linkage could possibly be improved by targeting regions within the M locus but attempts to integrate

into the *Nix* intron, for example, have not been successful and, to date, no transgenes have been inserted within the M locus in *Ae. aegypti* (Lutrat *et al.*, 2023).

### 4.2.3 Considerations for safe and efficient expression of transgenes integrated in the M locus

Identifying genetic "safe harbours" that promote efficient transgene expression and do not have a deleterious impact on fitness is challenging in eukaryotes in general, let alone in the heterochromatic M locus (Papapetrou and Schambach, 2016; Aznauryan *et al.*, 2022; Papapetrou *et al.*, 2011). Moreover, the regulatory context that leads to significant position effects of integrated transgenes is not fully understood, especially in *Ae. aegypti*, where repetitive elements are spread across its genome and have been proposed to cause "heterochromatisation" of open euchromatin and to have diverse functions with respect to gene regulation (Anderson *et al.*, 2023a; Timoshevskiy *et al.*, 2014; Sharakhov and Sharakhova, 2015).

Some criteria for safe harbours, originally used in the context of gene therapy, are outlined here. First, integration should not alter the regulation of neighbouring genes and the endogenous regulatory environment should not negatively affect expression of the inserted transgene promoter (Aznauryan *et al.*, 2022). Integrations upstream of a transcriptional start site or within the first or 5' intron of the transcript could be disruptive due to the presence of regulatory regions there (Papapetrou *et al.*, 2011; Anderson *et al.*, 2023b; Fyrberg *et al.*, 1998). Integration in a distal or 3' intron or downstream of the coding regions could be safer, while still benefiting from local enhancers and euchromatin that could upregulate transgene expression, especially of those found in gene rich regions (Frøkjær-Jensen *et al.*, 2008; Lombardo *et al.*, 2011)(Anderson *et al.*, 2023). Lastly, proximity to regulatory RNAs should also be considered when evaluating genetically neutral insertion sites because they play an important role in regulating gene expression; this could be especially important in the M locus, which has a considerable amount of lncRNAs with a putative role in male development (Papapetrou *et al.*, 2011; Papapetrou and Schambach, 2016; Fatica and Bozzoni, 2014; Matthews *et al.*, 2018).

In mosquitoes, candidates for genetic safe harbours have been explored through the development of population replacement homing gene drives that aim to introduce a

desirable trait, such as inability to spread disease, into a population (Nash *et al.*, 2019; Carballar-Lejarazú *et al.*, 2020; Hoermann *et al.*, 2021; Ellis *et al.*, 2022). These homing drives require genetically neutral sites that are also sequence conserved to spread efficiently, two often conflicting criteria (Beaghton *et al.*, 2017; Noble *et al.*, 2017). Integral gene drives aim to address this problem by fusing transgenes to the N- or C-terminal end of endogenous genes, taking advantage of the surrounding regulatory regions without having a negative impact on the expression of the host gene, similar to endogenous reporter genes (Nash *et al.*, 2019; Rojas-Fernandez *et al.*, 2015). Such a design would be challenging to adopt for integrating the SuperSIT construct in the M locus for several reasons, including a limited selection of genes with a desired expression profile and potential difficulties in expressing a large multi-cassette construct, and as such the integrations discussed in this chapter are all found in noncoding regions, using the criteria for genomic safe harbours discussed above.

### 4.2.4 Aims

The aim of this chapter was to identify loci within the noncoding region of the M-locus that would be amenable to transgenesis by CRISPR-Cas9 knock-in and would support efficient expression of a transgenic CRISPR-Cas9 construct in the male germline of *Ae. aegypti*. The previous chapters discussed two objectives of the SuperSIT project: one, to assess activity of male germline-expressing promoters when expressing Cas9; and two, to test sgRNAs targeting female-specific flight genes. They showed that a promising route for SuperSIT development was through paternal inheritance of dominant myo-fem mutations. This chapter will thus discuss the assembly of a SuperSIT CRISPR-Cas9 construct targeting *myo-fem* within the M locus to first achieve male-only inheritance, and second, to manipulate flight in non-inheriting female progeny. As position effects, where the chromosomal context in which a transgene is inserted affects expression levels, have been well documented here and elsewhere, this chapter will attempt integration of a SuperSIT CRISPR-Cas9 construct at multiple loci within the M locus (Anderson *et al.*, 2023a; Schotta *et al.*, 2003).

## 4.3 Methods

### 4.3.1 *In silico* analysis of the M locus for sites to integrate SuperSIT transgenes

#### 4.3.1.i Initial selection of possible regions to target in M locus and sequence retrieval

The M locus was first visualised in the VectorBase genome browser JBrowse using the AaegL5 genome assembly of the Liverpool *Aedes* Genome Working Group strain (LVP_AGWP) (Matthews *et al.*, 2018; Giraldo-Calderón *et al.*, 2022). Using available genome tracks, regions with a higher gene density, genomic complexity, and GC content, and a lower number of tandem repeats were highlighted for potential targeting. Sequences for seven regions were downloaded in fasta format from VectorBase using nearby genes as a reference point, given generic names, and imported into Benchling for further analysis and annotation (Appendix B Table 1). The seven regions in order of their location on the M locus were named: *myo-sex* intron, *myo-sex* downstream, intergenic region 1, intergenic region 2, Nix downstream, Nix intron, and Nix upstream; their genomic coordinates are found in Table 19. Together, approximately 710 kB of sequence was retrieved, over half of the available sequencing data for the M locus (1.27 Mb).

*Table 19 Genomic coordinates of the M locus sequences retrieved for each region.*

| Region name | Genomic coordinates |
|---|---|
| *myo-sex* intron | AaegL5_1:151,952,864..152,244,832 |
| *myo-sex* downstream | AaegL5_1:152,244,438..152,340,247 |
| Intergenic region 1 | AaegL5_1:152,349,018..152,403,537 |
| Intergenic region 2 | AaegL5_1:152,439,886..152,482,268 |
| *Nix* downstream | AaegL5_1:152,541,641..152,616,641 |
| *Nix* intron | AaegL5_1:152,613,641..152,721,167 |
| *Nix* upstream | AaegL5_1:152,721,170..152,766,374 |

#### 4.3.1.ii sgRNA design for targeting selected regions of M locus

The CHOPCHOP web tool (Labun *et al.*, 2019) was used to select target sites for CRISPR/Cas9-directed mutagenesis of the M locus regions. The imported VectorBase sequences were queried for sgRNAs against the AaegL5 genome of *Ae. aegypti,* the "knock-in" mode, and default settings on CHOPCHOP. 1213 sgRNAs with an efficiency score above 0.2 and fewer than two predicted off-targets with three mismatches only were selected for further analysis (Appendix B Table 2).

### 4.3.1.iii Repetitive DNA in M locus

Data on four repeat categories in the M locus—tandem repeats, low complexity repeats, transposable elements, and generic repeats—were requested from researchers at the Eukaryotic Pathogen, Vector and Host Informatics Resource centre (VEuPathDB, https://veupathdb.org), Gloria Giraldo-Calderon of Notre Dame University and John Brestelli of the University of Pennsylvania. To pick up on repeats that were not identified using data from VectorBase, select sequences were also queried using the Novoprolabs sequence finder (https://www.novoprolabs.com/tools/repeats-sequences-finder).

### 4.3.1.iv Consolidation of data and selection of smaller subregions to analyse further

The seven sequences retrieved from VectorBase were imported into Benchling and annotated with relevant genomic features. Thirty-two smaller subregions within the seven regions were selected for potential targeting based on having a GC content between 35-55%, the presence of sgRNAs predicted to be active and located outside of a transcript, RNA-Seq signal, or 5' end of the first intron, and the absence of tandem, low-complexity, and transposable element repeats (Table 20).

*Table 20 Subregions within each larger region that were selected for targeting. The number of subregions within each of the seven regions, their names, and the total number of sgRNAs per region covered by the subregions.*

| Region | No. subregions | Subregion names | No. sgRNAs |
|---|---|---|---|
| Myosex Intron | 7 | 1A; 3B; 4B; 4C; 5A; 6; 7 | 70 |
| Myosex Downstream | 7 | 1; 2; 2B; 4; 5A; 5B; 7 | 39 |
| Intergenic Region 1 | 1 | 1 | 3 |
| Intergenic Region 2 | 3 | 1-3 | 7 |
| Nix Downstream | 4 | 1-4 | 20 |
| Nix Intron | 6 | 1A; 3A; 3; 3B; 4; 5 | 30 |
| Nix Upstream | 4 | 1; 1A; 2B; 3 | 33 |
| Total | 32 | | 202 |

#### *4.3.1.iv.a Sequence verification of selected subregions*

Sequence variation in and around the sgRNA site could prevent Cas9 cutting or homology repair, and thus block or reduce the efficiency of homology-directed recombination (HDR) (Ang *et al.*, 2022). So, the 32 subregion sequences in the lab LVP strain were first compared to that of the AaegL5 genome assembly before sgRNAs could be selected for testing. PCR amplicons from 16 individual male samples were sequenced across all sgRNA sites found in the selected subregions to identify variants above a ~10% frequency.

Primer pairs were initially tested on pooled gDNA of 40 males before using successful primer pairs on individual samples (Appendix B Table 3 and

Appendix B Table 4). gDNA was extracted using the previously described method for single and pooled gDNA extraction using the MagMAX kit and KingFisher instrument. All sites were PCRed with Q5 polymerase and 20-50 ng of gDNA in 25 µL reactions according to the manufacturer's instructions. Annealing temperatures were determined using the NEB Tm calculator (https://tmcalculator.neb.com/#!/main) and extension times were calculated using a ratio of 20 s/kB. Sites that were successfully amplified from all 16 samples were PCR purified with AMPureXP beads using the Kingfisher Flex instrument. PCR purified samples were sent for sequencing with the sequencing primers listed in Appendix B Table 5.

### *4.3.2* Testing cleavage efficiency of sgRNAs *in vivo*

#### 4.3.2.i Method overview

To test the activity of sgRNAs targeting the M locus *in vivo*, sgRNAs were injected into mosquito eggs that were then harvested for mutation detection by amplicon sequencing of the target loci. sgRNA cleavage efficiency was then estimated according to the percentage of mutations at the target loci.

#### 4.3.2.ii sgRNAs selected for microinjection

To test sgRNA activity in a high throughput way, up to ten independent sgRNAs were injected at the same time and 11 injections were performed. To be able to compare sgRNA activity across injection groups, two independent positive control sgRNAs were included in each injection mix. It was reasoned that these would provide a measure of variation across the injection groups and help identify any sources of bias. The two previously verified control sgRNAs used were *kmo-499* and *Act4-6640* (Table 21).

*Table 21 sgRNA sequences of control sgRNAs.*

| sgRNA name | Sequence |
|---|---|
| Act4-6640 | TGTCTACCTTCCAAGCTATGTGG |
| kmo-499 | GGTTCCCTTCTACGGGCAGGG |

It was reasoned that the M locus sgRNAs had to be at least 20 kB apart to be included in the same injection mix to minimize any interference or bias in cutting or difficulty attributing cleavage activity between sgRNAs. Sequence verified sgRNAs were grouped

into eight bins according to their genomic coordinates and sorted by their CHOPCHOP rank. The top sgRNAs from each bin were then randomly assigned to 10 or 11 injection groups. In total, 83 M locus sgRNAs were selected to be tested for activity *in vivo* (Table 22).

*Table 22 Top sgRNAs chosen for testing* in vivo *and the injection groups they were assigned to. sgRNAs within one subregion or grouping of subregions were randomly assigned to 10 or 11 injection groups. *where sgRNAs from the same subregion were more than 20 kB apart, they could be assigned to the same injection group.*

| CHOPCHOP gRNA | Sequence | SgRNA name | Assigned injection group |
|---|---|---|---|
| **Bin 1: *myo-sex* intron 3B** | | | |
| 1 | CCAGTAGATCCAGTACAGCGGGG | MLoc_A_sgRNA_1 | 7 |
| 6 | TCTCCAGATACGGAGCTGGGAGG | MLoc_A_sgRNA_6 | 1 |
| 21 | CTCCAGTAGATCCAGTACAGCGG | MLoc_A_sgRNA_21 | 10 |
| 29 | CACACGGCGCACAGTAATGGTGG | MLoc_A_sgRNA_29 | 4 |
| 31 | AACTACAATTCGACCAAGGGTGG | MLoc_A_sgRNA_31 | 3 |
| 34 | TGCTCTCCAGATACGGAGCTGGG | MLoc_A_sgRNA_34 | 5 |
| 65 | TACAATTCGACCAAGGGTGGAGG | MLoc_A_sgRNA_65 | 9 |
| 71 | TCCAGTAGATCCAGTACAGCGGG | MLoc_A_sgRNA_71 | 2 |
| 76 | AAGATCCGTCAACCAGAAACAGG | MLoc_A_sgRNA_76 | 8 |
| 81 | TTGCTCTCCAGATACGGAGCTGG | MLoc_A_sgRNA_81 | 6 |
| **Bin 2: *myo-sex* intron 4B and 4C** | | | |
| 14 | CTATTGCTCGGATCACCGCGAGG | MLoc_B_sgRNA_14 | 6 |
| 15 | GACGGGACAGTATCGCGTCGAGG | MLoc_B_sgRNA_15 | 4 |
| 19 | CTTACCTGAGTCTGGGAGGGAGG | MLoc_B_sgRNA_19 | 8 |
| 93 | CGCATTACTTGCGGGCTGCCCGG | MLoc_B_sgRNA_93 | 7 |
| 125 | AACGCTGTCATGAATCCCGATGG | MLoc_B_sgRNA_125 | 10 |
| 29 | TGCGTCAGCACTATAGCGAATGG | MLoc_B_sgRNA_29 | 9 |
| 42 | GCAACCTTTGCTGTTGTACGGGG | MLoc_B_sgRNA_42 | 3 |
| 44 | GGGAATCATGTGCGGAAGGGTGG | MLoc_B_sgRNA_44 | 1 |
| 50 | CAAGTCTTGGTCAGACGTCTCGG | MLoc_B_sgRNA_50 | 5 |
| 82 | TGTAGTTGCAACATGTTTAGTGG | MLoc_B_sgRNA_82 | 2 |
| **Bin 3: *myo-sex* intron 5A** | | | |
| 14 | GGCTCTGAAAAACGTAAGCGTGG | MLoc_C_sgRNA_14 | 10 |
| 21 | ACCGAACTCGAACATCCCGTTGG | MLoc_C_sgRNA_21 | 7 |
| 45 | TACACTCATCGTATTTAGGCGGG | MLoc_C_sgRNA_45 | 1 |
| 53 | ACCAACGGGATGTTCGAGTTCGG | MLoc_C_sgRNA_53 | 5 |
| 75 | CTGTGGCCACAGTGTTGGTCCGG | MLoc_C_sgRNA_75 | 2 |
| 93 | GACGTTAAACAGAAAGTAACTGG | MLoc_C_sgRNA_93 | 8 |
| 101 | TTTCTTGACGCTGGAGAAACCGG | MLoc_C_sgRNA_101 | 9 |
| 139 | TCTGAAAAACGTAAGCGTGGAGG | MLoc_C_sgRNA_139 | 6 |
| 148 | GTACACTCATCGTATTTAGGCGG | MLoc_C_sgRNA_148 | 4 |
| 216 | TACTTGGATTACATGATCAATGG | MLoc_C_sgRNA_216 | 3 |
| **Bin 4: *myo-sex* downstream 1, 2, 2B, and 4** | | | |
| 1 | GCTTCGGTATTCGGAAACGACGG | MLoc_D_sgRNA_1 | 3 |
| 2 | GAAGGGATTGGGATTACCAGCGG | MLoc_D_sgRNA_2 | 10 |
| 6 | CAGGGCTAATAGAATTGCTGTGG | MLoc_D_sgRNA_6 | 4 |
| 10 | GTACGCAAATCTATTGGAGTAGG | MLoc_D_sgRNA_10 | 6 |
| 13 | TCGGCTAAGCAGGACGAGAAGGG | MLoc_D_sgRNA_13 | 1 |
| 17 | TTGTCGAATAACAGGGATTGTGG | MLoc_D_sgRNA_17 | 8 |
| 20 | TTCCTAGGGGGCTTAGCTAGGGG | MLoc_D_sgRNA_20 | 5 |
| 27 | TAATGGGGTTGGTTTAGTCGTGG | MLoc_D_sgRNA_27 | 9 |
| 43 | TCCTCTCTCTCAGAGCGGTACGG | MLoc_D_sgRNA_43 | 7 |
| 51 | TCTACGTCTTCCACTTCGCATGG | MLoc_D_sgRNA_51 | 2 |
| 56 | CGTGAGGTTTGGGCGAAGCTTGG | MLoc_D_sgRNA_56 | 11 |

| | | | |
|---|---|---|---|
| **Bin 5: *Nix* downstream 1-3** | | | |
| 5 | CATTCTGAATGGCCGTTGTGAGG | MLoc_F_sgRNA_5 | 6 |
| 9 | ATCACAAATCTGGTCACGATTGG | MLoc_F_sgRNA_9 | 1 |
| 11 | GACCCACAGGACGGCGATTGCGG | MLoc_F_sgRNA_11 | 8 |
| 15 | TATTTCACATGGACCATGCATGG | MLoc_F_sgRNA_15 | 2 |
| 19 | CCTGTTGGTGTGACAGGTTCAGG | MLoc_F_sgRNA_19 | 3 |
| 20 | AATCTTGCGCGCATCCAAACAGG | MLoc_F_sgRNA_20 | 9 |
| 29 | CATAAGGTTTGACGCCTGTTTGG | MLoc_F_sgRNA_29 | 4 |
| 39 | GAAAGGGGTACGGGCCAACTCGG | MLoc_F_sgRNA_39 | 5 |
| 41 | TCTCGGAAGCTGCATGGGGATGG | MLoc_F_sgRNA_41 | 10 |
| 44 | AAATTACCTGTTGGTGTGACAGG | MLoc_F_sgRNA_44 | 7 |
| **Bin 6: *Nix* intron 3B, 4, and 5** | | | |
| 6 | AAGACCCAGTCACCCAGTCACGG | MLoc_G_sgRNA_6 | 10 |
| 8 | TCCGGCTGCCGAGTCCTACGTGG | MLoc_G_sgRNA_8 | 7 |
| 10 | CCCTTCTACAGTTCTCAGCAAGG | MLoc_G_sgRNA_10 | 8 |
| 11 | CGGCGTTGAAGTGTTGTCAAGGG | MLoc_G_sgRNA_11 | 3 |
| 12 | TACCCGCGCATCAAGTCCGGTGG | MLoc_G_sgRNA_12 | 6 |
| 28 | ACTTGCCCAAGTGCCTTGGTGGG | MLoc_G_sgRNA_28 | 1 |
| 29 | TTCCTAATCAAAAAGGCGCATGG | MLoc_G_sgRNA_29 | 4 |
| 21 | GCGAGGCCCGATATTGCTGTTGG | MLoc_G_sgRNA_21 | 9 |
| 33 | AGCGTTCTCGGCTCCATCCTCGG | MLoc_G_sgRNA_33 | 5 |
| 43 | GGCAAGAGGCATTCATAAACAGG | MLoc_G_sgRNA_43 | 2 |
| 49 | GATCCCGAAACACAACAAGAGGG | MLoc_G_sgRNA_49 | 11 |
| **Bin 7: Nix intron 5, Myosex downstream 7, Nix intron 1A, and Nix upstream 2B and 3** | | | |
| 102 | TGATCCCGAAACACAACAAGAGG | MLoc_G_sgRNA_102 | 10 |
| 3 | CACTCTGGAGATGTAACCGTAGG | MLoc_E_sgRNA_3 | 1 |
| 85 | CTACGGTTACATCTCCAGAGTGG | MLoc_E_sgRNA_85 | 2 |
| 2 | GAAGTGGGATGATGATCACGCGG | MLoc_H_sgRNA_2 | 4 |
| 3 | AGTGGGATGATGATCACGCGGGG | MLoc_H_sgRNA_3 | 5 |
| 35 | ACCAACAGTATAGGAGCGATGGG | MLoc_H_sgRNA_35 | 8 |
| 68 | CATCAGCGACGATCATTTGTAGG | MLoc_H_sgRNA_68 | 7 |
| 74 | ATCAATTCTTCGCAGGTCGCTGG | MLoc_H_sgRNA_74 | 9 |
| 122 | TTCTTCGCAGGTCGCTGGTACGG | MLoc_H_sgRNA_122 | 6 |
| 35 | ACAGTGTACGCTCGTCATAGTGG | MLoc_I_sgRNA_35 | 11 |
| 47 | TACTAGGCTTAAAGAGTAAACGG | MLoc_I_sgRNA_47 | 3 |
| **Bin 8: Nix upstream 2B and 3** | | | |
| 1 | TAGACTACACTGTATCACGGTGG | MLoc_I_sgRNA_1 | 3 |
| 2 | CCTCAAATTCTCCTGCACTGAGG | MLoc_I_sgRNA_2 | 7 |
| 3 | AGTACATCATCATTCGCAGGGGG | MLoc_I_sgRNA_3 | 1 |
| 7 | CTGCTGGAATGCTGACGTGGAGG | MLoc_I_sgRNA_7 | 2 |
| 20 | TCGGTCTCGCGGCCAAAAAAGGG | MLoc_I_sgRNA_20 | 6 |
| 21 | TTATACATGCCGCAGTTTAGAGG | MLoc_I_sgRNA_21 | 5 |
| 23 | GACCATTACTTGTCCTATCCAGG | MLoc_I_sgRNA_23 | 10 |
| 12 | GCCCTGTACCCATTAATGGTTGG | MLoc_I_sgRNA_12 | 8 |
| 15 | GATTTGTATGTGTAGCCTCTCGG | MLoc_I_sgRNA_15 | 9 |
| 26 | GGTAAGTTATGTAGTCATGTAGG | MLoc_I_sgRNA_26 | 4 |

## 4.3.2.iii Microinjection of sgRNAs into mosquito embryos to detect *in vivo* cleavage efficiency

### *4.3.2.iii.a Injection mix components and assembly*

#### *4.3.2.iii.a.(i) Synthetic sgRNAs*

To standardize cutting and eliminate potential errors introduced during RNA synthesis and handling, commercially sourced Standard Synthetic SygRNA® (Merck, UK) sgRNAs

were used for genome editing. SygRNAs were modified with a stabilizing 2' 0 methyl & phosphothioate to increase stability toward exonucleases and increase genome editing efficiency (Hendel et al., 2015). Lyophilized SygRNAs were resuspended to a concentration of 30 µM (1500 ng/µL) in 10mM Tris containing buffer. Lyophilized sgRNAs were initially stored at 20°C, but after resuspension were stored at -80°C long term.

*4.3.2.iii.a.(ii) Recombinant SpCas9*

Recombinant Cas9 Protein (Merck, UK) from *Streptococcus pyogenes* expressed in *Escherichia coli* with an NLS was used. Lyophilized Cas9 protein was resuspended in nuclease free water to a stock concentration of 30 µM (5000 ng/µl) (glycerol was recommended but omitted here as a precaution against toxicity during microinjection). Lyophilized powder was dissolved by gently tapping on the tube, incubating on ice for 10 minutes, and spinning to bring material to the bottom of the tube.

*4.3.2.iii.a.(iii) Preparation of RNPs and injection mixes*

The Merck protocol for complexing RNPs was modified with recommended concentrations of CRISPR/Cas9 for mosquito microinjection. To generate knockout mutants, 10 ng/µl of each sgRNA (20 ng/µl in group 11), 300 ng/µl of purified NLS-SpCas9 protein and 1X injection buffer (final concentration 5mM KCl and 0.1mM NaH2PO4 at pH 6.8) was used. RNPs were first individually complexed to prevent preferential binding and limit competition in the injection mix. sgRNAs and Cas9 were complexed in a 4 µl reaction using concentrations of sgRNA and Cas9 that were calculated to allow subsequent transfer of 1 µl of each complexed RNP into a 12 µl injection mix. RNPs were incubated for 5-15 minutes at room temperature and then combined. Lastly, injection buffer was added to a final concentration of 1X in a 12 µl injection mix volume. Injection mixes were stored at -80°C until ready to inject.

*4.3.2.iii.a.(iv) RNAse free conditions*

All ribonucleoproteins (RNP) were prepared under nuclease free conditions. This included cleaning the workspace with RNaseZAP (Merck, UK), using barrier filter tips and RNAse-free plasticware, and keeping RNA on ice while handling and centrifuging.

*4.3.2.iii.b Microinjections and harvesting of embryonic gDNA*

Injection mixes were centrifuged at 14,000 x *g* and 4°C for at least 10 minutes prior to injection to minimize particulate matter that could have blocked the needle.

Approximately 120 embryos were injected for each guide group and left to rest in the temperature-controlled mosquito chambers overnight (Table 23). Eggs were harvested after 24 h by quick freezing in liquid nitrogen, after which the samples were stored at -80°C prior to gDNA extractions. gDNA was extracted using the Macherey-Nagel NucleoSpin genomic DNA extraction kit and quantified using a Nanodrop. DNA was sent by courier at room temperature to Michelle Anderson at York University for processing on the nanopore.

*Table 23 Number of eggs injected per guide group that were then harvested for Nanopore sequencing.*

| Injection group | Eggs injected |
|---|---|
| 1 | 107 |
| 2 | 109 |
| 3 | 105 |
| 4 | 105 |
| 5 | 104 |
| 6 | 122 |
| 7 | 122 |
| 8 | 123 |
| 9 | 123 |
| 10 | 107 |
| 11 | 123 |

## 4.3.2.iv Nanopore sequencing to detect CRISPR-editing events

The following part of these methods was performed by Michelle Anderson at York University and is adapted from the nanopore protocol (https://community.nanoporetech.com/docs). Nanopore sequencing was performed in two 96-sample runs. The first run was processed with *kmo*-targeted amplicons as a control and the second using the *Act4*-targeted amplicons. All sgRNAs within the same subregion amplicon were run in the same batch.

### *4.3.2.iv.a Amplicon library preparation*

#### *4.3.2.iv.a.(i) Adapter ligation PCR*

A first PCR was performed on gDNA from injected samples using specific primers tailed with universal sequences for adapter attachment. The primer sequences included the following adapter sequences (written from 5' to 3'): TTTCTGTTGGTGCTGATATTGC-[project-specific forward primer sequence] and ACTTGCCTGTCGCTCTATCTTC-[project-specific reverse primer sequence]. Amplicon specific primers are listed in Appendix B Table 6. The PCRs were performed using Q5 High-Fidelity polymerase and 100 ng of template gDNA in 50 µL reaction volumes. The thermocycler conditions were according

to the Q5 protocol, with the annealing temperature specified to 72°C. The PCR was allowed to run for 20 cycles before 37.5 µL was removed and the remainder run for an additional 20 cycles to be visualised by gel electrophoresis. The PCR products removed at cycle 20 were purified using AMPureXP beads and quantified and quality controlled using the Qubit HS assay according to the manufacturer's instructions. In total, 125 amplicons were PCRed: all 83 M locus sgRNA-targeted sample amplicons, 11 *kmo*-targeted positive control amplicons, 11 *Act4*-targeted positive control amplicons, and 20 un-injected negative control amplicons (18 M locus amplicons and 2 control amplicons).

*4.3.2.iv.a.(ii) Barcode ligation PCR*

A second PCR was then performed to incorporate the Oxford Nanopore barcode sequences into the amplicons (Appendix B Table 7). The PCR Barcoding Expansion Pack allowed barcoding of libraries with one of 96 PCR barcodes (BC1-BC96) such that up to 96 different libraries could be combined and loaded into a flow cell at the same time. PCRs were set up according to the Nanopore protocol using LongAmp Taq polymerase (New England Biolabs, UK), 100-200 fmol of each DNA sample, and 50 µL reaction volumes. Primers from the Barcode Expansion Pack were transferred to corresponding wells in a PCR plate. Wells were then mixed by pipetting, sealed with adhesive film, and briefly spun down in a plate spinner. Thermocycler conditions were set up according to the nanopore protocol. PCR products were then purified using AMPureXP beads as before and quantified using Qubit. Purified and barcoded libraries were diluted to 1nM and then pooled using 10ul of each sample in a 1.5 mL LoBind Eppendorf tube. The concentration of the pool was measured again on Qubit and then 1 µg was removed and diluted in 47 µL of nuclease-free water.

*4.3.2.iv.a.(iii) End repair and nanopore adapter ligation*

The pooled library was end-repaired and adapted for nanopore sequencing using NEBNext FFPE DNA Repair Mix and NEBNext Ultra II RNA Library Prep Kit for Illumina (New England Biolabs, UK) according to the manufacturer's instructions and the Nanopore protocol. Repaired and end-prepped pooled libraries were purified with AMPureXP beads and quantified on Qubit. DNA was then taken forward into the adapter ligation step, which was performed using NEBNext Quick Ligation Module and AMPureXP beads according to the manufacturer's instructions and Nanopore protocol. The final prepared library was quantified with Qubit.

*4.3.2.iv.b MinION sequencing and data analysis*

*4.3.2.iv.b.(i) Flow cell priming and loading of prepared library*

The flow cell was first primed according to the Nanopore protocol. The library was then prepared for loading by combining 3ul of 11nM (33 fmol) prepped library with Sequencing Buffer in a 75 µL volume. A higher amount of prepped library was used compared to the recommended 5-10 fmol because it was suggested that the newer versions of the flow cells perform better at higher ranges. Finally, the library was loaded into the MinION (MK1C) flow cell.

*4.3.2.iv.b.(ii) Basecalling and demultiplexing of raw reads*

The sequencing run and data conversion was performed by Sally James in the Genomics Laboratory at York University. Raw reads were converted to basecalled reads and demultiplexed using the MinKNOW software and finally shared with me in FASTQ format for downstream analysis.

## 4.3.2.v Analysis of amplicon sequencing data using CRISPResso2

CRISPResso2 was used to analyse the reads from the genome edited samples (Clement *et al.*, 2019). CRISPResso2 produces data on the frequency and efficiency of genome editing events by aligning reads to a reference amplicon sequence and quantifying mutations around a specified sgRNA site to determine whether a read is modified or unmodified by genome editing. CRISPResso2 accepts one sequence per sample, so the multiple FASTQ reads within each processed barcode folder were first concatenated into one FASTQ file per barcode and output into a single directory. CRISPResso2 was then run on each concatenated barcode read using the corresponding amplicon sequence and sgRNA site. CRISPResso2 was also run on the un-injected barcodes at each sgRNA site within the amplicon to get a baseline variation rate for each sgRNA site. This baseline value was then subtracted from the mutation frequency calculated for each M locus site. The script was run on the York University computing cluster.

### 4.3.3 Transgenesis of male mosquitoes with a SuperSIT construct at selected sgRNA target sites

#### 4.3.3.i Method overview

Transgenesis of male mosquitoes was performed by CRISPR-mediated knock-in at the selected M locus sgRNA sites. Constructs containing a *β2 tubulin* (*β2t,* AAEL019894)-driven CRISPR-Cas9 construct targeting *myo-fem* and site-specific homology arms to facilitate HDR were designed and cloned. Mosquito embryos were microinjected with CRISPR-Cas9 targeting a single selected sgRNA target site and its corresponding HDR donor plasmid. Surviving males were outcrossed and their offspring screened for transgenic males, which were outcrossed again and screened for male-biased inheritance and female-specific inability to fly.

#### 4.3.3.ii Design and synthesis of SuperSIT construct for HDR-based transgenesis

##### *4.3.3.ii.a Design of common backbone for HDR donor plasmids*

A common backbone plasmid for all M locus HDR plasmids was designed with the help of Michelle Anderson for synthesis (PL1133 in Figure 24). The backbone plasmid was made up of three expression cassettes: a Cas9 cassette, driven by germline promoter *β2t* (*β2t*-Cas9-SV40); the sgRNA cassette targeting *myo-fem* exon 2 tested in Chapter 3; and a transgenic marker cassette, driven by the Poly-Ubiquitin promoter (PUb-mCherry-K10). A construct containing cassettes *β2t*-Cas9-SV40 and PUb-mCherry-K10 was commercially synthesized and cloned into a pUC-GW-Amp backbone at the EcoRV restriction site (PL-1132) (Genewiz/Azenta); the pUC-GW-AMP plasmid backbone was shortened by 637 bp (404 bp on the 5' end and 233 bp on the 3' end) to increase the stability of the synthesized plasmid (a previous iteration of the plasmid also included a *proA* reporter construct, but this plasmid could not be synthesized due to issues with cloning). Plasmid 1132 was later modified in the lab to include an sgRNA cassette targeting *myo-fem* (see below) (PL-1133).

To reduce the number of downstream cloning steps required for each site, 5' and 3' homology arm sequences were designed to be inserted back-to-back, with the 3' HA preceding the 5' HA, at one cloning site in the plasmid. Theoretically, this would mean that

the entire plasmid would be inserted at each sgRNA site during HDR. An XhoI cloning site was inserted before the pUC backbone, such that upon integration, the construct would read pUC-GW-AMP, *β2t*-Cas9-SV40, U6-sgRNAs, and pUB-mCherry-K10. The plasmid also contained a loxP and attP site to allow site-specific integration via Cre-Lox or PhiC31, respectively. PhiC31 integrase recognizes attP sites and allows integration of DNA constructs carrying a compatible attB site (Thorpe and Smith, 1998). Additional restriction sites were added between each element of the cassettes, such that each part could easily be modified. The Cas9 was flanked by two NLS signals and had a 3X Flag tag for tagging. In total, plasmid 1133 was 16782 bp long.

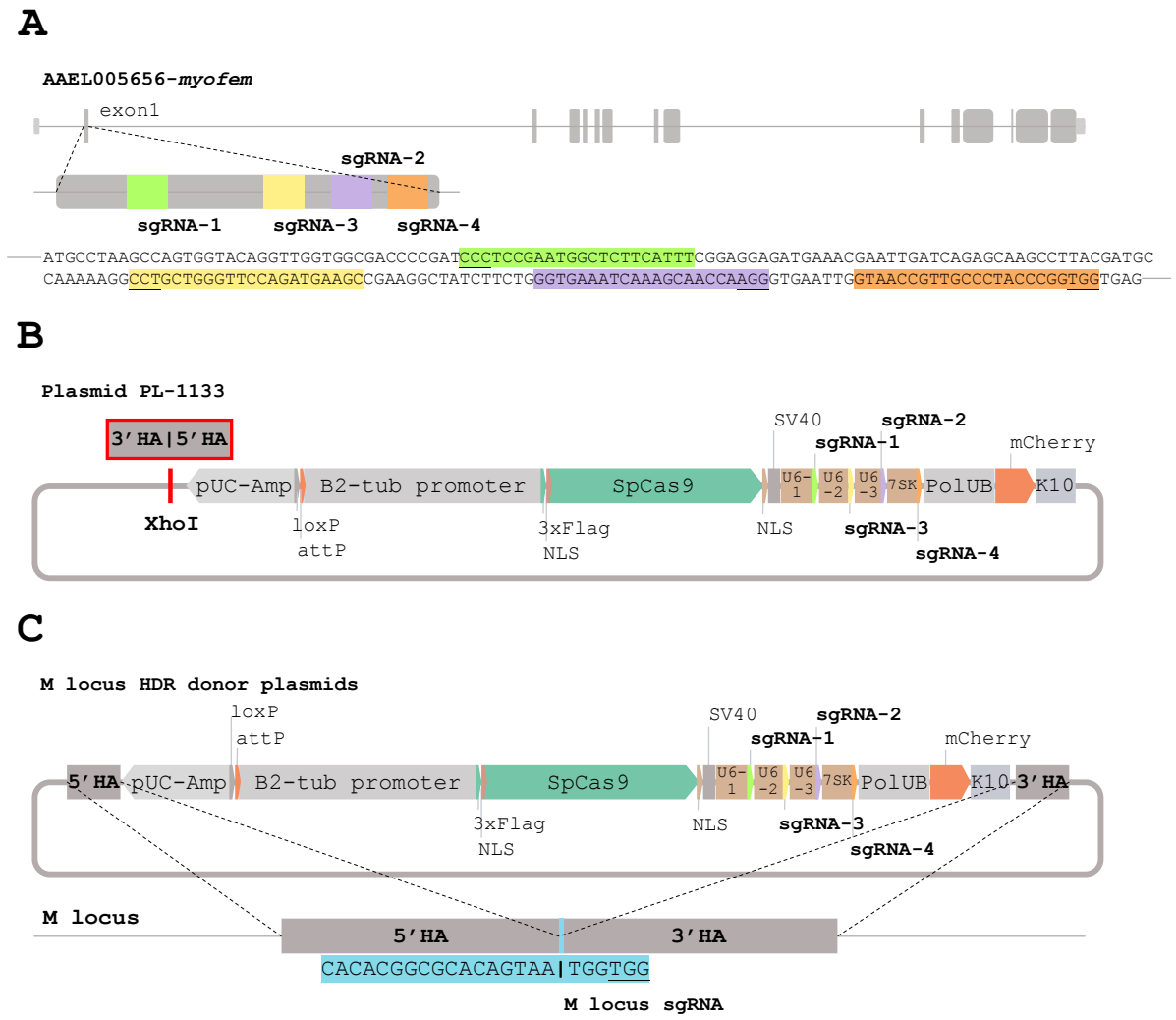



*Figure 24 A) Location of four sgRNA targets within exon 1 of the myo-fem gene (AAEL005656) in* Ae. aegypti, *disruption of which caused a 95% dominant loss of flight in females in preceding chapters. B) Schematic of plasmid PL-1133, the common backbone plasmid for all M locus HDR donor plasmids for the SuperSIT system. The backbone plasmid was made up of three expression cassettes: a Cas9 cassette, driven by germline promoter β2t (β2t-Cas9-SV40); the sgRNA cassette targeting myo-fem exon 1 tested in preceding chapters, and a transgenic marker cassette, driven by the Poly-Ubiquitin promoter (PUb-mCherry-K10). To reduce the number of downstream cloning steps required for each site, 5' and 3'*

*homology arm (HA) sequences were designed to be inserted back-to-back, with the 3' HA preceding the 5' HA, at the XhoI cloning site in the plasmid, marked in red. Gene cassettes are drawn to scale. Homology arms are 1000 bp each. C) Schematic of HDR donor plasmids with 5' and 3' HAs cloned in. Here, the 3' HA is shown at the end of the SuperSIT construct to emphasize that the entire intervening plasmid between the 5' and 3' HAs would be incorporated at the M locus target site, but the 3'HA is actually adjacent to the 5'HA. An example M locus sgRNA sequence is highlighted in blue to show that the HA contain the sgRNA sequence up to the cut site.*

#### *4.3.3.ii.a.(i) Integration of sgRNA cassette*

The sgRNA cassette targeting *myo-fem* was cloned into the Genewiz-synthesized backbone plasmid, PL1132, using NEBuilder HiFi DNA Assembly (NEB, UK). Plasmid 1132 was linearized overnight with RE BsaI-HFv2 (NEB, UK) using 3.5 µg of plasmid DNA, 3.5 µL of RE, 1 X rCS buffer, and nuclease-free water in a 60 µL reaction. The U6-sgRNA insert was amplified from plasmid AGG2199 using Q5 polymerase. HiFi primers were designed to create a minimum overlap of 30 bp between fragments and integrate RE sites AgeI and NotI on either side of cassette (Appendix B Table 8). Fragments were assembled in a 1 to 2 vector to insert molar ratio using a total of 0.03 pmols of fragments and incubated at 50°C for 1 hr. 2 µL of the HiFi assembly was then transformed in SURE 2 supercompetent cells according to the manufacturer's protocol (Agilent Technologies) and plated on LB-Ampicillin plates. Colonies were screened using primers in Appendix B Table 8 and DreamTaq polymerase (ThermoFischer, UK). Positive colonies were mini prepped using the NucleoSpin Plasmid Mini kit for plasmid DNA (Macherey-Nagel, UK), sequenced across both boundaries and the entire insert with primers in Appendix B Table 8. Two sequence-verified plasmids were midi prepped using NucleoBond Xtra Midi kit for endotoxin-free (EF) plasmid DNA (Macherey-Nagel, UK), eluted in EF water, and sequenced again using the same primers.

#### *4.3.3.ii.b Homology arm design and cloning of HDR donor plasmids*

Homology arms for HDR integration at the selected target sites were designed to be 1000 bp long on either end of the sgRNA cut site (except for the 3' HA of *Nix* intron 1A-68, which was 800 bp long). Homology arms were inserted at the XhoI site in backbone plasmid 1133, such that the 3' HA immediately preceded the 5' HA, both in the sense orientation. Another set of plasmids were designed that included the corresponding sgRNA site (including the PAM) between the homology arms so that the plasmid could be linearized during microinjection. These plasmids were cloned in case the method of back-to-back homology arms did not work without such linearisation. Some of the plasmids were cloned in the lab and the rest were commercially synthesized by TWIST Bioscience (USA)

(Table 24). The plasmids cloned by HiFi and by TWIST differed by 4bp due to TWIST employing a homology-based cloning method that preserved the RE site in contrast to the HiFi method.

*Table 24 List of HDR donor plasmids cloned in the lab or synthesized by TWIST*

| Plasmid no. | Subregion and sgRNA per plasmid | Cloned by |
|---|---|---|
| 1156 | Myosex downstream 2-20 | HiFi |
| 1157 | Myosex downstream 2-20-sgRNA | HiFi |
| 1136 | Myosex downstream 2B-6 | HiFi |
| 1158 | Myosex downstream 2B-sgRNA | HiFi |
| 1161 | Myosex downstream 7-85 | TWIST |
| 1164 | Myosex intron 3B-29 | TWIST |
| 1155 | Myosex intron 4B-125 | HiFi |
| 1160 | Myosex intron 4B-125-sgRNA | HiFi |
| 1134 | Myosex intron 4C-19 | HiFi |
| 1139 | Myosex intron 4C-19-sgRNA | HiFi |
| 1137 | Nix downstream 1-20 | HiFi |
| 1138 | Nix downstream 1-20-sgRNA | HiFi |
| 1165 | Nix downstream 2-39 | TWIST |
| 1166 | Nix downstream 3-44 | TWIST |
| 1162 | Nix intron 1A-68* | TWIST |
| | Nix intron 1A-68-sgRNA* | TWIST |
| 1167 | Nix intron 4-8 | TWIST |
| 1163 | Nix intron 5-102 | TWIST |
| | Nix intron 5-102-sgRNA | TWIST |
| 1168 | Nix upstream 2B-12 | HiFi |

*4.3.3.ii.b.(i) Final sequence verification of missing data for homology arms*

If the entire sequence 1000bp upstream and downstream of each selected sgRNA had not been sequenced (either during initial sequence verification of sgRNAs or nanopore sequencing), they were sequenced at this point (Appendix B Table 9 and Appendix B Table 10). The 5' homology arm of Nix Intron 3B sgRNA 43 could not be sequenced due to tandem repeats 800bp from the chosen sgRNA site and was thus excluded from downstream cloning.

*4.3.3.ii.b.(ii) Cloning of HDR donor plasmids by HiFi Assembly*

Plasmid 1133 was first linearized overnight with RE XhoI (NEB, UK) as above. The 5' and 3' HA were individually amplified from single male gDNA using Q5 polymerase and HiFi primers that were designed as above (Appendix C Table 11). Fragments were assembled in a 1 to 3 vector to insert molar ratio using a total of 0.06 pmols of fragments and incubated at 50°C for 1 hr. Transformation of SURE 2 cells was performed as before, and clones verified using primers in Appendix C Table 12. Midiprepped plasmids were eluted in EF water as before.

*4.3.3.ii.b.(iii) Synthesis of HDR donor plasmids at TWIST Bioscience*

Plasmid 1133 was shipped to TWIST Bioscience and onboarded there. A HA fragment consisting of the 3' HA followed by the 5' HA was synthesized for each HDR target site and cloned into 1133 at the XhoI site. Plasmids were ordered at midiprep scale in endotoxin free water. The concentrations of the midipreps were too low for the injection mixes, however, so the midipreps were concentrated in the lab. DNA was precipitated by adding 1/10 volume of 3 M Sodium Acetate pH 5.2 (Merck, UK) and 2.5 volumes of ice-cold 100% ethanol to the plasmid samples. The samples were mixed and stored overnight at -20°C before centrifuging at full speed in a microcentrifuge for 20 mins. The pellets were washed twice with 1 mL of room-temperature 70% ethanol, vortexed, and spun at 15,000 x g for 5 min. The pellet was allowed to air-dry and was then resuspended in EF water.

### 4.3.3.iii Microinjection of mosquito embryos for transgenesis

Injection mixes for transgenesis via HDR consisted of 100 ng/µl of sgRNA (Merck, UK), 300 ng/µl of recombinant Cas9 protein (Merck, UK), 700 ng/µl of HDR donor plasmid, and 1X Injection Buffer in 10 µL volumes. The Cas9 and sgRNA were pre-complexed at room-temperature for 5-15 mins in a 4 µL volume using 3000 ng of Cas9 and 1000 ng of sgRNA and RNAse free water. Injection buffer was added to 1X, followed by HDR donor plasmid, and EF water. Approximately 300 eggs were injected per construct. Microinjected eggs were incubated on wet cotton wool in the temperature-controlled mosquito chambers for five days before hatching.

### 4.3.3.iv Establishing SuperSIT lines

Eggs were hatched as usual in a solution of 3 drops of Liquify No1 (Interpet, UK) per 200 mL of deionized water in a vacuum hatcher. Surviving males were crossed in groups of fifteen to LVP using a ratio of 1 male to 5 females, while females were discarded since they lack the M locus. The $G_1$ progeny of each pool was screened for transgenics. Positive $G_1$ males were outcrossed again (females were discarded), and their progeny screened for male-biased inheritance and inability to fly in females. Correct integration of HDR donor plasmids was confirmed by PCR and Sanger sequencing of gDNA from transgenic $G_1$ males using the primers in Appendix B Table 13. gDNA was extracted using the MagMAX kit and KingFisher instrument.

### 4.3.4 Assessment of transgenic SuperSIT lines

#### 4.3.4.i Flight assessment

Flight ability was assessed using the flight test device described in Chapter 3 (see 3.3.5). The proportion of flightless individuals per tube was determined by dividing the number of individuals remaining in the tube at the end of the experiment by the total number of females added into it at the start of the experiment. The mean flightless rate was calculated using the total number of tubes that were used for the experiment.

A conservative estimate was then calculated by removing all females remaining in each tube at the end of the experiment and assessing flight ability in a cage. Flight ability was visually assessed by tapping the cage and gently agitating the adult females with a paintbrush. Individuals that could not fly in response to the trigger were deemed flightless. The proportion of flightless individuals was determined by dividing the number of flightless individuals by the total number of females that were put into each tube of the flight test device at the start of the experiment. This is a much more stringent–conservative–test. In the standard test, females that fail to reach the top chamber are considered non-flying (then normalised for the performance of wild type). In the conservative version, females must fail to leave the bottom chamber (or, more precisely, be found in it at the end of the test) and then additionally fail to fly when individually stimulated to do so.

