Imperial College London

Determining the landscape of resistance to gene drives in the malaria mosquito

Ioanna Morianou

Submitted on the 22 of December 2022

Department of Life Sciences

Imperial College London

Submitted in part fulfilment of the requirements for the degree of Doctor of Philosophy of Imperial College London and the Diploma of Imperial College London

Statement of Originality

The work presented in this thesis, as well as the interpretation of it is my own. Everything else is appropriately referenced or acknowledged.

Copyright Declaration

The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution-Non Commercial 4.0 International Licence (CC BY-NC).

Under this licence, you may copy and redistribute the material in any medium or format. You may also create and distribute modified versions of the work. This is on the condition that: you credit the author and do not use it, or any derivative works, for a commercial purpose.

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

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

Abstract

Gene drives are engineered selfish genetic elements with the potential to spread throughout entire insect populations for sustainable vector control. Recently, a gene drive was shown to eliminate caged populations of the malaria mosquito by targeting the highly conserved female-specific exon of the *doublesex* gene. This caused females, homozygous for the gene drive, to develop as sterile intersex individuals, leading to the observed population crash. However, target site resistant alleles that block gene drive activity, whilst encoding a functional copy of the target gene, may halt gene drive spread in the wild. These may be naturally occurring or generated by the gene drive itself. This thesis presents a pipeline for the discovery, genetic engineering, and testing of putative drive-resistant variants. First, to investigate the potential for natural resistance, existing population genomics data were interrogated for the presence of natural single nucleotide polymorphisms (SNPs) at the highly conserved gene drive target region. To investigate the potential for drive-induced resistance, a high-throughput assay was designed to generate a high volume of mutations at the gene drive target site and screen them for their ability to restore *dsx* function. These methods yielded three putatively resistant SNPs: one natural polymorphism and two rare Cas9-induced mutations. These were engineered in the mosquito genome for testing, using a novel method termed CRISPR-mediated cassette exchange (CriMCE). It was confirmed that all three polymorphisms are functional and offer full, partial or no resistance to gene drive. Importantly, partial resistance to gene drive is being demonstrated for the first time. To mitigate observed resistance, gene drive systems targeting multiple sites simultaneously were developed. These showed improved drive dynamics and caused rapid elimination of caged mosquito populations within 7-8 generations. The experimental pipeline described here can be applied to pre-empt and mitigate resistance against any gene drive strategy, prior to field testing.

Publications

The main findings of this PhD are currently being prepared for publication, whilst a methods paper that directly resulted from this work has been published in the CRISPR Journal:

- **Morianou, I.**, Crisanti, A., Nolan, T., Hammond A. M. (2022) CRISPR-mediated cassette exchange: a method to isolate precise marker-less edits. *The CRISPR Journal*, 5(6).

Collaborations and side-projects during the course of this PhD also resulted to the below publications:

- Hammond, A. M, Karlsson, X., **Morianou, I.**, Kyrou, K., Beaghton, A., Gribble, M., Kranjc, N., Galizi, R., Burt, A., Crisanti, A. (2021) Regulating the expression of gene drives is key to increasing their invasive potential and the mitigation of resistance. *PLoS Genetics*, 17(1): e1009321.

- Garrood, W. T., Kranjc, N., Petri, K., Kim, D. Y., Guo, J. A., Hammond, A. M., **Morianou, I.**, Pattanayak, V., Joung, J. K., Crisanti, A., Simoni, A. (2021) Analysis of off-target effects in CRISPR-based gene drives in the human malaria mosquito. *Proceedings of the National Academy of Sciences*, 118(22).

- Hammond, A.M., Pollegioni, P., Persampieri, T., North, A., Minuz, R., Trusso, A., Bucci, A., Kyrou, K., **Morianou, I.**, Simoni, A., Nolan, T., Muller, R., Crisanti, A. (2021) Gene-drive suppression of mosquito populations in large cages as a bridge between lab and field. *Nature Communications*, 12(1): 1-9.

- Kientega, M., Kranjc, N., Traore, N., Kabore, H., Soma, D. D., **Morianou, I.**, Namountougou, M., Belem, A. M. G., Diabate, A. (2022) Analysis of the genetic variation of the *fruitless* gene within the *Anopheles gambiae* (*Diptera*: *Culicidae*) complex populations in Africa. *Insects*, 13.

Acknowledgements

I would like to thank my PhD supervisors: **Prof Andrea Crisanti** who hosted me in his lab and supported my scientific and career development; and **Dr Tony Nolan**, the first person who introduced me to gene drive research, back in 2015, and who has mentored me on several occasions during my time at Imperial; his critical review of my work and feedback have been instrumental to my development as a scientist.