#### 4.3.4.ii Investigation of position effects on SuperSIT lines

Transgenic SuperSIT lines were provided with external sources of Cas9 and sgRNAs to investigate the influence of position effects on the expression of the SuperSIT CRISPR-Cas9 construct components. SuperSIT lines were bred with lines expressing germline promoter *shutdown*-driven Cas9 (*shu*-Cas9) and lines expressing U6-driven sgRNAs targeting *kmo* and *myo-fem*, respectively, to determine whether *β2t*-driven Cas9 expression, U6-driven sgRNA expression, or both, was limiting CRISPR activity. SuperSIT lines (PL-1134B1, PL-1136D3, and PL-1137A2) were crossed to a combination of transgenic lines expressing *shu*-Cas9 (AGG-1551B1) and two lines expressing *U6*-sgRNAs targeting *kmo* (AGG-1095) and *myo-fem* (AGG-2199C1), respectively. Previous embryonic assays had showed strong expression of Cas9 and sgRNA from these lines.

#### *4.3.4.ii.a Construct activity with external source of Cas9*

Male SuperSIT lines were crossed in groups to females from transgenic line AGG-1551B1 that expressed *shu*-Cas9 from an *m*-linked site. $F_0$ crosses were performed in a 1 to 2 male to female ratio and their $F_1$ progeny then screened for transgene inheritance under a fluorescence microscope. The transgenic marker in the SuperSIT lines and the *shu*-Cas9 line was PUb-mCherry in both, so $F_1$ males were genotyped based on the brightness of mCherry expression; trans-heterozygote males that carried both the SuperSIT and *shu*-Cas9 constructs were brighter, while single-heterozygote males that carried only the SuperSIT construct were discernibly dimmer in comparison. Double- and single-het $F_1$ males were then crossed separately to LVP in groups using the same ratio as before. Crosses using single-het males were performed to control for possible maternal deposition from *shu*-Cas9 $F_0$ mothers. $F_2$ progeny was then screened for transgene inheritance and females scored for their ability to fly. Flight was assessed in the flight test device using previously described protocols.

#### *4.3.4.ii.b Construct activity with external source of sgRNA*

Male SuperSIT lines were crossed in groups to females from transgenic line AGG-1095N1, which expressed sgRNAs targeting *kmo* from a knock-in integration in *kmo*, and AGG-2199C1, which expressed sgRNAs targeting *myo-fem* from a random integration site. $F_0$ crosses were performed and $F_1$ progeny screened for transgene inheritance as above (the transgenic markers for the sgRNA-expressing lines were both 3XP3-AmCyan, so inheritance of sgRNAs was easily differentiated from that of the SuperSIT construct). $F_1$ progeny from the *kmo*-sgRNA cross was also scored for eye mosaicism to test for paternal deposition of Cas9. $F_1$ females from the *myo-fem*-sgRNA cross were not tested for a paternal deposition effect because it was unclear what an incomplete flight phenotype (due to incomplete cutting of *myo-fem* from deposition) looked like; it may be that females can tolerate some degree of mosaicism if wild-type muscle cells, or wild-type nuclei in syncytial muscle cells, produce enough *myo-fem* protein to minimize the observed effect (Navarro-Payá *et al.*, 2020). $F_1$ transheterozygous males were crossed to LVP females and females that were homozygous for a knock-out (KO) mutation of *kmo* (*kmo*-KO). $F_2$ progeny from both crosses were scored for transgene inheritance and eye mosaicism. Finally, $F_2$ female progeny were assessed for flight using the flight test device to test for saturation of Cas9 with the presence of extra sgRNAs.

## 4.4 Results and Discussion

### 4.4.1 *In silico* analysis of M locus and selection of sgRNAs for cleavage efficiency test

The M locus was first analysed *in silico* for potential integration sites that would allow efficient expression of the SuperSIT construct and not impact fitness. Several criteria were assessed, including relative gene density, which was reasoned to positively impact transgene expression and be less vulnerable to silencing by heterochromatin, and general genetic complexity, which could improve the chances of knock-in by homology-directed repair (HDR). Additional data on chromatin accessibility was not available as a genome track for the updated AaegL5 genome assembly in VectorBase, so this was omitted from the initial analysis.

Using the AaegL5 genome assembly and data available from the bioinformatics research database VectorBase, seven regions were highlighted (Figure 25) (Matthews *et al.*, 2018; Giraldo-Calderón *et al.*, 2022). These regions were between the start of *myo-sex* (+ strand) to 50 kb upstream of *nix* (- strand) and made up approximately 710 kB of sequence, greater than half of the available sequencing data for the M locus (1.27 Mb). sgRNAs were designed to target sites with maximal on-target activity (sgRNA efficiency) and minimal cleavage in off-target genomic locations (sgRNA specificity) and 202 were selected for sequence verification (Labun *et al.*, 2019; Doench *et al.*, 2016).

#### 4.4.1.i Sequence validation of target sites

Sequence variation between the reference genome and lab strain is an important factor in determining the success of knock-in transgenesis. First, sequence variation around the sgRNA target site can block initial cleavage (Ang *et al.*, 2022; Sun *et al.*, 2022). Second, sequence divergence of more than 1.5% between the donor template and genomic sequence surrounding the sgRNA target site greatly reduces homology-directed repair (HDR) efficiency in mice, *D. melanogaster*, and *Ae. aegypti* (Do *et al.*, 2014; Elliott and Jasin, 2001; Elliott *et al.*, 1998; Ang *et al.*, 2022). Lastly, while the genomic assembly of the M locus is much improved, due to its highly repetitive nature, it was hypothesized that significant sequence divergence between the lab strain and reference may exist. Thus, sgRNA target sites were first validated by sequencing before proceeding with downstream steps.

The 202 sgRNAs were sorted by location into thirty-two smaller subregions (ranging from 1.7-5 kB) that were named, amplified, and sequenced. 23 of the 32 subregions produced a single amplicon at the expected size for all male individuals using the designed primer sets. Sequencing these amplicons showed that out of the 161 sgRNAs found in the 23 subregion amplicons, 134 from 18 subregions had the expected sequence (highlighted in blue in Figure 25). Together these results show that sequencing was an essential step.

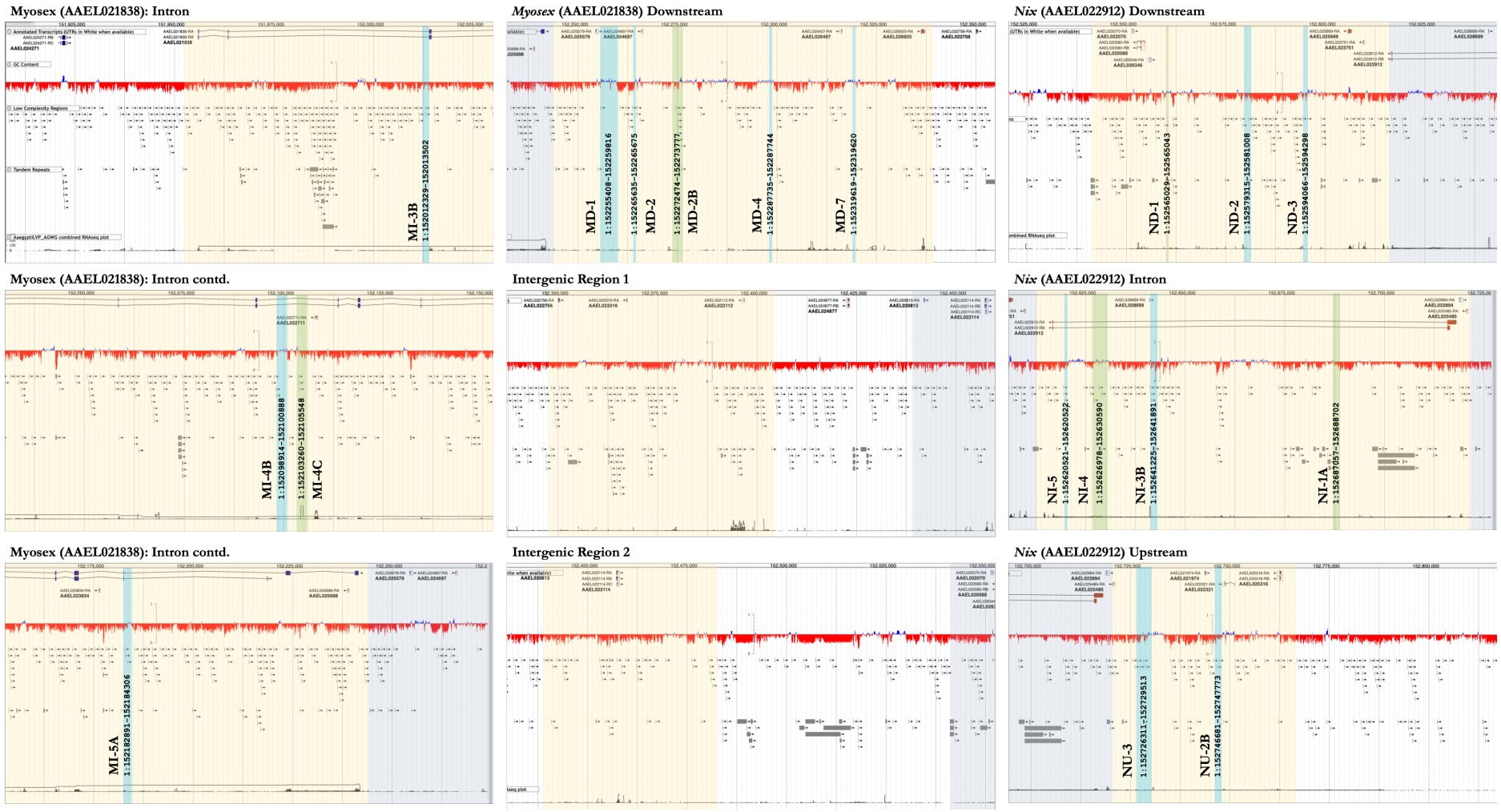



*Figure 25 The M locus was first visualised in the VectorBase genome browser JBrowse using the AaegL5 genome assembly of* Ae. aegypti*. Using available genome tracks, regions with a higher gene density, genomic complexity, and GC content, and a lower number of tandem repeats were highlighted for potential targeting. Screenshots are shown from the JBrowse viewer of the seven regions in the M locus for which sequences were retrieved and queried for sgRNAs in CHOPCHOP (regions are highlighted in yellow; grey highlighted regions indicate overlaps in adjacent images going from top to bottom, left to right). The seven regions in order of their location on the M locus were named:* myosex *intron,* myosex *downstream, intergenic region 1, intergenic region 2,* Nix *downstream,* Nix *intron, and* Nix *upstream; their genomic coordinates are found in Table 19. 18 smaller subregions are highlighted with their given names and genomic coordinates. Subregions in green indicate approximate knock-in integration sites that were achieved for the SuperSIT construct.*

### 4.4.1.ii *In vivo* cleavage efficiency of sgRNAs targeting the M locus

Predicted sgRNA efficiencies can vary considerably from measured ones, so testing sgRNA efficiencies *in vivo* is often recommended before large-scale mutagenesis experiments (Kistler, Vosshall and Matthews, 2015; Konstantakos *et al.*, 2022; Kotwica-Rolinska *et al.*, 2019; Labuhn *et al.*, 2018). Moreover, because chromatin accessibility has been found to correlate with efficiency of CRISPR-Cas9 mutagenesis, it was hypothesised that sgRNAs targeting the heterochromatic M locus may not cut as efficiently as expected (Uusi-Mäkelä et al., 2018). To test cleavage efficiency of sgRNAs in vivo, sgRNAs and Cas9 were injected into mosquito eggs that were then harvested for mutation detection by Oxford Nanopore sequencing of the target loci. CRISPResso2 was run on the long amplicon reads to determine the genome editing efficiency around each sgRNA site. The frequency of modified alleles was then used as a proxy for sgRNA cleavage efficiency (Clement *et al.*, 2019). Of the 134 sequence-verified sgRNAs, 83 were selected to test cleavage efficiency in mosquito embryos.

### 4.4.1.iii Control sgRNAs demonstrated the efficiency assessment was performed to a high degree of precision

sgRNA activity was screened in a high-throughput method: genomically distant sgRNAs were injected in groups of eight (or three for group 11) with two positive control sgRNAs included to measure variation across the injection groups. The two positive control sgRNAs targeting *kmo* and *Act4* showed low variability in their activity across the ten injection groups and were found to be normally distributed (Shapiro-Wilk test and data visualization in a QQ plot). The genome editing efficiency of the kmo sgRNA was about twice that of the Act4 sgRNA across all experiments (kmo-499 mean = 19.5 ± SE 1.5, Act4-6640 mean = 9.5 ± SE 0.8, ratio kmo-499: Act4-6640 = 2.1 ± SE 0.2, n = 10). Further, double the amount of RNPs were used for group 11 and the modification frequencies of the two control sgRNAs was about twice that of the means from groups 1-10 (ratio of group 11 to the mean in group 1-10 for *kmo-499* = 2.0 and *Act4-6640* = 2.6, ratio of *kmo-499*: *Act4-6640* in group 11 = 1.6). Lastly, the efficiencies of the positive control guides did not significantly correlate with injection group (Pearson correlation coefficient = 0.58, p-value = 0.08), suggesting similar amounts of injection mix were injected into all embryos. Together, these analyses showed the microinjection experiments were performed to a high degree of precision, allowing M locus sgRNA activities to be compared across groups.

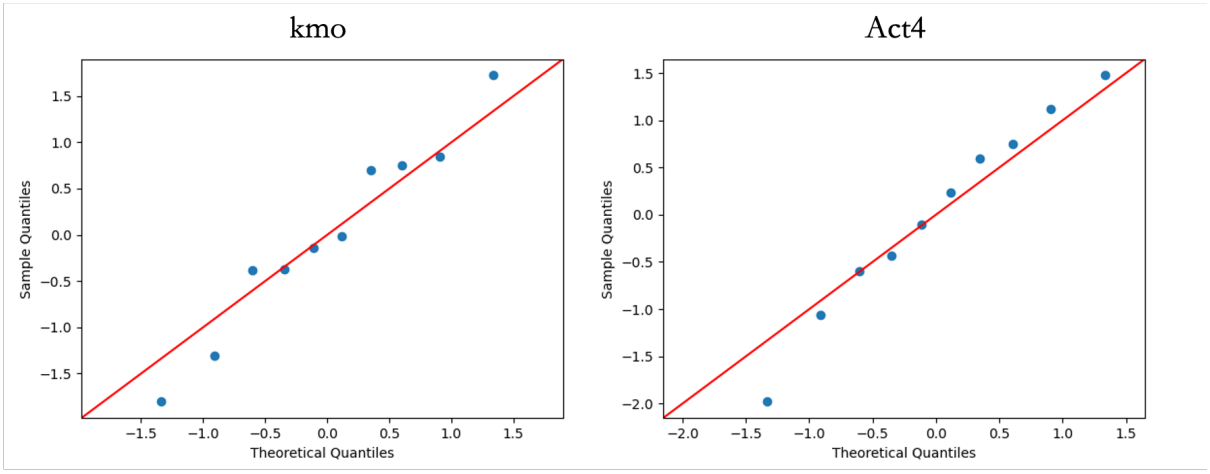



*Figure 26 QQ plots of the sgRNA efficiencies for* kmo *and* Act4 *across ten injections groups. Injection group 11 was omitted because a higher concentration of sgRNAs were used in that injection mix.*

### 4.4.1.iv M locus sgRNAs achieved high cleavage efficiency

The M locus sgRNAs achieved a high genome editing efficiency, up to 85.5% in some cases, which meant multiple sgRNAs could be chosen between for transgene knock-in (range 0-85.5%, mean = 34.4% ± SD 21.9) (Figure 27). 84% of tested sgRNAs showed a higher activity than the minimum recorded frequency for *kmo-499* (11.3%), an sgRNA that has been used to achieve a CRISPR-Cas9 mediated knock-in integration of a transgene. This suggests that these frequencies would be suitable for knock-in into the M locus.

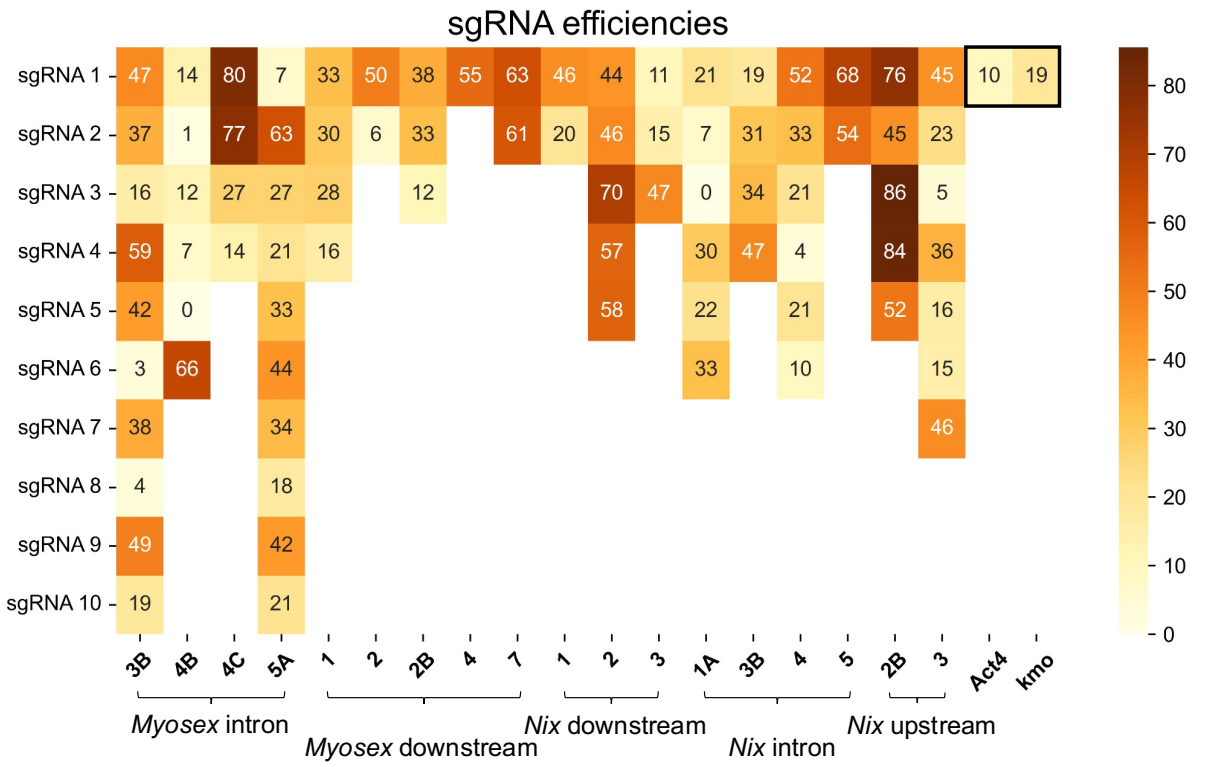



*Figure 27 In vivo sgRNA cleavage efficiencies of 82 sgRNAs targeting the M locus. sgRNAs and Cas9 were microinjected into mosquito eggs that were harvested 24 hours later for mutation detection by Oxford Nanopore sequencing of the target loci. sgRNA cleavage efficiency was estimated according to the frequency of mutated target loci minus the baseline variation rate of each sgRNA site in the un-injected control, determined by CRISPResso2. Cell values in the heatmap represent the genome editing efficiency of each sgRNA; 82 values are depicted (one sgRNA from myo-sex downstream 1, sgRNA 27, was not determined due to memory problems with CRISPResso2). The means of the Act4 and kmo control sgRNAs from the first 10 injection groups are shown on the right of the heatmap and are outlined in black. sgRNA numbers on the y axis each correspond to a specific sgRNA in each subregion; these are listed in Appendix B Table 14 for reference.*

Due to an error, the final concentration of each sgRNA was 10 ng/µl for groups 1-10 and 20 ng/µl for group 11 instead of the recommended 100 ng/µl concentration, but it was reasoned that because the ribonucleoproteins (RNPs) were pre-complexed in a separate step prior to combining them in the injection mix, that enough RNPs would have formed to continue with the injections. Additionally, the Merck protocol for complexing the synthetic SygRNAs recommends a 3: 1 molar ratio of sgRNA to Cas9, meaning the final concentration of each sgRNA would have been 18 ng/µl, not far from the concentration used. Indeed, the modification frequency was generally high across the M locus sgRNAs, so the lower concentration of sgRNAs was still acceptable. Further, doubling the concentration of sgRNAs in group 11 doubled the modification frequency, suggesting that injecting a single sgRNA at 100 ng/µl (the concentration that would be used for knock-in

injection mixes) would cut at much higher rates, close to completion for most of the sgRNAs.

### 4.4.1.v Conclusion

This high-throughput method for testing sgRNA libraries was an effective and useful step prior to attempting transgene knock-in at the respective loci in the M locus. Oxford Nanopore long-read sequencing facilitated this high-throughput method because amplicon libraries of sgRNAs that were relatively far apart (up to 4.4 kB here, but longer amplicons were possible) could be prepared using the same primers, reducing workload. Further, the use of commercially synthesized sgRNAs rather than sgRNAs *in vitro* transcribed in the lab standardized the experiment and likely even buffered against using a lower-than recommended concentration of sgRNAs. Due to the high efficiency of many of the sgRNAs used, the experiment could be further optimized to pool a greater number of sgRNAs per injection mix, such that up to 20-30 sgRNAs could be screened per injection.

Lastly, the CHOPCHOP predicted efficiencies for all injected sgRNAs did not correlate well with measured activity (Pearson correlation coefficient = 0.13, p-value = 0.24, R-squared = 0.018, Slope = 0.29) (Figure 28). Only sgRNAs with a predicted efficiency greater than 33% were chosen for testing *in vivo*, however, so conclusions cannot be made about the predictive capability of CHOPCHOP, since many of the sgRNAs were still active. One can conclude, though, that not all species- and experiment-specific variables are accounted for in the rules-based prediction approach used by CHOPCHOP and that verification of sgRNA activity remains an important step for CRISPR mutagenesis in *Ae. aegypti* and other mosquito species. Deep-learning tools could be trained on data sets such as this one to improve sgRNA activity predictions in future (Ham *et al.*, 2023; Konstantakos *et al.*, 2022).

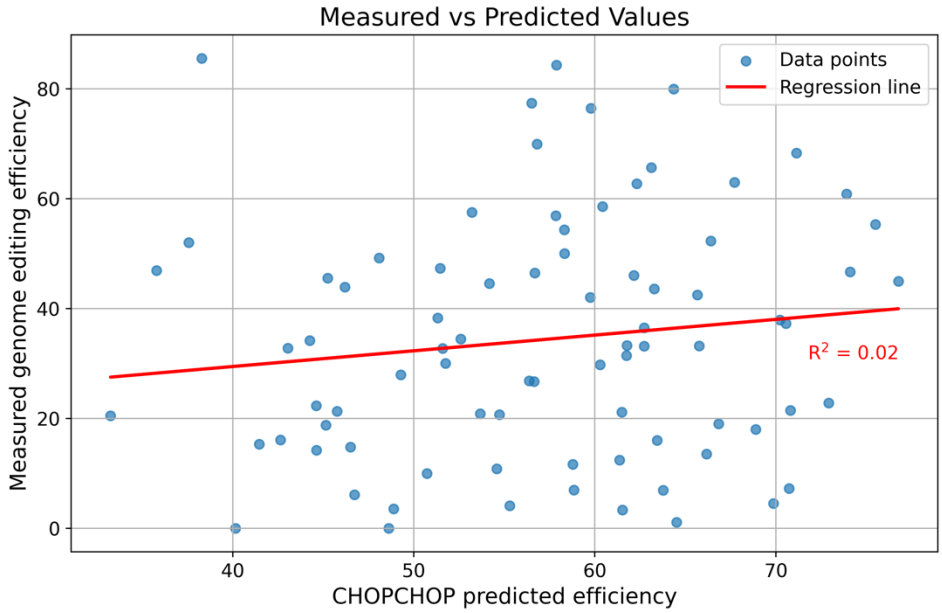



*Figure 28 Scatter plot showing the relationship between the CHOPCHOP predicted efficiency and the measured genome editing efficiency in terms of the percent of alleles modified. A linear regression model was fitted to the data points (red line); the coefficient of determination ($R^2$) for the model was 0.02, indicating a weak association between predicted and measured efficiencies.*

### 4.4.2 Transgenesis of male mosquitoes with a SuperSIT construct at selected sgRNA target sites

The first two steps towards successful knock-in of a SuperSIT construct in the M locus of *Ae. aegypti* by CRISPR-Cas9 mediated transgenesis, namely validation of target site sequence and sgRNA activity *in vivo*, produced a selection of sgRNAs that could be used for knock-in targeting. The aim of this next section was to integrate a SuperSIT construct expressing a CRISPR-Cas9 system targeting the female-specific flight gene *myo-fem* at selected sgRNA sites. Successful integration and expression of the system in the M locus would mean that only sons inherit the construct and daughters are unable to fly.

The SuperSIT CRISPR-Cas9 system was designed to include Cas9 driven by the germline promoter *β2t* and an sgRNA cassette targeting *myo-fem*. *β2t* was initially chosen to drive Cas9 expression because it induced a high degree of germline cutting and was robust to position effects in the embryonic assays discussed in Chapter 2. Other promoters could have been used to drive Cas9 expression instead, such as *shu* or *proA,* the former of which has been shown to induce the highest germline cleavage activity to date. SuperSIT lines

with these or other promoters could be tested in future, but as an initial assessment and proof of principle, it was reasoned that use of the *β2t* promoter would still be informative.

The sgRNA cassette incorporated into the SuperSIT construct targeted *myo-fem* at exon 2 and was shown in preceding chapters to induce dominant flightless mutations 95% of the time. If all female offspring inherited a mutated copy of *myo-fem*, this would come close to satisfying the second criteria of the SuperSIT system: that the fitness of all daughters of a SuperSIT male is disrupted. The best performing piggyBac integration (line 2199C1) disrupted flight in 73% of females when paired with a transgenic line expressing a *shu*-Cas9 construct from an M-linked site. However, strong position effects were observed in other *piggyBac* integrations of the same construct and similar constructs targeting *myo-fem* at exon 2 (97% dominant), where female flight was disrupted in 3-43% of females, suggesting this could be a risk upon integration in the M.

It was thus reasoned that multiple target sites within the M locus would need to be tested in anticipation of potential fitness costs related to disruption of the M locus and position effects that could impact transgene expression. Thirteen out of the 18 original subregions within the M locus were prioritized for knock-in of the SuperSIT system, each at the most active sgRNA target site for the region (Table 26). Relevant features of the subregions targeted are detailed in Table 25.

*Table 25 Relevant features of the 18 subregions within the M locus. Thirteen subregions (highlighted in yellow and marked with an "X" in the Chosen column) were prioritized for knock-in of the SuperSIT system. ? = unknown, O-/-X = subjective scale of complexity.*

| Region | Subregion | Description | <1 kB downstream of 3' coding region | <5kB upstream of 5' coding region | RNA-seq signal in amplicon | Intra-genic | Inter-genic | Complex-ity | Repeats | Chosen |
|---|---|---|---|---|---|---|---|---|---|---|
| **Myosex intron** | **Myosex intron 3B** | Intron 2 | | | | X | | / | | X |
| | **Myosex intron 4B** | Intron 5 | | | | X | | X | | X |
| | **Myosex intron 4C** | Intron 5 and 4kb downstream of AAEL022711 ncRNA | | | | X | | X | | X |
| | **Myosex intron 5A** | Intron 11 | | | | X | | / | | |
| **Myosex downstream** | **Myosex downstream 1** | Immediately downstream of AAEL025579 ncRNA | X | | | | X | X | sgRNAs flanked by 2 x 163 bp 400 bp apart and 2 x 103 bp 4kb apart | |
| | **Myosex downstream 2** | Immediately downstream of AAEL024697 ncRNA gene | X | | | | X | X | | X |
| | **Myosex downstream 2b** | RNA-seq signal in last 1kb of amplicon | ? | ? | X | | X | / | | X |
| | **Myosex downstream 4** | RNA-seq signal in last 1kb of amplicon | ? | ? | X | | X | X | sgRNA flanked by two 404 bp reverse complement repeats 12kb apart | |
| | **Myosex downstream 7** | 3 kb upstream of AAEL026407 ncRNA; 3.8 kb to RNA seq data | | X | | | X | X | | X |
| **Nix downstream** | **Nix downstream 1** | RNA-seq signal 1kb away | ? | ? | | | X | X | | X |
| | **Nix downstream 2** | RNA-seq signal 1kb away | ? | ? | | | X | O | | X |
| | **Nix downstream 3** | RNA-seq signal present in amplicon | ? | ? | X | | X | / | | X |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Nix intron** | **Nix intron 1A** | Close to tandem repeats | | | | X | | O | Tandem repeats nearby | X |
| | **Nix intron 3B** | Downstream of AAEL028699 ncRNA | X | | | X | | O | | X |
| | **Nix Intron 4** | No landmarks of note | | | | X | | X | | X |
| | **Nix intron 5** | No landmarks of note | | | | X | | X | | X |
| **Nix upstream** | **Nix upstream 2B** | Immediately downstream of AAEL023321 ncRNA and <2kb downstream of AAEL011974 ncRNA | X | | | | X | / | | X |
| | **Nix upstream 3** | RNA-seq signal present in amplicon | | | X | | X | X | | |
| | **18** | | | | | | | | | **13** |

*Table 26 Details of the most active sgRNA target sites chosen for each of the thirteen M locus subregions.*

| Site | sgRNA | Target sequence | Genomic location | Strand | GC content (%) | Self-comple-mentarity | MM0 | MM1 | MM2 | MM3 | Predicted efficiency | Measured efficiency |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Myosex intron 3B | **29** | CACACGGCGCACAGTAATGGTGG | 152013502 | - | 60 | 0 | 1 | 0 | 0 | 0 | 60.43 | 58.58 |
| Myosex intron 4B | **125** | AACGCTGTCATGAATCCCGATGG | 152099594 | + | 50 | 0 | 1 | 0 | 0 | 1 | 63.12 | 65.65 |
| Myosex intron 4C | **19** | CTTACCTGAGTCTGGGAGGGAGG | 152105544 | - | 60 | 1 | 1 | 0 | 0 | 0 | 64.35 | 79.95 |
| Myosex downstream 2 | **20** | TTCCTAGGGGGCTTAGCTAGGGG | 152265639 | - | 55 | 3 | 1 | 0 | 0 | 0 | 58.33 | 50.02 |
| Myosex downstream 2b | **6** | CAGGGCTAATAGAATTGCTGTGG | 152272480 | - | 45 | 1 | 1 | 0 | 0 | 0 | 65.77 | 33.18 |
| Myosex downstream 7 | **85** | CTACGGTTACATCTCCAGAGTGG | 152319619 | - | 50 | 0 | 1 | 0 | 0 | 1 | 73.91 | 60.87 |
| Nix downstream 1 | **20** | AATCTTGCGCGCATCCAAACAGG | 152565043 | + | 50 | 0 | 1 | 0 | 0 | 0 | 45.25 | 45.55 |
| Nix downstream 2 | **39** | GAAAGGGGTACGGGCCAACTCGG | 152579610 | - | 60 | 1 | 1 | 0 | 0 | 1 | 57.84 | 56.9 |
| Nix downstream 3 | **44** | AAATTACCTGTTGGTGTGACAGG | 152594072 | + | 40 | 2 | 1 | 0 | 0 | 1 | 51.46 | 47.34 |
| Nix intron 1A | **68** | CATCAGCGACGATCATTTGTAGG | 152687479 | - | 45 | 0 | 1 | 0 | 0 | 1 | 51.75 | 30.03 |
| Nix intron 3B | **43** | GGCAAGAGGCATTCATAAACAGG | 152641745 | - | 45 | 0 | 1 | 0 | 0 | 0 | 35.79 | 46.93 |
| Nix Intron 4 | **8** | TCCGGCTGCCGAGTCCTACGTGG | 152627878 | - | 70 | 1 | 1 | 0 | 0 | 0 | 66.41 | 52.31 |
| Nix intron 5 | **102** | TGATCCCGAAACACAACAAGAGG | 152620522 | + | 45 | 0 | 1 | 0 | 0 | 2 | 58.32 | 54.33 |
| Nix upstream 2B | **12** | GCCCTGTACCCATTAATGGTTGG | 152746781 | - | 50 | 1 | 1 | 0 | 0 | 0 | 59.78 | 76.45 |

### 4.4.2.i Knock-in efficiency at M locus target sites

Transgenic lines with male-only inheritance were attained for five of the thirteen regions targeted, including two in the *Nix* intron and one in the *myo-sex* intron. These were: *myo-sex* intron 4C at sgRNA 19 (PL-1134), *myo-sex* downstream 2B at sgRNA 6 (PL-1136), *Nix* downstream 1 at sgRNA 20 (PL-1137), *Nix* intron 4 at sgRNA 8 (PL-1167), and *Nix* intron 1A at sgRNA 68 (PL-1162) (these are highlighted in green in Figure 25). Canonical integration of the full donor plasmid was checked by PCR in all five lines, revealing that four were correctly integrated, while one, PL-1134, was not. 1134 seemed to have multiple copies of the *β2t* promoter, two of which were shorter than expected, and be integrated incorrectly or at the wrong site. Sequencing, e.g. of inverse PCR or Oxford Nanopore, to determine the integration site for PL-1134 and sequencing of all correctly amplified bands to verify integrity of all transgene cassettes could be performed in future.

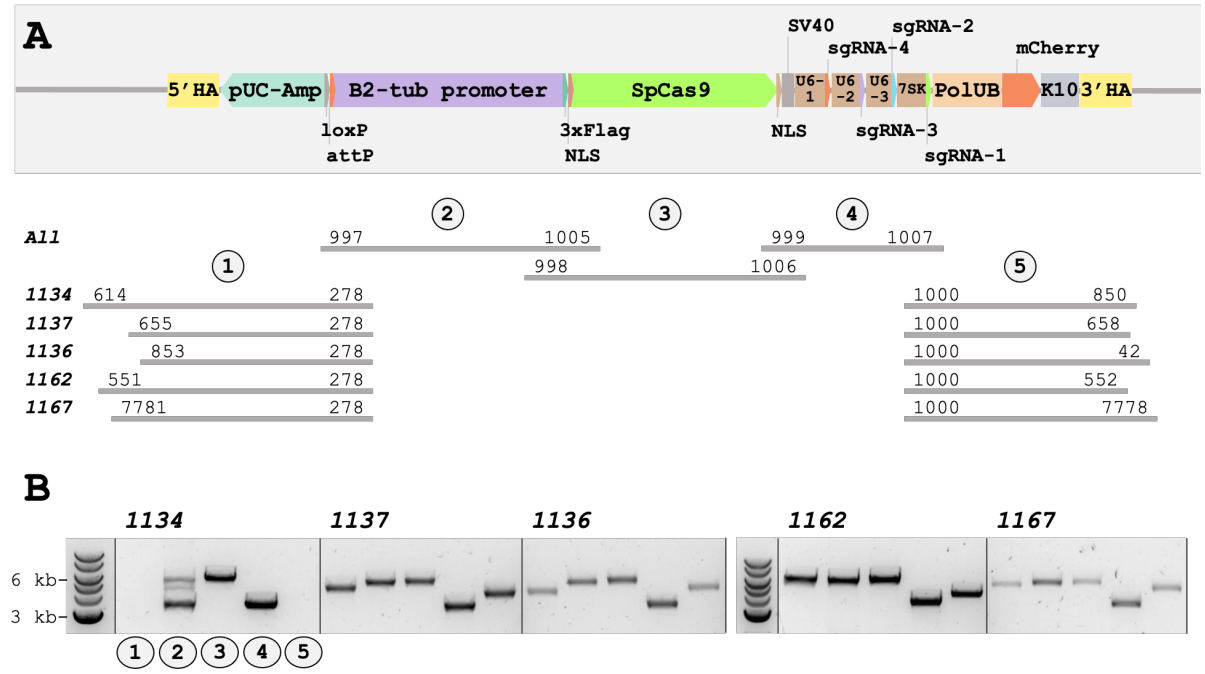



*Figure 29 A) Schematic of the SuperSIT construct integrated within an M locus target site. Correct integration of the entire transgenic construct was verified by PCR at five overlapping sites including: 1, outside each 5' HA into the β2-*tubulin (*β2*t) *promoter (expected size of fragments in each line:* 1134*: 5538 bp;* 1137*: 4671 bp;* 1136*: 4435 bp;* 1162*: 5253 bp;* 1167*; 4998 bp); 2, β2*t *promoter (expected size of fragment in all integrations: 5343 bp); 3, Cas9 (all: 5380 bp); 4, sgRNA cassette (all: 3479 bp); and 5, mCherry cassette out to outside of each 3' HA (in each line:* 1134*: 4440 bp;* 1137*: 4314 bp;* 1136*: 4700 bp;* 1162*: 4269 bp;* 1167*; 4843 bp). Lines represent sizes of amplicons and position relative to the construct; numbers indicate the PL number of each primer pair used. B) Gel of all five PCR products (1-5) for each transgenic SuperSIT line showing correct integration of the entire construct in 1137, 1136, 1162, and 1167; the molecular ladder used was the 1 kb Plus DNA Ladder (NEB) and its 3 and 6 kB bands are indicated.*

The knock-in method devised here, which used 3' to 5' back-to-back homology arms to integrate the entire donor plasmid, was mostly successful and showed that relatively large cargos (16.8 kB) could be integrated. Further, because approximately 2 kB of the donor template was plasmid backbone that was not relevant for the SuperSIT system, other relevant DNA could be swapped in if the 5' HA was cloned on the inside of the pUC backbone, for example if larger regulatory regions were needed or another sgRNA cassette.

A high survival rate for microinjected embryos was critical for knock-in of the SuperSIT construct because only half of the surviving $G_0$ embryos were taken forward for outcrossing (only males carry the M locus). Additionally, knock-in by HDR is generally a low efficiency process in *Ae. aegypti*, meaning many $G_0$ males, around 50-100, had to be outcrossed. The integration rates observed in this study ranged from 1-3.3%, which are similar to those reported in other studies (Ang *et al.*, 2022; Kistler, Vosshall and Matthews, 2015; Li *et al.*, 2017b). It is notable that the integration rates were still low when the measured cleavage by the sgRNAs was relatively high (30-79%) and likely improved compared to *in vitro* transcription of sgRNAs, suggesting that efficiency of HDR repair, maybe specifically in the M locus or due to the relatively long insert, is the limiting factor for knock-in generation.

*Table 27 Survival and integration rates of microinjected mosquito eggs, termed $G_0$, for transgenesis of males with the SuperSIT construct. Males reared to adulthood and females reared to pupal stage were used to calculate survival rates for microinjections. Minimum integration rates for each construct were calculated as a percentage by dividing the number of positive pools by the total number of surviving $G_0$s. This is the "minimum" rate because $G_1$s may arise from multiple independent integrations within a pool. Details for four sets of microinjections are listed. The first set was injected by me, the second set was injected with the help of the team in York, the third set was injected by myself and includes the constructs with the sgRNA between homology arms, and the fourth set was injected by me and includes constructs that were injected with a 10x dilution of required plasmid DNA in the injection mix due to an error.*

**Set 1**

| HDR donor plasmid # | Description | Embryos injected | G0 Survival | | Total | Survival (%) | M: F survival | Minimum canonical integration rate (%) |
|---|---|---|---|---|---|---|---|---|
| | | | Males | Females | | | | |
| 1134 | M-I-4C-19 | 290 | 60 | 64 | 124 | 42.8% | 0.94 | 1.7% |
| 1137 | N-D-1-20 | 405 | 96 | 88 | 184 | 45.4% | 1.09 | 1.0% |
| 1156 | M-D-2-20 | 315 | 84 | 91 | 175 | 55.6% | 0.92 | 0.0% |
| 1155 | M-I-4B-125 | 236 | 24 | 41 | 65 | 27.5% | 0.59 | 0.0% |
| 1155-redo | | 232 | 30 | nr | nr | nr | nr | 0.0% |
| 1136 | M-D-2B-6 | 322 | 59 | 57 | 116 | 36.0% | 1.04 | 1.7% |
| 1162 | N-I-1A-68 | 344 | 30 | 27 | 57 | 16.6% | 1.11 | 3.3% |
| 1164 | M_I_3B_29 | 358 | 79 | 95 | 174 | 48.6% | 0.83 | 0.0% |
| 1167 | N_I_4_8 | 314 | 56 | 56 | 112 | 35.7% | 1.00 | 1.8% |

**Set 2**

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 1161 | M-D-7-85 | 483 | 35 | 28 | 63 | 13.0% | 1.25 | 0.0% |
| 1163 | N-I-5-102 | 497 | 35 | 25 | 60 | 12.1% | 1.40 | 0.0% |
| 1165 | N_D_2_39 | 416 | 17 | 10 | 27 | 6.5% | 1.70 | 0.0% |
| 1166 | N_D_3_44 | 415 | 17 | 11 | 28 | 6.7% | 1.55 | 0.0% |
| 1168 | N_U_2B_12 | 558 | 11 | 12 | 23 | 4.1% | 0.92 | 0.0% |

**Set 3**

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 1139 | M-I-4C-19-sgRNA | 350 | 68 | 82 | 150 | 42.9% | 0.83 | 0.0% |
| 1138 | N-D-1-20-sgRNA | 333 | 64 | 74 | 138 | 41.4% | 0.86 | 0.0% |
| 1157 | M-D-2-20-sgRNA | 348 | 89 | 81 | 170 | 48.9% | 1.10 | 0.0% |

**Set 4**

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 1162 | N-I-1A-68 | 360 | 56 | 50 | 106 | 29.4% | 1.12 | 0.0% |
| 1163 | N-I-5-102 | 325 | 83 | 73 | 156 | 48.0% | 1.14 | 0.0% |

The target site of the surviving microinjected male embryos was expected to be modified by HDR or non-homologous end joining (NHEJ) mutations at frequencies equal to or greater than the ones measured in the *in vivo* cleavage efficiency test. Target site disruption did not seem to have a drastic impact on male survival, though, since male and female $G_0$ survival was generally equal across all chosen sites (Table 27), apart from PL-1155 where male survival was about half that of females (male to female $G_0$ survival ratio = 0.59). Male fitness in terms of the number of surviving $G_1$ offspring from male $G_0$ pools crossed to wild-type females was not recorded.

The first set of injections had comparatively higher survival rates (mean = 39.9%) than the second set (mean = 8.5%) (Table 27). Of the microinjections that were unsuccessful from the first set, namely *myo-sex* intron 4B at sgRNA 125 (PL-1155), *myo-sex* downstream 2 at sgRNA 20 (PL-1156), and *myo-sex* intron 3B at sgRNA 29 (PL-1164), and *myo-sex* downstream 7 at sgRNA 85 (PL-1161) from the second set, transgenics were recovered for all of them, but these were either carrying non-specific integrations (1156 and 1164, 1161 (potentially m-linked), did not produce offspring (1156), or did not survive to adulthood to determine the sex (1155 and 1161). Reattempting knock-in at these sites should not be prioritized because the target site could be found in unidentified repetitive DNA, heterochromatic DNA that could lead to more frequent off-target cutting, or be toxic if disrupted (Uusi-Mäkelä *et al.*, 2018). It is recommended, however, to reattempt knock-in at *Nix* intron 5 at sgRNA 102 (PL-1163), *Nix* downstream 2 at sgRNA 39 (PL-1165), *Nix* downstream 3 at sgRNA 44 (PL-1166), and *Nix* upstream 2B at sgRNA 12 (PL-1168) from the second set because these were likely unsuccessful due to poor quality microinjections rather than issues with the target site; male and female $G_0$ survival was generally equal, and no non-canonical integrations were found.

Lastly, the alternate donor plasmid design that incorporated the corresponding sgRNA target site between the homology arms such that the plasmid would be linearized *in situ* was unsuccessful (Table 27, set 3). Use of linear dsDNA template donors was thought to improve HDR efficiency, as they have been shown in human cells to be most efficient in causing both on- and off-target integration events when compared to ssDNA and circular templates (Li *et al.*, 2017a; Song and Stieger, 2017; Holkers *et al.*, 2014). Nevertheless, the alternate design, microinjected at three sites for which on-target integrations were recovered for two of them using the original HDR design (PL-1134 and PL-1137), only produced off-target integrations; the alternate design was tested at *myo-sex* intron 4C at sgRNA 19 (PL-1139), *Nix* downstream 1 at sgRNA 20 (PL-1138), and *myo-sex* downstream 2 at sgRNA 20 (PL-1157). These findings, while not conclusive, are in line with another study in *Ae. aegypti* that showed that circular donors were preferred for HDR-based knock-ins (Ang *et al.*, 2022).

### 4.4.2.ii Assessment of SuperSIT lines

All five SuperSIT lines transmitted the construct to sons only and two of the five lines impacted female flight, achieving proof of principle for this SuperSIT construct design. No female recombinants were recovered for any of the lines after screening around a thousand offspring for 1134, 1136, and 1137, and a smaller number for 1162 and 1167 (Appendix B Table 15). Larger studies could be performed in future to determine if recombination occurs and, if so, at what frequency and if the offspring would be fit, since some recombination in the M locus has been documented before (Krzywinska *et al.*, 2016b). Lastly, SuperSIT line PL-1136 showed the highest rate of non-flying females up to 50% at $G_3$, followed by PL-1137, which caused up to a 15% non-flying rate, assessed at $G_5$. SuperSIT lines PL-1134, PL-1162, and PL-1167 showed very low (0-1%) or no loss of flight in females (Figure 30, Table 28). These flightless rates reflect a conservative assessment of flight, so wildtype would be expected to be ~0%.

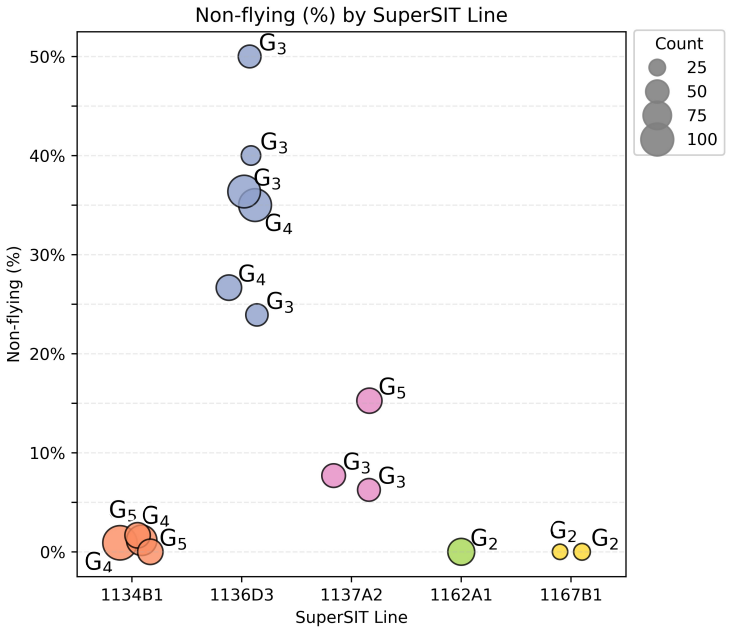



*Figure 30 Flightless rates of each SuperSIT line are shown. Data points are scaled by the number of females that were scored for flight and labelled with the generation number. Generation number is not an independent variable here because males inherit a mutated copy of myo-fem at the same rate as females, but do not show the phenotype since it is female specific. Data was collected from SuperSIT crosses to wild-type females and other experiments, where inheritance of constructs is believed to be neutral (see Table 28 below)*

*Table 28 Details of SuperSIT crosses used in Figure 29 to depict flight ratees achieved by each SuperSIT line that was generated. Crosses show the genotype of the male on the left and female on the right of the colon; brackets indicate the genotype of an individual and from which parent the constructs were inherited; if there is a "–" sign next to the construct, such as in (SuperSIT : 2199- ) : LVP, this indicates that the mother of the male used in the cross carried the 2199C1 construct, but the male did not inherit it. The genotype column indicates the genotype of the female in terms of whether it inherited the additional construct where there was one; "?" indicates the offspring were not screened for the extra marker. The SuperSIT column indicates the generation of the SuperSIT that females were screened at. LVP stands for the Liverpool strain of* Ae. aegypti *and indicates the wild-type strain.*

| SuperSIT line | Cross | Filial generation | Genotype | SuperSIT generation | Non-fliers (%) | Count |
|---|---|---|---|---|---|---|
| 1134B1 | SuperSIT : (1095 : LVP) | F1 | + | G4 | 1% | 110 |
| | SuperSIT : (1095 : LVP) | F1 | - | G4 | 1% | 86 |
| | SuperSIT : (1095 : LVP) | F1 | ? | G5 | 0% | 60 |
| | (SuperSIT : 2199- ) : LVP | F2 | | G5 | 2% | 60 |
| 1136D3 | SuperSIT : LVP | | | G3 | 50% | 48 |
| | SuperSIT : LVP | | | G4 | 35% | 100 |
| | SuperSIT : (1095 : LVP) | F1 | + | G3 | 24% | 46 |
| | SuperSIT : (1095 : LVP) | F1 | - | G3 | 40% | 35 |
| | SuperSIT : (1095 : LVP) | F1 | ? | G4 | 27% | 60 |
| | SuperSIT: (1551 : LVP) | F1 | - | G3 | 36% | 99 |
| 1137A2 | SuperSIT : (1095 : LVP) | F1 | + | G3 | 8% | 52 |
| | SuperSIT : (1095 : LVP) | F1 | - | G3 | 6% | 48 |
| | (SuperSIT : 2199- ) : LVP | F2 | | G5 | 15% | 59 |
| 1167B1 | SuperSIT : LVP | | | G2 | 0% | 22 |
| | SuperSIT : LVP | | | G2 | 0% | 26 |
| 1162A1 | SuperSIT : LVP | | | G2 | 0% | 67 |

Strong variability in the perceived brightness of the fluorescent marker was observed between and within the transgenic lines. Transgenic lines 1137, 1136, 1162, and 1167 were quite dim compared to 1134, which could be screened very easily at a low magnification under the fluorescence scope. It would be interesting to determine where and how 1134 is integrated to understand why this line appears brighter than the other four integrations. Transgenic line 1137 and, to a smaller degree, 1136 showed variability of expression between siblings of the same line (only a low number of offspring were screened for 1162 and 1167, so it is not known whether these are variable and 1134 consistently produced very "bright" males). From pooled crosses of "bright" and "dim" males, 14-22% had a visible marker in 1137 (n=2) with the remainder appearing negative, while 64-81% were "brighter" in 1136, with the remainder being much dimmer, but still distinguishably positive at the maximum magnification. The expression profile of the integrated fluorescent marker did not appear to be heritable, because when the offspring of a single "bright" male from 1137 was screened, most of the offspring were negative again. It was hypothesized that this variability was due to heterochromatic silencing of the construct in the M locus, which has been observed for transgene integrations in the Y chromosome of *Drosophila* (Buchman and Akbari, 2019).

Lastly, no comprehensive fitness tests were conducted, but a few males from transgenic line 1134 (about 1%) were observed to be impaired in their ability to fly, which was not seen in other lines. This is unlikely to be due to the *myo-fem* sgRNAs non-specifically targeting *myosex*, because the sgRNAs used are not found in the *myosex* gene and because the Cas9 and sgRNA cassettes did not appear to be active together. It is possible that the SuperSIT construct, which did not integrate at the expected site within the intron of *myosex*, instead integrated somewhere within its vicinity, negatively impacting it.

### 4.4.2.iii Investigation of position effects in SuperSIT lines

Transgenic SuperSIT lines were provided with external sources of Cas9 and sgRNAs to investigate the influence of position effects on the expression of the SuperSIT CRISPR-Cas9 construct and determine whether each site was limited by Cas9 or guide expression, or both. Three of the five SuperSIT lines (PL-1134, PL-1136, and PL-1137) were crossed to a combination of transgenic lines expressing *shu*-Cas9 (AGG-1551B1) and two lines expressing *U6*-sgRNAs targeting *kmo* (AGG-1095) and *myo-fem* (AGG-2199C1), respectively. Previous split drive assessments had showed strong expression of Cas9 and

sgRNA from these lines (Anderson *et al.*, 2023a). Due to the variable expression of the marker from site 1137, this transgenic line was only tested with an additional source of *kmo* guides, and not an additional source of *myo-fem* guides.

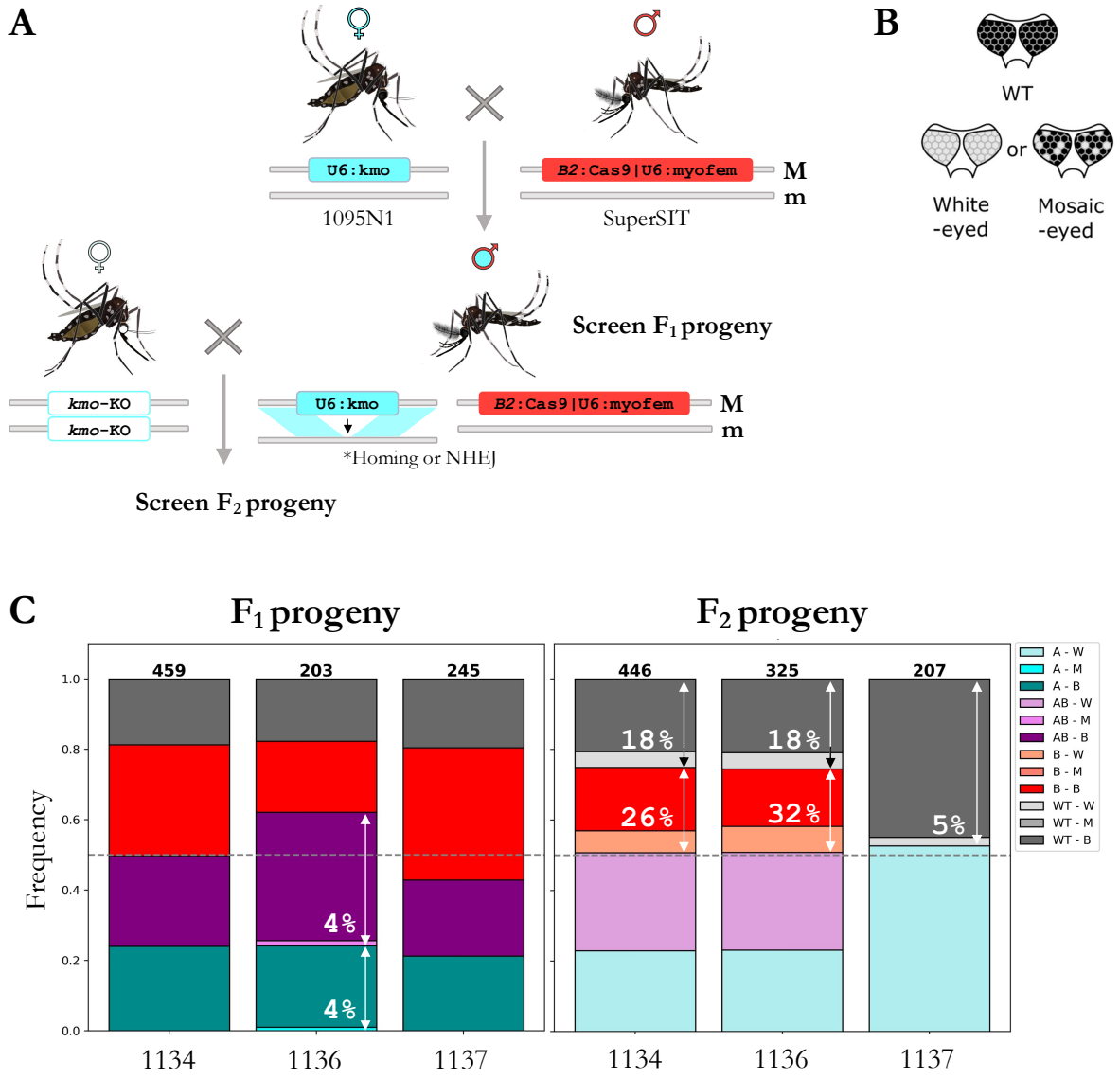
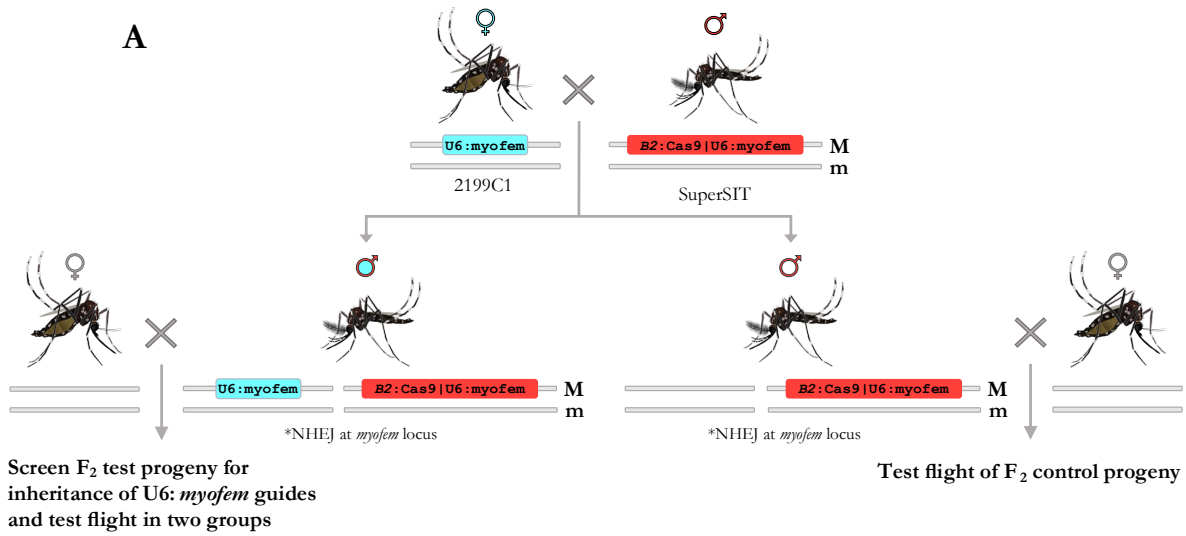



*Figure 31 A) Crossing scheme for determination of β2t-Cas9 activity in each SuperSIT transgenic line with guides targeting the* kmo *gene. Male SuperSIT lines were crossed in groups to females from transgenic line AGG-1095N1, which expressed sgRNAs targeting* kmo *from a knock-in integration. $F_1$ transheterozygous males carrying both the* kmo guide *knock-in allele and the SuperSIT construct were screened and crossed to females that were homozygous for a knock-out (KO) mutation of* kmo *(*kmo*-KO shown with white eyes). B) Illustration of the difference between the three eye-phenotypes that were scored: WT (B=black), mosaic (M), and white (W). C) Inheritance of four genotypes: A: kmo sgRNAs only (cyan); AB: kmo sgRNAs + SuperSIT construct (purple); B: SuperSIT construct only (red); and WT: no constructs inherited (grey) in $F_1$ and $F_2$ progeny. The grey dotted line at 50% frequency represents the expected inheritance of the kmo sgRNA allele. Due to strong variation in marker fluorescence in 1137 that made some offspring impossible to screen, the F2 progeny was not scored for inheritance of the SuperSIT construct. White values overlayed on 1136 in the $F_1$ progeny show the percentage of mosaic eyes within each genotype group. Percentages in the $F_2$ group show the percentage of white eyes within each*

*genotype group. NB all A and AB individuals in the $F_2$ group have white eyes because they inherit two disrupted copies of* kmo. *Figure A is adapted from (Anderson et al., 2023a) using publicly available male and female* Ae. aegypti *images from (Ramírez, 2019b; Ramírez, 2019a). Figure B is taken from (Anderson et al., 2023a).*

With an additional source of guides targeting *kmo*, the *β2t*-driven Cas9 cassette in the SuperSIT lines caused the white-eye phenotype in 22% of $F_2$ offspring in 1134, 24% in 1136, and 5% in 1137 (mean of offspring not inheriting the *kmo* guides from transheterozygous $F_1$ parents) (Figure 31). These results showed that the Cas9 expresses from all three integration sites. No inheritance bias to indicate homing was observed for the *kmo* knock-in allele, but this is unsurprising as this promoter is hypothesized to act late in the male germline. Mosaicism was observed in ~4% of the 1136 $F_1$ progeny that had inherited the *kmo* sgRNA-expressing allele, regardless of whether they had inherited the Cas9 (Figure 31C: $F_1$ progeny), indicating paternal deposition of Cas9 occurred in this line. Deposition of Cas9 can be differentiated from deposition of complexed Cas9 and sgRNA by screening the $F_2$ of transheterozygotes and single heterozygotes crossed to wild-type females; this cross was performed but offspring were not screened.



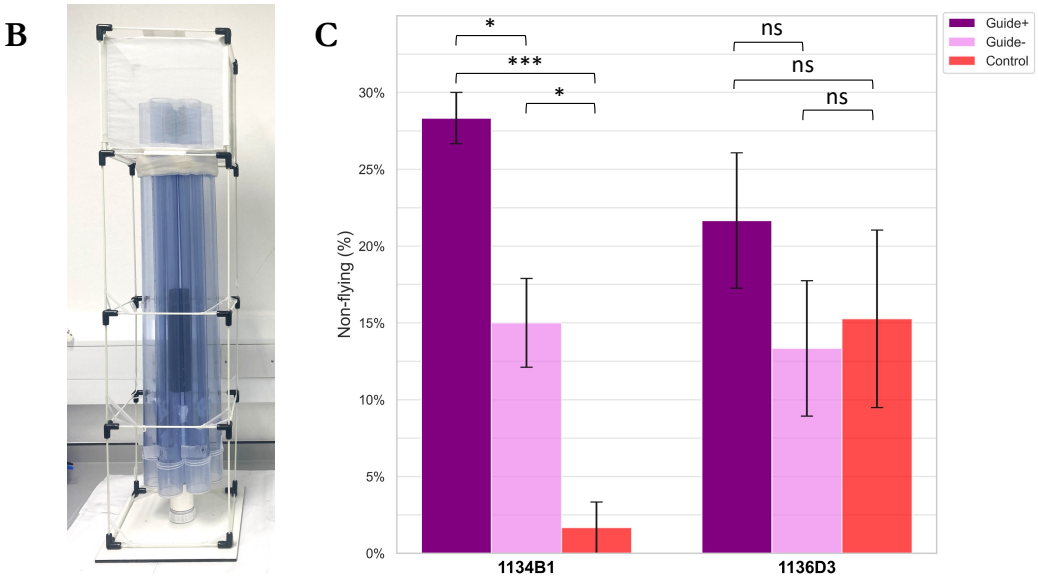




*Figure 32 A) Crossing scheme for determination of β2t-Cas9 activity in two SuperSIT transgenic lines with extra guides targeting the myo-fem gene. Male SuperSIT lines were crossed in groups to females from transgenic line AGG-2199C1, which expressed sgRNAs targeting myo-fem from a random integration site. $F_1$ transheterozygous males carrying both the myo-fem guide and the SuperSIT construct were screened and crossed to wild-type females. As a control, $F_1$ single heterozygous males carrying only the SuperSIT construct were also crossed to wild-type females. B) Flight in the $F_2$ progeny was tested using the flight test device. Each test group was tested in groups of twenty females per tube using three tubes. C) Conservative flight rates of $F_2$ progeny, estimated by dividing the number of females at the bottom of each tube in the flight test device at the end of a flight test (t=1h30) by the total number of females tested per tube (n=20). Purple bars represent the offspring from $F_1$ transheterozygous males, the darker shade is the group that inherited the sgRNA construct and the lighter shade is the group that did not inherit it. Red bars represent the control group for each transgenic line tested. Significance bars are included: *** if p < 0.001, ** if p < 0.01 and * if p < 0.05. Figure A is adapted from (Anderson et al., 2023a) using publicly available male and female* Ae. aegypti *images from (Ramírez, 2019b; Ramírez, 2019a).*

With an extra source of guides targeting *myo-fem*, flightless rates improved in 1134, but not 1136. For 1134, flightless rates significantly improved from 2% (mean, n=3) to 15% in females not carrying sgRNAs (p=0.016, Student's t-test) and 28% in females carrying

sgRNAs (p=0.0003). For 1136, there was no significant improvement with the extra source of guides (p=0.804 and 0.428 for those inheriting the sgRNAs and those not, respectively). Perhaps the *β2t*-Cas9 is saturated with guides expressed from this site already at the germline stage. A difference in the non-flying rate between those inheriting the sgRNAs and those that did would indicate paternal deposition of Cas9 into embryos. This difference was statistically significant in 1134 (p=0.016, Student's t-test) but not in 1136 (p=0.252). This contrasts with the previous experiment with additional *kmo* guides (which have different Pol III promoters), where paternal deposition was observed in 1136 but not in 1134, although the effect here was small (~4%).

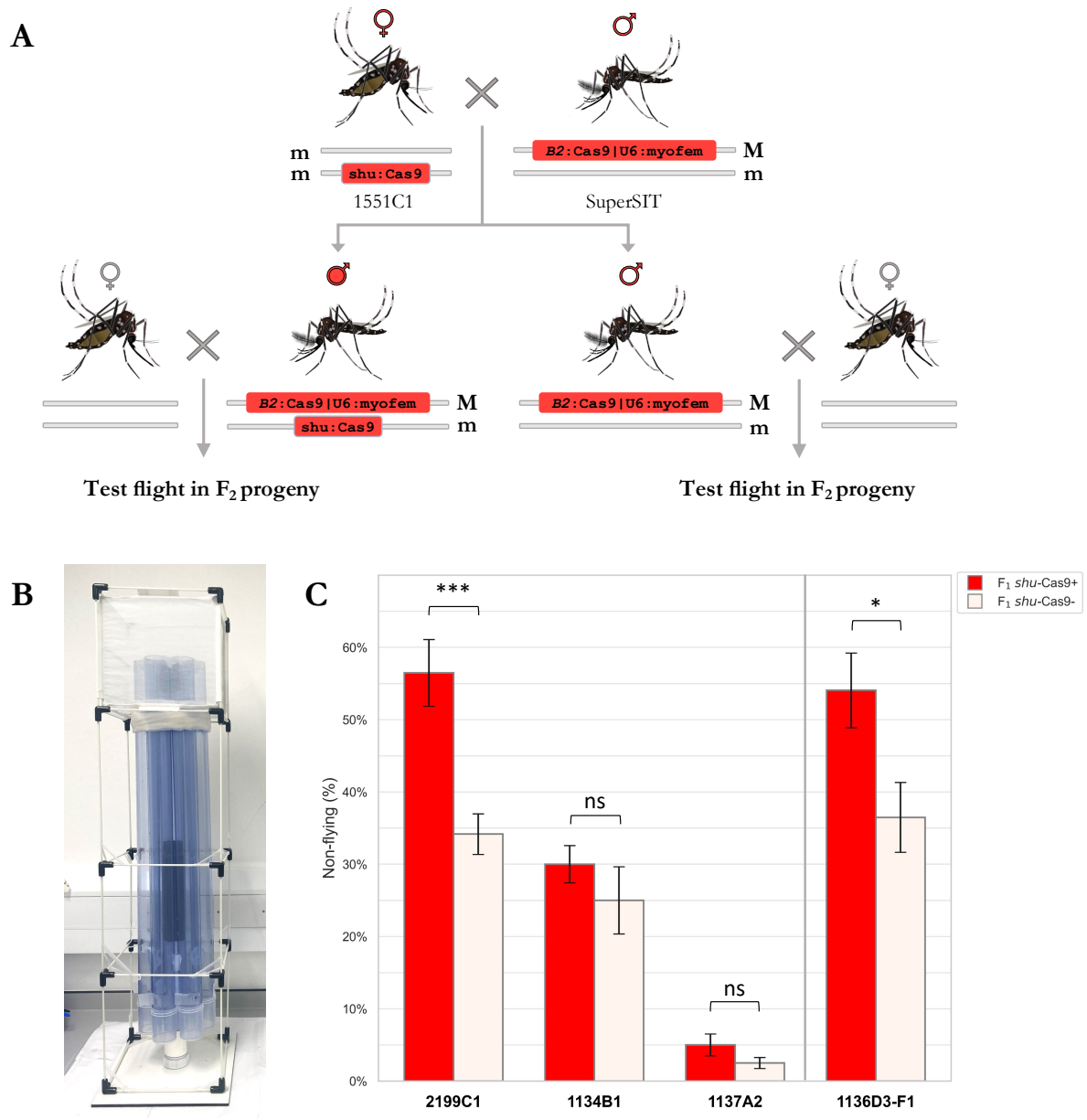



*Figure 33 A) Crossing scheme for determination of U6-sgRNA activity in three SuperSIT transgenic lines with an extra source of Cas9 driven by the* shu promoter. *Male SuperSIT lines were crossed in groups to females from transgenic line AGG-1551B1, which expresses* shu-*Cas9 from an m-linked integration site. $F_1$ transheterozygous males carrying both the* shu-*Cas9 and the SuperSIT construct were screened according to the brightness of the mCherry marker and crossed to wild-type females. To control for $F_0$ maternal deposition into $F_1$ embryos, $F_1$ single heterozygous males carrying only the SuperSIT construct were also crossed to wild-type females. B) Flight in the $F_2$ progeny was tested using the flight test device. Each test group was tested in groups of twenty females per tube using six tubes. C) Conservative flight rates of $F_2$ progeny, estimated by dividing the number of females at the bottom of each tube in the flight test device at the end of a flight test (t=1h30) by the total number of females tested per tube (n=20). Red and white bars represent the offspring from $F_1$ transheterozygous and single heterozygous males, respectively. Significance bars are included: *** if $p < 0.001$, ** if $p < 0.01$ and * if $p < 0.05$. Figure A is adapted from (Anderson et al., 2023a) using publicly available male and female* Ae. aegypti *images from (Ramírez, 2019b; Ramírez, 2019a).*

Lastly, with an extra source of Cas9 driven by the *shu* promoter, cleavage improved in 1134, but not in 1137; 1136 has not been determined yet and have not included

corresponding controls to show if shu-Cas9 did anything (Figure 33). In line 1134, flightless rates increased with an additional source of Cas9 from ~1% to 30% (mean, n=6) and 25% in female offspring of transheterozygous and single heterozygous males, respectively. This additional activity was predominantly due to *shu*-driven Cas9 maternally deposited into the $F_1$ male embryo since flightless rates between the two groups were similar and indicates no additional cutting from activity in the germline. In contrast, germline cutting was observed in the positive control cross, where males from the piggyBac integrated line carrying *myo-fem* sgRNAs (line 2199C1) were used instead of the SuperSIT lines, producing a mutated signal ~20% above the offspring from single heterozygous $F_1$ males (Figure 33C, 2199C1, $F_1$ *shu*-Cas9+ vs -). In line 1137, flightless rates did not improve in either group ($F_2$ of transheterozygotes = 5% and $F_2$ of single heterozygotes = 3%, mean n=6). Lastly, flight rates at $F_2$ were not determined for 1136 with *shu*-Cas9 due to an error, but $F_1$ females of 1136 crossed to *shu*-Cas9 were tested. $F_1$ daughters that had inherited *shu*-Cas9 seemed to have a higher flightless rate compared to those that had not (*shu*-Cas9+ = 54% vs *shu*-Cas9- = 36%, mean n=6), potentially suggesting paternal deposition of the U6-sgRNAs, although the mechanism by which this would work is unclear.

Together, these results depict a complex pattern of expression for the Cas9 and sgRNA cassettes at these sites. In 1134, both cassettes seem to express individually when provided with external sources of Cas9 and guides, but together they were not able to reduce flight in female descendants. Perhaps position effects at this site are changing the time of expression of the *myo-fem* guide cassette to a more embryonic stage, where maternally deposited *shu*-Cas9 would be present, and a less germline stage, where *β2t*-Cas9 seems to be active. In contrast, in 1136, expression of the *myo-fem* guide construct in the germline does seem to occur, which is like its expression pattern from the 2199C1 integration site, albeit lower, as was shown with both an M- and m-linked *shu*-Cas9 expressing line (2235 in chapter two and 1551C1 shown here, respectively). Between 1134 and 1136, expression of the *β2t*-Cas9 cassette seems to be similar, as was shown with the extra sources of U6 driven guides, suggesting this promoter is less affected by position effects, as was seen in chapter one. Further, signals of paternal deposition were observed for this promoter, which is the first time this has been engineered in *Ae. aegypti.*

#### 4.4.2.iv Conclusion

The aim of this section was to integrate a SuperSIT construct expressing a CRISPR-Cas9 system targeting the female-specific flight gene *myo-fem* within the M locus. Successful knock-in and expression of the system in the M locus would mean that only sons inherit the construct and daughters are unable to fly. Integration of the SuperSIT system was attempted across the M locus at thirteen target sites of sgRNAs verified to be active. Transgenic lines with male-only inheritance were attained for five of the thirteen regions targeted, four of which were correctly integrated. Two of the four correctly integrated lines impacted female flight, achieving proof of principle for this SuperSIT construct design. SuperSIT line PL-1136 showed the highest rate of non-flying females, up to 50%, followed by PL-1137, which caused up to a 15% non-flying rate. Expression of the fluorescent marker, Cas9, and sgRNA cassettes was variable across three integration sites that were investigated in greater depth and indicate position effects at these sites, especially at 1134 and 1137. Thus, the integration site of 1136, which is located downstream of *myo-fem* at site "2B", might be the best candidate for testing a similar SuperSIT construct design but with an alternate promoter that has shown higher rates of cutting, such as *shu*. Such a construct could also be tested at another previously validated integration site that was found to be tightly linked to the M locus and support high expression of a fluorescent marker (Lutrat *et al.*, 2023). Although none of the transgenic lines produced an effective SuperSIT system, together they showed proof of principle of the design and offer insight into possible avenues for optimizing the system, discussed below.

### 4.4.3 Future Directions

#### 4.4.3.i Improving expression of the SuperSIT system

The greatest limiting factor of building a SuperSIT system in *Ae. aegypti* seems to be achieving sufficient expression of the system from the heterochromatic M locus. It is notable that expression varied both between and within transgenic strains; in one line, the fluorescent marker was invisible in most of the male offspring, despite originating from a "bright" single parent. Position effects arise due to the influence of context-specific regulatory sequences such as enhancers and repressors of gene expression on the integration site (Wilson, Bellen and Gehring, 1990). Position-effect variegation (PEV),

used to describe varied expression between siblings of the same transgenic strain, occurs when a euchromatic transgene inserted into a region of heterochromatin is silenced due to the spread of heterochromatin associated marks into the transgene locus (Elgin and Reuter, 2013; Shatskikh, Abramov and Lavrov, 2017). It is possible that heterochromatin from the surrounding regions of the transgene integrated at some of the sites in the M locus is spreading into the transgene and silencing it.

*4.4.3.i.a Identifying new integration sites by analysing the chromatin environment of the M locus*

An improved understanding of the chromatin environment of the M locus during male developmental and reproductive stages could help identify integration sites that would promote SuperSIT expression. Chromatin structure plays a fundamental role in the dynamic regulatory machinery by coordinating access of proteins to a set of regulatory sequences termed *cis*-regulatory elements (CREs) (Li, Carey and Workman, 2007). Differences in chromatin accessibility at CREs can reflect transcriptional activity in a tissue- and stage-specific manner. Chromatin accessibility of *Ae. aegypti* has recently been studied in Aag2 cells, antennae, female brains, and embryos, but not yet in testes (Qu *et al.*, 2023; Matthews *et al.*, 2018; Nowling *et al.*, 2021; Behura *et al.*, 2016; Mysore, Li and Duman-Scheel, 2018). FAIRE-Seq (Formaldehyde-Assisted Identification of Regulatory Elements combined with next-generation sequencing) identifies open chromatin enriched with CREs that is then sequenced (Nowling *et al.*, 2021; Behura *et al.*, 2016). Data from a FAIRE-Seq experiment performed on 2-day old mixed sex whole embryos from *Ae. aegypti* was recently reprocessed using the AaegL5 genome assembly, which contains the M locus (Behura *et al.*, 2016; Nowling *et al.*, 2021). New integrations could be targeted to sites that correspond to peaks of open chromatin, but care should be taken because chromatin open at the embryonic stage may not be open in testes or during male germline development.

The Assay for Transposase Accessible Chromatin with sequencing (ATAC-seq) has emerged as the preferred tool for measuring chromatin accessibility due to its simplicity and scalability compared to FAIRE-Seq and other tools (Buenrostro *et al.*, 2015). ATAC-Seq has recently been used to identify optimal transgene integration sites in a non-model yeast organism and in Chinese Hamster Ovary (CHO) cells, where integration sites with high transgene expression correlated with larger regions that showed high

transcriptional activity and chromatin accessibly, but not necessarily with specific highly transcribed genes (Brady *et al.*, 2020; O'Brien *et al.*, 2018). Performing single cell ATAC-Seq on *Ae. aegypti* testes samples could not only help identify optimal integration sites for testes-specific transgene expression, but potentially map unknown CREs involved in male reproductive processes. Such a study would contribute to understanding the regulatory genome of the *Ae. aegypti* mosquito, which, compared to *Drosophila* and *Anopheles* mosquitoes, is relatively unexplored (Lezcano *et al.*, 2020; Ruiz, Ranford-Cartwright and Gómez-Díaz, 2021).

*4.4.3.i.b Insulators could mitigate position effects and gene silencing*

One suggested solution to mitigate position effects and insulate transgenes from their genomic context is with insulator elements and their associated proteins (Carballar-Lejarazú, Jasinskiene and James, 2013; Pikaart, Recillas-Targa and Felsenfeld, 1998; Recillas-Targa *et al.*, 2002; Roseman, Pirrotta and Geyer, 1993; Kellum and Schedl, 1991; Sarkar *et al.*, 2006; Yamamoto and Gerbi, 2022; She *et al.*, 2010; Barolo, Castro and Posakony, 2004; Markstein *et al.*, 2008; Buchman and Akbari, 2019). Insulators can block the effect of neighbouring enhancers and repressors, stimulate transcription, and form barriers protecting against encroaching heterochromatin (Bell, West and Felsenfeld, 2001; Kyrchanova and Georgiev, 2014; Schwartz *et al.*, 2012; Phillips-Cremins and Corces, 2013; Adelman *et al.*, 2008; Funikov *et al.*, 2020; Gaszner and Felsenfeld, 2006). A diverse range of insulators have been identified in the *Drosophila* genome, including the scs and scs' insulators found at the boundaries of the domain containing two *Hsp70* genes, and the best studied *gypsy*/*Su(Hw)* element discovered in the regulatory region of the *gypsy* retrotransposon (Kyrchanova and Georgiev, 2014). The *gypsy* insulator is thought to regulate gene expression by organizing chromatin into higher order domains of independent regulatory structures and blocking nearby enhancers and repressors (Gaszner and Felsenfeld, 2006; Markstein *et al.*, 2008).

Despite the widespread use of insulators in *Drosophila*, they are not routinely used in mosquito genetics. *gypsy* insulators from *D. melanogaster* were studied in the malaria mosquito *An. stephensi*, where they were used to flank transgenes integrated into random and site-specific loci in the genome (Carballar-Lejarazú, Jasinskiene and James, 2013). The insulators were found to significantly increase transgene expression across multiple sites, but it was unclear whether they stabilized expression as the authors claimed. There

was still substantial variation between the insulated constructs of the random integrations as well as between the two site-specific integrations, where even the uninsulated constructs seemed to be more similar (these p-values were not included, so it is difficult to make a conclusion). Further, while intergenerational variability of transgene expression was significantly lower between insulated constructs compared to uninsulated ones across several random integrations, suggesting the insulators reduced PEV, the effect was less pronounced between the site-specific integrations.

Recent studies in *Ae. aegypti* that investigated the effect of different germline promoters on Cas9 expression did not use insulators in the transposon-integrated constructs, despite strong position effects having been documented (Anderson *et al.*, 2023a; Verkuijl *et al.*, 2022b; Li *et al.*, 2020; Li *et al.*, 2017b; Reid *et al.*, 2022). Reasons for omitting insulators in these constructs likely included the inconclusive evidence that they stabilize expression, the effect of insulators on proximal genes that are not flanked, and the augmented risk of recombination due to the presence of two identical inverted repeats on either end of the integration (Carballar-Lejarazú, Jasinskiene and James, 2013; She *et al.*, 2010; Symington, Rothstein and Lisby, 2014). The latter risk could be mitigated by using two asymmetric insulators, as was done for a transgene integrated in a gene poor region of the Y chromosome in *D. melanogaster*, using a *gypsy* element on the 5' end and a CTCF binding element on the 3' end (CTCF is an insulator conserved across metazoan species, including mosquitoes) (Buchman and Akbari, 2019; Gamez *et al.*, 2021; Wu, Liu and Wang, 2020; Gray and Coates, 2005; Lezcano *et al.*, 2020; Ruiz, Ranford-Cartwright and Gómez-Díaz, 2021). Nevertheless, the impact of the insulators on the Y integration was debateable, since the authors noted varied expression of the transgene marker in siblings of the same line, a majority of which showed no transgene expression at all (Gamez *et al.*, 2021). Notably, the brightness of the marker transgene was not associated with the activity of a linked Cas9 (also documented elsewhere (Lutrat *et al.*, 2022)), complicating the picture of the regulatory environment at these sites. Unfortunately, the authors did not include an uninsulated construct for comparison.

The SuperSIT construct described here would benefit from enhanced expression at all five integration sites and protection from encroaching heterochromatin that is hypothesized to be a cause of both the PEV and silencing of the marker transgene observed in some of the lines. A study investigating the adaptation of essential genes

translocated from euchromatin into regions of repressive pericentric heterochromatin in *Drosophila* lineages found preserved binding sites for insulator protein BEAF-32 and an enrichment of insulator proteins BEAF-32, GAF (GAGA factor), and dCTCF (*Drosophila* homolog of CCCTC-binding factor) in the promoters of protein-coding genes (Funikov *et al.*, 2020). The authors' hypothesis arose from another study that demonstrated that heterochromatic genes do not have specifically adapted promoters, meaning genes translocated into heterochromatic environments must retain transcriptionally active euchromatin by other means (Yasuhara, DeCrease and Wakimoto, 2005; Funikov *et al.*, 2020). It would be interesting to investigate whether such sites are conserved and enriched in the M locus, especially around the male-determining gene *Nix*, and whether these *Aedes*- or mosquito-specific binding sites could be used to insulate the SuperSIT construct developed here or other transgenic constructs in general. In a brief scan of the untranslated 5' regulatory region of *Nix*, three BEAF-32 motifs were found in close succession to each other, however more sophisticated bioinformatics analyses would be needed to determine whether this is noteworthy.

*4.4.3.i.c Removal of pUC plasmid backbone that could be silencing the integrated construct*

There is approximately 2 kB of non-essential pUC plasmid backbone DNA that is incorporated into the genome that could be reducing or altering the expression of the SuperSIT construct. This was a consequence of the back-to-back homology arm design that enabled cloning of multiple HDR donor plasmids targeting the M locus in a high-throughput way. Research has shown, however, that standard pUC vectors >2 kB containing antibiotic resistance markers can cause transgene silencing in transfection studies (Boye *et al.*, 2022; Williams and Paez, 2023; Bruter, Kandarakov and Belyavsky, 2018; Lu *et al.*, 2012; Chen *et al.*, 2008; Suzuki, Kasai and Saeki, 2006). For example, researchers have found that reducing plasmid vector backbone can enhance gene expression levels in rat tenocytes *in vitro* and rat muscle cells *in vivo* (Boye *et al.*, 2022). Transgene inactivation is hypothesized to occur through heterochromatin formation on large regions (>1 kb) of unexpressed plasmid backbone (Chen *et al.*, 2008; Suzuki, Kasai and Saeki, 2006; Chen *et al.*, 2003; Williams and Paez, 2023; Riu *et al.*, 2007). It is possible that in the M locus, which is an area of dense heterochromatin, that the SuperSIT transgenes are particularly vulnerable to silencing. Future designs should omit this extra

2 kB of the pUC plasmid, as it offers no benefit, and could be negatively impacting the construct.

*4.4.3.i.d Alternative methods for sgRNA expression*

Efforts to improve activity of CRISPR-Cas9 systems in *Ae. aegypti* in general may need to overcome another technical challenge, which is ensuring colocalization of sgRNAs with Cas9 in germline cells. Spatiotemporal control of CRISPR activity is generally restricted by controlling Cas9 expression with Polymerase II regulatory elements from putative germline genes, as was described in Chapter 2. sgRNAs are instead expressed using a set of Pol III promoters from non-coding RNA genes, such as U6 and 7SK, that are thought to express ubiquitously and within nuclei where Cas9 genome targeting occurs (Anderson *et al.*, 2020; Konet *et al.*, 2007; Li *et al.*, 2020; Feng *et al.*, 2021; Gantz *et al.*, 2015; Hammond *et al.*, 2016). This type of Pol III gene is unique in that expression is only dependent on the presence of relatively short upstream regulatory elements, similar to the canonical architecture of Pol II genes (Murphy, Di Liegro and Melli, 1987). Further, expression is possible without the 5' and 3' end modification characteristic of pol II promoters, which can export sgRNAs from the nucleus and distort transcription initiation and termination sites that would lead to variation in the sgRNA sequence (Gao *et al.*, 2017; Nissim *et al.*, 2014; Ma *et al.*, 2014).

Several *Ae. aegypti*-derived Pol III promoters have been identified from genomic data, some of which have been shown to express well in *Ae. aegypti* Aag2 cells (Konet *et al.*, 2007; Rosado Rivera, 2018; Anderson *et al.*, 2020; Li *et al.*, 2020). These have been incorporated into single and multiplexed sgRNA constructs used throughout this thesis and in other studies to generate germline mutations and homing in *Ae. aegypti* transgenics (Li *et al.*, 2020; Reid *et al.*, 2022; Anderson *et al.*, 2024). As discussed in Chapter 2, split drive studies where the putative germline promoter driving Cas9 expression was varied have generally shown unintentionally high mutation rates where homing was the desired outcome or simply no activity at all (Li *et al.*, 2020; Verkuijl *et al.*, 2022b; Reid *et al.*, 2022; Anderson *et al.*, 2023a). Speculation about the causes of the low homing efficiencies have centred on limited or leaky expression of Cas9 during or outside of meiosis I, respectively, or fundamental differences in endogenous repair machinery/processes compared to species of the *Anopheles* genus where homing rates are among the highest (Champer *et al.*, 2017; Hammond *et al.*, 2021a; Verkuijl *et al.*,

2022a). However, an alternative or complementary hypothesis is that Pol III activity is also limited in meiotic germline cells, meaning sgRNAs would not colocalize with Cas9 even if Cas9 was expressed in the desired tissues. This would not only reduce homing efficiencies but also have significant implications for engineering a paternal effect mediated SuperSIT system, where it is optimal for Cas9 expression to occur during late spermatogenesis to increase the likelihood of carry-over to the embryo.

Determining the expression profile of U6 or 7SK ncRNA genes using RNA-seq studies is challenging, however, as most studies only select for polyadenylated messenger RNA (poly(A)-mRNA) (Akbari *et al.*, 2013a; Matthews *et al.*, 2016; Li *et al.*, 2022). Additionally, small RNA-seq studies often exclude Pol III-derived ncRNA molecules, too, as they are ~100-300 nucleotides long and small ncRNAs are typically size selected at <30 nucleotides (Ma *et al.*, 2021; Bryant, Ray and Mills, 2020). RNA datasets from extracted nuclei may be more likely to include Pol III ncRNA but, again, it would depend on the library preparation protocol that was used in terms of whether poly(A)-mRNA was selected for or not. Sequencing protocols like FAIRE-Seq or ATAC-Seq, that probe the chromatin state and help identify *cis*-regulatory elements could offer alternatives to assessing Pol III regulatory activity (Rosado Rivera, 2018; Behura *et al.*, 2016). As suggested before, a scATAC-seq dataset for differentiating cells of the testes (and ovaries) would thus be greatly beneficial in this context as well. Alternative methods for expressing and multiplexing sgRNAs from a single transcript have been investigated and are discussed below.

*4.4.3.i.d.(i) Alternate methods for expressing and multiplexing sgRNAs*

CRISRP-Cas12a (formerly known as Cpf1) is an alternative RNA-guided endonuclease with efficient genome editing capabilities in a broad range of organisms (Campa *et al.*, 2019; Lei *et al.*, 2022; Lewis, Yan and Finnigan, 2021; Port, Starostecka and Boutros, 2020; Dong *et al.*, 2020). An advantage of this system over Cas9 is that it is responsible for the processing and maturation of its own CRISPR RNA (crRNA) (Zetsche *et al.*, 2015; Fonfara *et al.*, 2016). The native crRNA-processing ability of Cas12a means that only a single promoter is required to drive several crRNAs, which allows multiplexing using a simplified molecular design (Zetsche *et al.*, 2017). While Pol III promoters have typically been used to express crRNA arrays independently of a Pol II-driven Cas12a, tandem expression of Cas12a and a multiplexed crRNA array has recently been demonstrated

from a single Pol II promoter using an mRNA stabilizing Triplex motif (Campa *et al.*, 2019). In the context of CRISPR-mediated genetic control systems, tandem expression could enable compatible spatiotemporal activity of sgRNAs and Cas protein, limiting leaky Cas9 expression and potentially improving paternal effect activity. Potential limitations of such a system include suboptimal ratios of the two expressed components that could lower genome editing efficiency. Further, Cas12a activity has been shown to be sensitive to temperatures below 25°C, however temperature tolerant variants have been engineered with substantially improve activity (Port, Starostecka and Boutros, 2020; Sanz Juste *et al.*, 2023; Kleinstiver *et al.*, 2019; Schindele and Puchta, 2020). CRISPR-Cas12a activity has been demonstrated in *Drosophila* and the silkworm *Bombyx mori* (Pan *et al.*, 2022; Dong *et al.*, 2020; Port, Starostecka and Boutros, 2020).

Alternative designs for expressing a multiplexed sgRNA array from a single Pol II promoter have been tested in several organisms. These employ RNA cleavage sequences to separate individual sgRNAs in an array, which is then processed using various mechanisms (McCarty *et al.*, 2020; Minkenberg, Wheatley and Yang, 2017). Strategies have included the use of virus-derived self-cleaving ribozymes, co-expression of a CRISPR system-derived Csy ribonuclease, and exploitation of the endogenous tRNA-processing machinery (Nissim *et al.*, 2014; Xu *et al.*, 2017; Kurata *et al.*, 2018; Ferreira *et al.*, 2018; McCarty *et al.*, 2019; Ding *et al.*, 2018; Port and Bullock, 2016; Xie, Minkenberg and Yang, 2015; Knapp *et al.*, 2019). The ribozyme and tRNA mechanisms have also been successfully used to express sgRNA arrays and Cas9 from a single transcriptional unit using poly-A linkers, spliced introns, and internal ribosome entry sites (Yoshioka *et al.*, 2015; Wang *et al.*, 2018; Tang *et al.*, 2016; Ding *et al.*, 2018). While transcript sharing may help to co-localize CRISPR components, tRNA sequences can harbour internal Pol III promoters that boost expression but also background activity. This could help to enhance cleavage efficiency but may be undesirable in contexts where strict control of CRISPR activity is needed (Xie, Minkenberg and Yang, 2015; Knapp *et al.*, 2019). CRISPR-Cas9 gene drives using tRNA-multiplexed sgRNAs in *Drosophila* showed high homing efficiencies that were greater than most similar single sgRNA designs (Champer *et al.*, 2020; Yang *et al.*, 2022). Finally, ribozymes have also been successfully tested in *Drosophila* with high editing efficiencies from a Pol II polyubiquitin promoter (Marshall *et al.*, 2017). Although use of the tRNA or ribozyme systems for multiplexing sgRNAs has not yet been reported in mosquitoes, the ubiquity of tRNA systems across organisms and

the demonstrated functionality of both systems in a broad range of hosts hold promise for future application in mosquitoes.

### 4.4.3.ii Conclusion

This chapter aimed to integrate a SuperSIT construct targeting female flight within the male-determining M locus of *Ae. aegypti*, thereby achieving male-only inheritance and disruption of daughters' fitness. First, loci within the M-locus that were hypothesized to be amenable to transgenesis by CRISPR-Cas9 knock-in, not have a deleterious impact on male fitness once disrupted, and which would support efficient expression of a transgenic CRISPR-Cas9 construct in the male germline of *Ae. aegypti* were identified. sgRNAs designed to target these putative regions in the M locus showed high cleavage efficiencies *in vivo* in a high-throughput experiment that was designed. Thirteen of these sgRNAs, targeting distinct loci in the M locus were then chosen for knock-in of the SuperSIT construct. Transgenic lines with male-only inheritance were attained for five of the thirteen regions targeted, including two in the *Nix* intron and one in the *myo-sex* intron. Four of these were correctly integrated, showing that the HDR donor plasmid design was successful, and that the M locus was amenable to knock-in, even of relatively large constructs. Two of the four correctly integrated lines impacted female flight (up to 50% using a conservative assessment in some crosses), achieving proof of principle for this SuperSIT construct design. This is the first description of a SuperSIT system in *Ae. aegypti* and will provide a base for future research, a few ideas of which were discussed above.

# Chapter 5: General Discussion

## 5.1 Summary of chapters

This thesis sought to engineer a proof-of-principle SuperSIT population suppression system in the arboviral mosquito vector *Ae. aegypti* for non-invasive and sustainable vector control. The first and second research chapters explored mechanisms of disrupting the fitness of non-inheriting SuperSIT daughters using CRISPR-Cas9. Chapter 2 tested regulatory elements of endogenous genes that could drive Cas9 expression in the male germline to disrupt the pre-meiotic sperm genome or induce a paternal effect in the embryo. The chapter showed that high rates of mutated sperm could be achieved with current CRISPR designs, while engineering a paternal effect was more challenging. Chapter 3 then explored the activity of guide RNAs when targeting two female-specific flight genes, *Act-4* and *myo-fem*, and analysed the dominance of the knock-out effect at the specific target sites. The chapter highlighted a significant finding that high rates of dominance exceeding 0.9 were achievable with a multiplexed sgRNA construct cleaving exon two of *myo-fem*. This dominance rate reflected a null phenotype, so would correspond to a substantial reduction in female fitness. When *Act-4* was targeted at its C-terminal end, female offspring consistently exhibited incomplete dominance, varying around 0.5 when independently mutagenized at five sgRNA target sites.

Together, the first two research chapters showed that a promising route for SuperSIT development was through paternal inheritance of dominant *myo-fem* mutations. The final research chapter, Chapter 4, thus sought to assemble and express a complete CRISPR-Cas9 construct from loci within the M locus, ensuring male-only inheritance and disruption of flight ability in daughters. Transgenic lines with canonical integrations of a 16 kb SuperSIT construct were established at four loci within the M locus that all demonstrated male-only inheritance. Satisfactory expression was achieved from one of these sites where the integrated construct caused flightlessness in up to 50% of female offspring. This is the first description of a proof-of-concept SuperSIT system in *Ae. aegypti,* which will be valuable for future optimization in this species, as well as for its development in other related species. Further, having produced a successful method for establishing active integrations within the M locus, together with additional improvements discussed in the text, makes possible other molecular designs with similar

or enhanced dynamics compared to the SuperSIT system. This final discussion will first identify technical challenges common to all chapters and then discuss broader implications of this research project.