I would also like to thank **Dr Andrew Hammond**, for devising the broader topic of my PhD, encouraging me to be the best scientist I could be, believing in me, and helping me push through when I wanted to quit the most.

Next, I would like to thank **Prof Austin Burt** who not only set the foundations for gene drive research with his 2003 paper, but was also always reachable and open to discussions, providing me with scientific guidance and modelling support along the way.

I also thank my progress review panel: **Prof Mark Isalan** and **Prof George Christophides** and my examiners: **Dr Philippos Papathanos** and **Dr Robert Weinzierl** for their useful comments and kind words that encouraged me to continue my career path in academic research, as well as my funders (**MRC DTP**) for supporting me financially.

I would like to thank the whole of **Crisanti lab** who became my second family, and particularly the people who supported my research starting from **Matt Gribble** who maintained all of my strains during the last and most challenging year of my PhD, **Louise Marston** without whom I would have generated zero mosquito strains, as she microinjected mosquito eggs to establish most of them, **Molly McGrath**, who I had the pleasure to supervise during her Master's, and whose work was instrumental in establishing some of my most treasured transgenics, **Dr Nace Kranjc** for putting in the time to teach the whole lab, and me in particular, how to use bioinfromatic tools that were essential to my data analysis, and **Dr Bhavin Khatri** for building all gene drive models presented in this thesis and having the patience to sit through extremely long meetings in order to achieve this.

I also thank **Dr Federica Bernardini** for useful discussions whether scientific or not, **Dr Chryssa Taxiarchi** for always listening and providing me with parental advice, and **Dr Kyros Kyrou** for being an amazing friend and sharing his data with me whenever needed. I would also like to thank my amazing PhD peers: **Anna Strampelli, Ignacio Tolosana, Matteo Vitale, Silvia Grilli, Dario Meacci** and my lovely

deskmate **Nicole Page**, who made work as fun as it could have been, but were also extremely reliable and willing to help at all times. I also thank **Giulia Morselli** for her constant warmth, and **Antonis Kriezis** for keeping me company until the late hours of the night when insectary work was the most intense. I finally thank **Xenia Karlsson**, without whom my start in the lab would have been impossible.

I wish to thank my whole family for supporting me throughout this PhD, for visiting me as often as they could, and always believing in me; as well as my friends without whom I could have never made it to the finishing line. In particular, I thank my sister, **Eleni**, who was the best flatmate, and my best friend, during the toughest year of my PhD, who ensured I was properly nourished at all times, provided me with a warm hug each night, after long hours of straight lab work, and put a smile on my face every day without fail; **George**, who treated me like family, and was there every single day to offer his support and delicious stir fries; my **mom** and **grandma**, for sharing their special recipes with me and keeping me fed from afar; my **dad** for providing me with film and music recommendations to take my mind off work when needed; and my lovely **aunts** and **uncles** and my **grandpa**, who each assisted me with their own piece of wisdom and word of encouragement.

I thank **Christos** for putting up with my PhD stress and showing me that there's more to life than work, and **Nikos** for becoming the most supportive friend and offering his help with everything he could think of. I also thank my two best friends in the whole world, who I know I can always count on no matter what, **Tatiana** and **Dafni**: through thick and thin I always return to our friendship. I also thank **Nefeli** and **Dimitra** for always including me, for listening and for helping me make sense of it all, **Jason A.** for helping me keep it together with his wise advice and encouragement, and **Jason M.** for keeping me company from afar when I felt the most lonely. Last but not least, I thank **John**, who is the best partner I could have asked for, whose friendship helped me navigate through the hardest times in the past year, and whose love and support is evident from everything that he does for me, including travelling for 8 hours straight to visit me every month in rural Italy.

I appreciate each and every one of you, so so much.