## 5.2 Technical challenges

### 5.2.1 Considerations for the current design

One of the greatest limiting factors to building a SuperSIT system in *Ae. aegypti* is achieving sufficient expression of the system from the heterochromatic M locus. While exchanging the regulatory elements that drive Cas9 with another set, like *shu,* will likely improve the rate of flightlessness at the tested loci, it may not be sufficient. Mitigating strategies were reviewed in Chapter 4 that highlighted the need to probe the chromatin state at putative integration sites with next-generation sequencing, to review the activity of Pol III promoters used to drive sgRNA expression—and simultaneously assess novel mechanisms for multiplexing and expressing sgRNAs from a single Pol II transcript—and, finally, to systematically investigate the impact of insulator sequences. It may otherwise be necessary to increase copy number of the SuperSIT construct or express Cas9 and sgRNA from different sites if sites compatible to the expression of both cannot be identified; the former may increase the risk of recombination if constructs are not sufficiently different, however, while the latter could possibly be avoided by expressing Cas9 and sgRNA from a single transcript.

Alternatively, integrating the transgene immediately outside of the M locus may improve expression at the expense of some, albeit low, recombination; rates of 0.01% were estimated for a locus 9 Mb from *Nix* that supported good expression of a genetic marker (Lutrat *et al.*, 2023). The bright expression allowed first instar larvae to be automatically sorted using a BioSorter, which would not have been possible for four out of the five transgenic lines made here (integration at the fifth site was non-canonical and will need to be verified). There is a risk that recombination could occur within the M locus anyway, though rarely, as evidence of *myo-sex* recombining to the m locus has been observed before (Hall *et al.*, 2014). However, rare recombination events of an M-linked construct to the m locus are not expected to be highly damaging to the SuperSIT system. The effect on release sizes and persistence in a population would be negligible at this low rate but would increase with increasing recombination rate and any male-associated fitness costs.

It would, however, necessitate special measures to ensure strain integrity when mass-rearing the strain, particularly if wild type males (recombinants that had lost the transgene) were substantially more fit. These measures will likely be necessary anyway due to the potential for loss-of-function mutations arising. The reverse, i.e. female recombinants that gain the transgene, would not be as problematic as these individuals would likely be non-viable due to sex lethal effects associated with the recombinant chromosome or fitness effects related to somatic expression of the transgene, and so would be selected out of the population (Krzywinska *et al.*, 2016b).

Further investigation into the associated fitness cost of disrupting either of the two flight genes, though especially *myo-fem* at exon two, will be needed. More sophisticated methods for assessing flight could help to discern the dominance of *myo-fem* more precisely than what the flight test device developed here could do. This will be important to detect subtle changes in wingbeat frequencies of mutants, which could have large impacts on reproductive success, especially in larger cages or the field. Importantly, though, even incomplete dominance, as long as the dominant effect is above 0.8, has been modelled to achieve sufficient population suppression within a reasonable time frame using a low release ratio (Burt and Deredec, 2018).

### 5.2.2 Limiting position effects

The primary transgenesis method utilized in this study for assessing the modular components of the SuperSIT system was *piggyBac* transposon-mediated integration. In insect genetics, *piggyBac* is frequently used due to its broad host range and high transformation efficiency that is about 5-10 times that of site-specific methods (Anderson *et al.*, 2024; Otte *et al.*, 2018; O'Brochta, 2022). However, due to its mostly random pattern of genome integration, a persistent technical challenge is the interpretation of transgene expression in the context of position effects. A common strategy to mitigate this is to generate several independent strains and compare transgene expression between them (Anderson *et al.*, 2023a). This can be useful for identifying homozygous viable and transcriptionally active transgene docking sites whose permissive endogenous regulatory environment (i.e. local promoters, enhancers, silencers, and chromatin structure) might otherwise be difficult to recapitulate, or in the context of genetic sensors for enhancer or gene trapping (Volohonsky *et al.*, 2015; Meredith *et al.*, 2011; Cherbas and Gong, 2014; Franz *et al.*, 2014; O'Brochta, 2022). As demonstrated here and

elsewhere, however, for the assessment of regulatory elements in the development and refinement of genetic control systems, including SuperSIT and gene drive, the usefulness of this strategy is limited compared to site-specific integration. Moreover, since many factors in the genetic designs of CRISPR-Cas9 systems are not independently varied, comparison across the relatively few studies in *Ae. aegypti* can make comparison even more challenging (Verkuijl *et al.*, 2022a).

Site-specific integration methods like CRISPR-mediated homology-directed repair (HDR) or phiC31-recombinase integration at docking sites offer alternatives to transposon integration. In *Ae. aegypti,* generating docking lines for phiC31 recombinase-mediated integration has not been routine practice, however, due to the low transformation efficiency (Haghighat-Khah *et al.*, 2015). This contrasts with *Drosophila* or *Anopheles,* where libraries of lines are available that can be shared between the research community, harmonizing findings. Recently, phiC31 integration and the associated Cre-*lox* system used for excision of undesired sequence were optimized for *Ae. aegypti*, expanding its usefulness for studies in this species (Carabajal Paladino et al., 2023). While this tool is a valuable resource, its transformation rates are comparable to those of CRISPR-mediated HDR, which may offer greater flexibility both in target site selection and final inserted transgene sequence.

Use of methods with low transformation efficiencies requires a higher initial investment in terms of the number of eggs that need to be microinjected. In Chapter 4, efforts were made to improve embryo survival rates from CRISPR-mediated HDR microinjection, as this is an additional bottleneck that reduces the efficiency of transformation experiments. Survival rates of 30-55% were achieved through incremental improvements in technique and methodology, surpassing the typical range of 5-15%. Additionally, the use of synthetic sgRNAs with stabilizing modifications was discussed as another potential modification to the CRISPR-mediated HDR protocol; other improvements have been reviewed elsewhere (Ang *et al.*, 2022; Finney, Romanowski and Adelman, 2022; Sun *et al.*, 2022). Such advancements could minimize the initial cost of site-specific methods, for which the downstream rearing process is generally less extensive than that of *piggyBac* integration (both in terms of isoline establishment and the number of transgene activity assays that must be performed to get an “average” readout). Further research and refinement of these techniques can advance the use of CRISPR-mediated HDR in genetic engineering

applications. For example, a novel method termed Receptor-Mediated Ovary Transduction of Cargo (ReMOT Control) circumvents the need for embryo microinjection by direct injection of CRISPR-Cas9 into the thorax of gravid females (Chaverra-Rodriguez *et al.*, 2018; Terradas *et al.*, 2022b; Li *et al.*, 2021b). This method can currently be used for germline editing, although mutation rates are relatively low. Research into delivering donor DNA for HDR is currently underway to adapt the technique for targeted knock-ins.

Novel transposase systems are being engineered that aim to increase the integration precision of *piggyBac* or other DNA transposons. Efforts have involved fusion of a secondary DNA-binding domain to the transposase protein that would direct the transposon to specific target sites (Luo *et al.*, 2017; Maragathavally *et al.*, 2006; Hew *et al.*, 2019; Owens *et al.*, 2013). More recently, a catalytically inactive "dead" Cas9 (dCas9) has been used to alter the insertional profile of various transposases (e.g. *Hsmar1*, *Himar, Sleeping Beauty,* and *piggyBac)*, with reports that it significantly increased specificity and allowed targeting of sites that were normally refractory to transposon insertions (Bhatt and Chalmers, 2019; Chen and Wang, 2019; Kovač *et al.*, 2020). While these systems are still being optimized, with the majority not yet tested in insects or insect cells, they could become important additions to the synthetic toolbox used to study insects.

Finally, identifying integration sites for efficient transgene expression is a challenge across eukaryotic systems, as was highlighted in Chapter 4 when targeting the M locus of *Ae. aegypti*. In gene drive and other genetic control systems, integration site selection will depend on the local regulatory environment and often, too, on unique genomic features like linkage to a sex-determining locus or chromosome, sequence conservation, endogenous genes with a phenotypic readout that is easily scorable (and viable) or causes a fitness effect, or simply a "neutral" site (e.g. within an intergenic region or an intron) (Anderson *et al.*, 2024; Li *et al.*, 2017b; Beaghton *et al.*, 2017; Noble *et al.*, 2017; Rojas-Fernandez *et al.*, 2015; Nash *et al.*, 2019). Choice of site may also be constrained by CRISPR-mediated knock-in efficiency, which can depend on availability of active sgRNAs and chromatin accessibility, as was discussed in Chapter 4 (Verkuijl and Rots, 2019). Furthermore, expression from some sites may be conducive to one transgene but not another (e.g. U6-sgRNA constructs that were particularly vulnerable to position effects (Chapter 3 and 4) or low CRISPR expression where antiviral effectors were highly active) or vary depending on the modular components of the integrated construct, which can

interfere with each other (Champer *et al.*, 2019; Terradas *et al.*, 2023; Reid *et al.*, 2022). Performing scATAC-Seq to detect regions of open chromatin will be critical in identifying multiple optimal integration sites. Such studies combined with transcriptomics have already led to the development of frameworks for enhancing heterologous gene expression in various eukaryotic organisms (O'Brien *et al.*, 2018; Brady *et al.*, 2020; Dhiman *et al.*, 2020; Boob *et al.*, 2024). A review of suitable target sites will ultimately help to harmonize research into regulatory elements and make research more transferrable between groups. Lastly, this will enable a solid exploration of tools like chromatin insulators or enhancer elements to promote robust transgene expression in general.

### 5.2.3 Resistance

Investigating the potential for resistance to evolve against the SuperSIT system will be critical. Strong selection pressures for resistance are expected to be exerted by any genetic intervention designed to suppress target populations (Alphey *et al.*, 2013; Esvelt *et al.*, 2014). Moreover, genetic systems that rely on nuclease cleavage can rapidly develop resistant alleles against themselves that are not already present in the target population due to the increased frequency of mutation via repair events (Hammond *et al.*, 2017; Unckless, Clark and Messer, 2017). This is especially risky for nuclease-based SuperSIT systems where NHEJ mutations are a desired outcome and the method by which it exerts its selective pressure on females. The extent of population suppression will thus be dependent on the probability of resistance developing at the target site. Notably, though, inheritance of the SuperSIT construct is independent of the reproductive load and resistance as the M (or Y) is not found in the females it targets.

Efforts should be made to minimize the risk of resistance evolving at the design stage. Mitigating strategies could include multiplexing sgRNAs, targeting multiple genes, targeting highly conserved sequences, and combining control strategies (Marshall *et al.*, 2017; Burt and Deredec, 2018; Geci, Willis and Burt, 2022). Already incorporated in this study was sgRNA multiplexing—which showed good activity could be achieved across individual guides—and investigation of other target sites that could be incorporated into the construct, namely the C-terminal region of *Act-4*. Although the site targeted in *Act-4* exhibited incomplete dominance, it could increase the chances of inheriting a dominant mutation and thus the reproductive load of the system (since *myo-fem* was not completely

haplo-insufficient), as well as further mitigate resistance. Care should be taken, however, as saturation of Cas9 could reduce cleavage at the primary target sites. Further, studies into the stoichiometric balance between *Act-4* and *myo-fem* will be required to ensure that function is not restored with two nulls, as was reported for *D. melanogaster*. However, preliminary results from the research group suggest this does not occur at least with the knock-in lines that were tested (personal communication from Michelle Anderson). Finally, target site conservation will need to be analysed, although strong selection pressures likely exist that maintain sequence integrity, at least in terms of protein coding potential, because of the risk of significant functional deficits if the mostly haplo-insufficient gene is altered.

### 5.2.4 Alternative mechanisms for engineering a dominant fitness effect in females

Exploring alternate strategies for delivering a dominant fitness effect to females will be important for future development of SuperSIT systems; for example, for use in combination with the current design, to increase the dominance effect (if a dominance rate of 0.9 is the maximum for targeting *myo-fem*), to mitigate resistance, or to apply the technology to other related species where orthologs of *myo-fem* are not found or are not sex-specific nor haplo-insufficient. Identifying other haplo-insufficient genes, particularly late-acting ones, is challenging, however. The initial model of the indirect flight muscles and the subsequent identification of the myofibril filament genes that were specialized to them emerged from forward genetics studies in *Drosophila melanogaster*. Forward genetics studies are challenging in *Ae. aegypti,* however, due to its large and repetitive genome and dearth of balancer chromosomes (Matthews and Vosshall, 2020). Thus, most of the work towards identifying gene function has involved reverse genetics from genes identified by homology to *Drosophila*, which could obscure some critically important genes necessary for mosquitoes. Developing these methods in *Ae. aegypti* has been proposed as an important goal for the field to better understand the underlying biological processes of mosquitoes, which diverged from *Drosophila* more than 260 million years ago (Arensburger *et al.*, 2010; Matthews and Vosshall, 2020). Single-cell and conventional omics studies will continue to be highly powerful tools, and coupled with REMOT control could make knock-out generation and phenotype screening easier, more modern, and high-throughput. Advancements in bioinformatics could also help identify putative haplo-

insufficient genes *in silico*, or help predict the likelihood and location of antimorphic mutations in female-specific genes (Shihab *et al.*, 2017; Yang *et al.*, 2021).

Preliminary results and evidence from the literature hold promise for the feasibility of engineering a paternal effect mediated SuperSIT system. Some of the challenges to depositing CRISPR-Cas9, such as the size of the molecule and difficulty in co-localization of Cas9 with sgRNAs, could be addressed by using other non-nuclease effectors in the paternal effect approach. Two such methods include the deposition of molecules (RNA or protein) that disrupt the sexual differentiation pathway in females but do not harm males, or that of toxins that are rescued by an antidote expressed zygotically (e.g. the Cif genes that have been implicated in *Wolbachia* cytoplasmic incompatibility); though this latter design is akin to another design called Maternal effect dominant embryonic arrest (Medea) (O'Leary and Adelman, 2020; Adelman and Tu, 2016; Burt and Deredec, 2018; Chen *et al.*, 2007; Akbari *et al.*, 2013b; McNamara *et al.*, 2024). Using effectors that do not cause double-stranded DNA breaks could be a way of mitigating resistance, too, as no error-prone repair events would be generated and thus may mitigate resistance more effectively. Challenges in generating enough mRNA or protein to cause a significant effect in the embryo remain, however.

A clear potential candidate for disrupting sexual differentiation in *Ae. aegypti* is the dominant male-determining factor, *Nix*. *Nix* is found within the M locus and is the primary signal in the sex determination cascade, determining the male-specific splice forms of the highly conserved genes *doublesex* (*dsx*), *transformer* (*tra*), and *fruitless* (*fru*) (Hall *et al.*, 2015). Knock-out of *Nix* results in feminized adults with developing ovaries, while ectopic expression of *Nix* in females can result in adults that are partially masculinized. Moreover, expression of *Nix* from an autosomal transgene was shown to be sufficient to convert genetic females into fertile phenotypic m/m males, while transgenic M/m males were fertile and had no significant differences in morphology (Aryan *et al.*, 2020). Non-genetic m/m males were flightless, however, due to the absence of another M-locus gene, *myo-sex*, which encodes the male-specific myosin gene of the indirect flight muscle. Thus, the phenotype of the *Nix*-converted males would be functionally sterile and flightless and be classed as a late-acting phenotype, similar to the disruption of *myo-fem* or *Act-4*. As discussed previously, lethality induced at later stages of mosquito development has been modelled to considerably increase the efficacy of vector control due to the preservation

of density-dependent effects that efficiently control larval survival (Phuc *et al.*, 2007; Burt and Deredec, 2018). While *Nix* is transcribed from the onset of zygotic genome activation (between 2-4h post oviposition)—presenting an opportunity for triggering this male-differentiation cascade in females through male deposited protein or mRNA—the gene is also expressed throughout development, so it is not known whether a single paternal affect signal would be enough, though ectopic expression did have a significant effect.

Other suggested designs rely on the homing of a dominant lethal allele in the male germline such that all females that inherit a single copy are non-viable whilst males develop normally; these are generally called fsRIDL-with-drive from the self-limiting female-specific Release of Insects carrying a Dominant Lethal design, though it does not "drive" as the name suggests (Thomas *et al.*, 2000). While these systems are limited by homing efficiencies, which are currently inadequate in *Ae. aegypti,* such a system could facilitate enhanced effector gene expression in transgenic females. Further, highly effective late acting dominant lethal genes have already been characterized and developed for fsRIDL systems against the mosquitoes *Ae. aegypti, Ae. albopictus,* and *Anopheles gambiae* (Labbé *et al.*, 2012; Thomas *et al.*, 2000; Phuc *et al.*, 2007; Smidler *et al.*, 2023). Dominant lethal effects could also be imposed by expression of dominant antimorphic variants of flight genes or other sex-specific genes. As discussed above, transgenic expression of the male factor *Nix* in *Ae. aegypti* females would also lead to a dominant effect (Adelman and Tu, 2016). Male factors have also been discovered in *An. gambiae, An. stephensi*, and *Ae. albopictus* (Criscione *et al.*, 2013; Criscione, Qi and Tu, 2016; Krzywinska *et al.*, 2016a; Lutrat *et al.*, 2022). Remarkably, in *Ae. albopictus,* transgenic expression of the endogenous *Nix* orthologue in female embryos resulted in fertile males that could also fly (Lutrat *et al.*, 2022). This effectively results in an autosomal dominant M locus that could be used to create a similar SuperSIT design by targeting the *Ae. albopictus myo-fem* orthologue or a population suppressing homing drive by sex conversion, which may or may not be desirable; such a drive could potentially be created in *Ae. aegypti* if *myo-sex* was included in the construct.

## 5.3 Conclusion

The SuperSIT population suppression system has the potential to be an effective technology for localized, sustainable, and environmentally friendly vector control. A SuperSIT system against *Ae. aegypti* has immense potential for application in the field and for use as a model to target other closely related mosquito vectors of arboviral diseases. Proof-of-concept of other self-limiting sex ratio distorters has now been demonstrated in *D. melanogaster*, the global agricultural pest *Ceratitis capitata*, and the malaria vector *An. gambiae,* for which different mechanisms of exerting a female-specific fitness effect are being intensely pursued (Fasulo *et al.*, 2020; Meccariello *et al.*, 2021; Tolosana *et al.*, 2024; Lawler *et al.*, 2024; Haber *et al.*, 2024). These developments highlight the rapid advancements in genetic engineering and underscore the demand for these types of genetic systems for the control of harmful species. Nevertheless, these genetic systems will certainly need to be tailored to individual species, as the same approach may not be possible nor appropriate in all cases. For *Ae. aegypti*, aside from the feasibility of engineering this system against it, certain life history traits, like adaptation to urban environments, limited dispersal, and desiccation resistant eggs, would make SuperSIT particularly suitable to its efficient control. Interventions could be centred around towns and cities where disease burden is highest, which simultaneously offers practical advantages in terms of reduced transport and handling and logistical complexity.

The realisation of these self-limiting strategies could cause a significant shift from first generation SIT and fsRIDL systems within the next decade. Conventional SIT and fsRIDL designs and *Wolbachia* have already demonstrated the possibility of successful suppression or modification of *Ae. aegypti* populations and built operational capacity that will be critical for future deployment of these designs. The successes of the past two decades have also reduced uncertainty around genetic vector control among regulators and other stakeholders and will likely improve acceptance of more efficient systems, like SuperSIT and others, in future (Geci, Willis and Burt, 2022; Burt and Deredec, 2018). As with many grand challenges, successful control of arboviral diseases will require cross-sector collaboration and use of these promising new tools in complement with existing vector control methods. Such synergy within integrated pest management approaches will be vital, especially at a time when the frequency of arboviral disease outbreaks is increasing.

Engagement and collaboration with governments and local communities continues to be crucial prior to and throughout these genetic control interventions. Fortunately, there is a strong drive from within the research community to proactively ensure fair and ethical governance of both driving and non-driving technologies (Long *et al.*, 2020). Greater global collaboration and sharing of knowledge and resources will be needed to build partnerships with and capacity in low- and middle-income countries where these technologies are expected to have the highest positive impact. There is a unique opportunity here to benefit entire communities regardless of considerable infrastructure or social and economic inequities. It is hoped that this will continue to be a driving principle behind open and transparent research as the pace of development, demand for new and sustainable interventions, and competition for patents inevitably increase. Biotechnology has greatly expanded the societal challenges that can be sustainably addressed within a reasonable time frame, from human health to agricultural productivity to biodiversity conservation, and progress is expected to continue dramatically over the next decade.

# Appendix

## Appendix A: Supplementary material for Chapter 2

*Appendix A Table 1 Primers used in PCR reactions to clone plasmids AGG-2141 and AGG-2142. Blue text denotes primers used to build both plasmids*

| Primer number | Primer name | Fragment | Template | Oligo sequence |
|---|---|---|---|---|
| To clone precursor plasmids AGG2083 and 2084 | | | | |
| 6377 | 2083_F01 | B2-DsR | AGG-1918 | TACCACTGGCGCGTTTGGGGGCCGGCCGGAGTTCTTTCACAAACGTATACTACAAAGCTTTTAAAAAAC |
| 6376 | 2083_R01 | | | CACGAAAATCTGCGGGTCCTCCACCTTCCGCT |
| 6372 | 2084_F01 | Prot-DsR | AGG-1920 | TACCACTGGCGCGTTTGGGGGCCGGCCATTTGCAATCCTATAAGAAACTTCAAGTGTAAGAGAAATTATTTTGTGAC |
| 6376 | 2083_R01 | | | CACGAAAATCTGCGGGTCCTCCACCTTCCGCT |
| 6375 | 2083_F02 | UbS-Cas9 | AGG-2139 | GGTGGAGGACCCGCAGATTTTCGTGAAGACCCTG |
| 6374 | 2083_R02 | | | GATCAATTAATTACTTTTTCTTTTTTGCCTGGC |
| 6373 | 2083_F03 | SV40 | AGG-1918 | AAAAAGAAAAAGTAATTAATTGATCATAATCAGCCATAC |
| 6279 | 2082_R2 | | | AAGTTATCCTAGGACTAGTGGCGCGCCTAAGATACATTGATGAGTTTG |
| To replace marker cassettes and clone final plasmids AGG2141 and 2142 | | | | |
| 6583 | attB-3xP3_F | 3xP3-mCherry-K10(AflII) | AGG-1927 | GTCACGGCGGGCATGTCGACTTAATTAACCCGGGGATCTAATTCAATTAG |
| 6584 | K10-AflIII_R | | | ATGGATTGTTGGCCAAACTTGTCGCTTAAGTGGCTGGGAATTGTAACCGTAG |
| 6583 | attB-3xP3_F | 3xP3-mCherry-K10(Bsu36I) | AGG-1927 | GTCACGGCGGGCATGTCGACTTAATTAACCCGGGGATCTAATTCAATTAG |
| 6585 | K10-Bsu36I_R | | | TTTGCAATTAAATTACTTTGCTACCTGAGGGTTAAATCTTCCCCATTGATAATAATAATTC |
| Sequencing primers used to verify sequence of plasmids AGG2141 and AGG2142. | | | | |
| 7384 | M13-48REV | | AGG-2141 | CGGATAACAATTTCACACAG |
| 6819 | LA674 | | | GTGCTATGTTTATGGCGCTATTAAC |
| 3473 | LA3473_mCherry_SbfI_F1 | | | CTCCCTGCAGGACGGCGAGT |
| 6098 | K10_F | | | ACTAAGCGCGGTGAAACCCAATTGAAC |
| 5641 | Beta2-F | | | ACTCTCAAGAATTGCTCCAAATCCA |
| 5699 | Beta2-R4 | | | AAGGACGTGCTGAAGTGGTATGCGG |
| 5700 | Beta2-R5 | | | GCGACGTATCCACACAGTAACCATGA |
| 6718 | 979.B2-tubulin-R.colony | | | TACGTCTCCTTTGTGGCAGAGCATG |
| 6716 | 2566.ddDsRED R | | | GGAAGGACAGCTTCTTGTAGTC |
| 6068 | DsRed_fwd_1 | | | CGAGAACGTCATCACCGAGTTCATGC |
| 5160 | 5160.NLS.F | | | GACCCAAGAAAAAGCGGAAGGTGG |
| 6713 | 1757.hCas9_Fw_339--> | | | GGAAGAGGATAAGAAGCACGAGC |
| 6711 | 1759.hCas9_Fw_1341--> | | | AAGATCCTGACCTTCCGCATCC |
| 4212 | hSpCas9_qPCR_R1 | | | GCCTTATCCAGTTCGCTCAG |
| 4210 | hSpCas9_qPCR_F1 | | | CCGAAGAGGTCGTGAAGAAG |
| 3189 | M13F-Phos | | | CGCCAGGGTTTTCCCAGTCACGAC |

| | | | |
|---|---|---|---|
| 7384 | M13-48REV | AGG-2142 | CGGATAACAATTTCACACAG |
| 6819 | LA674 | | GTGCTATGTTTATGGCGCTATTAAC |
| 3473 | LA3473_mCherry_SbfI_F1 | | CTCCCTGCAGGACGGCGAGT |
| 6098 | K10_F | | ACTAAGCGCGGTGAAACCCAATTGAAC |
| 6378 | Prot-F2 | | GTTGATGAAGTTGTACTGGTCCAT |
| 5702 | Prot-R3 | | TAACCGCAAGGCTACGAAACCCACC |
| 5639 | Prot-R1 | | ATATGCTTGGACGCGGTCTGGTTA |
| 5522 | Prot prom_r | | GATCTCCCATCTTGATCCAATAAGTGTGTAAAG |
| 1393 | 1393.DSRED.QPCR | | CAGCCCATTGTCTTCTTCTGCATCAC |
| 6560 | 5160.NLS.R | | CCACCTTCCGCTTTTTCTTGGGTC |
| 5160 | 5160.NLS.F | | GACCCAAGAAAAAGCGGAAGGTGG |
| 6713 | 1757.hCas9_Fw_339--> | | GGAAGAGGATAAGAAGCACGAGC |
| 4212 | hSpCas9_qPCR_R1 | | GCCTTATCCAGTTCGCTCAG |
| 4210 | hSpCas9_qPCR_F1 | | CCGAAGAGGTCGTGAAGAAG |
| 4198 | aCas9_qPCR_F1 | | CAGCACAAGCACTACCTGGA |

# Appendix B: Supplementary material for Chapter 4

*Appendix B Table 1 The sequence of each M locus region was retrieved from VectorBase using the specified reference gene and the number of bases relative to the gene start and stop. The total length of the sequence retrieved is also listed.*

| Region name | VectorBase gene used for sequence retrieval | Region of retrieved sequence relative to gene | Length (bp) |
|---|---|---|---|
| *Myosex* intron | AAEL021838(+) | geneStart-3000 to geneEnd+3000 | 291,969 |
| *Myosex* downstream | AAEL025579(+) | geneStart-10000 to geneEnd+85000 | 95,810 |
| Intergenic region 1 | AAEL022756(+) | geneStart-1000 to geneEnd+53000 | 54,520 |
| Intergenic region 2 | AAEL020813(+) | geneStart-2000 to geneEnd+40000 | 42,383 |
| *Nix* downstream | AAEL022912(-) | geneCodeEnd+0 to geneEnd+75000 | 75,001 |
| *Nix* intron | AAEL022912(-) | geneStart-3000 to geneEnd+3000 | 107,503 |
| *Nix* upstream | AAEL022321(-) | geneStart-15000 to geneEnd+27000 | 45,207 |

*Appendix B Table 2 No. of sgRNAs available in each M locus region with an efficiency score above 0.2 and fewer than two predicted off-targets with three mismatches only*

| Region | sgRNAs in region |
|---|---|
| *Myosex* Intron (sequence split into three searches) | 196; 236; 234 |
| *Myosex* Downstream | 149 |
| Intergenic Region 1 | 79 |
| Intergenic Region 2 | 26 |
| *Nix* Downstream | 55 |
| *Nix* Intron | 138 |
| *Nix* Upstream | 100 |
| Total | 1213 |

*Appendix B Table 3 Primer pairs used to successfully amplify each of the 32 subregions from individual male gDNA. Subregions that could not be amplified are indicated.*

| Region | Subregion | F primer | Number | Sequence | R primer | Number | Sequence | bp |
|---|---|---|---|---|---|---|---|---|
| *Myosex* Intron | 1A | Could not be amplified; 2 sets of primers tested | | | | | | |
| | 3B | Myosex_Int3B_F2 | 603 | GACTCTTCCTGGTATTAACTCG | Myosex_Int3B_R2 | 605 | GACTCGTAATGTGTACAAAGGC | 2239 |
| | 4B | Myosex_Int4B_F2 | 609 | CTGTACCGAATCTCGAAATC | Myosex_Int4B_R1 | 610 | CGTTACAGCTAATGTCAAGTGA | 2854 |
| | 4C | Myosex_Int4C_F1 | 614 | TCTCTCTCTAAGCCTCATCCTA | Myosex_Int4C_R2 | 616 | GACCACTCATCAACTCTTACGT | 2761 |
| | 5A | Myosex_Int5A_F1 | 617 | GAGTGGAGTATGACGTTCAACT | Myosex_Int5A_R1 | 619 | CTATATGTTGAACAGCATCCC | 2499 |
| | 6 | Myosex_6_F2 | 624 | TAGAACATCGAAATGTGGAGAG | Myosex_6_R2 | 626 | TGGCTCATAGAGCTTATAGCTT | 3483 |
| | 7 | Myosex_7_F1 | 631 | CTTCGGAAGACATTTACACTCT | Myosex_7_R2 | 634 | ATAGCTTATCACAGAGTGGTCC | 3482 |
| *Myosex* Downstream | 1 | Myosex_DS_F1 | 10 | GTAGAACCTGGAAGAGCTTTGT | Myosex_DS_R1 | 13 | ATAGATGACCTCTATGACCTCC | 5008 |
| | 2 | Myosex_DS2_F1 | 20 | CAGATTCGAATGCAGACAGTAC | Myosex_DS2_R2 | 24 | CGAAGTGGTAGTTGTGAAGAG | 2330 |
| | 2B | Myosex_DS2B_F1 | 37 | AAGTCCCTTTCCTCTTGTAGTC | Myosex_DS2B_R1 | 39 | CATAATGGTGACTCACACACAG | 2041 |
| | 4 | Myosex_DS4_F1 | 26 | GTGCTATCCTGTCTTGATGTG | Myosex_DS4_R1 | 28 | TGAACTATCTGTGGTTGCTACC | 1740 |
| | 5A | Could not be amplified; 2 sets of primers tested | | | | | | |
| | 5B | Could not be amplified; 2 sets of primers tested | | | | | | |
| | 7 | Myosex_DS7_F1 | 31 | TTAGTCTTCAGCTTGTAGACCC | Myosex_DS7_R1 | 33 | GAGAGACGGTTACATGGTACTT | 2636 |
| Intergenic Region 1 | 1 | Could not be amplified; 1 set of primers tested | | | | | | |
| Intergenic Region 2 | 1 | IG2_1_F1 | 661 | GATAGACTTTGGATGTCATTCC | IG2_1_R2 | 663 | CTGATTACATACGTGAAACAGC | 2534 |
| | 2 | IG2_2_F1 | 666 | AAAGTCTCCACACAGCAGTAAC | IG2_2_R2 | 669 | CTTAGAAAACTCCCCATCTTAC | 2178 |
| | 3 | IG2_3_F1 | 672 | ATGAGGAAGTCATGTCTGTGTT | IG2_3_R2 | 674 | GTCTCGTAAAATGGCTTATACC | 2101 |
| *Nix* Downstream | 1 | Nix_DS1_F1 | 655 | CTAACTTGGTGGGAATCTCTAG | Nix_DS1_R2 | 658 | CACTTCTCCTACACAGACCAAT | 2644 |
| | 2 | Nix_DS2_F1 | 649 | GACTCAAGGCTTTAAGGTAAAG | Nix_DS2_R2 | 652 | GTCAATATTCGGAGCTACGTAC | 2740 |
| | 3 | Nix_DS3_F2 | 644 | GGTTGTTACCACTGATATGTCC | Nix_DS3_R2 | 646 | TAATCTCTGGAGCGTCTAGATT | 2329 |
| | 4 | Nix_DS4_F1 | 638 | CGAATGGTCACATCTAGTATCC | Nix_DS4_R2 | 641 | CAGAAGAATCACAACCTGTCTT | 2692 |
| Nix Intron | 1A | NI1_F2 | 551 | AATCCACACACATTCTCCAC | NI1_R1 | 552 | AGACAGTCGATAGTTCTCGTGT | 2981 |
| | 3A | NR3A_F2 | 523 | TGTACTGGAACCCTATGTCTTC | NR3A_R2 | 525 | AGTTTATCTGACTACGACGAGG | 2248 |
| | 3 | NR3_F2 | 517 | GATGAGTACACACACGCAGTAG | NR3_R2 | 519 | CGTACACTTTCACTGACGTGTA | 2249 |
| | 3B | NR3B_F2 | 510 | GATAGGCAGACATTGTACGAGT | NR3B_R1 | 511 | GAGATTCACGGAAGAACTACTG | 2985 |
| | 4 | NixInt4_F4 | 7788 | CTACACTACCCGCCATAAGTATAG | NixInt4_R3 | 7790 | GGACAGCTGTAAGAAGAGAGACAG | 4442 |
| | 5 | NI5_F1 | 562 | GTAGTTCCTCATCAGTTCCAGA | NI5_R1 | 564 | AGTCTACTTGAGTATTCCGTGC | 2448 |
| *Nix* Upstream | 1 | NixUp1_F2 | 8032 | TAACTGGGCTCTAAGGTTAGTC | NixUp1_R2 | 8033 | GGATAGCGATATGGAGTCTATC | 4462 |
| | 1A | NixUp1A_F2 | 545 | GGACTGGACTAGATGATGAGAA | NixUp1A_R1 | 546 | GACTAACCTTAGAGCCCAGTTA | 2494 |
| | 2B | NixUp2B_F1 | 556 | AAGTACCATGTAACCGTCTCTC | NixUp2B_R1 | 557 | AGTGTACGGGGATAGACTATTG | 1778 |
| | 3 | NixUp3_F3 | 8058 | CTGGTCACTAAACATACCCTTC | NixUp3_R3 | 8061 | ACTGTTCTTAGCGTGTAGAGCT | 4227 |

*Appendix B Table 4 Primers that failed to amplify a PCR product in the specified M locus subregions.*

| Subregion | F primer | Number | Sequence | R primer | Number | Sequence | bp |
|---|---|---|---|---|---|---|---|
| *Myosex* Intron 1A | Myosex_Int1A_F1 | 536 | TTCTAAGAGTGCCACACCTTAC | Myosex_Int1A_R1 | 538 | GTTACAGAACCGTGTAGTGTCC | 3218 |
| | Myosex_Int1A_F1 | 536 | TTCTAAGAGTGCCACACCTTAC | Myosex_Int1A_R2 | 598 | ACAGAGAGATCAGGCATCTGT | 3255 |
| | Myosex_Int1A_F2 | 537 | GCAAAGCAGGTAGTTCATG | Myosex_Int1A_R1 | 538 | GTTACAGAACCGTGTAGTGTCC | 3099 |
| | Myosex_Int1A_F2 | 537 | GCAAAGCAGGTAGTTCATG | Myosex_Int1A_R2 | 598 | ACAGAGAGATCAGGCATCTGT | 3136 |
| *Myosex* downstream 5A | Myosex_DS_5A_F1 | | AGGTCCACAGCTACATTTAGTG | Myosex_DS_5A_R1 | | TGAGTCTACTAGAAGAGCCGAC | 2343 |
| | Myosex_DS_5A_F1 | | AGGTCCACAGCTACATTTAGTG | Myosex_DS_5A_R2 | | GATGCTAAGTTCCTGGAGTTAG | 2506 |
| | Myosex_DS_5A_F2 | | ACTAGGGATCTCAATCAGGATC | Myosex_DS_5A_R1 | | TGAGTCTACTAGAAGAGCCGAC | 2268 |
| | Myosex_DS_5A_F2 | | ACTAGGGATCTCAATCAGGATC | Myosex_DS_5A_R2 | | GATGCTAAGTTCCTGGAGTTAG | 2431 |
| *Myosex* downstream 5B | Myosex_DS_5B_F1 | | AGTAGATGCTTCTCCACATCAC | Myosex_DS_5B_R1 | | ACTGATCTTAGGCTAGTCGATG | 1398 |
| | Myosex_DS_5B_F1 | | AGTAGATGCTTCTCCACATCAC | Myosex_DS_5B_R2 | | CTGTTGGAGAGAAAAGTACCAG | 1706 |
| | Myosex_DS_5B_F2 | | GATACCTGTACATTGGACGACT | Myosex_DS_5B_R1 | | ACTGATCTTAGGCTAGTCGATG | 1974 |
| | Myosex_DS_5B_F2 | | GATACCTGTACATTGGACGACT | Myosex_DS_5B_R2 | | CTGTTGGAGAGAAAAGTACCAG | 2282 |
| Intergenic region 1 | IG1_1_F1 | 677 | TACCACAGACAAACAGACGTAA | IG1_1_R1 | 678 | GTCACTGGGATATGAAGCTAAT | 2138 |
| | IG1_1_F1 | 677 | TACCACAGACAAACAGACGTAA | IG1_1_R2 | 679 | ACATTGCTATATCTCGAGATCC | 1937 |

*Appendix B Table 5 Primers used to sequence each subregion successfully amplified from all 16 individual male gDNA samples.*

| Subregion | Name | No. | Sequence | sgRNA covered |
|---|---|---|---|---|
| *Myosex* intron | | | | |
| 3B | Myosex_Int3B_Seq_F1 | 606 | GTCGTGTAAGGAAAGTGACAAG | 1, 71, 21, 81, 34, 6, 147 |
| | Myosex_Int3B_Seq_F2 | 607 | AGTGCATAGAAAACTGTGGC | 81, 34, 6, 147, 184, 76, 31, 65, 29 |
| 4B | Myosex_Int4B_Seq_F1 | 612 | CTGTACCGAATCTCGAAATC | 14, 29, 200 |
| | Myosex_Int4B_Seq_F2 | 613 | GATATGAGTCATGAGTCTCGCT | 125, 147, 217, 158, 15, 26, 175 |
| | Myosex_Int4B_R1 | 610 | CGTTACAGCTAATGTCAAGTGA | 42, 93 |
| 4C | Myosex_Int4C_F1 | 614 | TCTCTCTCTAAGCCTCATCCTA | 44, 204, 50, 82 |
| | Myosex_Int4C_R1 | 615 | CTGGTCTTGCTCTTCTAGACAT | 19, 20 |
| 5A | Myosex_Int5A_Seq_F1 | 621 | AAAACTCTACGAGAAATCGG | 45, 148, 101, 216 |
| | Myosex_Int5A_Seq_F2 | 622 | GTGAAGCTTTTCTTCGAAG | 93, 53, 21, 14, 139, 75 |
| 6 | Not sent for sequencing-non-specific band in gel | | | |
| 7 | Not sent for sequencing-not all samples amplified | | | |
| *Myosex* downstream | | | | |
| 1 | Myosex_DS_F3 | 12 | GCTCAATAAGCTATCCTGGA | 27, 61, 56 |
| | Myosex_DS_Seq_F1 | 16 | GTAATCTTCTCTCGGTGAGTCA | 105, 118, 100 |
| | Myosex_DS_Seq_F2 | 17 | CTTCTACAGAGCAGCCTGAC | 45, 13, 43, 52 |
| | Myosex_DS_Seq_F3 | 19 | CAGCAAACACATCGATACC | 10, 59, 134 |
| 2 | Myosex_DS2_F2 | 21 | GAGTGGCAGATGTGAGAATC | 55, 65, 67, 21 |

| | | | | |
|---|---|---|---|---|
| | Myosex_DS2_Seq_F1 | 25 | CAAAGGACCCCAGTAGATAC | 130, 30, 42, 20, 66, 75, 51 |
| 2B | Myosex_DS2B_F1 | 37 | AAGTCCCTTTCCTCTTGTAGTC | 6, 29 |
| | Myosex_DS2B_Seq_F1 | 41 | TGTGGCACTATTGAACTACG | 146, 17, 2 |
| 4 | Myosex_DS4_R1 | 28 | TGAACTATCTGTGGTTGCTACC | 1, 142 |
| | Myosex_DS4_Seq_F1 | 30 | CTTCCAGAATATTCGGCTC | 1, 142 |
| 7 | Myosex_DS7_Seq_F1 | 35 | GAACTTGAAAGGCTGACGT | 3, 85 |
| | Myosex_DS7_Seq_R1 | 36 | GGACTTACTCAAACCACTGTCA | 3, 85 |
| *Nix* downstream | | | | |
| 1 | Nix_DS1_Seq_F1 | 659 | CTTACCTTGCACGCTAAAG | 20, 29 |
| | Nix_DS1_Seq_R1 | 660 | TTAGACAGTTCCCGTATGTCAT | 20, 29 |
| 2 | Nix_DS2_F2 | 650 | AAGGCAAAGGTGAGTACTAGAG | 9, 42, 32, 2, 4 |
| | Nix_DS2_Seq_F1 | 653 | GTTTTTCCTCTCGGATGAT | 41, 5, 39, 11 |
| 3 | Nix_DS3_Seq_F1 | 647 | ACACGTTTCACCGAGAGA | 15, 19, 44 |
| | Nix_DS3_Seq_R1 | 648 | CAATAGCCAGAATCAGCTCTAC | 15, 19, 44 |
| 4 | Not sent for sequencing-non-specific band in gel | | | |
| *Nix* intron | | | | |
| 1A | NI1_Seq_F1 | 553 | ACTCAATACTACCTCGAGAACG | 35 |
| | NI1_Seq_F2 | 554 | CAACATTCGTAACAACTCCG | 2, 3, 68, 122, 74 |
| 3A | NR3A_Seq_F1 | 526 | GTCGAAGATGGTAGGCTTTT | 5 |
| | NR3A_Seq_R1 | 527 | CACCGGATGGATCATTAC | 5 |
| 3 | NR3_Seq_F1 | 520 | ATTCAGAACCTGAGGCATTG | 76,88,18 |
| | NR3_Seq_R1 | 521 | CACTGCTAGACGGTTGAAAC | 76,88,18 |
| 3B | NR3B_Seq_F1 | 514 | CAAGAAACATACCTGCTGC | 43,29,6,22,11,66,79,80, 103, 136 |
| | NR3B_Seq_R2 | 515 | GGGCGCGATTAGTACTTA | 43,29,6,22,11,66,79,80 |
| 4 | NixInt4_F_Seq4 | 7782 | GTTGTCCTCGTGAGATCG | 33 |
| | NixInt4_F_Seq7 | 7785 | GTACACATCGTTGAGCGTAAC | 23, 32, 14, 12, 82, 10, 104, 110, 115, 132 |
| | NixInt4_R_Seq9 | 8045 | GAAATCCGACTCATCTTCC | 15, 8, 23, 32, 14, 12 |
| | NixInt4_R_Seq10 | 8047 | ACAGTTGACAGCACAACCTT | 28, 40, 21 |
| 5 | NI5_SEQ_F1 | 565 | GAGCCGTACTACCTGAACTAAG | 49, 102 |
| | NI5_SEQ_R1 | 566 | AGTACTTCTCCGTATCAGAC | 49, 102 |
| *Nix* upstream | | | | |
| 1 | NixUp1_F_Seq3 | 8036 | TTCTGAGAAGTAGTTGGCGTAC | 22, 19, 4, 5, 6, 45 |
| | NixUp1_F_Seq4 | 8037 | CAAGGAGAGGCTTTAGATTACC | 4, 5, 6, 45, 10 |
| | NixUp1_F_Seq5 | 8038 | CGGAAGATTGTTTAGTGCTG | 10 |
| | NixUp1_F_Seq8 | 8041 | GTGGTGAATTTACGATCAGC | 13 |
| 1A | NixUp1A_Seq_F1 | 547 | GGATACGAACCGTTCAGA | 46, 39 |
| | NixUp1A_Seq_F2 | 548 | ACTACAAAGGTCACGCAGTC | 46, 39 |
| | NixUp1A_Seq_F3 | 549 | GTGGTTGTGGAAGAACTAGAAG | 14, 44 |
| 2B | NixUp2B_Seq_F1 | 559 | CTATTCAGCTGTCAGTGTGCT | 15, 47, 23 |

| | | | | |
|---|---|---|---|---|
| | NixUp2B_Seq_F2 | 561 | CTTTGTTGTGTTGTCAAGGC | 12, 25, 16, 35 |
| 3 | NixUp3_R2 | 8060 | GAGACTCGCTAGTATTTCGTCT | 1, 8, 21, 36, 3, 26, 40 |
| | NixUp3_F_Seq1 | 8062 | ACACAAGGAAGCAACGAC | 20, 32, 1, 8, 21, 36 |
| | NixUp3_R_Seq1 | 8065 | GAGTTTTAGGCCTTCCAAAC | 20, 32, 1, 8, 21, 36 |
| | NixUp3_F_Seq4 | 8066 | ATGATCCTGCAATGGACC | 7, 9, 11, 37, 2 |
| Intergenic region 2 | | | | |
| 1 | IG2_1_R1 | 662 | GGATGAAGTAGTAAAGCAAAGC | 3 |
| | IG2_1_Seq_F1 | 664 | GTTTTTCCGGTCATTCAAC | 17 |
| | IG2_1_Seq_R1 | 665 | CCTTCAGATCTTCCATTTTC | 23 |
| 2 | Not sent for sequencing-non-specific band in gel | | | |
| 3 | Not sent for sequencing-non-specific band in gel | | | |

*Appendix B Table 6 Primers used to amplify each amplicon in the adapter ligation PCR.*

| Amplicon | Primer pair | Primer name | LA no. | Sequence | Length (bp) | Samples used on | No. of samples |
|---|---|---|---|---|---|---|---|
| Myosex intron 3B | 1 | Myosex_Int3B_F2_ONP | 8121 | TTTCTGTTGGTGCTGATATTGCGACTCTTCCTGGTATTAACTCG | 2239 | 1-10 and uninj | 11 |
| | | Myosex_Int3B_R2_ONP | 8122 | ACTTGCCTGTCGCTCTATCTTCGACTCGTAATGTGTACAAAGGC | | | |
| Myosex intron 4C | 2 | Myosex_Int4C_F1_ONP | 8123 | TTTCTGTTGGTGCTGATATTGCTCTCTCTCTAAGCCTCATCCTA | 2761 | 1, 2, 5, 8 and uninj | 5 |
| | | Myosex_Int4C_R2_ONP | 8124 | ACTTGCCTGTCGCTCTATCTTCGACCACTCATCAACTCTTACGT | | | |
| Myosex intron 4B | 3 | Myosex_Int4B_F2_ONP | 8125 | TTTCTGTTGGTGCTGATATTGCCTGTACCGAATCTCGAAATC | 2504 | 3, 4, 6, 7, 9, 10 and uninj | 7 |
| | | Myosex_Int4B_R1_ONP | 8126 | ACTTGCCTGTCGCTCTATCTTCCGTTACAGCTAATGTCAAGTGA | | | |
| Myosex intron 5A | 4 | Myosex_Int5A_F1_ONP | 8127 | TTTCTGTTGGTGCTGATATTGCGAGTGGAGTATGACGTTCAACT | 2499 | 1-10 and uninj | 11 |
| | | Myosex_Int5A_R1_ONP | 8128 | ACTTGCCTGTCGCTCTATCTTCCTATATGTTGAACAGCATCCC | | | |
| Myosex downstream 1 | 5 | Myosex_DS_F1_ONP | 8129 | TTTCTGTTGGTGCTGATATTGCGTAGAACCTGGAAGAGCTTTGT | 5008 | 1, 6, 7, 9, 11 and uninj | 6 |
| | | Myosex_DS_R1_ONP | 8130 | ACTTGCCTGTCGCTCTATCTTCATAGATGACCTCTATGACCTCC | | | |
| Myosex downstream 2 | 6 | Myosex_DS2_F1_ONP | 8131 | TTTCTGTTGGTGCTGATATTGCCAGATTCGAATGCAGACAGTAC | 2330 | 2, 5 and uninj | 3 |
| | | Myosex_DS2_R2_ONP | 8132 | ACTTGCCTGTCGCTCTATCTTCCGAAGTGGTAGTTGTGAAGAG | | | |
| Myosex downstream 4 | 7 | Myosex_DS4_F1_ONP | 8133 | TTTCTGTTGGTGCTGATATTGCGTGCTATCCTGTCTTGATGTG | 1740 | 3 only and uninj | 2 |
| | | Myosex_DS4_R1_ONP | 8134 | ACTTGCCTGTCGCTCTATCTTCTGAACTATCTGTGGTTGCTACC | | | |
| Myosex downstream 2b | 8 | Myosex_DS2B_F1_ONP | 8135 | TTTCTGTTGGTGCTGATATTGCAAGTCCCTTTCCTCTTGTAGTC | 2041 | | 4 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | Myosex_DS2B_R1_ONP | 8136 | ACTTGCCTGTCGCTCTATCTTCCATAATGGTGACTCACACACAG | | 4, 8, 10 and uninj | |
| Myosex downstream 7 | 9 | Myosex_DS7_F1_ONP | 8137 | TTTCTGTTGGTGCTGATATTGCTTAGTCTTCAGCTTGTAGACCC | 2636 | 1, 2 and uninj | 3 |
| | | Myosex_DS7_R1_ONP | 8138 | ACTTGCCTGTCGCTCTATCTTCGAGAGACGGTTACATGGTACTT | | | |
| Nix downstream 2 | 10 | Nix_DS2_F1_ONP | 8139 | TTTCTGTTGGTGCTGATATTGCGACTCAAGGCTTTAAGGTAAAG | 2740 | 1, 5, 6, 8, 10 and uninj | 6 |
| | | Nix_DS2_R2_ONP | 8140 | ACTTGCCTGTCGCTCTATCTTCGTCAATATTCGGAGCTACGTAC | | | |
| Nix downstream 3 | 11 | Nix_DS3_F2_ONP | 8141 | TTTCTGTTGGTGCTGATATTGCGGTTGTTACCACTGATATGTCC | 2329 | 2, 3, 7 and uninj | 4 |
| | | Nix_DS3_R2_ONP | 8142 | ACTTGCCTGTCGCTCTATCTTCTAATCTCTGGAGCGTCTAGATT | | | |
| Nix downstream 1 | 12 | Nix_DS1_F1_ONP | 8143 | TTTCTGTTGGTGCTGATATTGCCTAACTTGGTGGGAATCTCTAG | 2644 | 4, 9 and uninj | 3 |
| | | Nix_DS1_R2_ONP | 8144 | ACTTGCCTGTCGCTCTATCTTCCACTTCTCCTACACAGACCAAT | | | |
| Nix Intron 4 | 13 | NixInt4_F4_ONP | 8145 | TTTCTGTTGGTGCTGATATTGCCTACACTACCCGCCATAAGTATAG | 4442 | 1, 5-9 and uninj | 7 |
| | | NixInt4_R3_ONP | 8146 | ACTTGCCTGTCGCTCTATCTTCGGACAGCTGTAAGAAGAGAGACAG | | | |
| Nix intron 3B | 14 | NR3B_F2_ONP | 8147 | TTTCTGTTGGTGCTGATATTGCGATAGGCAGACATTGTACGAGT | 2985 | 2, 3, 4, 10 and uninj | 5 |
| | | NR3B_R1_ONP | 8148 | ACTTGCCTGTCGCTCTATCTTCGAGATTCACGGAAGAACTACTG | | | |
| Nix intron 1A | 15 | NI1_F2_ONP | 8149 | TTTCTGTTGGTGCTGATATTGCAATCCACACACATTCTCCAC | 2981 | 4-9 and uninj | 7 |
| | | NI1_R1_ONP | 8150 | ACTTGCCTGTCGCTCTATCTTCAGACAGTCGATAGTTCTCGTGT | | | |
| Nix intron 5 | 16 | NI5_F1_ONP | 8151 | TTTCTGTTGGTGCTGATATTGCGTAGTTCCTCATCAGTTCCAGA | 2448 | 10-11 and uninj | 3 |
| | | NI5_R1_ONP | 8152 | ACTTGCCTGTCGCTCTATCTTCAGTCTACTTGAGTATTCCGTGC | | | |
| Nix upstream 3 | 17 | NixUp3_F3_ONP | 8153 | TTTCTGTTGGTGCTGATATTGCCTGGTCACTAAACATACCCTTC | 4227 | 1-7 and uninj | 8 |
| | | NixUp3_R3_ONP | 8154 | ACTTGCCTGTCGCTCTATCTTCACTGTTCTTAGCGTGTAGAGCT | | | |
| Nix upstream 2B | 18 | NixUp2B_F1_ONP | 8155 | TTTCTGTTGGTGCTGATATTGCAAGTACCATGTAACCGTCTCTC | 1778 | 3, 8, 9, 10, 11 and uninj | 6 |
| | | NixUp2B_R1_ONP | 8156 | ACTTGCCTGTCGCTCTATCTTCAGTGTACGGGGATAGACTATTG | | | |
| Act4 | 19 | Act4_Cterm_F1_ONP | 8157 | TTTCTGTTGGTGCTGATATTGCTCCGAGAAGTCTTATGAACTTC | 623 | All samples and uninj | 12 |
| | | Act4 3'UTR_R01_ONP | 8158 | ACTTGCCTGTCGCTCTATCTTCATAACGTGATTTGATGTTCGATGCA | | | |
| Kmo | 20 | kmo miseq F_ONP | 8159 | TTTCTGTTGGTGCTGATATTGCttatgatgatcgccctgcccaatcaggatc | 500 | All samples and uninj | 12 |
| | | kmo miseq R_ONP | 8160 | ACTTGCCTGTCGCTCTATCTTCTACGGTACCTACAAAACGACCGCATACA | | | |
| | | | | | | | 125 |

*Appendix B Table 7 Barcodes assigned to each sample amplicon in each nanopore run.*

| Run 1 | | | Run 2 | | |
|---|---|---|---|---|---|
| Sample | Amplicon | Barcode | Sample | Amplicon | Barcode |
| 2 | Myosex downstream 2 | BC01 | 1 | Myosex intron 3B | BC01 |
| 5 | Myosex downstream 2 | BC02 | 2 | Myosex intron 3B | BC02 |
| uninj | Myosex downstream 2 | BC03 | 3 | Myosex intron 3B | BC03 |
| 3 | Myosex downstream 4 | BC04 | 4 | Myosex intron 3B | BC04 |
| uninj | Myosex downstream 4 | BC05 | 5 | Myosex intron 3B | BC05 |
| 4 | Myosex downstream 2b | BC06 | 6 | Myosex intron 3B | BC06 |
| 8 | Myosex downstream 2b | BC07 | 7 | Myosex intron 3B | BC07 |
| 10 | Myosex downstream 2b | BC08 | 8 | Myosex intron 3B | BC08 |
| uninj | Myosex downstream 2b | BC09 | 9 | Myosex intron 3B | BC09 |
| 1 | Myosex downstream 7 | BC10 | 10 | Myosex intron 3B | BC10 |
| 2 | Myosex downstream 7 | BC11 | uninj | Myosex intron 3B | BC11 |
| uninj | Myosex downstream 7 | BC12 | 1 | Myosex intron 5A | BC12 |
| 1 | Nix downstream 2 | BC13 | 2 | Myosex intron 5A | BC13 |
| 5 | Nix downstream 2 | BC14 | 3 | Myosex intron 5A | BC14 |
| 6 | Nix downstream 2 | BC15 | 4 | Myosex intron 5A | BC15 |
| 8 | Nix downstream 2 | BC16 | 5 | Myosex intron 5A | BC16 |
| 10 | Nix downstream 2 | BC17 | 6 | Myosex intron 5A | BC17 |
| uninj | Nix downstream 2 | BC18 | 7 | Myosex intron 5A | BC18 |
| 4 | Nix downstream 1 | BC19 | 8 | Myosex intron 5A | BC19 |
| 9 | Nix downstream 1 | BC20 | 9 | Myosex intron 5A | BC20 |
| uninj | Nix downstream 1 | BC21 | 10 | Myosex intron 5A | BC21 |
| 2 | Nix intron 3B | BC22 | uninj | Myosex intron 5A | BC22 |
| 3 | Nix intron 3B | BC23 | 1 | Myosex downstream 1 | BC23 |
| 4 | Nix intron 3B | BC24 | 6 | Myosex downstream 1 | BC24 |
| 10 | Nix intron 3B | BC25 | 7 | Myosex downstream 1 | BC25 |
| uninj | Nix intron 3B | BC26 | 9 | Myosex downstream 1 | BC26 |
| 10 | Nix intron 5 | BC27 | 11 | Myosex downstream 1 | BC27 |
| 11 | Nix intron 5 | BC28 | uninj | Myosex downstream 1 | BC28 |
| uninj | Nix intron 5 | BC29 | 1 | Nix Intron 4 | BC29 |
| 3 | Nix upstream 2B | BC30 | 5 | Nix Intron 4 | BC30 |
| 8 | Nix upstream 2B | BC31 | 6 | Nix Intron 4 | BC31 |
| 9 | Nix upstream 2B | BC32 | 7 | Nix Intron 4 | BC32 |
| 10 | Nix upstream 2B | BC33 | 8 | Nix Intron 4 | BC33 |
| 11 | Nix upstream 2B | BC34 | 9 | Nix Intron 4 | BC34 |
| uninj | Nix upstream 2B | BC35 | uninj | Nix Intron 4 | BC35 |
| 1 | Kmo | BC36 | 1 | Nix upstream 3 | BC36 |
| 2 | Kmo | BC37 | 2 | Nix upstream 3 | BC37 |
| 3 | Kmo | BC38 | 3 | Nix upstream 3 | BC38 |
| 4 | Kmo | BC39 | 4 | Nix upstream 3 | BC39 |
| 5 | Kmo | BC40 | 5 | Nix upstream 3 | BC40 |
| 6 | Kmo | BC41 | 6 | Nix upstream 3 | BC41 |
| 7 | Kmo | BC42 | 7 | Nix upstream 3 | BC42 |
| 8 | Kmo | BC43 | uninj | Nix upstream 3 | BC43 |
| 9 | Kmo | BC44 | 1 | Act4 | BC44 |
| 10 | Kmo | BC45 | 2 | Act4 | BC45 |
| 11 | Kmo | BC46 | 3 | Act4 | BC46 |
| THS1 | Kmo | BC47 | 4 | Act4 | BC47 |
| THS2 | Kmo | BC48 | 5 | Act4 | BC48 |
| THS3 | Kmo | BC49 | 6 | Act4 | BC49 |
| THS4 | Kmo | BC50 | 7 | Act4 | BC50 |
| THS5 | Kmo | BC51 | 8 | Act4 | BC51 |
| THS6 | Kmo | BC52 | 9 | Act4 | BC52 |
| uninj | Kmo | BC53 | 10 | Act4 | BC53 |
| 1 | Myosex intron 4C | BC74 | 11 | Act4 | BC54 |
| 2 | Myosex intron 4C | BC75 | THS1 | Act4 | BC55 |
| 5 | Myosex intron 4C | BC76 | THS2 | Act4 | BC56 |
| 8 | Myosex intron 4C | BC77 | THS3 | Act4 | BC57 |

| | | | | | |
|---|---|---|---|---|---|
| uninj | Myosex intron 4C | BC78 | THS4 | Act4 | BC58 |
| 3 | Myosex intron 4B | BC79 | THS5 | Act4 | BC59 |
| 4 | Myosex intron 4B | BC80 | THS6 | Act4 | BC60 |
| 6 | Myosex intron 4B | BC81 | uninj | Act4 | BC61 |
| 7 | Myosex intron 4B | BC82 | | | |
| 9 | Myosex intron 4B | BC83 | | | |
| 10 | Myosex intron 4B | BC84 | | | |
| uninj | Myosex intron 4B | BC85 | | | |
| 2 | Nix downstream 3 | BC86 | | | |
| 3 | Nix downstream 3 | BC87 | | | |
| 7 | Nix downstream 3 | BC88 | | | |
| uninj | Nix downstream 3 | BC89 | | | |
| 4 | Nix intron 1A | BC90 | | | |
| 5 | Nix intron 1A | BC91 | | | |
| 6 | Nix intron 1A | BC92 | | | |
| 7 | Nix intron 1A | BC93 | | | |
| 8 | Nix intron 1A | BC94 | | | |
| 9 | Nix intron 1A | BC95 | | | |
| uninj | Nix intron 1A | BC96 | | | |

*Appendix B Table 8 Primers used to clone 1133. Bold are the AgeI and NotI sites that were added back into the plasmid.*

| PCR | Primer name | PL no. | Sequence | Length (bp) |
|---|---|---|---|---|
| sgRNA cassette from 2199 | MLOC_hifi_f2 | | TCCAAACTCATCAATGTATCTTAGGTCTCGACCGGT GGCCGGCCGTTTCCAGACTTTCCCTCCC | 2727 |
| | MLOC_hifi_r2 | | ATTCAATGCACAAGCTACATGTAAAGATAGCGGCC GCAAAAAAAGCACCGACTCGGTGCCCTGCATTC | |
| Colony screen | 463_Cas13bF | 463 | atatacccgaTGCGTTGCTCTTC | 1285 |
| | 183_LA981.Pub-932 | 183 | TGTTCGGTTGGATGGCTCTA | |
| Sequencing positive plasmids | L346R | 149 | tttcactgcattctagttgtggtttgtcc | |
| | LA981.Pub-932 | 183 | TGTTCGGTTGGATGGCTCTA | |
| | AePub-R2-nestedPCR | 293 | AGTGACGTCAACCCTTCTAA | |
| | Cas13bF | 463 | atatacccgaTGCGTTGCTCTTC | |
| | oJW2012_tracr | 464 | GGTGCCACTTTTTCAAGTTGATAAC | |
| | AE026white | 465 | TGATTTGTCCGTGCGGTATG | |
| | sgRNA.seq | 468 | TGACAAGTTCATCCACTTCTCCTG | |
| | TBP.seq11 | 479 | gcatgataacggacttgccttattcc | |

*Appendix B Table 9 PCR primers used to produce amplicons for sequencing of missing HA data.*

| Target | Primer no. | F primer | Sequence | Primer no. | R primer | Sequence | Amplicon (bp) |
|---|---|---|---|---|---|---|---|
| Myosex intron 4B 5' HA | 867 | M_I_4B_125_F1 | TACCCTCTTTATTCGTAGGTTT | 869 | M_I_4B_125_R1 | ACAGTTCTCTTCATGGAACAAC | 1769 |
| | 868 | M_I_4B_125_F2 | GGTAGTTTGTTTGACCAAAGAT | 869 | M_I_4B_125_R1 | ACAGTTCTCTTCATGGAACAAC | 1675 |
| Myosex intron 4C 19 3' HA | 849 | M_I_4C_19_F1 | ATACGACTGTTATGCGATTACA | 850 | M_I_4C_19_R1 | GTACTGACCTTTGGAAAACAAT | 2088 |
| | 849 | M_I_4C_19_F1 | ATACGACTGTTATGCGATTACA | 851 | M_I_4C_19_R2 | CTATCAGTACACACTTTGCCAA | 2357 |
| Myosex downstream 2B 6 5' HA | 853 | M_D_2B_6_F1 | AGGATGACCTGAAAGTAAGACA | 855 | M_D_2B_6_R1 | AGCGATTATAGAACTAAAACCC | 1952 |
| Myosex downstream 2B 2 3' HA | 899 | M_D_2B_2_F1 | ATATAACGCTAAATTAGGGTCC | 900 | M_D_2B_2_R1 | AAGTACATAATCAGGGAACGAC | 1366 |
| | 899 | M_D_2B_2_F1 | ATATAACGCTAAATTAGGGTCC | 901 | M_D_2B_2_R2 | TTCAGCTCACCGATACATATTA | 1799 |
| Nix intron 3B 43 3' HA | 856 | N_I_3B_43_F1 | ATTATCCTTTTACATGGGTGTC | 858 | N_I_3B_43_R1 | GTTCTCGTTCGATTACTGAAGT | 2167 |
| | 856 | N_I_3B_43_F1 | ATTATCCTTTTACATGGGTGTC | 859 | N_I_3B_43_R2 | TCATCTTTACTCCAAGGTCAGT | 2468 |
| | 857 | N_I_3B_43_F2 | CACCATATGAATTAATACACGG | 858 | N_I_3B_43_R1 | GTTCTCGTTCGATTACTGAAGT | 1887 |
| | 857 | N_I_3B_43_F2 | CACCATATGAATTAATACACGG | 859 | N_I_3B_43_R2 | TCATCTTTACTCCAAGGTCAGT | 2188 |
| Nix upstream 2B 12 3'HA | 864 | N_U_2B_12_F1 | GAAGCAGAAGAGTGTCAGTTTT | 866 | N_U_2B_12_R1 | TTCATTTGTGCCACATACTAAC | 1767 |
| | 865 | N_U_2B_12_F2 | AGCTGTCCAGAATGAATATCAT | 866 | N_U_2B_12_R1 | TTCATTTGTGCCACATACTAAC | 1590 |
| Nix intron 4 | 911 | N-I-4-F1 | ATTGACCACCTACCATATTGTC | 914 | N-I-4-R1 | ATTGAAGAAATCGATCCAGTAC | 2880 |
| | 911 | N-I-4-F1 | ATTGACCACCTACCATATTGTC | 915 | N-I-4-R2 | GGTAGTTCCTGAAAAGAAGAAA | 2828 |
| | 912 | N-I-4-F2 | ACAGCGCACTATTCACATATTA | 914 | N-I-4-R1 | ATTGAAGAAATCGATCCAGTAC | 2695 |
| | 912 | N-I-4-F2 | ACAGCGCACTATTCACATATTA | 915 | N-I-4-R2 | GGTAGTTCCTGAAAAGAAGAAA | 2747 |
| | 913 | N-I-4-F3 | ATCACAAATCTTACCTACGGAC | 916 | N-I-4-R3 | TACAGGTATGTGCAGCATAAAT | 2125 |
| | 913 | N-I-4-F3 | ATCACAAATCTTACCTACGGAC | 917 | N-I-4-R4 | GAAGAGGAAAAATATCAACGTC | 2064 |
| Myosex intron 3b | 923 | M-I-3B-F1 | TAACAAGCTCTGCTGTTTCTTA | 925 | M-I-3B-R1 | CGTGACATTACTTGCATTATGT | 1693 |
| | 924 | M-I-3B-F2 | GAAGATGGAAGATCGTATAAGC | 925 | M-I-3B-R1 | CGTGACATTACTTGCATTATGT | 1912 |
| | 928 | M-I-3B-F3 | ATAATGCAAGTAATGTCACGTG | 929 | M-I-3B-R2 | TAATCCCTTTACAGATCCTGAC | 1989 |

*Appendix B Table 10 Primers used to sequence homology arm amplicons in Appendix B Table 9.*

| Target | Primer no. | Primer name | Sequence |
|---|---|---|---|
| Myosex intron 4B 5' HA | 869 | M_I_4B_125_R1 | ACAGTTCTCTTCATGGAACAAC |
| | 870 | M_I_4B_125_SeqF1 | GCATGAATGTCATGGTAGTCTA |
| | 871 | M_I_4B_125_SeqF2 | ATCCTCTGAACTTTTCTATTGC |
| Myosex intron 4C 19 3' HA | 616 | 616_Myosex_Int4C_R2 | GACCACTCATCAACTCTTACGT |
| | 837 | M_I_4C_19_3HA_S_R1 | CGATGAGTCTATCGTGTGTTAC |
| | 874 | M_I_4C_19_F1 | ACGTAAGAGTTGATGAGTGGTC |
| | 875 | M_I_4C_19_F2 | TTCATCTCTGTCTGGTAGAACC |
| | 850 | M_I_4C_19_R1 | GTACTGACCTTTGGAAAACAAT |
| Myosex downstream 2B 6 5' HA | 853 | M_D_2B_6_F1 | AGGATGACCTGAAAGTAAGACA |
| | 843 | M_D_2B_6_5HA_S_F1 | GGAAGCACCTGTCTATGTAAGA |
| | 844 | M_D_2B_6_5HA_S_R1 | CCTTTGTTGCTGTTTCTAATAC |
| | 37 | 37_Myosex_DS2B_F1 | AAGTCCCTTTCCTCTTGTAGTC |
| Myosex downstream 2B 2 3' HA | 899 | M_D_2B_2_F1 | ATATAACGCTAAATTAGGGTCC |
| | 38 | 39_Myosex_DS2B_R1 | CATAATGGTGACTCACACACAG |
| | 900 | M_D_2B_2_R1 | AAGTACATAATCAGGGAACGAC |
| Nix intron 3B 43 3' HA | 510 | | GATAGGCAGACATTGTACGAGT |
| | 845 | N_I_3B_43_3HA_S_F1 | GATTGCTGACGATTTAATGATG |
| | 846 | N_I_3B_43_3HA_S_R1 | GTTCTCGTTCGATTACTGAAGT |
| | 894 | N_I_3B_43_F2 | AGTACGATACATGAAAGGGATC |
| | 859 | N_I_3B_43_R2 | TCATCTTTACTCCAAGGTCAGT |
| Nix upstream 2B 12 3'HA | 561 | 561_NixUp2B_Seq_F2 | CTTTGTTGTGTTGTCAAGGC |
| | 897 | N_U_2B_12_F3 | TTGATGGGATACATTATAGGC |
| | 558 | 558_NixUp2B_R2 | AGGACGTTATCTTGAGTGTAGG |
| | 866 | N_U_2B_12_R1 | TTCATTTGTGCCACATACTAAC |
| | 898 | N_U_2B_12_R2 | GCGGTAGTGATTTAGATAAACTG |
| Nix intron 4 | 914 | N-I-4-R1 | ATTGAAGAAATCGATCCAGTAC |
| | | N-I-4-S-R1 | CTGTGAAAAGAGTGCTGAGTG |
| | | N-I-4-S-R2 | GAGCTGATCCGTGTCTTTAC |
| | 7781 | | GTCGACATAGACCAGCAAATAC |
| | | N-I-4-S-R3 | GATACGGAGACTCTGAAGGTAA |
| | | N-I-4-S-R4 | GAAAGCCATCACCTACTGTG |
| | | N-I-4-S-R5 | CACCTCTCAAACGTTTTTAA |
| | 911 | N-I-4-F1 | ATTGACCACCTACCATATTGTC |
| | 7786 | NixInt4_F_Seq8 | GTGTCTTCTACAGCCAAACC |
| | 7775 | NixInt4_R1 | GGACAGCTGTAAGAAGAGAGAC |
| | 7778 | NixInt4_R2 | CTTCTTGTCTCGATAAGGACAG |
| | 917 | N-I-4-R4 | GAAGAGGAAAAATATCAACGTC |
| Myosex intron 3b | 923 | M-I-3B-F1 | TAACAAGCTCTGCTGTTTCTTA |
| | 925 | M-I-3B-R1 | CGTGACATTACTTGCATTATGT |
| | | M-I-3B-S-R1 | ATTTGGTCTTATGTGTGGAGTG |
| | | M-I-3B-S-F1 | CCAAACAAAGGACCTTACTAAT |
| Myosex intron 3b | | M-I-3B-R2 | TAATCCCTTTACAGATCCTGAC |
| | 605 | | GACTCGTAATGTGTACAAAGGC |
| | 604 | | GGAAGAGAGCGATTATCTGA |
| | 607 | | AGTGCATAGAAAACTGTGGC |

*Appendix B Table 11 Primers used to amplify 5' and 3' homology arms for HiFi assembly of HDR donor plasmids.*

| Target | Primer no. | F primer | Sequence | Primer no. | R primer | Sequence | Amplicon (bp) |
|---|---|---|---|---|---|---|---|
| Myosex intron 4B 125 3' HA | 800 | M_I_4B_125_3HA_F1 | tggcgcgtttgggggcggcgcgccccagccgatggagtggcaactgccaga | 801 | M_I_4B_125_3HA_R1 | acttctgcattgcaagacgaagaatgccataggtttatggattgtc | 1045 |
| Myosex intron 4B 125 5' HA | 803 | M_I_4B_125_5HA_F1 | atggcattcttcgtcttgcaatgcagaagtccggaaac | 805 | M_I_4B_125_5HA_R1 | tttttacggttcctggccccgcggggcaccggattcatgacagcgttcacag | 1045 |
| Myosex intron 4B 125 3' HA-sgRNA | 800 | M_I_4B_125_3HA_F1 | tggcgcgtttgggggcggcgcgccccagccgatggagtggcaactgccaga | 802 | M_I_4B_125_3HA_R2s | acttctgcattgcaaccatcgggattcatgacagcgttgacgaagaatgccataggtttatggattgtc | 1068 |
| Myosex intron 4B 125 sgRNA-5' HA | 804 | M_I_4B_125_5HA_F2s | atggcattcttcgtcaacgctgtcatgaatcccgatggttgcaatgcagaagtccggaaac | 805 | M_I_4B_125_5HA_R1 | tttttacggttcctggccccgcggggcaccggattcatgacagcgttcacag | 1068 |
| Myosex intron 4C 19 3' HA | 818 | M_I_4C_19_3HA_F1 | tggcgcgtttgggggcggcgcgccccagctcccagactcaggtaagtcagatg | 827 | M_I_4C_19_3HA_R1 | aattccagcctcaaactcgactgtttttgaaaacagtcc | 1045 |
| Myosex intron 4C 19 5' HA | 819 | M_I_4C_19_5HA_F1 | tcaaaaacagtcgagtttgaggctggaattttgtacaac | 828 | M_I_4C_19_5HA_R1 | tttttacggttcctggccccgcggggcaccgggaggcaccgattacacac | 1045 |
| Myosex intron 4C 19 3' HA-sgRNA | 818 | M_I_4C_19_3HA_F1 | tggcgcgtttgggggcggcgcgccccagctcccagactcaggtaagtcagatg | 829 | M_I_4C_19_3HA_R2s | aattccagcctcaaacctccctcccagactcaggtaagctcgactgtttttgaaaacagtcc | 1068 |
| Myosex intron 4C 19 sgRNA 5' HA | 820 | M_I_4C_19_5HA_F2s | tcaaaaacagtcgagcttacctgagtctgggagggaggtttgaggctggaattttgtacaac | 828 | M_I_4C_19_5HA_R1 | tttttacggttcctggccccgcggggcaccgggaggcaccgattacacac | 1068 |
| Myosex downstream 2 20 3' HA | 794 | M_D_2_20_3HA_F1 | tggcgcgtttgggggcggcgcgccccagcgctaagccccctaggaagtcttc | 795 | M_D_2_20_3HA_R1 | aataattctcccggcgttgtgaagagcaaccacatcg | 1045 |
| Myosex downstream 2 20 5' HA | 797 | M_D_2_20_5HA_F1 | gttgctcttcacaacgccgggagaattattctacgtgggcg | 799 | M_D_2_20_5HA_R1 | tttttacggttcctggccccgcggggcacctaggggggaggaggctggctcct | 1045 |
| Myosex downstream 2 20 3' HA-sgRNA | 794 | M_D_2_20_3HA_F1 | tggcgcgtttgggggcggcgcgccccagcgctaagccccctaggaagtcttc | 796 | M_D_2_20_3HA_R2s | aataattctcccggcccccctagctaagcccccctaggaagttgtgaagagcaaccacatcg | 1068 |
| Myosex downstream 2 20 sgRNA-5' HA | 798 | M_D_2_20_5HA_F2s | gttgctcttcacaacttcctaggggggcttagctaggggggccggagaattattctacgtgggcg | 799 | M_D_2_20_5HA_R1 | tttttacggttcctggccccgcggggcacctaggggggaggaggctggctcct | 1068 |
| Myosex downstream 2B 6 3' HA | 821 | M_D_2B_6_3HA_F1 | tggcgcgtttgggggcggcgcgccccagccaattctattagccctgtcactctgc | 830 | M_D_2B_6_3HA_R1 | tgtgtcggagttactgccctacctcgaaaggtttggc | 1045 |
| Myosex downstream 2B 6 5' HA | 822 | M_D_2B_6_5HA_F1 | ctttcgaggtagggcagtaactccgacacattgccag | 831 | M_D_2B_6_5HA_R1 | tttttacggttcctggccccgcggggcaccctgtggatgcggctaaagcact | 1045 |
| Myosex downstream 2B 6 3' HA-sgRNA | 821 | M_D_2B_6_3HA_F1 | tggcgcgtttgggggcggcgcgccccagccaattctattagccctgtcactctgc | 832 | M_D_2B_6_3HA_R2s | tgtgtcggagttactccacagcaattctattagccctggccctacctcgaaaggtttggc | 1068 |
| Myosex downstream 2B 6 sgRNA-5' HA | 823 | M_D_2B_6_5HA_F2s | ctttcgaggtagggccagggctaatagaattgctgtggagtaactccgacacattgccag | 831 | M_D_2B_6_5HA_R1 | tttttacggttcctggccccgcggggcaccctgtggatgcggctaaagcact | 1068 |
| Myosex downstream 2B 2 3' HA | 902 | M_D_2B_2_3HA_F1 | tggcgcgtttgggggcggcgcgccccagcGTAATCCCAATCCCTTCTTATAAAGGG | 904 | M_D_2B_2_3HA_R1 | cacgctgcaatcagtACACATCCGGTTCTCAATGGGAG | 1045 |
| Myosex downstream 2B 2 5' HA | 903 | M_D_2B_2_5HA_F1 | gagaaccggatgtgtACTGATTGCAGCGTGGTACGTTTATG | 905 | M_D_2B_2_5HA_R1 | tttttacggttcctggccccgcggggcaccCAGCGGCGCATCTAGGCGATAA | 1045 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Nix downstream 1 20 3' HA | 806 | N_D_1_20_3 HA_F1 | tggcgcgtttgggggggcggcgcgccccagcaacaggcgtca aaccttatgag | 807 | N_D_1_20_3H A_R1 | cagctagctaactcccgagcaccaaccactttggaag | 1045 |
| Nix downstream 1 20 5' HA | 809 | N_D_1_20_5 HA_F1 | agtggttggtgctcgggagttagctagctgtactaactttcta g | 811 | N_D_1_20_5H A_R1 | tttttacggttcctggccccgcggggcacctggatgcgcgcaa gattaacgaaag | 1045 |
| Nix downstream 1 20 3' HA-sgRNA | 806 | N_D_1_20_3 HA_F1 | tggcgcgtttgggggggcggcgcgccccagcaacaggcgtca aaccttatgag | 808 | N_D_1_20_3H A_R2s | cagctagctaactcccctgtttggatgcgcgcaagattcgagc accaaccactttggaag | 1068 |
| Nix downstream 1 20 sgRNA-5' HA | 810 | N_D_1_20_5 HA_F2s | agtggttggtgctcgaatcttgcgcgcatccaaacagggga gttagctagctgtactaactttctag | 811 | N_D_1_20_5H A_R1 | tttttacggttcctggccccgcggggcacctggatgcgcgcaa gattaacgaaag | 1068 |
| Nix upstream 2B 12 3' HA | 950 | N-U-2B-12-3HA-F1 | tggcgcgtttgggggggcggcgcgccccagcATTAATGG GTACAGGGCCTTTAC | 953 | N-U-2B-12-3HA-R1 | tgtagcctctcggctCACAGAGATGCAATACAGACCA ATATTG | 1045 |
| Nix upstream 2B 12 5' HA | 951 | N-U-2B-12-5HA-F1 | tattgcatctctgtgAGCCGAGAGGCTACACATACA AATC | 954 | N-U-2B-12-5HA-R1 | tttttacggttcctggccccgcggggcaccGGTTGGTATCC ACTTCGTAAGTAC | 1045 |
| Colony PCR of plasmids | 872 | K10_3utr_F1 | ccaacaatccatctaatagcta | 873 | pUC_Amp_R1 | gacaggtatccggtaagc | 2392-2415 |

*Appendix B Table 12 Primers used to sequence mini and midi preps of HDR donor plasmids.*

| Target | Primer no. | F primer | Sequence |
|---|---|---|---|
| Myosex intron 4B 125 | 872 | K10_3utr_F1 | ccaacaatccatctaatagcta |
| Myosex intron 4B 125 sgRNA | 877 | M_I_4B_125_F1 | GATGCTCATACAGCTTCGTAC |
| | 878 | M_I_4B_125_F2 | GAGTTTAGGTTTCGAAAATCTG |
| | 879 | M_I_4B_125_F3 | ATCCTCTGAACTTTTCTATTGC |
| | 873 | pUC_Amp_R1 | gacaggtatccggtaagc |
| Myosex intron 4C 19 | 872 | K10_3utr_F1 | ccaacaatccatctaatagcta |
| Myosex intron 4C 19 sgRNA | 874 | M_I_4C_19_F1 | ACGTAAGAGTTGATGAGTGGTC |
| | 875 | M_I_4C_19_F2 | TTCATCTCTGTCTGGTAGAACC |
| | 876 | M_I_4C_19_F3 | GAATTTGGAAAACTTTGGTC |
| | 873 | pUC_Amp_R1 | gacaggtatccggtaagc |
| Myosex downstream 2 20 | 872 | K10_3utr_F1 | ccaacaatccatctaatagcta |
| Myosex downstream 2 20 sgRNA | 880 | M_D_2_20_F1 | CGTGGTCGTTTAATTGATT |
| | 881 | M_D_2_20_F2 | GGGAATGTAGAAACTGAATTGT |
| | 882 | M_D_2_20_F3 | GTCAGTTAGTTTAGAATGTAGGGAA |
| | 873 | pUC_Amp_R1 | gacaggtatccggtaagc |
| Myosex downstream 2B 6 | 872 | K10_3utr_F1 | ccaacaatccatctaatagcta |
| Myosex downstream 2B 6 sgRNA | 890 | M_D_2B_6_F1 | CGTGGTACGTTTATGTTTCTAT |
| | 891 | M_D_2B_6_F2 | GAAAATTGATCTCAATCGAGTG |
| | 892 | M_D_2B_6_F3 | ATGTAAGATGTGCTCTGGTTCT |
| | 873 | pUC_Amp_R1 | gacaggtatccggtaagc |
| Myosex downstream 2B 2 | 872 | K10_3utr_F1 | ccaacaatccatctaatagcta |
| | 908 | M_D_2B_2_F1 | GATCCATTCTTGTAGGAGGTC |
| | 909 | M_D_2B_2_F2 | ATCGTCACAATAGACAGGAAGT |
| | 910 | M_D_2B_2_F3 | GGGTTTTAGTTCTATAATCGCT |
| | 873 | pUC_Amp_R1 | gacaggtatccggtaagc |
| Nix downstream 1 20 | 872 | K10_3utr_F1 | ccaacaatccatctaatagcta |
| Nix downstream 1 20 sgRNA | 896 | N_D_1_20_F1 | GAGAGGACAGATGGTGTACG |
| | 888 | N_D_1_20_F2 | TTTTACGCAGTATGGACAGTT |
| | 889 | N_D_1_20_F3 | TTATGTTTGTCTTGTACCCACA |
| | 873 | pUC_Amp_R1 | gacaggtatccggtaagc |
| Nix upstream 2B 12 | 872 | K10_3utr_F1 | ccaacaatccatctaatagcta |
| | 897 | N_U_2B_12_F3 | TTGATGGGATACATTATAGGC |
| | 973 | N-U-2B-12-F1 | GAACTTTTTGGATTTCCTTG |
| | 974 | N-U-2B-12-F2 | CAATGTTACGAAATCAGCTCTA |
| | 873 | pUC_Amp_R1 | gacaggtatccggtaagc |

*Appendix B Table 13 Primers used to verify integration of SuperSIT construct in the five transgenic lines that showed male-only inheritance (1134, 1136, 1137, 1162, and 1167).*

| Amplicon | Target | Primer no. | F primer | Sequence | Primer no. | R primer | Sequence3 | Ta (Q5) | Amplicon (bp) |
|---|---|---|---|---|---|---|---|---|---|
| 2 | B2 | 997 | pUC-GW_F1 | cagattgtactgagagtgcaca | 1005 | Cas9_R1 | cttatccacatccgagttatct | 62 | 5343 |
| 3 | Cas9 | 998 | B2_F1 | CCACAATGTAAATCAATCACAG | 1006 | U6_R1 | gataacggactagccttatttt | 60 | 5380 |
| 4 | U6-guides | 999 | Cas9_F1 | atgaaatcatcgaacagatctc | 1007 | AePUB_R1 | AATCTGATAACCAAATCCAGAG | 60 | 3479 |
| 1 | 1134-5' junction | 614 | Myosex_Int4C_F1 | TCTCTCTCTAAGCCTCATCCTA | 278 | B2-prom-seq-R1 | GCAGTGTTGCATGATATACTTC | 62 | 5538 |
| 5 | 1134-3' junction | 1000 | U6_F1 | TGTGTGGCTTTTATACATCCTA | 850 | M_I_4C_19_R1 | GTACTGACCTTTGGAAAACAAT | 61 | 4440 |
| 1 | 1137-5' junction | 655 | Nix_DS1_F1 | CTAACTTGGTGGGAATCTCTAG | 278 | B2-prom-seq-R1 | GCAGTGTTGCATGATATACTTC | 62 | 4671 |
| 5 | 1137-3' junction | 1000 | U6_F1 | TGTGTGGCTTTTATACATCCTA | 658 | Nix_DS1_R2 | CACTTCTCCTACACAGACCAAT | 62 | 4314 |
| 1 | 1136-5' junction | 853 | M_D_2B_6_F1 | AGGATGACCTGAAAGTAAGACA | 278 | B2-prom-seq-R1 | GCAGTGTTGCATGATATACTTC | 62 | 4435 |
| 5 | 1136-3' junction | 1000 | U6_F1 | TGTGTGGCTTTTATACATCCTA | 42 | Myosex_DS2B_Seq_R1 | CGTCATGGTTGAGATTGTC | 61 | 4700 |
| 1 | 1162-5' junction | 551 | NI1_F2 | AATCCACACACATTCTCCAC | 278 | B2-prom-seq-R1 | GCAGTGTTGCATGATATACTTC | 62 | 5253 |
| 5 | 1162-3' junction | 1000 | U6_F1 | TGTGTGGCTTTTATACATCCTA | 552 | NI1_R1 | AGACAGTCGATAGTTCTCGTGT | 62 | 4269 |
| 1 | 1167-5' junction-extended | 7781 | NixInt4_F_Seq3 | GTCGACATAGACCAGCAAATAC | 278 | B2-prom-seq-R1 | GCAGTGTTGCATGATATACTTC | 62 | 4998 |
| 5 | 1167-3' junction-extended | 1000 | U6_F1 | TGTGTGGCTTTTATACATCCTA | 7778 | NixInt4_R2 | CTTCTTGTCTCGATAAGGACAG | 62 | 4843 |

*Appendix B Table 14 Reference table used to generate Figure 27, which shows the in vivo sgRNA cleavage efficiencies of 82 sgRNAs targeting the M locus. This table shows the original sgRNA number per subregion targeted ("Amplicon") and its corresponding sgRNA # in the heatmap (labelled 1-10). sgRNAs and Cas9 were microinjected into mosquito eggs that were harvested 24 hours later for mutation detection by Oxford Nanopore sequencing of the target loci. sgRNA cleavage efficiency was estimated according to the frequency of mutated target loci minus the baseline variation rate of each sgRNA site in the un-injected control, determined by CRISPResso2. Cleavage efficiency represents the genome editing efficiency of each sgRNA; 82 values are shown (one sgRNA from myo-sex downstream 1, sgRNA 27, was not determined due to memory problems with CRISPResso2).*

| Amplicon | sgRNA | sgRNA # in heatmap | Cleavage efficiency |
|---|---|---|---|
| M_D_1 | 10 | 1 | 33.16 |
| M_D_1 | 13 | 2 | 29.78 |
| M_D_1 | 43 | 3 | 27.96 |
| M_D_1 | 56 | 4 | 16.25 |
| M_D_2 | 20 | 1 | 50.02 |
| M_D_2 | 51 | 2 | 6.09 |
| M_D_2B | 2 | 1 | 37.92 |
| M_D_2B | 6 | 2 | 33.18 |
| M_D_2B | 17 | 3 | 11.63 |
| M_D_4 | 1 | 1 | 55.30 |
| M_D_7 | 3 | 1 | 62.99 |
| M_D_7 | 85 | 2 | 60.87 |
| M_I_3B | 1 | 1 | 46.70 |
| M_I_3B | 6 | 2 | 37.29 |
| M_I_3B | 21 | 3 | 16.01 |
| M_I_3B | 29 | 4 | 58.58 |
| M_I_3B | 31 | 5 | 42.04 |
| M_I_3B | 34 | 6 | 3.35 |
| M_I_3B | 65 | 7 | 38.29 |
| M_I_3B | 71 | 8 | 3.53 |
| M_I_3B | 76 | 9 | 49.20 |
| M_I_3B | 81 | 10 | 18.79 |
| M_I_4B | 14 | 1 | 13.52 |
| M_I_4B | 15 | 2 | 1.09 |
| M_I_4B | 29 | 3 | 12.43 |
| M_I_4B | 42 | 4 | 6.97 |
| M_I_4B | 93 | 5 | -7.40 |
| M_I_4B | 125 | 6 | 65.65 |
| M_I_4C | 19 | 1 | 79.95 |
| M_I_4C | 44 | 2 | 77.38 |
| M_I_4C | 50 | 3 | 26.83 |
| M_I_4C | 82 | 4 | 14.20 |
| M_I_5A | 14 | 1 | 6.91 |
| M_I_5A | 21 | 2 | 62.74 |
| M_I_5A | 45 | 3 | 26.74 |
| M_I_5A | 53 | 4 | 20.87 |
| M_I_5A | 75 | 5 | 32.74 |
| M_I_5A | 93 | 6 | 43.91 |
| M_I_5A | 101 | 7 | 34.19 |
| M_I_5A | 139 | 8 | 17.99 |
| M_I_5A | 148 | 9 | 42.50 |
| M_I_5A | 216 | 10 | 21.31 |
| N_D_1 | 20 | 1 | 45.55 |
| N_D_1 | 29 | 2 | 20.48 |
| N_D_2 | 5 | 1 | 43.59 |
| N_D_2 | 9 | 2 | 46.48 |
| N_D_2 | 11 | 3 | 69.95 |
| N_D_2 | 39 | 4 | 56.90 |
| N_D_2 | 41 | 5 | 57.51 |
| N_D_3 | 15 | 1 | 10.85 |

| | | | |
|---|---|---|---|
| N_D_3 | 19 | 2 | 14.80 |
| N_D_3 | 44 | 3 | 47.34 |
| N_I_1A | 2 | 1 | 21.47 |
| N_I_1A | 3 | 2 | 7.25 |
| N_I_1A | 35 | 3 | -0.52 |
| N_I_1A | 68 | 4 | 30.03 |
| N_I_1A | 74 | 5 | 22.31 |
| N_I_1A | 122 | 6 | 32.77 |
| N_I_3B | 6 | 1 | 19.04 |
| N_I_3B | 11 | 2 | 31.45 |
| N_I_3B | 29 | 3 | 34.46 |
| N_I_3B | 43 | 4 | 46.93 |
| N_I_4 | 8 | 1 | 52.31 |
| N_I_4 | 10 | 2 | 33.26 |
| N_I_4 | 12 | 3 | 21.15 |
| N_I_4 | 21 | 4 | 4.11 |
| N_I_4 | 28 | 5 | 20.70 |
| N_I_4 | 33 | 6 | 9.96 |
| N_I_5 | 49 | 1 | 68.29 |
| N_I_5 | 102 | 2 | 54.33 |
| N_U_2B | 12 | 1 | 76.45 |
| N_U_2B | 15 | 2 | 44.55 |
| N_U_2B | 23 | 3 | 85.54 |
| N_U_2B | 35 | 4 | 84.33 |
| N_U_2B | 47 | 5 | 52.02 |
| N_U_3 | 1 | 1 | 44.98 |
| N_U_3 | 2 | 2 | 22.83 |
| N_U_3 | 3 | 3 | 4.54 |
| N_U_3 | 7 | 4 | 36.48 |
| N_U_3 | 20 | 5 | 16.07 |
| N_U_3 | 21 | 6 | 15.30 |
| N_U_3 | 26 | 7 | 46.01 |

*Appendix B Table 15 Recombination frequencies of the SuperSIT transgenes (1134, 1136, and 1167) from the M locus. Recombination frequency is defined as the proportion of males that were positive and of females that were negative, due to recombination from the M to the m, as a percentage. Parentheses indicate the parental cross; the left side of the colon specifies the genotype of the father, and the right side specifies the genotype of the mother.*

| Generation/cross | (+) males | (-) males | (-) females | (+) females | Total offspring | Recombination frequency |
|---|---|---|---|---|---|---|
| 1134B1 | | | | | | |
| 1134B1 : (1095 : LVP) | 263 | 0 | 196 | 0 | 459 | 0 |
| (1134B1 : 1095) : KMO-KO | 232 | 0 | 214 | 0 | 446 | 0 |
| (1134B1 : 2199) : LVP | 194 | 0 | 190 | 0 | 384 | 0 |
| (1134B1 : LVP) : LVP | 200 | 0 | 154 | 0 | 354 | 0 |
| | | | | Total | 1643 | 0 |
| 1136D3 | | | | | | |
| G4 | 219 | 0 | 202 | 0 | 421 | 0 |
| 1136D3 : (1095 : LVP) | 118 | 0 | 85 | 0 | 203 | 0 |
| (1136D3 : 1095) : KMO-KO | 167 | 0 | 158 | 0 | 325 | 0 |
| | | | | Total | 949 | 0 |
| 1167B1 | | | | | | |
| G2 | 47 | 0 | 26 | 0 | 73 | 0 |
| | | | | Total | 73 | 0 |

# References


Abas, M. F. B., Rafie, A. S. B. M., Yusoff, H. B. and Ahmad, K. A. B. (2016) 'Flapping wing micro-aerial-vehicle: Kinematics, membranes, and flapping mechanisms of ornithopter and insect flight', *Chinese Journal of Aeronautics,* 29(5), pp. 1159-1177.

Adelman, Z. N., Anderson, M. A., Morazzani, E. M. and Myles, K. M. (2008) 'A transgenic sensor strain for monitoring the RNAi pathway in the yellow fever mosquito, Aedes aegypti', *Insect biochemistry and molecular biology,* 38(7), pp. 705-713.

Adelman, Z. N. and Tu, Z. (2016) 'Control of mosquito-borne infectious diseases: sex and gene drive', *Trends in parasitology,* 32(3), pp. 219-229.