Contents

Figures

Tables

Abbreviations

ACT artemisin-based combination therapy

ALE autosome-linked editor

CDS coding sequence

CriMCE CRISPR-mediated cassette exchange

COPAS complex object parametric analyser and sorter

CRISPR clustered regularly interspersed short palindromic repeats

DL docking line

DSB double-stranded break

EJ end-joining

GDS gene drive system

HDR homology-directed repair

HEG homing endonuclease gene

IIT incompatible insect technique

indel insertion/deletion

IRS indoor residual spraying

ITN insecticide-treated net

MDFS male-drive female sterile

MHC major histocompatibility complex

MMEJ microhomology-mediated end-joining

MSCI meiotic sex chromosome inactivation

NGS next-generation sequencing

NHEJ non-homologous end-joining

PAM proto-spacer adjacent motif

PBM post-bloodmeal mortality

RBCs red blood cells

RIDL release of insects carrying dominant lethals

RMCE recombinase-mediated cassette exchange

SGE selfish genetic element

SIT sterile insect technique

SMC seasonal malaria chemoprevenetion

SNP single nucleotide polymorphism

ssODN single-stranded oligo-deoxyribonucleotide

TALEN transcription activator-like effector nuclease

TE transposable element

UTR untranslated region

WGS whole genome sequencing

WHO World Health Organisation

ZFN zinc-finger nuclease

Chapter 1

Introduction

1.1 The Malaria Burden

1.1.1 The disease, its vector, and life cycle

Malaria is the deadliest vector-borne disease globally. It is a disease endemic to 85 countries in the world, mostly part of sub-Saharan Africa and Asian-Pacific regions. It is caused by *Plasmodium* parasites and transmitted through the bites of infected female *Anopheles* mosquitoes. Its symptoms include fever, headaches and chills that usually appear 10-15 days post-infection. Symptoms can be mild and difficult to be recognised as malaria, but left untreated the disease can progress severely and lead to death. At higher risk from dying are infants and children, pregnant women and immunocompromised individuals (WHO 2021*a*).

There are 5 species of Appicomplexan *Plasmodium* parasites that cause human malaria, two of which pose the greatest threat: *P. falciparum*, which is the deadliest of all and most prevalent in sub-Saharan Africa (**Figure 1.1A**), and *P. vivax*, which is the dominant parasite in most other endemic countries (WHO 2021*a*).

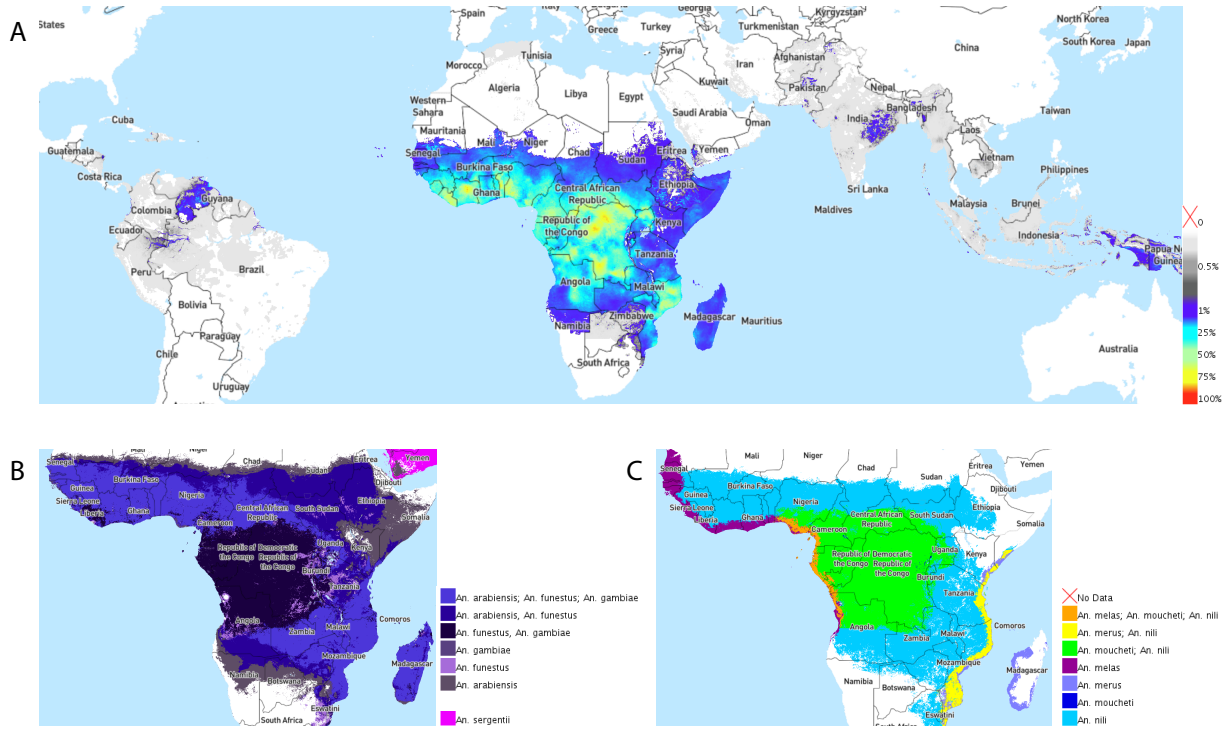


Figure 1.1: The world distribution of malaria and its vectors. (A) The world distribution of the most deadly malaria parasite: *Plasmodium falciparum*. **(B)** The distribution of dominant mosquito vector species of malaria in Africa. **(C)** The distribution of secondary malaria vectors of Malaria in Africa. All maps were generated using the Malaria Atlas Project interactive map tool.