Afify, A., Betz, J. F., Riabinina, O., Lahondère, C. and Potter, C. J. (2019) 'Commonly used insect repellents hide human odors from Anopheles mosquitoes', *Current Biology,* 29(21), pp. 3669-3680. e5.

Agudelo, J., Toro, M., Ramírez-Sánchez, L. F., Barrientos, L. M., Alfonso-Parra, C. and Avila, F. W. (2022) 'Putative degradation of non-stored sperm in the female reproductive tract of the dengue vector mosquito Aedes aegypti', *Frontiers in Tropical Diseases,* 3, pp. 816556.

Akbari, O. S., Antoshechkin, I., Amrhein, H., Williams, B., Diloreto, R., Sandler, J. and Hay, B. A. (2013a) 'The developmental transcriptome of the mosquito Aedes aegypti, an invasive species and major arbovirus vector', *G3: Genes, Genomes, Genetics,* 3(9), pp. 1493-1509.

Akbari, O. S., Matzen, K. D., Marshall, J. M., Huang, H., Ward, C. M. and Hay, B. A. (2013b) 'A synthetic gene drive system for local, reversible modification and suppression of insect populations', *Current biology,* 23(8), pp. 671-677.

Alcalay, Y., Fuchs, S., Galizi, R., Bernardini, F., Haghighat-Khah, R. E., Rusch, D. B., Adrion, J. R., Hahn, M. W., Tortosa, P. and Rotenberry, R. (2021) 'The potential for a released autosomal X-shredder becoming a driving-Y chromosome and invasively suppressing wild populations of malaria mosquitoes', *Frontiers in Bioengineering and Biotechnology,* 9, pp. 752253.

Aldersley, A. and Cator, L. J. (2019) 'Female resistance and harmonic convergence influence male mating success in Aedes aegypti', *Scientific reports,* 9(1), pp. 2145.

Alphey, L. (2014) 'Genetic control of mosquitoes', *Annual review of entomology,* 59, pp. 205-224.

Alphey, L., Benedict, M., Bellini, R., Clark, G. G., Dame, D. A., Service, M. W. and Dobson, S. L. (2010) 'Sterile-insect methods for control of mosquito-borne diseases: an analysis', *Vector-borne and zoonotic diseases,* 10(3), pp. 295-311.

Alphey, L., McKemey, A., Nimmo, D., Neira Oviedo, M., Lacroix, R., Matzen, K. and Beech, C. (2013) 'Genetic control of Aedes mosquitoes', *Pathogens and global health,* 107(4), pp. 170-179.

Alphey, L. S., Crisanti, A., Randazzo, F. and Akbari, O. S. (2020) 'Standardizing the definition of gene drive', *Proceedings of the National Academy of Sciences,* 117(49), pp. 30864-30867.

An, H. and Mogami, K. (1996) 'Isolation of 88F Actin Mutants ofDrosophila melanogasterand Possible Alterations in the Mutant Actin Structures', *Journal of molecular biology,* 260(4), pp. 492-505.

Anderson, M. A., Gonzalez, E., Ang, J. X., Shackleford, L., Nevard, K., Verkuijl, S. A., Edgington, M. P., Harvey-Samuel, T. and Alphey, L. (2023a) 'Closing the gap to effective gene drive in Aedes aegypti by exploiting germline regulatory elements', *Nature Communications,* 14(1), pp. 338.

Anderson, M. A., Gonzalez, E., Edgington, M. P., Ang, J. X., Purusothaman, D.-K., Shackleford, L., Nevard, K., Verkuijl, S. A., Harvey-Samuel, T. and Leftwich, P. T. (2022) 'A multiplexed,

confinable CRISPR/Cas9 gene drive propagates in caged Aedes aegypti populations', *BioRxiv*, pp. 2022.08. 12.503466.

Anderson, M. A., Leftwich, P. T., Wilson, R., Paladino, L. Z. C., Basu, S., Rooney, S., Adelman, Z. N. and Alphey, L. (2023b) 'AePUb promoter length modulates gene expression in Aedes aegypti', *Scientific Reports,* 13(1), pp. 20352.

Anderson, M. A., Purcell, J., Verkuijl, S. A., Norman, V. C., Leftwich, P. T., Harvey-Samuel, T. and Alphey, L. S. (2020) 'Expanding the CRISPR toolbox in culicine mosquitoes: in vitro validation of Pol III promoters', *ACS synthetic biology,* 9(3), pp. 678-681.

Anderson, M. A. E., Gonzalez, E., Edgington, M. P., Ang, J. X. D., Purusothaman, D.-K., Shackleford, L., Nevard, K., Verkuijl, S. A. N., Harvey-Samuel, T., Leftwich, P. T., Esvelt, K. and Alphey, L. (2024) 'A multiplexed, confinable CRISPR/Cas9 gene drive can propagate in caged Aedes aegypti populations', *Nature Communications,* 15(1), pp. 729.

Anderson, M. E., Mavica, J., Shackleford, L., Flis, I., Fochler, S., Basu, S. and Alphey, L. (2019) 'CRISPR/Cas9 gene editing in the West Nile Virus vector, Culex quinquefasciatus Say', *PLoS One,* 14(11), pp. e0224857.

Andrés, M., Su, M. P., Albert, J. and Cator, L. J. (2020) 'Buzzkill: targeting the mosquito auditory system', *Current opinion in insect science,* 40, pp. 11-17.

Ang, J. X. D., Nevard, K., Ireland, R., Purusothaman, D.-K., Verkuijl, S. A., Shackleford, L., Gonzalez, E., Anderson, M. A. and Alphey, L. (2022) 'Considerations for homology-based DNA repair in mosquitoes: Impact of sequence heterology and donor template source', *PLoS genetics,* 18(2), pp. e1010060.

Annas, G. J., Beisel, C. L., Clement, K., Crisanti, A., Francis, S., Galardini, M., Galizi, R., Grünewald, J., Immobile, G. and Khalil, A. S. (2021) 'A code of ethics for gene drive research', *The CRISPR journal,* 4(1), pp. 19-24.

Ant, T. H., Mancini, M. V., McNamara, C. J., Rainey, S. M. and Sinkins, S. P. (2023) 'Wolbachia-Virus interactions and arbovirus control through population replacement in mosquitoes', *Pathogens and Global Health,* 117(3), pp. 245-258.

Arensburger, P., Megy, K., Waterhouse, R. M., Abrudan, J., Amedeo, P., Antelo, B., Bartholomay, L., Bidwell, S., Caler, E. and Camara, F. (2010) 'Sequencing of Culex quinquefasciatus establishes a platform for mosquito comparative genomics', *Science,* 330(6000), pp. 86-88.

Artieri, C. G. and Fraser, H. B. (2014) 'Transcript length mediates developmental timing of gene expression across Drosophila', *Molecular biology and evolution,* 31(11), pp. 2879-2889.

Aryan, A., Anderson, M. A., Biedler, J. K., Qi, Y., Overcash, J. M., Naumenko, A. N., Sharakhova, M. V., Mao, C., Adelman, Z. N. and Tu, Z. (2020) 'Nix alone is sufficient to convert female Aedes aegypti into fertile males and myo-sex is needed for male flight', *Proceedings of the National Academy of Sciences,* 117(30), pp. 17702-17709.

Aubrey, B. J., Kelly, G. L., Kueh, A. J., Brennan, M. S., O'Connor, L., Milla, L., Wilcox, S., Tai, L., Strasser, A. and Herold, M. J. (2015) 'An inducible lentiviral guide RNA platform enables the identification of tumor-essential genes and tumor-promoting mutations in vivo', *Cell reports,* 10(8), pp. 1422-1432.

Awe, S. and Renkawitz-Pohl, R. (2010) 'Histone H4 acetylation is essential to proceed from a histone- to a protamine-based chromatin structure in spermatid nuclei of Drosophila melanogaster', *Systems biology in reproductive medicine,* 56(1), pp. 44-61.

Aznauryan, E., Yermanos, A., Kinzina, E., Devaux, A., Kapetanovic, E., Milanova, D., Church, G. M. and Reddy, S. T. (2022) 'Discovery and validation of human genomic safe harbor sites for gene and cell therapies', *Cell Reports Methods,* 2(1).

Bachtrog, D. (2013) 'Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration', *Nature Reviews Genetics,* 14(2), pp. 113-124.

Balestrino, F., Puggioli, A., Carrieri, M., Bouyer, J. and Bellini, R. (2017) 'Quality control methods for Aedes albopictus sterile male production', *PLoS neglected tropical diseases,* 11(9), pp. e0005881.

Barckmann, B., Chen, X., Kaiser, S., Jayaramaiah-Raja, S., Rathke, C., Dottermusch-Heidel, C., Fuller, M. T. and Renkawitz-Pohl, R. (2013) 'Three levels of regulation lead to protamine and Mst77F expression in Drosophila', *Developmental biology,* 377(1), pp. 33-45.

Barolo, S., Castro, B. and Posakony, J. W. (2004) 'New Drosophila transgenic reporters: insulated P-element vectors expressing fast-maturing RFP', *Biotechniques,* 36(3), pp. 436-442.

Bartussek, J., Mutlu, A. K., Zapotocky, M. and Fry, S. N. (2013) 'Limit-cycle-based control of the myogenic wingbeat rhythm in the fruit fly Drosophila', *Journal of the royal society interface,* 10(80), pp. 20121013.

Bassett, A. R., Tibbit, C., Ponting, C. P. and Liu, J.-L. (2013) 'Highly efficient targeted mutagenesis of Drosophila with the CRISPR/Cas9 system', *Cell reports,* 4(1), pp. 220-228.

Beaghton, A., Beaghton, P. J. and Burt, A. (2016) 'Gene drive through a landscape: reaction–diffusion models of population suppression and elimination by a sex ratio distorter', *Theoretical population biology,* 108, pp. 51-69.

Beaghton, A., Hammond, A., Nolan, T., Crisanti, A., Godfray, H. C. J. and Burt, A. (2017) 'Requirements for driving antipathogen effector genes into populations of disease vectors by homing', *Genetics,* 205(4), pp. 1587-1596.

Beall, C. J., Sepanski, M. A. and Fyrberg, E. A. (1989) 'Genetic dissection of Drosophila myofibril formation: effects of actin and myosin heavy chain null alleles', *Genes & development,* 3(2), pp. 131-140.

Behura, S. K., Sarro, J., Li, P., Mysore, K., Severson, D. W., Emrich, S. J. and Duman-Scheel, M. (2016) 'High-throughput cis-regulatory element discovery in the vector mosquito Aedes aegypti', *BMC genomics,* 17(1), pp. 1-11.

Bell, A. C., West, A. G. and Felsenfeld, G. (2001) 'Insulators and boundaries: versatile regulatory elements in the eukaryotic genome', *Science,* 291(5503), pp. 447-450.

Benedict, M. Q., Charlwood, J. D., Harrington, L. C., Lounibos, L. P., Reisen, W. K. and Tabachnick, W. J. (2018) 'Guidance for evaluating the safety of experimental releases of mosquitoes, emphasizing mark-release-recapture techniques', *Vector-Borne and Zoonotic Diseases,* 18(1), pp. 39-48.

Benedict, M. Q., Levine, R. S., Hawley, W. A. and Lounibos, L. P. (2007) 'Spread of the tiger: global risk of invasion by the mosquito Aedes albopictus', *Vector-borne and zoonotic Diseases,* 7(1), pp. 76-85.

Berg, J. S., Powell, B. C. and Cheney, R. E. (2001) 'A millennial myosin census', *Molecular biology of the cell,* 12(4), pp. 780-794.

Bernardini, F., Galizi, R., Menichelli, M., Papathanos, P.-A., Dritsou, V., Marois, E., Crisanti, A. and Windbichler, N. (2014) 'Site-specific genetic engineering of the Anopheles gambiae Y chromosome', *Proceedings of the National Academy of Sciences,* 111(21), pp. 7600-7605.

Bernardini, F., Galizi, R., Wunderlich, M., Taxiarchi, C., Kranjc, N., Kyrou, K., Hammond, A., Nolan, T., Lawniczak, M. N. and Papathanos, P. A. (2017) 'Cross-species Y chromosome function between malaria vectors of the Anopheles gambiae species complex', *Genetics,* 207(2), pp. 729-740.

Bernstein, S. I. and Milligan, R. A. 1997. Fine tuning a molecular motor: the location of alternative domains in the Drosophila myosin head. Elsevier.

Bernstein, S. I., Mogami, K., Donady, J. J. and Emerson Jr, C. P. (1983) 'Drosophila muscle myosin heavy chain encoded by a single gene in a cluster of muscle mutations', *Nature,* 302(5907), pp. 393-397.

Bernstein, S. I., O'Donnell, P. T. and Cripps, R. M. (1993) 'Molecular genetic analysis of muscle development, structure, and function in Drosophila', *International review of cytology,* 143, pp. 63-152.

Bhatt, S. and Chalmers, R. (2019) 'Targeted DNA transposition in vitro using a dCas9-transposase fusion protein', *Nucleic acids research,* 47(15), pp. 8126-8135.

Bhatt, S., Gething, P. W., Brady, O. J., Messina, J. P., Farlow, A. W., Moyes, C. L., Drake, J. M., Brownstein, J. S., Hoen, A. G. and Sankoh, O. (2013) 'The global distribution and burden of dengue', *Nature,* 496(7446), pp. 504-507.

Biedler, J. K., Hu, W., Tae, H. and Tu, Z. (2012) 'Identification of early zygotic genes in the yellow fever mosquito Aedes aegypti and discovery of a motif involved in early zygotic genome activation', *PloS one,* 7(3), pp. e33933.

Boerke, A., Dieleman, S. and Gadella, B. (2007) 'A possible role for sperm RNA in early embryo development', *Theriogenology,* 68, pp. S147-S155.

Bond, J. G., Osorio, A. R., Avila, N., Gómez-Simuta, Y., Marina, C. F., Fernández-Salas, I., Liedo, P., Dor, A., Carvalho, D. O. and Bourtzis, K. (2019) 'Optimization of irradiation dose to Aedes aegypti and Ae. albopictus in a sterile insect technique program', *PloS one,* 14(2), pp. e0212520.

Boob, A. G., Zhu, Z., Intasian, P., Jain, M., Petrov, Vassily A., Lane, S. T., Tan, S.-I., Xun, G. and Zhao, H. (2024) 'CRISPR-COPIES: an in silico platform for discovery of neutral integration sites for CRISPR/Cas-facilitated gene integration', *Nucleic Acids Research,* 52(6), pp. e30-e30.

Bourc'his, D. and Voinnet, O. (2010) 'A small-RNA perspective on gametogenesis, fertilization, and early zygotic development', *Science,* 330(6004), pp. 617-622.

Bouyer, J. (2023) 'When less is more: accounting for overcompensation in mosquito SIT projects', *Trends in Parasitology*.

Bouyer, J., Culbert, N. J., Dicko, A. H., Pacheco, M. G., Virginio, J., Pedrosa, M. C., Garziera, L., Pinto, A. M., Klaptocz, A. and Germann, J. (2020a) 'Field performance of sterile male mosquitoes released from an uncrewed aerial vehicle', *Science Robotics,* 5(43), pp. eaba6251.

Bouyer, J., Yamada, H., Pereira, R., Bourtzis, K. and Vreysen, M. J. (2020b) 'Phased conditional approach for mosquito management using sterile insect technique', *Trends in Parasitology,* 36(4), pp. 325-336.

Boye, C., Arpag, S., Francis, M., DeClemente, S., West, A., Heller, R. and Bulysheva, A. (2022) 'Reduction of plasmid vector backbone length enhances reporter gene expression', *Bioelectrochemistry,* 144, pp. 107981.

Brady, J. R., Tan, M. C., Whittaker, C. A., Colant, N. A., Dalvie, N. C., Love, K. R. and Love, J. C. (2020) 'Identifying improved sites for heterologous gene integration using ATAC-seq', *ACS synthetic biology,* 9(9), pp. 2515-2524.

Braun, R. E., Peschon, J. J., Behringer, R. R., Brinster, R. L. and Palmiter, R. D. (1989) 'Protamine 3'-untranslated sequences regulate temporal translational control and subcellular localization of growth hormone in spermatids of transgenic mice', *Genes & development,* 3(6), pp. 793-802.

Bruter, A. V., Kandarakov, O. F. and Belyavsky, A. V. (2018) 'Persistence of plasmid-mediated expression of transgenes in human mesenchymal stem cells depends primarily on CpG levels of both vector and transgene', *The Journal of Gene Medicine,* 20(2-3), pp. e3009.

Bryant, W. B., Ray, S. and Mills, M. K. (2020) 'Global analysis of small non-coding RNA populations across tissues in the malaria vector, Anopheles gambiae', *Insects,* 11(7), pp. 406.

Buchman, A. and Akbari, O. S. (2019) 'Site-specific transgenesis of the Drosophila melanogaster Y-chromosome using CRISPR/Cas9', *Insect Molecular Biology,* 28(1), pp. 65-73.

Buenrostro, J. D., Wu, B., Chang, H. Y. and Greenleaf, W. J. (2015) 'ATAC-seq: a method for assaying chromatin accessibility genome-wide', *Current protocols in molecular biology,* 109(1), pp. 21.29. 1-21.29. 9.

Burki, T. (2020) 'Genetically modified mosquitoes', *The Lancet Infectious Diseases,* 20(12), pp. 1375-1376.

Burt, A. (2003) 'Site-specific selfish genes as tools for the control and genetic engineering of natural populations', *Proceedings of the Royal Society of London. Series B: Biological Sciences,* 270(1518), pp. 921-928.

Burt, A. and Deredec, A. (2018) 'Self-limiting population genetic control with sex-linked genome editors', *Proceedings of the Royal Society B: Biological Sciences,* 285(1883), pp. 20180776.

Burtis, K. C. and Baker, B. S. (1989) 'Drosophila doublesex gene controls somatic sexual differentiation by producing alternatively spliced mRNAs encoding related sex-specific polypeptides', *Cell,* 56(6), pp. 997-1010.

Campa, C. C., Weisbach, N. R., Santinha, A. J., Incarnato, D. and Platt, R. J. (2019) 'Multiplexed genome engineering by Cas12a and CRISPR arrays encoded on single transcripts', *Nature Methods,* 16(9), pp. 887-893.

Campbell, C. L., Dickson, L. B., Lozano-Fuentes, S., Juneja, P., Jiggins, F. M. and Black, W. C. (2017) 'Alternative patterns of sex chromosome differentiation in Aedes aegypti (L)', *BMC genomics,* 18(1), pp. 1-11.

Carabajal Paladino, L. Z., Wilson, R., Tng, P. Y., Dhokiya, V., Keen, E., Cuber, P., Larner, W., Rooney, S., Nicholls, M. and Uglow, A. (2023) 'Optimizing CRE and PhiC31 mediated recombination in Aedes aegypti', *Frontiers in Bioengineering and Biotechnology,* 11, pp. 1254863.

Carballar-Lejarazú, R., Jasinskiene, N. and James, A. A. (2013) 'Exogenous gypsy insulator sequences modulate transgene expression in the malaria vector mosquito, Anopheles stephensi', *Proceedings of the National Academy of Sciences,* 110(18), pp. 7176-7181.

Carballar-Lejarazú, R., Ogaugwu, C., Tushar, T., Kelsey, A., Pham, T. B., Murphy, J., Schmidt, H., Lee, Y., Lanzaro, G. C. and James, A. A. (2020) 'Next-generation gene drive for population modification of the malaria vector mosquito, Anopheles gambiae', *Proceedings of the National Academy of Sciences,* 117(37), pp. 22805-22814.

Carpenetti, T. L., Aryan, A., Myles, K. M. and Adelman, Z. N. (2012) 'Robust heat-inducible gene expression by two endogenous hsp70-derived promoters in transgenic Aedes aegypti', *Insect molecular biology,* 21(1), pp. 97-106.

Carpenter, J., Blomefield, T. and Vreysen, M. (2012) 'A flight cylinder bioassay as a simple, effective quality control test for Cydia pomonella', *Journal of Applied Entomology,* 136(9), pp. 711-720.

Carvalho, A. B. (2002) 'Origin and evolution of the Drosophila Y chromosome', *Current opinion in genetics & development,* 12(6), pp. 664-668.

Carvalho, A. B., Dobo, B. A., Vibranovski, M. D. and Clark, A. G. (2001) 'Identification of five new genes on the Y chromosome of Drosophila melanogaster', *Proceedings of the National Academy of Sciences,* 98(23), pp. 13225-13230.

Carvalho, D. O., McKemey, A. R., Garziera, L., Lacroix, R., Donnelly, C. A., Alphey, L., Malavasi, A. and Capurro, M. L. (2015) 'Suppression of a field population of Aedes aegypti in Brazil by sustained release of transgenic male mosquitoes', *PLoS neglected tropical diseases,* 9(7), pp. e0003864.

Casas-Tintó, S. and Ferrús, A. (2021) 'The haplolethality paradox of the wupA gene in Drosophila', *PLOS Genetics,* 17(3), pp. e1009108.

Cassone, B. J., Kay, R. G., Daugherty, M. P. and White, B. J. (2017) 'Comparative transcriptomics of malaria mosquito testes: function, evolution, and linkage', *G3: Genes, Genomes, Genetics,* 7(4), pp. 1127-1136.

Castillo, J., Jodar, M. and Oliva, R. (2018) 'The contribution of human sperm proteins to the development and epigenome of the preimplantation embryo', *Human Reproduction Update,* 24.

Cator, L. J., Arthur, B. J., Harrington, L. C. and Hoy, R. R. (2009) 'Harmonic convergence in the love songs of the dengue vector mosquito', *Science,* 323(5917), pp. 1077-1079.

Cator, L. J., Arthur, B. J., Ponlawat, A. and Harrington, L. C. (2011) 'Behavioral observations and sound recordings of free-flight mating swarms of Ae. aegypti (Diptera: Culicidae) in Thailand', *Journal of medical entomology,* 48(4), pp. 941-946.

Cator, L. J. and Harrington, L. C. (2011) 'The harmonic convergence of fathers predicts the mating success of sons in Aedes aegypti', *Animal behaviour,* 82(4), pp. 627-633.

Cator, L. J. and Zanti, Z. (2016) 'Size, sounds and sex: interactions between body size and harmonic convergence signals determine mating success in Aedes aegypti', *Parasites & vectors,* 9, pp. 1-11.

Catteruccia, F., Benton, J. P. and Crisanti, A. (2005) 'An Anopheles transgenic sexing strain for vector control', *Nature biotechnology,* 23(11), pp. 1414-1417.

Celestino-Montes, A., Hernández-Martínez, S., Rodríguez, M. H., Cázares-Raga, F. E., Vázquez-Calzada, C., Lagunes-Guillén, A., Chávez-Munguía, B., Rubio-Miranda, J. Á., Hernández-Cázares, F. d. J., Cortés-Martínez, L. and Hernández-Hernández, F. d. l. C. (2021) 'Development of the indirect flight muscles of Aedes aegypti, a main arbovirus vector', *BMC Developmental Biology,* 21(1), pp. 11.

Chakravorty, S., Tanner, B. C., Foelber, V. L., Vu, H., Rosenthal, M., Ruiz, T. and Vigoreaux, J. O. (2017) 'Flightin maintains myofilament lattice organization required for optimal flight power and courtship song quality in Drosophila', *Proceedings of the Royal Society B: Biological Sciences,* 284(1854), pp. 20170431.

Chakravorty, S., Vu, H., Foelber, V. and Vigoreaux, J. O. (2014) 'Mutations of the Drosophila Myosin Regulatory Light Chain Affect Courtship Song and Reduce Reproductive Success', *PLOS ONE,* 9(2), pp. e90077.

Chakravorty, S., Wajda, M. P. and Vigoreaux, J. O. (2012) 'Courtship song analysis of Drosophila muscle mutants', *Methods,* 56(1), pp. 87-94.

Champer, J., Chung, J., Lee, Y. L., Liu, C., Yang, E., Wen, Z., Clark, A. G. and Messer, P. W. (2019) 'Molecular safeguarding of CRISPR gene drive experiments', *Elife,* 8, pp. e41439.

Champer, J., Liu, J., Oh, S. Y., Reeves, R., Luthra, A., Oakes, N., Clark, A. G. and Messer, P. W. (2018) 'Reducing resistance allele formation in CRISPR gene drive', *Proceedings of the National Academy of Sciences,* 115(21), pp. 5522-5527.

Champer, J., Reeves, R., Oh, S. Y., Liu, C., Liu, J., Clark, A. G. and Messer, P. W. (2017) 'Novel CRISPR/Cas9 gene drive constructs reveal insights into mechanisms of resistance allele formation and drive efficiency in genetically diverse populations', *PLoS genetics,* 13(7), pp. e1006796.

Champer, S. E., Oh, S. Y., Liu, C., Wen, Z., Clark, A. G., Messer, P. W. and Champer, J. (2020) 'Computational and experimental performance of CRISPR homing gene drive strategies with multiplexed gRNAs', *Science Advances,* 6(10), pp. eaaz0525.

Charlesworth, B. (1991) 'The evolution of sex chromosomes', *Science,* 251(4997), pp. 1030-1033.

Charlwood, J. and Jones, M. (1980) 'Mating in the mosquito, Anopheles gambiae sl: II. Swarming behaviour', *Physiological Entomology,* 5(4), pp. 315-320.

Chassin, H., Müller, M., Tigges, M., Scheller, L., Lang, M. and Fussenegger, M. (2019) 'A modular degron library for synthetic circuits in mammalian cells', *Nature communications,* 10(1), pp. 2013.

Chaverra-Rodriguez, D., Macias, V. M., Hughes, G. L., Pujhari, S., Suzuki, Y., Peterson, D. R., Kim, D., McKeand, S. and Rasgon, J. L. (2018) 'Targeted delivery of CRISPR-Cas9 ribonucleoprotein into arthropod ovaries for heritable germline gene editing', *Nature communications,* 9(1), pp. 3008.

Chen, C., Compton, A., Nikolouli, K., Wang, A., Aryan, A., Sharma, A., Qi, Y., Dellinger, C., Hempel, M., Potters, M., Augustinos, A., Severson, D. W., Bourtzis, K. and Tu, Z. (2022) 'Marker-assisted mapping enables forward genetic analysis in Aedes aegypti, an arboviral vector with vast recombination deserts', *Genetics,* 222(3).

Chen, C.-H., Huang, H., Ward, C. M., Su, J. T., Schaeffer, L. V., Guo, M. and Hay, B. A. (2007) 'A synthetic maternal-effect selfish genetic element drives population replacement in Drosophila', *science,* 316(5824), pp. 597-600.

Chen, J., Luo, J., Wang, Y., Gurav, A. S., Li, M., Akbari, O. S. and Montell, C. (2021) 'Suppression of female fertility in Aedes aegypti with a CRISPR-targeted male-sterile mutation', *Proceedings of the National Academy of Sciences,* 118(22), pp. e2105075118.

Chen, S. P. and Wang, H. H. (2019) 'An engineered Cas-transposon system for programmable and site-directed DNA transpositions', *The CRISPR journal,* 2(6), pp. 376-394.

Chen, W., McKenna, A., Schreiber, J., Haeussler, M., Yin, Y., Agarwal, V., Noble, W. S. and Shendure, J. (2019) 'Massively parallel profiling and predictive modeling of the outcomes of CRISPR/Cas9-mediated double-strand break repair', *Nucleic Acids Research,* 47(15), pp. 7989-8003.

Chen, Z.-Y., He, C.-Y., Ehrhardt, A. and Kay, M. A. (2003) 'Minicircle DNA vectors devoid of bacterial DNA result in persistent and high-level transgene expression in vivo', *Molecular therapy,* 8(3), pp. 495-500.

Chen, Z.-Y., Riu, E., He, C.-Y., Xu, H. and Kay, M. A. (2008) 'Silencing of episomal transgene expression in liver by plasmid bacterial backbone DNA is independent of CpG methylation', *Molecular Therapy,* 16(3), pp. 548-556.

Cherbas, L. and Gong, L. (2014) 'Cell lines', *Methods,* 68(1), pp. 74-81.

Chin, D. D. and Lentink, D. (2016) 'Flapping wing aerodynamics: from insects to vertebrates', *Journal of Experimental Biology,* 219(7), pp. 920-932.

Chun, M. and Falkenthal, S. (1988) 'Ifm(2)2 is a myosin heavy chain allele that disrupts myofibrillar assembly only in the indirect flight muscle of Drosophila melanogaster', *Journal of Cell Biology,* 107(6), pp. 2613-2621.

Clement, K., Rees, H., Canver, M. C., Gehrke, J. M., Farouni, R., Hsu, J. Y., Cole, M. A., Liu, D. R., Joung, J. K. and Bauer, D. E. (2019) 'CRISPResso2 provides accurate and rapid genome editing sequence analysis', *Nature biotechnology,* 37(3), pp. 224-226.

Coates, C. J., Jasinskiene, N., Miyashiro, L. and James, A. A. (1998) 'Mariner transposition and transformation of the yellow fever mosquito, Aedes aegypti', *Proceedings of the National Academy of Sciences,* 95(7), pp. 3748-3751.

Cochet, A., Calba, C., Jourdain, F., Grard, G., Durand, G. A., Guinard, A., Noël, H., Paty, M.-C. and Franke, F. (2022) 'Autochthonous dengue in mainland France, 2022: Geographical extension and incidence increase', *Eurosurveillance,* 27(44), pp. 2200818.

Condon, K. C., Condon, G. C., Dafa'alla, T. H., Fu, G., Phillips, C. E., Jin, L., Gong, P. and Alphey, L. (2007) 'Genetic sexing through the use of Y-linked transgenes', *Insect biochemistry and molecular biology,* 37(11), pp. 1168-1176.

Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P. D., Wu, X., Jiang, W. and Marraffini, L. A. (2013) 'Multiplex genome engineering using CRISPR/Cas systems', *Science,* 339(6121), pp. 819-823.

Conley, K. E. and Lindstedt, S. L. (2002) 'Energy-saving mechanisms in muscle: the minimization strategy', *Journal of experimental biology,* 205(15), pp. 2175-2181.

Courchesne-Smith, C. L. and Tobin, S. L. (1989) 'Tissue-specific expression of the 79B actin gene during Drosophila development', *Developmental biology,* 133(2), pp. 313-321.

Cridland, J. M. and Begun, D. J. (2023) 'Male-derived transcripts isolated from the mated female reproductive tract in Drosophila melanogaster', *G3 Genes|Genomes|Genetics,* 13(11).

Cripps, R. M., Ball, E., Stark, M., Lawn, A. and Sparrow, J. C. (1994) 'Recovery of dominant, autosomal flightless mutants of Drosophila melanogaster and identification of a new gene required for normal muscle structure and function', *Genetics,* 137(1), pp. 151-164.

Criscione, F., Qi, Y., Saunders, R., Hall, B. and Tu, Z. (2013) 'A unique Y gene in the A sian malaria mosquito A nopheles stephensi encodes a small lysine-rich protein and is transcribed at the onset of embryonic development', *Insect molecular biology,* 22(4), pp. 433-441.

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

Culbert, N. J., Balestrino, F., Dor, A., Herranz, G. S., Yamada, H., Wallner, T. and Bouyer, J. (2018) 'A rapid quality control test to foster the development of genetic control in mosquitoes', *Scientific Reports,* 8(1), pp. 16179.

Culbert, N. J., Somda, N. S. B., Hamidou, M., Soma, D. D., Caravantes, S., Wallner, T., Wadaka, M., Yamada, H. and Bouyer, J. (2020) 'A rapid quality control test to foster the development of the sterile insect technique against Anopheles arabiensis', *Malaria Journal,* 19(1), pp. 44.

Dadoune, J. P. (2009) 'Spermatozoal RNAs: what about their functions?', *Microscopy research and technique,* 72(8), pp. 536-551.

Daniels, R. W., Rossano, A. J., Macleod, G. T. and Ganetzky, B. (2014) 'Expression of multiple transgenes from a single construct using viral 2A peptides in Drosophila', *PloS one,* 9(6), pp. e100637.

Davis, S., Bax, N. and Grewe, P. (2001) 'Engineered underdominance allows efficient and economical introgression of traits into pest populations', *Journal of theoretical biology,* 212(1), pp. 83-98.

de Clare, M., Pir, P. and Oliver, S. G. (2011) 'Haploinsufficiency and the sex chromosomes from yeasts to humans', *BMC Biology,* 9(1), pp. 15.

Degarege, A., Fennie, K., Degarege, D., Chennupati, S. and Madhivanan, P. (2019) 'Improving socioeconomic status may reduce the burden of malaria in sub Saharan Africa: A systematic review and meta-analysis', *PloS one,* 14(1), pp. e0211205.

Deora, T., Gundiah, N. and Sane, S. P. (2017) 'Mechanics of the thorax in flies', *Journal of Experimental Biology,* 220(8), pp. 1382-1395.

Deredec, A., Burt, A. and Godfray, H. C. J. (2008) 'The population genetics of using homing endonuclease genes in vector and pest management', *Genetics,* 179(4), pp. 2013-2026.

Deredec, A., Godfray, H. C. J. and Burt, A. (2011) 'Requirements for effective malaria control with homing endonuclease genes', *Proceedings of the National Academy of Sciences,* 108(43), pp. E874-E880.

Dhiman, H., Campbell, M., Melcher, M., Smith, K. D. and Borth, N. (2020) 'Predicting favorable landing pads for targeted integrations in Chinese hamster ovary cell lines by learning stability characteristics from random transgene integrations', *Computational and Structural Biotechnology Journal,* 18, pp. 3632-3648.

Dhole, S., Vella, M. R., Lloyd, A. L. and Gould, F. (2018) 'Invasion and migration of spatially self-limiting gene drives: A comparative analysis', *Evolutionary applications,* 11(5), pp. 794-808.

Diao, F. and White, B. H. (2012) 'A novel approach for directing transgene expression in Drosophila: T2A-Gal4 in-frame fusion', *Genetics,* 190(3), pp. 1139-1144.

Dickinson, M. (2006) 'Insect flight', *Current Biology,* 16(9), pp. R309-R314.

Dickinson, M. H. and Tu, M. S. (1997) 'The function of dipteran flight muscle', *Comparative Biochemistry and Physiology Part A: Physiology,* 116(3), pp. 223-238.

Dickson, B. J. (2008) 'Wired for sex: the neurobiology of Drosophila mating decisions', *Science,* 322(5903), pp. 904-909.

Ding, D., Chen, K., Chen, Y., Li, H. and Xie, K. (2018) 'Engineering introns to express RNA guides for Cas9-and Cpf1-mediated multiplex genome editing', *Molecular plant,* 11(4), pp. 542-552.

Do, A. T., Brooks, J. T., Le Neveu, M. K. and LaRocque, J. R. (2014) 'Double-strand break repair assays determine pathway choice and structure of gene conversion events in Drosophila melanogaster', *G3: Genes, Genomes, Genetics,* 4(3), pp. 425-432.

Doench, J. G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E. W., Donovan, K. F., Smith, I., Tothova, Z., Wilen, C. and Orchard, R. (2016) 'Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9', *Nature biotechnology,* 34(2), pp. 184-191.

Dohn, T. E. and Cripps, R. M. (2018) 'Absence of the Drosophila jump muscle actin Act79B is compensated by up-regulation of Act88F', *Developmental Dynamics,* 247(4), pp. 642-649.

Dong, Z., Qin, Q., Hu, Z., Zhang, X., Miao, J., Huang, L., Chen, P., Lu, C. and Pan, M. (2020) 'CRISPR/Cas12a Mediated Genome Editing Enhances Bombyx mori Resistance to BmNPV', *Frontiers in Bioengineering and Biotechnology,* 8.

Dor, A., Maggiani-Aguilera, A. M., Valle-Mora, J., Bond, J. G., Marina, C. F. and Liedo, P. (2020) 'Assessment of Aedes aegypti (Diptera: Culicidae) males flight ability for SIT application: Effect

of device design, duration of test, and male Age', *Journal of Medical Entomology,* 57(3), pp. 824-829.

Dos Santos, G. R., Durovni, B., Saraceni, V., Riback, T. I. S., Pinto, S. B., Anders, K. L., Moreira, L. A. and Salje, H. (2022) 'Estimating the effect of the wMel release programme on the incidence of dengue and chikungunya in Rio de Janeiro, Brazil: a spatiotemporal modelling study', *The Lancet Infectious Diseases,* 22(11), pp. 1587-1595.

Downes, J. (1969) 'The swarming and mating flight of Diptera', *Annual review of entomology,* 14(1), pp. 271-298.

Drummond, D. R., Hennessey, E. S. and Sparrow, J. C. (1991) 'Characterisation of missense mutations in the Act88F gene of Drosophila melanogaster', *Molecular and General Genetics MGG,* 226, pp. 70-80.

Dyck, V. A., Hendrichs, J. and Robinson, A. S. (2021) *Sterile insect technique: principles and practice in area-wide integrated pest management*. Taylor & Francis.

Dye, C. (1984) 'Models for the population dynamics of the yellow fever mosquito, Aedes aegypti', *The Journal of Animal Ecology*, pp. 247-268.

Edenborough, K. M., Flores, H. A., Simmons, C. P. and Fraser, J. E. (2021) 'Using Wolbachia to eliminate dengue: Will the virus fight back?', *Journal of virology,* 95(13), pp. e02203-20.

Elgin, S. C. and Reuter, G. (2013) 'Position-effect variegation, heterochromatin formation, and gene silencing in Drosophila', *Cold Spring Harbor perspectives in biology,* 5(8), pp. a017780.

Elliott, B. and Jasin, M. (2001) 'Repair of double-strand breaks by homologous recombination in mismatch repair-defective mammalian cells', *Molecular and cellular biology,* 21(8), pp. 2671-2682.

Elliott, B., Richardson, C., Winderbaum, J., Nickoloff, J. A. and Jasin, M. (1998) 'Gene conversion tracts from double-strand break repair in mammalian cells', *Molecular and cellular biology,* 18(1), pp. 93-101.

Ellis, D. A., Avraam, G., Hoermann, A., Wyer, C. A., Ong, Y. X., Christophides, G. K. and Windbichler, N. (2022) 'Testing non-autonomous antimalarial gene drive effectors using self-eliminating drivers in the African mosquito vector Anopheles gambiae', *PLoS Genetics,* 18(6), pp. e1010244.

Emerson Jr, C. P. and Bernstein, S. I. (1987) 'Molecular genetics of myosin', *Annual review of biochemistry,* 56(1), pp. 695-726.

Esvelt, K. M., Smidler, A. L., Catteruccia, F. and Church, G. M. (2014) 'Concerning RNA-guided gene drives for the alteration of wild populations', *elife,* 3, pp. e03401.

Evans, K. G., Neale, Z. R., Holly, B., Canizela, C. C. and Juliano, S. A. (2022) 'Survival-larval density relationships in the field and their implications for control of container-dwelling Aedes mosquitoes', *Insects,* 14(1), pp. 17.

Ewing, A. W. (1977) 'The neuromuscular basis of courtship song in Drosophila: the role of the indirect flight muscles', *Journal of comparative physiology,* 119, pp. 249-265.

Fabian, L. and Brill, J. A. (2012) 'Drosophila spermiogenesis: Big things come from little packages', *Spermatogenesis,* 2(3), pp. 197-212.

Faria, N. R., Quick, J., Claro, I. M., Theze, J., de Jesus, J. G., Giovanetti, M., Kraemer, M. U., Hill, S. C., Black, A. and da Costa, A. C. (2017) 'Establishment and cryptic transmission of Zika virus in Brazil and the Americas', *Nature,* 546(7658), pp. 406-410.

Fasulo, B., Meccariello, A., Morgan, M., Borufka, C., Papathanos, P. A. and Windbichler, N. (2020) 'A fly model establishes distinct mechanisms for synthetic CRISPR/Cas9 sex distorters', *PLoS Genetics,* 16(3), pp. e1008647.

Fatica, A. and Bozzoni, I. (2014) 'Long non-coding RNAs: new players in cell differentiation and development', *Nature Reviews Genetics,* 15(1), pp. 7-21.

Feng, X., López Del Amo, V., Mameli, E., Lee, M., Bishop, A. L., Perrimon, N. and Gantz, V. M. (2021) 'Optimized CRISPR tools and site-directed transgenesis towards gene drive development in Culex quinquefasciatus mosquitoes', *Nature communications,* 12(1), pp. 2960.

Ferreira, R., Skrekas, C., Nielsen, J. and David, F. (2018) 'Multiplexed CRISPR/Cas9 genome editing and gene regulation using Csy4 in Saccharomyces cerevisiae', *ACS synthetic biology,* 7(1), pp. 10-15.

Finney, M., Romanowski, J. and Adelman, Z. N. (2022) 'Strategies to improve homology-based repair outcomes following CRISPR-based gene editing in mosquitoes: lessons in how to keep any repair disruptions local', *Virology Journal,* 19(1), pp. 128.

Fischer, B. E., Wasbrough, E., Meadows, L. A., Randlet, O., Dorus, S., Karr, T. L. and Russell, S. (2012) 'Conserved properties of Drosophila and human spermatozoal mRNA repertoires', *Proceedings of the Royal Society B: Biological Sciences,* 279(1738), pp. 2636-2644.

Focks, D. A. (1981) 'An improved separator for the developmental stages, sexes and species of mosquitoes (Diptera, Culicidae)'.

Fonfara, I., Richter, H., Bratovič, M., Le Rhun, A. and Charpentier, E. (2016) 'The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA', *Nature,* 532(7600), pp. 517-521.

Fontaine, A., Filipović, I., Fansiri, T., Hoffmann, A. A., Cheng, C., Kirkpatrick, M., Rašić, G. and Lambrechts, L. (2017) 'Extensive genetic differentiation between homomorphic sex chromosomes in the mosquito vector, Aedes aegypti', *Genome biology and evolution,* 9(9), pp. 2322-2335.

Franklinos, L. H., Jones, K. E., Redding, D. W. and Abubakar, I. (2019) 'The effect of global change on mosquito-borne disease', *The Lancet infectious diseases,* 19(9), pp. e302-e312.

Franz, A. W., Sanchez-Vargas, I., Raban, R. R., Black IV, W. C., James, A. A. and Olson, K. E. (2014) 'Fitness impact and stability of a transgene conferring resistance to dengue-2 virus following introgression into a genetically diverse Aedes aegypti strain', *PLoS neglected tropical diseases,* 8(5), pp. e2833.

Frøkjær-Jensen, C., Wayne Davis, M., Hopkins, C. E., Newman, B. J., Thummel, J. M., Olesen, S.-P., Grunnet, M. and Jorgensen, E. M. (2008) 'Single-copy insertion of transgenes in Caenorhabditis elegans', *Nature genetics,* 40(11), pp. 1375-1383.

Fu, G., Lees, R. S., Nimmo, D., Aw, D., Jin, L., Gray, P., Berendonk, T. U., White-Cooper, H., Scaife, S. and Kim Phuc, H. (2010) 'Female-specific flightless phenotype for mosquito control', *Proceedings of the National Academy of Sciences,* 107(10), pp. 4550-4554.

Funikov, S. Y., Rezvykh, A. P., Kulikova, D. A., Zelentsova, E. S., Protsenko, L. A., Chuvakova, L. N., Tyukmaeva, V. I., Arkhipova, I. R. and Evgen'ev, M. B. (2020) 'Adaptation of gene loci to heterochromatin in the course of Drosophila evolution is associated with insulator proteins', *Scientific Reports,* 10(1), pp. 11893.

Fyrberg, E. A., Fyrberg, C. C., Biggs, J. R., Saville, D., Beall, C. J. and Ketchum, A. (1998) 'Functional Nonequivalence of Drosophila Actin Isoforms', *Biochemical Genetics,* 36(7), pp. 271-287.

Fyrberg, E. A., Mahaffey, J. W., Bond, B. J. and Davidson, N. (1983) 'Transcripts of the six Drosophila actin genes accumulate in a stage-and tissue-specific manner', *Cell,* 33(1), pp. 115-123.

Galizi, R., Doyle, L. A., Menichelli, M., Bernardini, F., Deredec, A., Burt, A., Stoddard, B. L., Windbichler, N. and Crisanti, A. (2014) 'A synthetic sex ratio distortion system for the control of the human malaria mosquito', *Nature communications,* 5(1), pp. 3977.

Galizi, R., Hammond, A., Kyrou, K., Taxiarchi, C., Bernardini, F., O'Loughlin, S. M., Papathanos, P.-A., Nolan, T., Windbichler, N. and Crisanti, A. (2016) 'A CRISPR-Cas9 sex-ratio distortion system for genetic control', *Scientific reports,* 6(1), pp. 31139.

Gallup, J. L. and Sachs, J. D. (2000) 'The economic burden of malaria', *CID Working Paper Series*.

Gamez, S., Chaverra-Rodriguez, D., Buchman, A., Kandul, N. P., Mendez-Sanchez, S. C., Bennett, J. B., Sánchez C, H. M., Yang, T., Antoshechkin, I. and Duque, J. E. (2021) 'Exploiting a Y chromosome-linked Cas9 for sex selection and gene drive', *Nature communications,* 12(1), pp. 7202.

Gantz, V. M., Jasinskiene, N., Tatarenkova, O., Fazekas, A., Macias, V. M., Bier, E. and James, A. A. (2015) 'Highly efficient Cas9-mediated gene drive for population modification of the malaria vector mosquito Anopheles stephensi', *Proceedings of the National Academy of Sciences,* 112(49), pp. E6736-E6743.

Gao, Z., Harwig, A., Berkhout, B. and Herrera-Carrillo, E. (2017) 'Mutation of nucleotides around the+ 1 position of type 3 polymerase III promoters: The effect on transcriptional activity and start site usage', *Transcription,* 8(5), pp. 275-287.

Garcia Castillo, S. S., Pritts, K. S., Krishnan, R. S., Harrington, L. C. and League, G. P. (2021) 'Harmonic convergence coordinates swarm mating by enhancing mate detection in the malaria mosquito Anopheles gambiae', *Scientific Reports,* 11(1), pp. 24102.

Garlovsky, M. D., Sandler, J. A. and Karr, T. L. (2022) 'Functional diversity and evolution of the Drosophila sperm proteome', *Molecular & Cellular Proteomics,* 21(10).

Gaszner, M. and Felsenfeld, G. (2006) 'Insulators: exploiting transcriptional and epigenetic mechanisms', *Nature Reviews Genetics,* 7(9), pp. 703-713.

Geci, R., Willis, K. and Burt, A. (2022) 'Gene drive designs for efficient and localisable population suppression using Y-linked editors', *PLOS Genetics,* 18(12), pp. e1010550.

Gibson, G. and Russell, I. (2006) 'Flying in tune: sexual recognition in mosquitoes', *Current Biology,* 16(13), pp. 1311-1316.

Gibson, G., Warren, B. and Russell, I. J. (2010) 'Humming in tune: sex and species recognition by mosquitoes on the wing', *Journal of the Association for Research in Otolaryngology,* 11, pp. 527-540.

Gill, R. J., Ramos-Rodriguez, O. and Raine, N. E. (2012) 'Combined pesticide exposure severely affects individual-and colony-level traits in bees', *Nature,* 491(7422), pp. 105-108.

Giraldo-Calderón, G. I., Harb, O. S., Kelly, S. A., Rund, S. S., Roos, D. S. and McDowell, M. A. (2022) 'VectorBase. org updates: bioinformatic resources for invertebrate vectors of human pathogens and related organisms', *Current opinion in insect science,* 50, pp. 100860.

Gorman, K., Young, J., Pineda, L., Márquez, R., Sosa, N., Bernal, D., Torres, R., Soto, Y., Lacroix, R. and Naish, N. (2016) 'Short-term suppression of Aedes aegypti using genetic control does not facilitate Aedes albopictus', *Pest management science,* 72(3), pp. 618-628.

Gramates, L. S., Agapite, J., Attrill, H., Calvi, B. R., Crosby, M. A., dos Santos, G., Goodman, J. L., Goutte-Gattat, D., Jenkins, V. K., Kaufman, T., Larkin, A., Matthews, B. B., Millburn, G.,

Strelets, V. B. and Consortium, t. F. (2022) 'FlyBase: a guided tour of highlighted features', *Genetics,* 220(4).

Gray, C. E. and Coates, C. J. (2005) 'Cloning and characterization of cDNAs encoding putative CTCFs in the mosquitoes, Aedes aegypti and Anopheles gambiae', *BMC Molecular Biology,* 6(1), pp. 1-9.

Grimaldi, D. and Engel, M. S. (2005) *Evolution of the Insects*. Cambridge University Press.

Grobbelaar, A. A., Weyer, J., Moolla, N., van Vuren, P. J., Moises, F. and Paweska, J. T. (2016) 'Resurgence of yellow fever in Angola, 2015–2016', *Emerging infectious diseases,* 22(10), pp. 1854.

Guerra, C. A., Reiner, R. C., Perkins, T. A., Lindsay, S. W., Midega, J. T., Brady, O. J., Barker, C. M., Reisen, W. K., Harrington, L. C. and Takken, W. (2014) 'A global assembly of adult female mosquito mark-release-recapture data to inform the control of mosquito-borne pathogens', *Parasites & vectors,* 7, pp. 1-15.

Gunasekaran, K., Sadanandane, C., Panneer, D., Kumar, A., Rahi, M., Dinesh, S., Vijayakumar, B., Krishnaraja, M., Subbarao, S. K. and Jambulingam, P. (2022) 'Sensitivity of wMel and wAlbB Wolbachia infections in Aedes aegypti Puducherry (Indian) strains to heat stress during larval development', *Parasites & Vectors,* 15(1), pp. 1-10.

Gur, Y. and Breitbart, H. (2006) 'Mammalian sperm translate nuclear-encoded proteins by mitochondrial-type ribosomes', *Genes & development,* 20(4), pp. 411-416.

Haber, D. A., Arien, Y., Lamdan, L. B., Alcalay, Y., Zecharia, C., Krsticevic, F., Yonah, E. S., Avraham, R. D., Krzywinska, E., Krzywinski, J., Marois, E., Windbichler, N. and Papathanos, P. A. (2024) 'Targeting mosquito X-chromosomes reveals complex transmission dynamics of sex ratio distorting gene drives', *bioRxiv*, pp. 2023.08.20.554004.

Haghighat-Khah, R. E., Scaife, S., Martins, S., St John, O., Matzen, K. J., Morrison, N. and Alphey, L. (2015) 'Site-specific cassette exchange systems in the Aedes aegypti mosquito and the Plutella xylostella moth', *PLoS One,* 10(4), pp. e0121097.

Haigh, S. E., Salvi, S. S., Sevdali, M., Stark, M., Goulding, D., Clayton, J. D., Bullard, B., Sparrow, J. C. and Nongthomba, U. (2010) 'Drosophila indirect flight muscle specific Act88F actin mutants as a model system for studying congenital myopathies of the human ACTA1 skeletal muscle actin gene', *Neuromuscular Disorders,* 20(6), pp. 363-374.

Hall, A. B., Basu, S., Jiang, X., Qi, Y., Timoshevskiy, V. A., Biedler, J. K., Sharakhova, M. V., Elahi, R., Anderson, M. A. and Chen, X.-G. (2015) 'A male-determining factor in the mosquito Aedes aegypti', *Science,* 348(6240), pp. 1268-1270.

Hall, A. B., Papathanos, P.-A., Sharma, A., Cheng, C., Akbari, O. S., Assour, L., Bergman, N. H., Cagnetti, A., Crisanti, A. and Dottorini, T. (2016) 'Radical remodeling of the Y chromosome in a recent radiation of malaria mosquitoes', *Proceedings of the National Academy of Sciences,* 113(15), pp. E2114-E2123.

Hall, A. B., Timoshevskiy, V. A., Sharakhova, M. V., Jiang, X., Basu, S., Anderson, M. A. E., Hu, W., Sharakhov, I. V., Adelman, Z. N. and Tu, Z. (2014) 'Insights into the Preservation of the Homomorphic Sex-Determining Chromosome of Aedes aegypti from the Discovery of a Male-Biased Gene Tightly Linked to the M-Locus', *Genome Biology and Evolution,* 6(1), pp. 179-191.

Ham, D. T., Browne, T. S., Banglorewala, P. N., Wilson, T. L., Michael, R. K., Gloor, G. B. and Edgell, D. R. (2023) 'A generalizable Cas9/sgRNA prediction model using machine transfer learning with small high-quality datasets', *Nature Communications,* 14(1), pp. 5514.

Hammond, A., Galizi, R., Kyrou, K., Simoni, A., Siniscalchi, C., Katsanos, D., Gribble, M., Baker, D., Marois, E. and Russell, S. (2016) 'A CRISPR-Cas9 gene drive system targeting female

reproduction in the malaria mosquito vector Anopheles gambiae', *Nature biotechnology,* 34(1), pp. 78-83.

Hammond, A., Karlsson, X., Morianou, I., Kyrou, K., Beaghton, A., Gribble, M., Kranjc, N., Galizi, R., Burt, A. and Crisanti, A. (2021a) 'Regulating the expression of gene drives is key to increasing their invasive potential and the mitigation of resistance', *PLoS genetics,* 17(1), pp. e1009321.

Hammond, A., Pollegioni, P., Persampieri, T., North, A., Minuz, R., Trusso, A., Bucci, A., Kyrou, K., Morianou, I. and Simoni, A. (2021b) 'Gene-drive suppression of mosquito populations in large cages as a bridge between lab and field', *Nature communications,* 12(1), pp. 4589.

Hammond, A. M., Kyrou, K., Bruttini, M., North, A., Galizi, R., Karlsson, X., Kranjc, N., Carpi, F. M., D'Aurizio, R. and Crisanti, A. (2017) 'The creation and selection of mutations resistant to a gene drive over multiple generations in the malaria mosquito', *PLoS genetics,* 13(10), pp. e1007039.

Han, Q., Calvo, E., Marinotti, O., Fang, J., Rizzi, M., James, A. and Li, J. (2003) 'Analysis of the wild-type and mutant genes encoding the enzyme kynurenine monooxygenase of the yellow fever mosquito, Aedes aegypti', *Insect molecular biology,* 12(5), pp. 483-490.

Harris, A. F., McKemey, A. R., Nimmo, D., Curtis, Z., Black, I., Morgan, S. A., Oviedo, M. N., Lacroix, R., Naish, N. and Morrison, N. I. (2012) 'Successful suppression of a field mosquito population by sustained release of engineered male mosquitoes', *Nature biotechnology,* 30(9), pp. 828-830.

Hastings, G. A. and Emerson Jr, C. P. (1991) 'Myosin functional domains encoded by alternative exons are expressed in specific thoracic muscles of Drosophila', *The Journal of cell biology,* 114(2), pp. 263-276.

Hawley, W., Pumpuni, C., Brady, R. and Craig Jr, G. (1989) 'Overwintering survival of Aedes albopictus (Diptera: Culicidae) eggs in Indiana', *Journal of Medical Entomology,* 26(2), pp. 122-129.

Hecht, N. (1996) 'Posttranscriptional regulation of postmeiotic gene expression', *Signal Transduction in Testicular Cells: Basic and Clinical Aspects*: Springer, pp. 123-140.

Hedenström, A. (2014) 'How insect flight steering muscles work', *PLoS biology,* 12(3), pp. e1001822.

Helinski, M. E., Parker, A. G. and Knols, B. G. (2009) 'Radiation biology of mosquitoes', *Malaria journal,* 8(2), pp. 1-13.

Hemingway, J., Ranson, H., Magill, A., Kolaczinski, J., Fornadel, C., Gimnig, J., Coetzee, M., Simard, F., Roch, D. K. and Hinzoumbe, C. K. (2016) 'Averting a malaria disaster: will insecticide resistance derail malaria control?', *The Lancet,* 387(10029), pp. 1785-1788.

Hendel, A., Bak, R. O., Clark, J. T., Kennedy, A. B., Ryan, D. E., Roy, S., Steinfeld, I., Lunstad, B. D., Kaiser, R. J. and Wilkens, A. B. (2015) 'Chemically modified guide RNAs enhance CRISPR-Cas genome editing in human primary cells', *Nature biotechnology,* 33(9), pp. 985-989.

Heuschen, A.-K., Lu, G., Razum, O., Abdul-Mumin, A., Sankoh, O., Von Seidlein, L., D'Alessandro, U. and Müller, O. (2021) 'Public health-relevant consequences of the COVID-19 pandemic on malaria in sub-Saharan Africa: a scoping review', *Malaria journal,* 20, pp. 1-16.

Hew, B. E., Sato, R., Mauro, D., Stoytchev, I. and Owens, J. B. (2019) 'RNA-guided piggyBac transposition in human cells', *Synthetic Biology,* 4(1).

Hien, N. T., Anh, D. D., Le, N. H., Yen, N. T., Phong, T. V., Nam, V. S., Duong, T. N., Nguyen, N. B., Huong, D. T. and Hung, L. Q. (2021) 'Environmental factors influence the local establishment of Wolbachia in Aedes aegypti mosquitoes in two small communities in central Vietnam', *Gates Open Research,* 5.

Hill, R. W., Wyse, G. A. and Anderson, M. (2012) *Animal physiology* Sunderland, Mass.: Sinauer Associates, Inc. Publishers Sunderland, Mass. Available at: https://archive.org/details/animalphysiology0000hill.

Hiromi, Y., Okamoto, H., Gehring, W. J. and Hotta, Y. (1986) 'Germline transformation with Drosophila mutant actin genes induces constitutive expression of heat shock genes', *Cell,* 44(2), pp. 293-301.

Hoermann, A., Tapanelli, S., Capriotti, P., Del Corsano, G., Masters, E. K., Habtewold, T., Christophides, G. K. and Windbichler, N. (2021) 'Converting endogenous genes of the malaria mosquito into simple non-autonomous gene drives for population replacement', *Elife,* 10, pp. e58791.

Hoffmann, A. A., Montgomery, B., Popovici, J., Iturbe-Ormaetxe, I., Johnson, P., Muzzi, F., Greenfield, M., Durkan, M., Leong, Y. and Dong, Y. (2011) 'Successful establishment of Wolbachia in Aedes populations to suppress dengue transmission', *Nature,* 476(7361), pp. 454-457.

Holkers, M., Maggio, I., Henriques, S. F., Janssen, J. M., Cathomen, T. and Gonçalves, M. A. (2014) 'Adenoviral vector DNA for accurate genome editing with engineered nucleases', *Nature methods,* 11(10), pp. 1051-1057.

Hosken, D. J. and Hodgson, D. J. (2014) 'Why do sperm carry RNA? Relatedness, conflict, and control', *Trends in ecology & evolution,* 29(8), pp. 451-455.

Iwamoto, H. (2011) 'Structure, function and evolution of insect flight muscle', *Biophysics,* 7, pp. 21-28.

Iwamoto, H. (2017) 'The earliest molecular response to stretch of insect flight muscle as revealed by fast X-ray diffraction recording', *Scientific Reports,* 7(1), pp. 42272.

Izquierdo, A., Fahrenberger, M., Persampieri, T., Benedict, M. Q., Giles, T., Catteruccia, F., Emes, R. D. and Dottorini, T. (2019) 'Evolution of gene expression levels in the male reproductive organs of Anopheles mosquitoes', *Life science alliance,* 2(1).

James, S., Collins, F. H., Welkhoff, P. A., Emerson, C., Godfray, H. C. J., Gottlieb, M., Greenwood, B., Lindsay, S. W., Mbogo, C. M. and Okumu, F. O. (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', *The American journal of tropical medicine and hygiene,* 98(6 Suppl), pp. 1.

Jangir, P. K. and Prasad, A. (2022) 'Spatial distribution of insecticide resistance and susceptibility in Aedes aegypti and Aedes albopictus in India', *International Journal of Tropical Insect Science,* 42(2), pp. 1019-1044.

Jiang, X., Biedler, J. K., Qi, Y., Hall, A. B. and Tu, Z. (2015) 'Complete dosage compensation in Anopheles stephensi and the evolution of sex-biased genes in mosquitoes', *Genome biology and evolution,* 7(7), pp. 1914-1924.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A. and Charpentier, E. (2012) 'A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity', *science,* 337(6096), pp. 816-821.

Jové, V., Venkataraman, K., Gabel, T. M. and Duvall, L. B. (2020) 'Feeding and quantifying animal-derived blood and artificial meals in Aedes aegypti mosquitoes', *JoVE (Journal of Visualized Experiments),* (164), pp. e61835.

Juneja, P., Osei-Poku, J., Ho, Y. S., Ariani, C. V., Palmer, W. J., Pain, A. and Jiggins, F. M. (2014) 'Assembly of the genome of the disease vector Aedes aegypti onto a genetic linkage map allows mapping of genes affecting disease transmission', *PLoS neglected tropical diseases,* 8(1), pp. e2652.

Kaiser, V. B. and Bachtrog, D. (2010) 'Evolution of sex chromosomes in insects', *Annual review of genetics,* 44, pp. 91-112.

Karlik, C. C., Saville, D. L. and Fyrberg, E. A. (1987) 'Two missense alleles of the Drosophila melanogaster act88F actin gene are strongly antimorphic but only weakly induce synthesis of heat shock proteins', *Molecular and cellular biology,* 7(9), pp. 3084-3091.

Kashir, J., Nomikos, M. and Lai, F. A. (2018) 'Phospholipase C zeta and calcium oscillations at fertilisation: the evidence, applications, and further questions', *Advances in biological regulation,* 67, pp. 148-162.

Kaur, R., Leigh, B. A., Ritchie, I. T. and Bordenstein, S. R. (2022) 'The Cif proteins from Wolbachia prophage WO modify sperm genome integrity to establish cytoplasmic incompatibility', *PLoS biology,* 20(5), pp. e3001584.

Kaur, R., Meier, C. J., McGraw, E. A., Hillyer, J. F. and Bordenstein, S. R. (2024) 'The mechanism of cytoplasmic incompatibility is conserved in Wolbachia-infected Aedes aegypti mosquitoes deployed for arbovirus control', *PLOS Biology,* 22(3), pp. e3002573.

Kellum, R. and Schedl, P. (1991) 'A position-effect assay for boundaries of higher order chromosomal domains', *Cell,* 64(5), pp. 941-950.

Kemphues, K. J., Kaufman, T. C., Raff, R. A. and Raff, E. C. (1982) 'The testis-specific β-tubulin subunit in Drosophila melanogaster has multiple functions in spermatogenesis', *Cell,* 31(3), pp. 655-670.

Kim, D., DeBriere, T. J., Cherukumalli, S., White, G. S. and Burkett-Cadena, N. D. (2021) 'Infrared light sensors permit rapid recording of wingbeat frequency and bioacoustic species identification of mosquitoes', *Scientific Reports,* 11(1), pp. 10042.

Kingan, S. B., Heaton, H., Cudini, J., Lambert, C. C., Baybayan, P., Galvin, B. D., Durbin, R., Korlach, J. and Lawniczak, M. K. (2019) 'A high-quality de novo genome assembly from a single mosquito using PacBio sequencing', *Genes,* 10(1), pp. 62.

Kistler, K. E., Vosshall, L. B. and Matthews, B. J. (2015) 'Genome engineering with CRISPR-Cas9 in the mosquito Aedes aegypti', *Cell reports,* 11(1), pp. 51-60.

Kleene, K. C. (1989) 'Poly (A) shortening accompanies the activation of translation of five mRNAs during spermiogenesis in the mouse', *Development,* 106(2), pp. 367-373.

Kleinschmidt, I., Bradley, J., Knox, T. B., Mnzava, A. P., Kafy, H. T., Mbogo, C., Ismail, B. A., Bigoga, J. D., Adechoubou, A. and Raghavendra, K. (2018) 'Implications of insecticide resistance for malaria vector control with long-lasting insecticidal nets: a WHO-coordinated, prospective, international, observational cohort study', *The Lancet infectious diseases,* 18(6), pp. 640-649.

Kleinschmidt, I. and Rowland, M. (2021) 'Insecticides and malaria', *Ecology and Control of Vector-borne Diseases*: Wageningen Academic Publishers, pp. 266.

Kleinstiver, B. P., Sousa, A. A., Walton, R. T., Tak, Y. E., Hsu, J. Y., Clement, K., Welch, M. M., Horng, J. E., Malagon-Lopez, J. and Scarfò, I. (2019) 'Engineered CRISPR–Cas12a variants with increased activities and improved targeting ranges for gene, epigenetic and base editing', *Nature biotechnology,* 37(3), pp. 276-282.

Knapp, D. J. H. F., Michaels, Y. S., Jamilly, M., Ferry, Q. R. V., Barbosa, H., Milne, T. A. and Fulga, T. A. (2019) 'Decoupling tRNA promoter and processing activities enables specific Pol-II Cas9 guide RNA expression', *Nature Communications,* 10(1), pp. 1490.

Konet, D., Anderson, J., Piper, J., Akkina, R., Suchman, E. and Carlson, J. (2007) 'Short-hairpin RNA expressed from polymerase III promoters mediates RNA interference in mosquito cells', *Insect molecular biology,* 16(2), pp. 199-206.

Konkon, A. K., Padonou, G. G., Osse, R., Salako, A. S., Zoungbédji, D. M., Sina, H., Sovi, A., Tokponnon, F., Aïkpon, R. and Noukpo, H. (2023) 'Insecticide resistance status of Aedes aegypti and Aedes albopictus mosquitoes in southern Benin, West Africa', *Tropical Medicine and Health,* 51(1), pp. 22.

Konstantakos, V., Nentidis, A., Krithara, A. and Paliouras, G. (2022) 'CRISPR–Cas9 gRNA efficiency prediction: an overview of predictive tools and the role of deep learning', *Nucleic Acids Research,* 50(7), pp. 3616-3637.

Kotwica-Rolinska, J., Chodakova, L., Chvalova, D., Kristofova, L., Fenclova, I., Provaznik, J., Bertolutti, M., Wu, B. C.-H. and Dolezel, D. (2019) 'CRISPR/Cas9 genome editing introduction and optimization in the non-model insect Pyrrhocoris apterus', *Frontiers in physiology,* 10, pp. 891.

Kovač, A., Miskey, C., Menzel, M., Grueso, E., Gogol-Döring, A. and Ivics, Z. (2020) 'RNA-guided retargeting of S leeping Beauty transposition in human cells', *Elife,* 9, pp. e53868.

Kraemer, M. U., Reiner Jr, R. C., Brady, O. J., Messina, J. P., Gilbert, M., Pigott, D. M., Yi, D., Johnson, K., Earl, L. and Marczak, L. B. (2019) 'Past and future spread of the arbovirus vectors Aedes aegypti and Aedes albopictus', *Nature microbiology,* 4(5), pp. 854-863.

Krzywinska, E., Dennison, N. J., Lycett, G. J. and Krzywinski, J. (2016a) 'A maleness gene in the malaria mosquito Anopheles gambiae', *Science,* 353(6294), pp. 67-69.

Krzywinska, E., Kokoza, V., Morris, M., De La Casa-Esperon, E., Raikhel, A. and Krzywinski, J. (2016b) 'The sex locus is tightly linked to factors conferring sex-specific lethal effects in the mosquito Aedes aegypti', *Heredity,* 117(6), pp. 408-416.

Kuno, G. (2014) 'Early history of laboratory breeding of Aedes aegypti (Diptera: Culicidae) focusing on the origins and use of selected strains', *Journal of medical entomology,* 47(6), pp. 957-971.

Kurata, M., Wolf, N. K., Lahr, W. S., Weg, M. T., Kluesner, M. G., Lee, S., Hui, K., Shiraiwa, M., Webber, B. R. and Moriarity, B. S. (2018) 'Highly multiplexed genome engineering using CRISPR/Cas9 gRNA arrays', *PLoS one,* 13(9), pp. e0198714.

Kwasnieski, J. C., Orr-Weaver, T. L. and Bartel, D. P. (2019) 'Early genome activation in Drosophila is extensive with an initial tendency for aborted transcripts and retained introns', *Genome research,* 29(7), pp. 1188-1197.

Kyrchanova, O. and Georgiev, P. (2014) 'Chromatin insulators and long-distance interactions in Drosophila', *FEBS letters,* 588(1), pp. 8-14.

Kyrou, K., Hammond, A. M., Galizi, R., Kranjc, N., Burt, A., Beaghton, A. K., Nolan, T. and Crisanti, A. (2018) 'A CRISPR–Cas9 gene drive targeting doublesex causes complete population suppression in caged Anopheles gambiae mosquitoes', *Nature biotechnology,* 36(11), pp. 1062-1066.

Labbé, G. M. C., Scaife, S., Morgan, S. A., Curtis, Z. H. and Alphey, L. (2012) 'Female-Specific Flightless (fsRIDL) Phenotype for Control of Aedes albopictus', *PLOS Neglected Tropical Diseases,* 6(7), pp. e1724.

Labuhn, M., Adams, F. F., Ng, M., Knoess, S., Schambach, A., Charpentier, E. M., Schwarzer, A., Mateo, J. L., Klusmann, J.-H. and Heckl, D. (2018) 'Refined sgRNA efficacy prediction improves large-and small-scale CRISPR–Cas9 applications', *Nucleic acids research,* 46(3), pp. 1375-1385.

Labun, K., Montague, T. G., Krause, M., Torres Cleuren, Y. N., Tjeldnes, H. and Valen, E. (2019) 'CHOPCHOP v3: expanding the CRISPR web toolbox beyond genome editing', *Nucleic Acids Research,* 47(W1), pp. W171-W174.

Landmann, F., Orsi, G. A., Loppin, B. and Sullivan, W. (2009) 'Wolbachia-Mediated Cytoplasmic Incompatibility Is Associated with Impaired Histone Deposition in the Male Pronucleus', *PLOS Pathogens,* 5(3), pp. e1000343.

Lawler, C. D., Nuñez, A. K. P., Hernandes, N., Bhide, S., Lohrey, I., Baxter, S. and Robin, C. (2024) 'The haplolethal gene wupA of Drosophila exhibits potential as a target for an X-poisoning gene drive', *G3: Genes, Genomes, Genetics,* 14(4), pp. jkae025.

League, G. P., Harrington, L. C., Pitcher, S. A., Geyer, J. K., Baxter, L. L., Montijo, J., Rowland, J. G., Johnson, L. M., Murdock, C. C. and Cator, L. J. (2021) 'Sexual selection theory meets disease vector control: Testing harmonic convergence as a "good genes" signal in Aedes aegypti mosquitoes', *PLoS Neglected Tropical Diseases,* 15(7), pp. e0009540.

Lees, R. S., Gilles, J. R., Hendrichs, J., Vreysen, M. J. and Bourtzis, K. (2015) 'Back to the future: the sterile insect technique against mosquito disease vectors', *Current Opinion in Insect Science,* 10, pp. 156-162.

Leftwich, P. T., Purcell, J. C., Anderson, M. A. E., Fragkoudis, R., Basu, S., Lycett, G. and Alphey, L. (2023) 'A synthetic biology approach to transgene expression', *bioRxiv*, pp. 2023.08.31.555539.

Lei, R., Li, Y., Li, L., Wang, J., Cui, Z., Ju, R., Jiang, L., Liao, X., Wu, P. and Wang, X. (2022) 'A CRISPR/Cas12a-based portable platform for rapid detection of Leptosphaeria maculans in Brassica crops', *Frontiers in Plant Science,* 13, pp. 976510.

Lemke, S. B. and Schnorrer, F. (2017) 'Mechanical forces during muscle development', *Mechanisms of Development,* 144, pp. 92-101.

Lewis, I. C., Yan, Y. and Finnigan, G. C. (2021) 'Analysis of a Cas12a-based gene-drive system in budding yeast', *Access microbiology,* 3(12), pp. 000301.

Lezcano, Ó. M., Sánchez-Polo, M., Ruiz, J. L. and Gómez-Díaz, E. (2020) 'Chromatin structure and function in mosquitoes', *Frontiers in Genetics,* 11, pp. 602949.

Li, B., Carey, M. and Workman, J. L. (2007) 'The role of chromatin during transcription', *Cell,* 128(4), pp. 707-719.

Li, H. (2011) 'A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data', *Bioinformatics,* 27(21), pp. 2987-2993.

Li, H., Beckman, K. A., Pessino, V., Huang, B., Weissman, J. S. and Leonetti, M. D. (2017a) 'Design and specificity of long ssDNA donors for CRISPR-based knock-in', *BioRxiv*, pp. 178905.

Li, H., Janssens, J., De Waegeneer, M., Kolluru, S. S., Davie, K., Gardeux, V., Saelens, W., David, F. P., Brbić, M. and Spanier, K. (2022) 'Fly Cell Atlas: A single-nucleus transcriptomic atlas of the adult fruit fly', *Science,* 375(6584), pp. eabk2432.

Li, M., Bui, M., Yang, T., Bowman, C. S., White, B. J. and Akbari, O. S. (2017b) 'Germline Cas9 expression yields highly efficient genome engineering in a major worldwide disease vector, Aedes aegypti', *Proceedings of the National Academy of Sciences,* 114(49), pp. E10540-E10549.

Li, M., Yang, T., Bui, M., Gamez, S., Wise, T., Kandul, N. P., Liu, J., Alcantara, L., Lee, H. and Edula, J. R. (2021a) 'Suppressing mosquito populations with precision guided sterile males', *Nature Communications,* 12(1), pp. 5374.

Li, M., Yang, T., Kandul, N. P., Bui, M., Gamez, S., Raban, R., Bennett, J., Sánchez C, H. M., Lanzaro, G. C. and Schmidt, H. (2020) 'Development of a confinable gene drive system in the human disease vector Aedes aegypti', *Elife,* 9, pp. e51701.

Li, X., Xu, Y., Zhang, H., Yin, H., Zhou, D., Sun, Y., Ma, L., Shen, B. and Zhu, C. (2021b) 'ReMOT Control Delivery of CRISPR-Cas9 Ribonucleoprotein Complex to Induce Germline Mutagenesis in the Disease Vector Mosquitoes Culex pipiens pallens (Diptera: Culicidae)', *Journal of Medical Entomology,* 58(3), pp. 1202-1209.

Liang, X., Tan, C. H., Sun, Q., Zhang, M., Wong, P. S. J., Li, M. I., Mak, K. W., Martín-Park, A., Contreras-Perera, Y. and Puerta-Guardo, H. (2022) 'Wolbachia w AlbB remains stable in Aedes aegypti over 15 years but exhibits genetic background-dependent variation in virus blocking', *PNAS nexus,* 1(4), pp. pgac203.

Liu, W.-M., Pang, R. T., Chiu, P. C., Wong, B. P., Lao, K., Lee, K.-F. and Yeung, W. S. (2012) 'Sperm-borne microRNA-34c is required for the first cleavage division in mouse', *Proceedings of the National Academy of Sciences,* 109(2), pp. 490-494.

Lobo, N. F., Clayton, J. R., Fraser, M. J., Kafatos, F. C. and Collins, F. H. (2006) 'High efficiency germ-line transformation of mosquitoes', *Nature protocols,* 1(3), pp. 1312-1317.

Lombardo, A., Cesana, D., Genovese, P., Di Stefano, B., Provasi, E., Colombo, D. F., Neri, M., Magnani, Z., Cantore, A. and Riso, P. L. (2011) 'Site-specific integration and tailoring of cassette design for sustainable gene transfer', *Nature methods,* 8(10), pp. 861-869.

Long, K. C., Alphey, L., Annas, G. J., Bloss, C. S., Campbell, K. J., Champer, J., Chen, C.-H., Choudhary, A., Church, G. M. and Collins, J. P. (2020) 'Core commitments for field trials of gene drive organisms', *Science,* 370(6523), pp. 1417-1419.

Loppin, B., Dubruille, R. and Horard, B. (2015) 'The intimate genetics of Drosophila fertilization', *Open biology,* 5(8), pp. 150076.

Lu, J., Zhang, F., Xu, S., Fire, A. Z. and Kay, M. A. (2012) 'The extragenic spacer length between the 5′ and 3′ ends of the transgene expression cassette affects transgene silencing from plasmid-based vectors', *Molecular Therapy,* 20(11), pp. 2111-2119.

Luo, W., Galvan, D. L., Woodard, L. E., Dorset, D., Levy, S. and Wilson, M. H. (2017) 'Comparative analysis of chimeric ZFP-, TALE-and Cas9-piggyBac transposases for integration into a single locus in human cells', *Nucleic acids research,* 45(14), pp. 8411-8422.

Lutrat, C., Burckbuchler, M., Olmo, R. P., Beugnon, R., Fontaine, A., Akbari, O. S., Argilés-Herrero, R., Baldet, T., Bouyer, J. and Marois, E. (2023) 'Combining two genetic sexing strains allows sorting of non-transgenic males for Aedes genetic control', *Communications Biology,* 6(1), pp. 646.

Lutrat, C., Olmo, R. P., Baldet, T., Bouyer, J. and Marois, E. (2022) 'Transgenic expression of Nix converts genetic females into males and allows automated sex sorting in Aedes albopictus', *Communications biology,* 5(1), pp. 210.

Ma, H., Wu, Y., Dang, Y., Choi, J.-G., Zhang, J. and Wu, H. (2014) 'Pol III promoters to express small RNAs: delineation of transcription initiation', *Molecular Therapy-Nucleic Acids,* 3.

Ma, Q., Srivastav, S. P., Gamez, S., Dayama, G., Feitosa-Suntheimer, F., Patterson, E. I., Johnson, R. M., Matson, E. M., Gold, A. S. and Brackney, D. E. (2021) 'A mosquito small RNA genomics resource reveals dynamic evolution and host responses to viruses and transposons', *Genome research,* 31(3), pp. 512-528.

Madeira, F., Park, Y. M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., Basutkar, P., Tivey, A. R., Potter, S. C. and Finn, R. D. (2019) 'The EMBL-EBI search and sequence analysis tools APIs in 2019', *Nucleic acids research,* 47(W1), pp. W636-W641.

Magori, K. and Gould, F. (2006) 'Genetically engineered underdominance for manipulation of pest populations: a deterministic model', *Genetics,* 172(4), pp. 2613-2620.

Mahaffey, J. W., Coutu, M. D., Fyrberg, E. A. and Inwood, W. (1985) 'The flightless Drosophila mutant raised has two distinct genetic lesions affecting accumulation of myofibrillar proteins in flight muscles', *Cell,* 40(1), pp. 101-110.

Maïga, H., Lu, D., Mamai, W., Bimbilé Somda, N. S., Wallner, T., Bakhoum, M. T., Bueno Masso, O., Martina, C., Kotla, S. S. and Yamada, H. (2022) 'Standardization of the FAO/IAEA flight test for quality control of sterile mosquitoes', *Frontiers in Bioengineering and Biotechnology,* 10, pp. 876675.

Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., DiCarlo, J. E., Norville, J. E. and Church, G. M. (2013) 'RNA-guided human genome engineering via Cas9', *Science,* 339(6121), pp. 823-826.

Manier, M. K., Belote, J. M., Berben, K. S., Novikov, D., Stuart, W. T. and Pitnick, S. (2010) 'Resolving mechanisms of competitive fertilization success in Drosophila melanogaster', *science,* 328(5976), pp. 354-357.

Manigart, O., Ouedraogo, I., Ouedraogo, H. S., Sow, A. and Lokossou, V. K. (2024) 'Dengue epidemic in Burkina Faso: how can the response improve?', *The Lancet,* 403(10425), pp. 434-435.

Maragathavally, K., Kaminski, J. M., Coates, C., Maragathavally, K., Kaminski, J. and Coates, C. (2006) 'Chimeric Mos1 and piggyBac transposases result in site-directed integration', *The FASEB journal,* 20(11), pp. 1880-1882.

Markstein, M., Pitsouli, C., Villalta, C., Celniker, S. E. and Perrimon, N. (2008) 'Exploiting position effects and the gypsy retrovirus insulator to engineer precisely expressed transgenes', *Nature genetics,* 40(4), pp. 476-483.

Marshall, D. J. and Uller, T. (2007) 'When is a maternal effect adaptive?', *Oikos,* 116(12), pp. 1957-1963.

Marshall, J. M., Buchman, A., Sánchez C, H. M. and Akbari, O. S. (2017) 'Overcoming evolved resistance to population-suppressing homing-based gene drives', *Scientific Reports,* 7(1), pp. 3776.

Marshall, J. M. and Hay, B. A. (2012a) 'Confinement of gene drive systems to local populations: a comparative analysis', *Journal of Theoretical Biology,* 294, pp. 153-171.

Marshall, J. M. and Hay, B. A. (2012b) 'General principles of single-construct chromosomal gene drive', *Evolution,* 66(7), pp. 2150-2166.

Marygold, S. J., Roote, J., Reuter, G., Lambertsson, A., Ashburner, M., Millburn, G. H., Harrison, P. M., Yu, Z., Kenmochi, N. and Kaufman, T. C. (2007) 'The ribosomal protein genes and Minute loci of Drosophila melanogaster', *Genome biology,* 8, pp. 1-26.

Matthews, B. J., Dudchenko, O., Kingan, S. B., Koren, S., Antoshechkin, I., Crawford, J. E., Glassford, W. J., Herre, M., Redmond, S. N. and Rose, N. H. (2018) 'Improved reference genome of Aedes aegypti informs arbovirus vector control', *Nature,* 563(7732), pp. 501-507.

Matthews, B. J., McBride, C. S., DeGennaro, M., Despo, O. and Vosshall, L. B. (2016) 'The neurotranscriptome of the Aedes aegypti mosquito', *BMC genomics,* 17(1), pp. 1-20.

Matthews, B. J. and Vosshall, L. B. (2020) 'How to turn an organism into a model organism in 10 'easy'steps', *Journal of Experimental Biology,* 223(Suppl_1), pp. jeb218198.

Matthews, B. J., Younger, M. A. and Vosshall, L. B. (2019) 'The ion channel ppk301 controls freshwater egg-laying in the mosquito Aedes aegypti', *Elife,* 8, pp. e43963.

Mayer, S. V., Tesh, R. B. and Vasilakis, N. (2017) 'The emergence of arthropod-borne viral diseases: A global prospective on dengue, chikungunya and zika fevers', *Acta tropica,* 166, pp. 155-163.

Mayr, C. (2019) 'What are 3′ UTRs doing?', *Cold Spring Harbor perspectives in biology,* 11(10), pp. a034728.

McCarty, N. S., Graham, A. E., Studená, L. and Ledesma-Amaro, R. (2020) 'Multiplexed CRISPR technologies for gene editing and transcriptional regulation', *Nature communications,* 11(1), pp. 1281.

McCarty, N. S., Shaw, W. M., Ellis, T. and Ledesma-Amaro, R. (2019) 'Rapid assembly of gRNA arrays via modular cloning in yeast', *ACS synthetic biology,* 8(4), pp. 906-910.

McCullough, E. L., Whittington, E., Singh, A., Pitnick, S., Wolfner, M. F. and Dorus, S. (2022) 'The life history of Drosophila sperm involves molecular continuity between male and female reproductive tracts', *Proceedings of the National Academy of Sciences,* 119(11), pp. e2119899119.

McDonough-Goldstein, C. E., Pitnick, S. and Dorus, S. (2021) 'Drosophila oocyte proteome composition covaries with female mating status', *Scientific Reports,* 11(1), pp. 3142.

McNamara, C. J., Ant, T. H., Harvey-Samuel, T., White-Cooper, H., Martinez, J., Alphey, L. and Sinkins, S. P. (2024) 'Transgenic expression of cif genes from Wolbachia strain w AlbB recapitulates cytoplasmic incompatibility in Aedes aegypti', *Nature Communications,* 15(1), pp. 869.

Meccariello, A., Krsticevic, F., Colonna, R., Del Corsano, G., Fasulo, B., Papathanos, P. A. and Windbichler, N. (2021) 'Engineered sex ratio distortion by X-shredding in the global agricultural pest Ceratitis capitata', *BMC biology,* 19, pp. 1-14.

Mei, W., Jin, Z., Lai, F., Schwend, T., Houston, D. W., King, M. L. and Yang, J. (2013) 'Maternal Dead-End1 is required for vegetal cortical microtubule assembly during Xenopus axis specification', *Development,* 140(11), pp. 2334-2344.

Melo, E. S. d. and Wallau, G. L. (2020) 'Mosquito genomes are frequently invaded by transposable elements through horizontal transfer', *PLoS genetics,* 16(11), pp. e1008946.

Meredith, J. M., Basu, S., Nimmo, D. D., Larget-Thiery, I., Warr, E. L., Underhill, A., McArthur, C. C., Carter, V., Hurd, H. and Bourgouin, C. (2011) 'Site-specific integration and expression of an anti-malarial gene in transgenic Anopheles gambiae significantly reduces Plasmodium infections', *PLoS One,* 6(1), pp. e14587.

Messina, J. P., Brady, O. J., Scott, T. W., Zou, C., Pigott, D. M., Duda, K. A., Bhatt, S., Katzelnick, L., Howes, R. E. and Battle, K. E. (2014) 'Global spread of dengue virus types: mapping the 70 year history', *Trends in microbiology,* 22(3), pp. 138-146.

Messina, J. P., Kraemer, M. U., Brady, O. J., Pigott, D. M., Shearer, F. M., Weiss, D. J., Golding, N., Ruktanonchai, C. W., Gething, P. W. and Cohn, E. (2016) 'Mapping global environmental suitability for Zika virus', *elife,* 5, pp. e15272.

Miller, B. M., Bloemink, M. J., Nyitrai, M., Bernstein, S. I. and Geeves, M. A. (2007) 'A variable domain near the ATP-binding site in Drosophila muscle myosin is part of the communication pathway between the nucleotide and actin-binding sites', *Journal of molecular biology,* 368(4), pp. 1051-1066.

Minkenberg, B., Wheatley, M. and Yang, Y. (2017) 'Chapter Seven - CRISPR/Cas9-Enabled Multiplex Genome Editing and Its Application', in Weeks, D.P. and Yang, B. (eds.) *Progress in Molecular Biology and Translational Science*: Academic Press, pp. 111-132.

Mogami, K. and Hotta, Y. (1981) 'Isolation of Drosophila flightless mutants which affect myofibrillar proteins of indirect flight muscle', *Molecular and General Genetics MGG,* 183(3), pp. 409-417.

Morgan, M., Kumar, L., Li, Y. and Baptissart, M. (2021) 'Post-transcriptional regulation in spermatogenesis: all RNA pathways lead to healthy sperm', *Cellular and Molecular Life Sciences,* 78, pp. 8049-8071.

Morrill, S. A. (2020) *The persistence of haploinsufficiency and its role in genome evolution*. Doctor of Philosophy in Biology, Massachusetts Institute of Technology, Boston.

Morris, A. C., Eggleston, P. and Crampton, J. M. (1989) 'Genetic transformation of the mosquito Aedes aegypti by micro-injection of DNA', *Medical and veterinary entomology,* 3(1), pp. 1-7.

Morrison, N. I. (2022) 'Self-Limiting Insects for Pest Management', *Transgenic Insects: Techniques and Applications*: CABI GB, pp. 459-473.

Mousseau, T. A. and Fox, C. W. (1998) 'The adaptive significance of maternal effects', *Trends in ecology & evolution,* 13(10), pp. 403-407.

Mukundarajan, H., Hol, F. J. H., Castillo, E. A., Newby, C. and Prakash, M. (2017) 'Using mobile phones as acoustic sensors for high-throughput mosquito surveillance', *elife,* 6, pp. e27854.

Muller, H. J. (1964) 'The relation of recombination to mutational advance', *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis,* 1(1), pp. 2-9.

Munn, K. and Steward, R. (2000) 'The shut-down gene of Drosophila melanogaster encodes a novel FK506-binding protein essential for the formation of germline cysts during oogenesis', *Genetics,* 156(1), pp. 245-256.

Muñoz, D., Jimenez, A., Marinotti, O. and James, A. A. (2004) 'The AeAct-4 gene is expressed in the developing flight muscles of female Aedes aegypti', *Insect molecular biology,* 13(5), pp. 563-568.

Murphy, S., Di Liegro, C. and Melli, M. (1987) 'The in vitro transcription of the 7SK RNA gene by RNA polymerase III is dependent only on the presence of an upstream promoter', *Cell,* 51(1), pp. 81-87.

Mysore, K., Hapairai, L. K., Li, P., Roethele, J. B., Sun, L., Igiede, J., Misenti, J. K. and Duman-Scheel, M. (2021) 'A functional requirement for sex-determination M/m locus region lncRNA genes in Aedes aegypti female larvae', *Scientific Reports,* 11(1), pp. 10657.

Mysore, K., Li, P. and Duman-Scheel, M. (2018) 'Identification of Aedes aegypti cis-regulatory elements that promote gene expression in olfactory receptor neurons of distantly related dipteran insects', *Parasites & vectors,* 11(1), pp. 1-11.

Namekawa, S. H. and Lee, J. T. (2009) 'XY and ZW: is meiotic sex chromosome inactivation the rule in evolution?', *PLoS genetics,* 5(5), pp. e1000493.

Nash, A., Urdaneta, G. M., Beaghton, A. K., Hoermann, A., Papathanos, P. A., Christophides, G. K. and Windbichler, N. (2019) 'Integral gene drives for population replacement', *Biology open,* 8(1), pp. bio037762.

Navarro-Payá, D., Flis, I., Anderson, M. A., Hawes, P., Li, M., Akbari, O. S., Basu, S. and Alphey, L. (2020) 'Targeting female flight for genetic control of mosquitoes', *PLoS neglected tropical diseases,* 14(12), pp. e0008876.

Newton, M., Wood, R. and Southern, D. (1978) 'Cytological mapping of the M and D loci in the mosquito, Aedes aegypti (L.)', *Genetica,* 48(2), pp. 137-143.

Nissim, L., Perli, S. D., Fridkin, A., Perez-Pinera, P. and Lu, T. K. (2014) 'Multiplexed and programmable regulation of gene networks with an integrated RNA and CRISPR/Cas toolkit in human cells', *Molecular cell,* 54(4), pp. 698-710.

Noble, C., Min, J., Olejarz, J., Buchthal, J., Chavez, A., Smidler, A. L., DeBenedictis, E. A., Church, G. M., Nowak, M. A. and Esvelt, K. M. (2019) 'Daisy-chain gene drives for the alteration of local populations', *Proceedings of the National Academy of Sciences,* 116(17), pp. 8275-8282.

Noble, C., Olejarz, J., Esvelt, K. M., Church, G. M. and Nowak, M. A. (2017) 'Evolutionary dynamics of CRISPR gene drives', *Science advances,* 3(4), pp. e1601964.

Noguchi, T. Q., Gomibuchi, Y., Murakami, K., Ueno, H., Hirose, K., Wakabayashi, T. and Uyeda, T. Q. (2010) 'Dominant negative mutant actins identified in flightless Drosophila can be classified into three classes', *Journal of Biological Chemistry,* 285(7), pp. 4337-4347.

Nolan, T. and Hammond, A. M. (2022) 'Sex-, tissue-and stage-specific transgene expression', *Transgenic Insects: Techniques and Applications,* 10, pp. 42.

Nongthomba, U., Ansari, M., Thimmaiya, D., Stark, M. and Sparrow, J. (2007) 'Aberrant Splicing of an Alternative Exon in the Drosophila Troponin-T Gene Affects Flight Muscle Development', *Genetics,* 177(1), pp. 295-306.

Nongthomba, U., Pasalodos-Sanchez, S., Clark, S., Clayton, J. D. and Sparrow, J. C. (2001) 'Expression and function of the Drosophila ACT88F actin isoform is not restricted to the indirect flight muscles', *Journal of Muscle Research & Cell Motility,* 22, pp. 111-119.

North, A. R., Burt, A. and Godfray, H. C. J. (2020) 'Modelling the suppression of a malaria vector using a CRISPR-Cas9 gene drive to reduce female fertility', *BMC biology,* 18(1), pp. 1-14.

Nowling, R. J., Behura, S. K., Halfon, M. S., Emrich, S. J. and Duman-Scheel, M. (2021) 'PeakMatcher facilitates updated Aedes aegypti embryonic cis-regulatory element map', *Hereditas,* 158(1), pp. 1-8.

Nsoesie, E. O., Kraemer, M. U., Golding, N., Pigott, D. M., Brady, O. J., Moyes, C. L., Johansson, M. A., Gething, P. W., Velayudhan, R. and Khan, K. (2016) 'Global distribution and environmental suitability for chikungunya virus, 1952 to 2015', *Eurosurveillance,* 21(20), pp. 30234.

O'Brien, S. A., Lee, K., Fu, H. Y., Lee, Z., Le, T. S., Stach, C. S., McCann, M. G., Zhang, A. Q., Smanski, M. J. and Somia, N. V. (2018) 'Single copy transgene integration in a transcriptionally active site for recombinant protein synthesis', *Biotechnology journal,* 13(10), pp. 1800226.

O'Brochta, D. A. (2022) 'Transposon-Based Technologies for Insects', *Transgenic Insects: Techniques and Applications*: CABI GB, pp. 1-22.

O'Donnell, P. T., Collier, V. L., Mogami, K. and Bernstein, S. I. (1989) 'Ultrastructural and molecular analyses of homozygous-viable Drosophila melanogaster muscle mutants indicate there is a complex pattern of myosin heavy-chain isoform distribution', *Genes Dev,* 3(8), pp. 1233-1246.

O'Leary, S. and Adelman, Z. N. (2020) 'CRISPR/Cas9 knockout of female-biased genes AeAct-4 or myo-fem in Ae. aegypti results in a flightless phenotype in female, but not male mosquitoes', *PLOS Neglected Tropical Diseases,* 14(12), pp. e0008971.

O'Reilly, K. M., Hendrickx, E., Kharisma, D. D., Wilastonegoro, N. N., Carrington, L. B., Elyazar, I. R., Kucharski, A. J., Lowe, R., Flasche, S. and Pigott, D. M. (2019) 'Estimating the burden of dengue and the impact of release of wMel Wolbachia-infected mosquitoes in Indonesia: a modelling study', *BMC medicine,* 17(1), pp. 1-14.

Okamoto, H., Hiromi, Y., Ishikawa, E., Yamada, T., Isoda, K., Maekawa, H. and Hotta, Y. (1986) 'Molecular characterization of mutant actin genes which induce heat-shock proteins in Drosophila flight muscles', *The EMBO Journal,* 5(3), pp. 589-596.

Ostermeier, G. C., Miller, D., Huntriss, J. D., Diamond, M. P. and Krawetz, S. A. (2004) 'Delivering spermatozoan RNA to the oocyte', *Nature,* 429(6988), pp. 154-154.

Otte, M., Netschitailo, O., Kaftanoglu, O., Wang, Y., Page Jr, R. and Beye, M. (2018) 'Improving genetic transformation rates in honeybees', *Scientific reports,* 8(1), pp. 16534.

Oviedo-Pastrana, M., Méndez, N., Mattar, S., Arrieta, G. and Gomezcaceres, L. (2017) 'Epidemic outbreak of Chikungunya in two neighboring towns in the Colombian Caribbean: a survival analysis', *Archives of Public Health,* 75, pp. 1-8.

Owens, J. B., Mauro, D., Stoytchev, I., Bhakta, M. S., Kim, M.-S., Segal, D. J. and Moisyadi, S. (2013) 'Transcription activator like effector (TALE)-directed piggyBac transposition in human cells', *Nucleic acids research,* 41(19), pp. 9197-9207.

Page, N., Taxiarchi, C., Tonge, D., Kuburic, J., Chesters, E., Kriezis, A., Kyrou, K., Game, L., Nolan, T. and Galizi, R. (2023) 'Single-cell profiling of Anopheles gambiae spermatogenesis defines the onset of meiotic silencing and premeiotic overexpression of the X chromosome', *Communications Biology,* 6(1), pp. 850.

Paixão, E. S., Teixeira, M. G. and Rodrigues, L. C. (2018) 'Zika, chikungunya and dengue: the causes and threats of new and re-emerging arboviral diseases', *BMJ global health,* 3(Suppl 1).

Paiz-Reyes, C., Taracena, M. L., Flores-Ayuso, P., Hunt, C., Benedict, A. C., Illescas, K., Morales, M., Padilla, N., Dotson, E. M. and Pennington, P. M. (2020) 'Oral Delivery of dsRNA Targeting a Female-Biased Flight Muscle Actin Impairs Flight in the Malaria Vector, Anopheles Albimanus'.

Pan, X., Luo, Y., Liao, N., Zhang, Y., Xiao, M., Chen, P., Lu, C. and Dong, Z. (2022) 'CRISPR/Cpf1 multiplex genome editing system increases silkworm tolerance to BmNPV', *International Journal of Biological Macromolecules,* 200, pp. 566-573.

Papapetrou, E. P., Lee, G., Malani, N., Setty, M., Riviere, I., Tirunagari, L. M., Kadota, K., Roth, S. L., Giardina, P. and Viale, A. (2011) 'Genomic safe harbors permit high β-globin transgene expression in thalassemia induced pluripotent stem cells', *Nature biotechnology,* 29(1), pp. 73-78.

Papapetrou, E. P. and Schambach, A. (2016) 'Gene insertion into genomic safe harbors for human gene therapy', *Molecular Therapy,* 24(4), pp. 678-684.

Papp, B., Pál, C. and Hurst, L. D. (2003) 'Dosage sensitivity and the evolution of gene families in yeast', *Nature,* 424(6945), pp. 194-197.

Pérez-Mojica, J. E., Enders, L., Walsh, J., Lau, K. H. and Lempradl, A. (2023) 'Continuous transcriptome analysis reveals novel patterns of early gene expression in Drosophila embryos', *Cell genomics,* 3(3).

Perkins, A. D. and Tanentzapf, G. (2014) 'An ongoing role for structural sarcomeric components in maintaining Drosophila melanogaster muscle function and structure', *PLoS One,* 9(6), pp. e99362.

Perrin, B. J. and Ervasti, J. M. (2010) 'The actin gene family: function follows isoform', *Cytoskeleton,* 67(10), pp. 630-634.

Pfeiffer, B. D., Truman, J. W. and Rubin, G. M. (2012) 'Using translational enhancers to increase transgene expression in Drosophila', *Proceedings of the National Academy of Sciences,* 109(17), pp. 6626-6631.

Pfitzner, C., White, M. A., Piltz, S. G., Scherer, M., Adikusuma, F., Hughes, J. N. and Thomas, P. Q. (2020) 'Progress toward zygotic and germline gene drives in mice', *The CRISPR Journal,* 3(5), pp. 388-397.

Phillips-Cremins, J. E. and Corces, V. G. (2013) 'Chromatin insulators: linking genome organization to cellular function', *Molecular cell,* 50(4), pp. 461-474.

Phuc, H. K., Andreasen, M. H., Burton, R. S., Vass, C., Epton, M. J., Pape, G., Fu, G., Condon, K. C., Scaife, S. and Donnelly, C. A. (2007) 'Late-acting dominant lethal genetic systems and mosquito control', *BMC biology,* 5(1), pp. 1-11.

Pikaart, M. J., Recillas-Targa, F. and Felsenfeld, G. (1998) 'Loss of transcriptional activity of a transgene is accompanied by DNA methylation and histone deacetylation and is prevented by insulators', *Genes & development,* 12(18), pp. 2852-2862.

Pinto, S. B., Riback, T. I., Sylvestre, G., Costa, G., Peixoto, J., Dias, F. B., Tanamas, S. K., Simmons, C. P., Dufault, S. M. and Ryan, P. A. (2021) 'Effectiveness of Wolbachia-infected mosquito deployments in reducing the incidence of dengue and other Aedes-borne diseases in Niterói, Brazil: A quasi-experimental study', *PLoS neglected tropical diseases,* 15(7), pp. e0009556.

Port, F. and Bullock, S. L. (2016) 'Augmenting CRISPR applications in Drosophila with tRNA-flanked sgRNAs', *Nature methods,* 13(10), pp. 852-854.

Port, F., Starostecka, M. and Boutros, M. (2020) 'Multiplexed conditional genome editing with Cas12a in Drosophila', *Proceedings of the National Academy of Sciences,* 117(37), pp. 22890-22899.

Powell, J. R. and Tabachnick, W. J. (2013) 'History of domestication and spread of Aedes aegypti-a review', *Memórias do Instituto Oswaldo Cruz,* 108, pp. 11-17.

Preall, J. B., Czech, B., Guzzardo, P. M., Muerdter, F. and Hannon, G. J. (2012) 'shutdown is a component of the Drosophila piRNA biogenesis machinery', *Rna,* 18(8), pp. 1446-1457.

Price, D. C., Egizi, A. and Fonseca, D. M. (2015) 'The ubiquity and ancestry of insect doublesex', *Scientific reports,* 5(1), pp. 13068.

Purusothaman, D.-K., Shackleford, L., Anderson, M. A., Harvey-Samuel, T. and Alphey, L. (2021) 'CRISPR/Cas-9 mediated knock-in by homology dependent repair in the West Nile Virus vector Culex quinquefasciatus Say', *Scientific Reports,* 11(1), pp. 14964.

Qu, J., Betting, V., van Iterson, R., Kwaschik, F. M. and van Rij, R. P. (2023) 'Chromatin profiling identifies transcriptional readthrough as a conserved mechanism for piRNA biogenesis in mosquitoes', *Cell Reports,* 42(3).

Raja, S. J. and Renkawitz-Pohl, R. (2005) 'Replacement by Drosophila melanogaster protamines and Mst77F of histones during chromatin condensation in late spermatids and role of sesame in the removal of these proteins from the male pronucleus', *Molecular and cellular biology*.

Ramírez, A. L. 2019a. Aedes aegypti mosquito. figshare.

Ramírez, A. L. 2019b. (male) Aedes aegypti. figshare.

Ranson, H. and Lissenden, N. (2016) 'Insecticide resistance in African Anopheles mosquitoes: a worsening situation that needs urgent action to maintain malaria control', *Trends in parasitology,* 32(3), pp. 187-196.

Rathke, C., Baarends, W. M., Jayaramaiah-Raja, S., Bartkuhn, M., Renkawitz, R. and Renkawitz-Pohl, R. (2007) 'Transition from a nucleosome-based to a protamine-based chromatin configuration during spermiogenesis in Drosophila', *Journal of cell science,* 120(9), pp. 1689-1700.

Rathke, C., Barckmann, B., Burkhard, S., Jayaramaiah-Raja, S., Roote, J. and Renkawitz-Pohl, R. (2010) 'Distinct functions of Mst77F and protamines in nuclear shaping and chromatin condensation during Drosophila spermiogenesis', *European Journal of Cell Biology,* 89(4), pp. 326-338.

Rausch, T., Fritz, M. H.-Y., Untergasser, A. and Benes, V. (2020) 'Tracy: basecalling, alignment, assembly and deconvolution of sanger chromatogram trace files', *BMC Genomics,* 21(1), pp. 230.

Raz, A. A., Vida, G. S., Stern, S. R., Mahadevaraju, S., Fingerhut, J. M., Viveiros, J. M., Pal, S., Grey, J. R., Grace, M. R., Berry, C. W., Li, H., Janssens, J., Saelens, W., Shao, Z., Hu, C., Yamashita, Y. M., Przytycka, T., Oliver, B., Brill, J. A., Krause, H., Matunis, E. L., White-Cooper, H., DiNardo, S. and Fuller, M. T. (2023) 'Emergent dynamics of adult stem cell lineages from single nucleus and single cell RNA-Seq of Drosophila testes', *eLife,* 12, pp. e82201.

Recillas-Targa, F., Pikaart, M. J., Burgess-Beusse, B., Bell, A. C., Litt, M. D., West, A. G., Gaszner, M. and Felsenfeld, G. (2002) 'Position-effect protection and enhancer blocking by the chicken β-globin insulator are separable activities', *Proceedings of the National Academy of Sciences,* 99(10), pp. 6883-6888.

Reedy, M. C., Bullard, B. and Vigoreaux, J. O. (2000) 'Flightin is essential for thick filament assembly and sarcomere stability in Drosophila flight muscles', *The Journal of cell biology,* 151(7), pp. 1483-1500.

Reid, W., Williams, A. E., Sanchez-Vargas, I., Lin, J., Juncu, R., Olson, K. E. and Franz, A. W. (2022) 'Assessing single-locus CRISPR/Cas9-based gene drive variants in the mosquito Aedes aegypti via single-generation crosses and modeling', *G3,* 12(12), pp. jkac280.

Reitmayer, C. M., Pathak, A. K., Harrington, L. C., Brindley, M. A., Cator, L. J. and Murdock, C. C. (2021) 'Sex, age, and parental harmonic convergence behavior affect the immune performance of Aedes aegypti offspring', *Communications Biology,* 4(1), pp. 723.

Rijal, K. R., Adhikari, B., Ghimire, B., Dhungel, B., Pyakurel, U. R., Shah, P., Bastola, A., Lekhak, B., Banjara, M. R. and Pandey, B. D. (2021) 'Epidemiology of dengue virus infections in Nepal, 2006–2019', *Infectious diseases of poverty,* 10, pp. 1-10.

Riu, E., Chen, Z.-Y., Xu, H., He, C.-Y. and Kay, M. A. (2007) 'Histone modifications are associated with the persistence or silencing of vector-mediated transgene expression in vivo', *Molecular Therapy,* 15(7), pp. 1348-1355.

Rohde, B. B., Staunton, K. M., Zeak, N. C., Beebe, N., Snoad, N., Bondarenco, A., Liddington, C., Anderson, J. A., Xiang, W., Mankin, R. W. and Ritchie, S. A. (2019) 'Waterproof, low-cost, long-battery-life sound trap for surveillance of male Aedes aegypti for rear-and-release mosquito control programmes', *Parasites & Vectors,* 12(1), pp. 417.

Rojas-Fernandez, A., Herhaus, L., Macartney, T., Lachaud, C., Hay, R. T. and Sapkota, G. P. (2015) 'Rapid generation of endogenously driven transcriptional reporters in cells through CRISPR/Cas9', *Scientific Reports,* 5(1), pp. 9811.

Rosado Rivera, H. M. (2018) *Identification of U6 promoters for CRISPR-Cas9 genome engineering in Aedes aegypti*. Johns Hopkins University.

Roseman, R. R., Pirrotta, V. and Geyer, P. K. (1993) 'The su (Hw) protein insulates expression of the Drosophila melanogaster white gene from chromosomal position-effects', *The EMBO journal,* 12(2), pp. 435-442.

Ross, P. A., Axford, J. K., Yang, Q., Staunton, K. M., Ritchie, S. A., Richardson, K. M. and Hoffmann, A. A. (2020) 'Heatwaves cause fluctuations in w Mel Wolbachia densities and frequencies in Aedes aegypti', *PLoS neglected tropical diseases,* 14(1), pp. e0007958.

Rowley, W. A. and Graham, C. L. (1968) 'The effect of age on the flight performance of female Aedes aegypti mosquitoes', *Journal of insect physiology,* 14(5), pp. 719-728.

Rozek, C. E. and Davidson, N. (1983) 'Drosophila has one myosin heavy-chain gene with three developmentally regulated transcripts', *Cell,* 32(1), pp. 23-34.

Rudolph, J. E., Kimble, M., Hoyle, H. D., Subler, M. A. and Raff, E. C. (1987) 'Three Drosophila Beta-Tubulin Sequences: A Developmentally Regulated Isoform (β 3), the Testis-Specific Isoform (β 2), and an Assembly-Defective Mutation of the Testis-Specific Isoform (B2T8) Reveal Both an

Ancient Divergence in Metazoan Isotypes and Structural Constraints for Beta-Tubulin Function', *Molecular and cellular biology*.

Ruiz, J. L., Ranford-Cartwright, L. C. and Gómez-Díaz, E. (2021) 'The regulatory genome of the malaria vector Anopheles gambiae: integrating chromatin accessibility and gene expression', *NAR Genomics and Bioinformatics,* 3(1).

Ryan, P. A., Turley, A. P., Wilson, G., Hurst, T. P., Retzki, K., Brown-Kenyon, J., Hodgson, L., Kenny, N., Cook, H. and Montgomery, B. L. (2019) 'Establishment of wMel Wolbachia in Aedes aegypti mosquitoes and reduction of local dengue transmission in Cairns and surrounding locations in northern Queensland, Australia', *Gates open research,* 3.

Sachs, J. and Malaney, P. (2002) 'The economic and social burden of malaria', *Nature,* 415(6872), pp. 680-685.

Sæbøe-Larssen, S., Lyamouri, M., Merriam, J., Oksvold, M. P. and Lambertsson, A. (1998) 'Ribosomal protein insufficiency and the minute syndrome in Drosophila: a dose-response relationship', *Genetics,* 148(3), pp. 1215-1224.

Santel, A., Kaufmann, J., Hyland, R. and Renkawitz-Pohl, R. (2000) 'The initiator element of the Drosophila β 2 tubulin gene core promoter contributes to gene expression in vivo but is not required for male germ-cell specific expression', *Nucleic acids research,* 28(6), pp. 1439-1446.

Sanz Juste, S., Okamoto, E. M., Nguyen, C., Feng, X. and López Del Amo, V. (2023) 'Next-generation CRISPR gene-drive systems using Cas12a nuclease', *Nature Communications,* 14(1), pp. 6388.

Sarkar, A., Atapattu, A., Belikoff, E. J., Heinrich, J. C., Li, X., Horn, C., Wimmer, E. A. and Scott, M. J. (2006) 'Insulated piggyBac vectors for insect transgenesis', *BMC biotechnology,* 6(1), pp. 1-9.

Sawadogo, P., Namountougou, M., Toé, K., Rouamba, J., Maïga, H., Ouédraogo, K., Baldet, T., Gouagna, L. C., Kengne, P. and Simard, F. (2014) 'Swarming behaviour in natural populations of Anopheles gambiae and An. coluzzii: review of 4 years survey in rural areas of sympatry, Burkina Faso (West Africa)', *Acta tropica,* 132, pp. S42-S52.

Scali, C., Catteruccia, F., Li, Q. and Crisanti, A. (2005) 'Identification of sex-specific transcripts of the Anopheles gambiae doublesex gene', *Journal of experimental biology,* 208(19), pp. 3701-3709.

Schindele, P. and Puchta, H. (2020) 'Engineering CRISPR/LbCas12a for highly efficient, temperature-tolerant plant gene editing', *Plant Biotechnology Journal,* 18(5), pp. 1118.

Schotta, G., Ebert, A., Dorn, R. and Reuter, G. 'Position-effect variegation and the genetic dissection of chromatin regulation in Drosophila'. *Seminars in cell & developmental biology*: Elsevier, 67-75.

Schrauf, S., Tschismarov, R., Tauber, E. and Ramsauer, K. (2020) 'Current efforts in the development of vaccines for the prevention of Zika and Chikungunya virus infections', *Frontiers in immunology,* 11, pp. 592.

Schwartz, Y. B., Linder-Basso, D., Kharchenko, P. V., Tolstorukov, M. Y., Kim, M., Li, H.-B., Gorchakov, A. A., Minoda, A., Shanower, G. and Alekseyenko, A. A. (2012) 'Nature and function of insulator protein binding sites in the Drosophila genome', *Genome research,* 22(11), pp. 2188-2198.

Scolari, F., Schetelig, M. F., Bertin, S., Malacrida, A. R., Gasperi, G. and Wimmer, E. A. (2008) 'Fluorescent sperm marking to improve the fight against the pest insect Ceratitis capitata (Wiedemann; Diptera: Tephritidae)', *New biotechnology,* 25(1), pp. 76-84.

Sene, N. M., Mavridis, K., Ndiaye, E. H., Diagne, C. T., Gaye, A., Ngom, E. H. M., Ba, Y., Diallo, D., Vontas, J. and Dia, I. (2021) 'Insecticide resistance status and mechanisms in Aedes aegypti populations from Senegal', *PLoS Neglected Tropical Diseases,* 15(5), pp. e0009393.

Sette, C., Bevilacqua, A., Bianchini, A., Mangia, F., Geremia, R. and Rossi, P. (1997) 'Parthenogenetic activation of mouse eggs by microinjection of a truncated c-kit tyrosine kinase present in spermatozoa', *Development,* 124(11), pp. 2267-2274.

Sette, C., Paronetto, M. P., Barchi, M., Bevilacqua, A., Geremia, R. and Rossi, P. (2002) 'Tr-kit-induced resumption of the cell cycle in mouse eggs requires activation of a Src-like kinase', *The EMBO Journal*.

Sevdali, M., Kumar, V., Peckham, M. and Sparrow, J. (2013) 'Human congenital myopathy actin mutants cause myopathy and alter Z-disc structure in Drosophila flight muscle', *Neuromuscular Disorders,* 23(3), pp. 243-255.

Shadrin, I., Khodabukus, A. and Bursac, N. (2016) 'Striated muscle function, regeneration, and repair', *Cellular and molecular life sciences,* 73, pp. 4175-4202.

Sharakhov, I. V. and Sharakhova, M. V. (2015) 'Heterochromatin, histone modifications, and nuclear architecture in disease vectors', *Current opinion in insect science,* 10, pp. 110-117.

Shatskikh, A. S., Abramov, Y. A. and Lavrov, S. A. (2017) 'Trans-inactivation: Repression in a wrong place', *Fly,* 11(2), pp. 96-103.

Shaul, O. (2017) 'How introns enhance gene expression', *The international journal of biochemistry & cell biology,* 91, pp. 145-155.

She, W., Lin, W., Zhu, Y., Chen, Y., Jin, W., Yang, Y., Han, N., Bian, H., Zhu, M. and Wang, J. (2010) 'The gypsy Insulator of Drosophila melanogaster, Together With Its Binding Protein Suppressor of Hairy-Wing, Facilitate High and Precise Expression of Transgenes in Arabidopsis thaliana', *Genetics,* 185(4), pp. 1141-1150.

Sherrard-Smith, E., Hogan, A. B., Hamlet, A., Watson, O. J., Whittaker, C., Winskill, P., Ali, F., Mohammad, A. B., Uhomoibhi, P. and Maikore, I. (2020) 'The potential public health consequences of COVID-19 on malaria in Africa', *Nature medicine,* 26(9), pp. 1411-1416.

Shihab, H. A., Rogers, M. F., Campbell, C. and Gaunt, T. R. (2017) 'HIPred: an integrative approach to predicting haploinsufficient genes', *Bioinformatics,* 33(12), pp. 1751-1757.

Simoni, A., Hammond, A. M., Beaghton, A. K., Galizi, R., Taxiarchi, C., Kyrou, K., Meacci, D., Gribble, M., Morselli, G. and Burt, A. (2020) 'A male-biased sex-distorter gene drive for the human malaria vector Anopheles gambiae', *Nature biotechnology,* 38(9), pp. 1054-1060.

Singh, S. H., Kumar, P., Ramachandra, N. B. and Nongthomba, U. (2014) 'Roles of the troponin isoforms during indirect flight muscle development in Drosophila', *Journal of genetics,* 93, pp. 379-388.

Sinka, M., Pironon, S., Massey, N., Longbottom, J., Hemingway, J., Moyes, C. and Willis, K. (2020) 'A new malaria vector in Africa: predicting the expansion range of Anopheles stephensi and identifying the urban populations at risk', *Proceedings of the National Academy of Sciences,* 117(40), pp. 24900-24908.

Sinkins, S. P. and Gould, F. (2006) 'Gene drive systems for insect disease vectors', *Nature Reviews Genetics,* 7(6), pp. 427-435.

Smidler, A. L., Pai, J. J., Apte, R. A., Sánchez C., H. M., Corder, R. M., Jeffrey Gutiérrez, E., Thakre, N., Antoshechkin, I., Marshall, J. M. and Akbari, O. S. (2023) 'A confinable female-lethal population suppression system in the malaria vector, Anopheles gambiae', *Science Advances,* 9(27), pp. eade8903.

Smith, M. K. and Wakimoto, B. T. (2007) 'Complex regulation and multiple developmental functions of misfire, the Drosophila melanogaster ferlin gene', *BMC developmental biology,* 7, pp. 1-15.

Smith, R., Walter, M., Hice, R., O'Brochta, D. and Atkinson, P. (2007) 'Testis-specific expression of the β2 tubulin promoter of Aedes aegypti and its application as a genetic sex-separation marker', *Insect molecular biology,* 16(1), pp. 61-71.

Song, F. and Stieger, K. (2017) 'Optimizing the DNA donor template for homology-directed repair of double-strand breaks', *Molecular Therapy-Nucleic Acids,* 7, pp. 53-60.

Spinner, S. A., Barnes, Z. H., Puinean, A. M., Gray, P., Dafa'alla, T., Phillips, C. E., Nascimento de Souza, C., Frazon, T. F., Ercit, K. and Collado, A. (2022) 'New self-sexing Aedes aegypti strain eliminates barriers to scalable and sustainable vector control for governments and communities in dengue-prone environments', *Frontiers in Bioengineering and Biotechnology,* 10.

Stanaway, J. D., Shepard, D. S., Undurraga, E. A., Halasa, Y. A., Coffeng, L. E., Brady, O. J., Hay, S. I., Bedi, N., Bensenor, I. M. and Castañeda-Orjuela, C. A. (2016) 'The global burden of dengue: an analysis from the Global Burden of Disease Study 2013', *The Lancet infectious diseases,* 16(6), pp. 712-723.

Stoeckius, M., Grün, D. and Rajewsky, N. (2014) 'Paternal RNA contributions in the Caenorhabditis elegans zygote', *The EMBO Journal,* 33(16), pp. 1740-1750.

Stork, N. E. (2018) 'How many species of insects and other terrestrial arthropods are there on Earth?', *Annual review of entomology,* 63, pp. 31-45.

Su, M. P., Georgiades, M., Bagi, J., Kyrou, K., Crisanti, A. and Albert, J. T. (2020) 'Assessing the acoustic behaviour of Anopheles gambiae (s.l.) dsxF mutants: implications for vector control', *Parasites & Vectors,* 13(1), pp. 507.

Su, M. P. and Kamikouchi, A. (2023) 'Acoustic Communication in Fruit Flies and Mosquitoes', in Seki, Y. (ed.) *Acoustic Communication in Animals: From Insect Wingbeats to Human Music (Bioacoustics Series Vol.1)*. Singapore: Springer Nature Singapore, pp. 27-45.

Sun, C., Shao, Y. and Iqbal, J. (2023) 'Insect Insights at the Single-Cell Level: Technologies and Applications', *Cells,* 13(1), pp. 91.

Sun, R., Li, M., McMeniman, C. J. and Akbari, O. S. (2022) 'CRISPR-Mediated Genome Engineering in Aedes aegypti', *piRNA: Methods and Protocols*: Springer, pp. 23-51.

Suzuki, M., Kasai, K. and Saeki, Y. (2006) 'Plasmid DNA sequences present in conventional herpes simplex virus amplicon vectors cause rapid transgene silencing by forming inactive chromatin', *Journal of virology,* 80(7), pp. 3293-3300.

Swank, D. M., Bartoo, M. L., Knowles, A. F., Iliffe, C., Bernstein, S. I., Molloy, J. E. and Sparrow, J. C. (2001) 'Alternative exon-encoded regions of Drosophila myosin heavy chain modulate ATPase rates and actin sliding velocity', *Journal of Biological Chemistry,* 276(18), pp. 15117-15124.

Swank, D. M., Knowles, A. F., Suggs, J. A., Sarsoza, F., Lee, A., Maughan, D. W. and Bernstein, S. I. (2002) 'The myosin converter domain modulates muscle performance', *Nature Cell Biology,* 4(4), pp. 312-317.

Swank, D. M., Kronert, W. A., Bernstein, S. I. and Maughan, D. W. (2004) 'Alternative N-terminal regions of Drosophila myosin heavy chain tune muscle kinetics for optimal power output', *Biophysical journal,* 87(3), pp. 1805-1814.

Swank, D. M., Wells, L., Kronert, W. A., Morrill, G. E. and Bernstein, S. I. (2000) 'Determining structure/function relationships for sarcomeric myosin heavy chain by genetic and transgenic manipulation of Drosophila', *Microscopy research and technique,* 50(6), pp. 430-442.

Symington, L. S., Rothstein, R. and Lisby, M. (2014) 'Mechanisms and regulation of mitotic recombination in Saccharomyces cerevisiae', *Genetics,* 198(3), pp. 795-835.

Tadros, W. and Lipshitz, H. D. (2009) 'The maternal-to-zygotic transition: a play in two acts', *Development,* 136(18), pp. 3033-3042.

Takken, W., Koenraadt, C. J., Spitzen, J. and Velayudhan, R. (2021) 'Global Vector Control Response–supporting the pillars', *Ecology and Control of Vector-borne Diseases*: Wageningen Academic Publishers, pp. 456-465.

Tan, L., Hu, Y., Li, Y., Yang, L., Cai, X., Liu, W., He, J., Wu, Y., Liu, T. and Wang, N. (2020) 'Investigation of the molecular biology underlying the pronounced high gene targeting frequency at the Myh9 gene locus in mouse embryonic stem cells', *Plos one,* 15(3), pp. e0230126.

Tang, X., Zheng, X., Qi, Y., Zhang, D., Cheng, Y., Tang, A., Voytas, D. F. and Zhang, Y. (2016) 'A single transcript CRISPR-Cas9 system for efficient genome editing in plants', *Molecular plant,* 9(7), pp. 1088-1091.

Taxiarchi, C., Kranjc, N., Kriezis, A., Kyrou, K., Bernardini, F., Russell, S., Nolan, T., Crisanti, A. and Galizi, R. (2019) 'High-resolution transcriptional profiling of Anopheles gambiae spermatogenesis reveals mechanisms of sex chromosome regulation', *Scientific reports,* 9(1), pp. 14841.

Terradas, G., Bennett, J. B., Li, Z., Marshall, J. M. and Bier, E. (2023) 'Genetic conversion of a split-drive into a full-drive element', *Nature Communications,* 14(1), pp. 191.

Terradas, G., Hermann, A., James, A. A., McGinnis, W. and Bier, E. (2022a) 'High-resolution in situ analysis of Cas9 germline transcript distributions in gene-drive Anopheles mosquitoes', *G3,* 12(1), pp. jkab369.

Terradas, G., Macias, V. M., Peterson, H., McKeand, S., Krawczyk, G. and Rasgon, J. L. (2022b) 'Receptor-Mediated Ovary Transduction of Cargo–ReMOT Control: a Comprehensive Review and Detailed Protocol for Implementation', *Transgenic Insects: Techniques and Applications*, pp. 125-148.

Thi Hue Kien, D., Edenborough, K., da Silva Goncalves, D., Thuy Vi, T., Casagrande, E., Thi Le Duyen, H., Thi Long, V., Thi Dui, L., Thi Tuyet Nhu, V. and Thi Giang, N. (2023) 'Genome evolution of dengue virus serotype 1 under selection by Wolbachia pipientis in Aedes aegypti mosquitoes', *Virus Evolution,* 9(1), pp. vead016.

Thizy, D., Carter, L., Coche, I., Delborne, J. A., Emerson, C., Kormos, A., Mankad, A., Toe, L. P., Roberts, A. and Rohwer, Y. (2022) 'Public Acceptability and Stakeholder Engagement for Genetic Control Technologies', *Transgenic Insects: Techniques and Applications*: CABI GB, pp. 474-492.

Thizy, D., Toe, L. P., Mbogo, C., Matoke-Muhia, D., Alibu, V. P., Barnhill-Dilling, S. K., Chantler, T., Chongwe, G., Delborne, J. and Kapiriri, L. (2021) 'Proceedings of an expert workshop on community agreement for gene drive research in Africa-Co-organised by KEMRI, PAMCA and Target Malaria', *Gates open research,* 5.

Thomas, D. D., Donnelly, C. A., Wood, R. J. and Alphey, L. S. (2000) 'Insect population control using a dominant, repressible, lethal genetic system', *Science,* 287(5462), pp. 2474-2476.

Thorpe, H. M. and Smith, M. C. (1998) 'In vitro site-specific integration of bacteriophage DNA catalyzed by a recombinase of the resolvase/invertase family', *Proceedings of the National Academy of Sciences,* 95(10), pp. 5505-5510.

Timoshevskiy, V. A., Kinney, N. A., deBruyn, B. S., Mao, C., Tu, Z., Severson, D. W., Sharakhov, I. V. and Sharakhova, M. V. (2014) 'Genomic composition and evolution of Aedes aegyptichromosomes revealed by the analysis of physically mapped supercontigs', *BMC biology,* 12(1), pp. 1-13.

Toé, H. K., Zongo, S., Guelbeogo, M. W., Kamgang, B., Viana, M., Tapsoba, M., Sanou, A., Traoré, A., McCall, P. J. and Sagnon, N. F. (2022) 'Multiple insecticide resistance and first evidence of V410L kdr mutation in Aedes (Stegomyia) aegypti (Linnaeus) from Burkina Faso', *Medical and Veterinary Entomology,* 36(3), pp. 309-319.

Tolosana, I., Willis, K., Burt, A., Gribble, M., Nolan, T., Crisanti, A. and Bernardini, F. (2024) 'A Y chromosome-linked genome editor for efficient population suppression in the malaria vector Anopheles gambiae', *bioRxiv*, pp. 2024.05.14.594116.

Tomaszkiewicz, M., Medvedev, P. and Makova, K. D. (2017) 'Y and W chromosome assemblies: approaches and discoveries', *Trends in genetics,* 33(4), pp. 266-282.

Tørresen, O. K., Star, B., Mier, P., Andrade-Navarro, M. A., Bateman, A., Jarnot, P., Gruca, A., Grynberg, M., Kajava, A. V. and Promponas, V. J. (2019) 'Tandem repeats lead to sequence assembly errors and impose multi-level challenges for genome and protein databases', *Nucleic acids research,* 47(21), pp. 10994-11006.

Turner, J., Krishna, R., Van't Hof, A. E., Sutton, E. R., Matzen, K. and Darby, A. C. (2018) 'The sequence of a male-specific genome region containing the sex determination switch in Aedes aegypti', *Parasites & vectors,* 11, pp. 1-5.

Turner, J. M. (2007) 'Meiotic sex chromosome inactivation'.

Unckless, R. L., Clark, A. G. and Messer, P. W. (2017) 'Evolution of resistance against CRISPR/Cas9 gene drive', *Genetics,* 205(2), pp. 827-841.

Utarini, A., Indriani, C., Ahmad, R. A., Tantowijoyo, W., Arguni, E., Ansari, M. R., Supriyati, E., Wardana, D. S., Meitika, Y. and Ernesia, I. (2021) 'Efficacy of Wolbachia-infected mosquito deployments for the control of dengue', *New England Journal of Medicine,* 384(23), pp. 2177-2186.

Uusi-Mäkelä, M. I., Barker, H. R., Bäuerlein, C. A., Häkkinen, T., Nykter, M. and Rämet, M. (2018) 'Chromatin accessibility is associated with CRISPR-Cas9 efficiency in the zebrafish (Danio rerio)', *PloS one,* 13(4), pp. e0196238.

van Overbeek, M., Capurso, D., Carter, M. M., Thompson, M. S., Frias, E., Russ, C., Reece-Hoyes, J. S., Nye, C., Gradia, S. and Vidal, B. (2016) 'DNA repair profiling reveals nonrandom outcomes at Cas9-mediated breaks', *Molecular cell,* 63(4), pp. 633-646.

Varshavsky, A. (1996) 'The N-end rule: functions, mysteries, uses', *Proceedings of the National Academy of Sciences,* 93(22), pp. 12142-12149.

Varshavsky, A. (2011) 'The N-end rule pathway and regulation by proteolysis', *Protein science,* 20(8), pp. 1298-1345.

Vastenhouw, N. L., Cao, W. X. and Lipshitz, H. D. (2019) 'The maternal-to-zygotic transition revisited', *Development,* 146(11), pp. dev161471.

Verkhusha, V. V., Kuznetsova, I. M., Stepanenko, O. V., Zaraisky, A. G., Shavlovsky, M. M., Turoverov, K. K. and Uversky, V. N. (2003) 'High stability of Discosoma DsRed as compared to Aequorea EGFP', *Biochemistry,* 42(26), pp. 7879-7884.

Verkuijl, S. A., Ang, J. X., Alphey, L., Bonsall, M. B. and Anderson, M. A. (2022a) 'The challenges in developing efficient and robust synthetic homing endonuclease gene drives', *Frontiers in Bioengineering and Biotechnology,* 10, pp. 856981.

Verkuijl, S. A., Gonzalez, E., Li, M., Ang, J. X., Kandul, N. P., Anderson, M. A., Akbari, O. S., Bonsall, M. B. and Alphey, L. (2022b) 'A CRISPR endonuclease gene drive reveals distinct mechanisms of inheritance bias', *Nature Communications,* 13(1), pp. 7145.

Verkuijl, S. A. and Rots, M. G. (2019) 'The influence of eukaryotic chromatin state on CRISPR–Cas9 editing efficiencies', *Current opinion in biotechnology,* 55, pp. 68-73.

Vibranovski, M. D., Chalopin, D. S., Lopes, H. F., Long, M. and Karr, T. L. (2010) 'Direct Evidence for Postmeiotic Transcription During Drosophila melanogaster Spermatogenesis', *Genetics,* 186(1), pp. 431-433.

Vigoreaux, J. O., Hernandez, C., Moore, J., Ayer, G. and Maughan, D. (1998) 'A Genetic Deficiency That Spans the Flightin Gene of Drosophila Melanogaster affects the Ultrastructure and Function of the Flight Muscles', *Journal of Experimental Biology,* 201(13), pp. 2033-2044.

Volohonsky, G., Terenzi, O., Soichot, J., Naujoks, D. A., Nolan, T., Windbichler, N., Kapps, D., Smidler, A. L., Vittu, A., Costa, G., Steinert, S., Levashina, E. A., Blandin, S. A. and Marois, E. (2015) 'Tools for Anopheles gambiae Transgenesis', *G3 Genes|Genomes|Genetics,* 5(6), pp. 1151-1163.

Vyazunova, I. and Lan, Q. (2004) 'Stage-specific expression of two actin genes in the yellow fever mosquito, Aedes aegypti', *Insect molecular biology,* 13(3), pp. 241-249.

Walker, T., Johnson, P., Moreira, L., Iturbe-Ormaetxe, I., Frentiu, F., McMeniman, C., Leong, Y. S., Dong, Y., Axford, J. and Kriesner, P. (2011) 'The w Mel Wolbachia strain blocks dengue and invades caged Aedes aegypti populations', *Nature,* 476(7361), pp. 450-453.

Waltz, E. (2021) 'First genetically modified mosquitoes released in the United States', *Nature,* 593(7858), pp. 175-176.

Waltz, E. 2022. First results from US trial of genetically modified mosquitos. NATURE PORTFOLIO HEIDELBERGER PLATZ 3, BERLIN, 14197, GERMANY.

Wang, M., Mao, Y., Lu, Y., Wang, Z., Tao, X. and Zhu, J. K. (2018) 'Multiplex gene editing in rice with simplified CRISPR-Cpf1 and CRISPR-Cas9 systems', *Journal of integrative plant biology,* 60(8), pp. 626-631.

Ward, C. M., Su, J. T., Huang, Y., Lloyd, A. L., Gould, F. and Hay, B. A. (2011) 'Medea selfish genetic elements as tools for altering traits of wild populations: a theoretical analysis', *Evolution,* 65(4), pp. 1149-1162.

Warmke, J. W., Kreuz, A. J. and Falkenthal, S. (1989) 'Co-localization to chromosome bands 99E1-3 of the Drosophila melanogaster myosin light chain-2 gene and a haplo-insufficient locus that affects flight behavior', *Genetics,* 122(1), pp. 139-151.

Waterhouse, R. M., Aganezov, S., Anselmetti, Y., Lee, J., Ruzzante, L., Reijnders, M. J., Feron, R., Bérard, S., George, P. and Hahn, M. W. (2020) 'Evolutionary superscaffolding and chromosome anchoring to improve Anopheles genome assemblies', *BMC biology,* 18, pp. 1-20.

Weaver, S. C., Charlier, C., Vasilakis, N. and Lecuit, M. (2018) 'Zika, chikungunya, and other emerging vector-borne viral diseases', *Annual review of medicine,* 69, pp. 395-408.

Wells, L., Edwards, K. A. and Bernstein, S. I. (1996) 'Myosin heavy chain isoforms regulate muscle function but not myofibril assembly', *The EMBO journal,* 15(17), pp. 4454-4459.

Weng, S.-C., Masri, R. A. and Akbari, O. S. (2023) 'Advances and challenges in synthetic biology for mosquito control', *Trends in Parasitology*.

Werren, J. H., Baldo, L. and Clark, M. E. (2008) 'Wolbachia: master manipulators of invertebrate biology', *Nature Reviews Microbiology,* 6(10), pp. 741-751.

Wilder-Smith, A. and Monath, T. P. (2017) 'Responding to the threat of urban yellow fever outbreaks', *The Lancet Infectious Diseases,* 17(3), pp. 248-250.

Wilkie, A. (1994) 'The molecular basis of genetic dominance', *Journal of medical genetics,* 31(2), pp. 89.

Williams, J. A. and Paez, P. A. (2023) 'Improving Cell and Gene Therapy Safety and Performance using Next-Generation NanoplasmidTM Vectors', *Molecular Therapy-Nucleic Acids*.

Wilson, C., Bellen, H. J. and Gehring, W. J. (1990) 'Position effects on eukaryotic gene expression', *Annual review of cell biology,* 6(1), pp. 679-714.

Wilson, K. L., Fitch, K. R., Bafus, B. T. and Wakimoto, B. T. (2006) 'Sperm plasma membrane breakdown during Drosophila fertilization requires sneaky, an acrosomal membrane protein'.

Windbichler, N., Papathanos, P. A. and Crisanti, A. (2008) 'Targeting the X chromosome during spermatogenesis induces Y chromosome transmission ratio distortion and early dominant embryo lethality in Anopheles gambiae', *PLoS genetics,* 4(12), pp. e1000291.

World Health Organization (2017) *Global vector control response 2017–2030. Licence: CC BY-NC-SA 3.0 IGO*, Geneva: World Health Organization (978-92-4-151297-8.

World Health Organization (2018) 'Global report on insecticide resistance in malaria vectors: 2010–2016'.

World Health Organization (2019) 'Dengue vaccine: WHO position paper, September 2018-recommendations', *Vaccine,* 37(35), pp. 4848-4849.

World Health Organization (2021a) *Guidance framework for testing genetically modified mosquitoes, second edition. Licence: CC BY-NC-SA 3.0 IGO.*, Geneva: World Health Organization.

World Health Organization (2021b) *Thirteenth meeting of the WHO Vector Control Advisory Group. Licence: CC BY-NC-SA 3.0 IGO*, Geneva: World Health Organization.

World Health Organization (2022) *Manual for monitoring insecticide resistance in mosquito vectors and selecting appropriate interventions. Licence: CC BY-NC-SA 3.0 IGO*, Geneva: World Health Organization.

World Health Organization (2023a) 'Dengue – Global situation', *Disease Outbreak News*, *https://www.who.int/emergencies/disease-outbreak-news/item/2023-DON498*.

World Health Organization (2023b) *World malaria report 2023. Licence: CC BY-NC-SA 3.0 IGO.* Geneva: World Health Organization.

Wu, A. T., Sutovsky, P., Manandhar, G., Xu, W., Katayama, M., Day, B. N., Park, K.-W., Yi, Y.-J., Xi, Y. W. and Prather, R. S. (2007) 'PAWP, a Sperm-specific WW Domain-binding Protein, Promotes Meiotic Resumption and Pronuclear Development during Fertilization*♦', *Journal of Biological Chemistry,* 282(16), pp. 12164-12175.

Wu, Q., Liu, P. and Wang, L. (2020) 'Many facades of CTCF unified by its coding for three-dimensional genome architecture', *Journal of Genetics and Genomics,* 47(8), pp. 407-424.

Xie, K., Minkenberg, B. and Yang, Y. (2015) 'Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-processing system', *Proceedings of the National Academy of Sciences,* 112(11), pp. 3570-3575.

Xu, J., Bi, H., Chen, R., Aslam, A., Li, Z., Ling, L., Zeng, B., Huang, Y. and Tan, A. (2015) 'Transgenic characterization of two testis-specific promoters in the silkworm, B ombyx mori', *Insect molecular biology,* 24(2), pp. 183-190.

Xu, L., Zhao, L., Gao, Y., Xu, J. and Han, R. (2017) 'Empower multiplex cell and tissue-specific CRISPR-mediated gene manipulation with self-cleaving ribozymes and tRNA', *Nucleic acids research,* 45(5), pp. e28-e28.

Yamada, H., Maiga, H., Juarez, J., De Oliveira Carvalho, D., Mamai, W., Ali, A., Bimbile-Somda, N. S., Parker, A. G., Zhang, D. and Bouyer, J. (2019) 'Identification of critical factors that significantly affect the dose-response in mosquitoes irradiated as pupae', *Parasites & vectors,* 12(1), pp. 1-13.

Yamamoto, D. S., Sumitani, M. and Matsuoka, H. 'Identification of the beta2-tubulin gene promoter regulating testis-specific expression from Anopheles stephensi'. *Proc. Arthropod. Embryol. Soc. Jpn*, 21.

Yamamoto, Y. and Gerbi, S. A. (2022) 'Development of transformation for genome editing of an emerging model organism', *Genes,* 13(7), pp. 1108.

Yang, E., Metzloff, M., Langmüller, A. M., Xu, X., Clark, A. G., Messer, P. W. and Champer, J. (2022) 'A homing suppression gene drive with multiplexed gRNAs maintains high drive conversion efficiency and avoids functional resistance alleles', *G3,* 12(6), pp. jkac081.

Yang, Y., Li, S., Wang, Y., Ma, Z., Wong, K.-C. and Li, X. (2021) 'Identification of haploinsufficient genes from epigenomic data using deep forest', *Briefings in Bioinformatics,* 22(5).

Yasuhara, J. C., DeCrease, C. H. and Wakimoto, B. T. (2005) 'Evolution of heterochromatic genes of Drosophila', *Proceedings of the National Academy of Sciences,* 102(31), pp. 10958-10963.

Yoshioka, S., Fujii, W., Ogawa, T., Sugiura, K. and Naito, K. (2015) 'Development of a mono-promoter-driven CRISPR/Cas9 system in mammalian cells', *Scientific reports,* 5(1), pp. 18341.

Zamyatin, A., Avdeyev, P., Liang, J., Sharma, A., Chen, C., Lukyanchikova, V., Alexeev, N., Tu, Z., Alekseyev, M. A. and Sharakhov, I. V. (2021) 'Chromosome-level genome assemblies of the malaria vectors Anopheles coluzzii and Anopheles arabiensis', *Gigascience,* 10(3), pp. giab017.

Zelinger, E., Brumfeld, V., Rechav, K., Waiger, D., Kossovsky, T. and Heifetz, Y. (2024) 'Three-dimensional correlative microscopy of the Drosophila female reproductive tract reveals modes of communication in seminal receptacle sperm storage', *Communications Biology,* 7(1), pp. 155.

Zetsche, B., Gootenberg, J. S., Abudayyeh, O. O., Slaymaker, I. M., Makarova, K. S., Essletzbichler, P., Volz, S. E., Joung, J., Van Der Oost, J. and Regev, A. (2015) 'Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system', *cell,* 163(3), pp. 759-771.

Zetsche, B., Heidenreich, M., Mohanraju, P., Fedorova, I., Kneppers, J., DeGennaro, E. M., Winblad, N., Choudhury, S. R., Abudayyeh, O. O., Gootenberg, J. S., Wu, W. Y., Scott, D. A., Severinov, K., van der Oost, J. and Zhang, F. (2017) 'Multiplex gene editing by CRISPR–Cpf1 using a single crRNA array', *Nature Biotechnology,* 35(1), pp. 31-34.

Zhang, J., Houston, D. W., King, M. L., Payne, C., Wylie, C. and Heasman, J. (1998) 'The role of maternal VegT in establishing the primary germ layers in Xenopus embryos', *Cell,* 94(4), pp. 515-524.

Zhang, Q., Sun, K., Chinazzi, M., Pastore y Piontti, A., Dean, N. E., Rojas, D. P., Merler, S., Mistry, D., Poletti, P. and Rossi, L. (2017) 'Spread of Zika virus in the Americas', *Proceedings of the national academy of sciences,* 114(22), pp. E4334-E4343.

Zhang, S. and Bernstein, S. I. (2001) 'Spatially and temporally regulated expression of myosin heavy chain alternative exons during Drosophila embryogenesis', *Mechanisms of development,* 101(1-2), pp. 35-45.

Zhang, Z.-Q. (2013a) 'Animal biodiversity: An update of classification and diversity in 2013. In: Zhang, z.-q.(ed.) animal biodiversity: An outline of higher-level classification and survey of taxonomic richness (addenda 2013)', *Zootaxa,* 3703(1), pp. 5–11-5–11.

Zhang, Z.-Q. (2013b) 'Phylum Arthropoda. In: Zhang, Z.-Q.(Ed.) Animal biodiversity: an outline of higher-level classification and survey of taxonomic richness (Addenda 2013)', *Zootaxa,* 3703(1), pp. 17–26-17–26.

Zhao, C. and Swank, D. M. (2017) 'The Drosophila indirect flight muscle myosin heavy chain isoform is insufficient to transform the jump muscle into a highly stretch-activated muscle type', *American Journal of Physiology-Cell Physiology,* 312(2), pp. C111-C118.

Zheng, X., Zhang, D., Li, Y., Yang, C., Wu, Y., Liang, X., Liang, Y., Pan, X., Hu, L. and Sun, Q. (2019) 'Incompatible and sterile insect techniques combined eliminate mosquitoes', *Nature,* 572(7767), pp. 56-61.

Zhong, J., Peters, A. H., Kafer, K. and Braun, R. E. (2001) 'A highly conserved sequence essential for translational repression of the protamine 1 messenger rna in murine spermatids', *Biology of reproduction,* 64(6), pp. 1784-1789.

Zimowska, G. J., Nirmala, X. and Handler, A. M. (2009) 'The β2-tubulin gene from three tephritid fruit fly species and use of its promoter for sperm marking', *Insect biochemistry and molecular biology,* 39(8), pp. 508-515.

Zug, R. and Hammerstein, P. (2012) 'Still a host of hosts for Wolbachia: analysis of recent data suggests that 40% of terrestrial arthropod species are infected', *PloS one,* 7(6), pp. e38544.