Malaria is transmitted to humans and other mammals by the infective bites of anopheline mosquitoes. Currently, there are 3,591 recognised species of mosquitoes in the world, of which 483 belong to the *Anopheles* genus (*Valid Species — Mosquito Taxonomic Inventory* 2021). Of those, approximately 70 are able

to transmit human malaria and 41 are considered to be vector species or species complexes (Sinka et al. 2012). However, only three species complexes are primary vectors of the disease in Africa, where over 95% of all malaria cases occur. These are *An. gambiae*, *An. arabiensis*, and *An. funestus*.

The malaria parasite has a complex life cycle that requires both a mosquito and human host to be completed (**Figure 1.2**). After a female infected mosquito injects *Plasmodium* sporozoites into the bloodstream of a human host, they migrate to the liver and mature into merozoites (1-2 asymptomatic weeks). Merozoites can then infect red blood cells (RBCs), where the parasite can reproduce asexually (symptomatic stage, high fever). The parasite reproduces asexually multiple times, and its population size can increase from 10 to 100 parasites at the time of infection, to 10^8-10^{13} parasite within a few weeks (Chang et al. 2013).

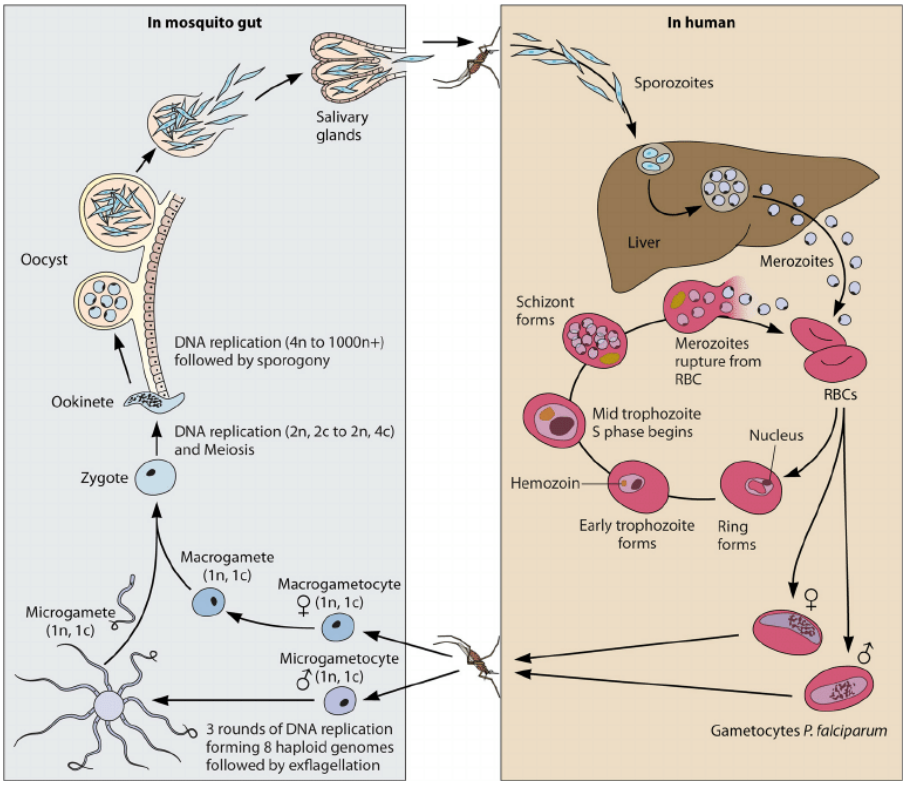


Figure 1.2: The malaria parasite life cycle. Clockwise, from the top: after an infected female mosquito blood-feeds on a human host, it releases sporozoites, through its saliva into the bloodstream, that migrate to the liver and form cysts while replicating asexually into merozoites. The cysts burst releasing merozoites into the bloodstream, that enter RBCs, and go through more asexual cell divisions leading to the release of more merozoites into the bloodstream. Merozoites can also develop into gametocytes in RBCs that are picked up the next time a female mosquito ingests a bloodmeal from an infected host. The parasite then completes its sexual life cycle inside the gut of the mosquito host, leading to the development of ookinetes that form oocysts on the gut lining that burst releasing more ookinetes that migrate into the salivary glands of the mosquito, so they can be transported into the bloodstream of their next host. Figure reproduced from Lee et al. (2014) with permission from the corresponding author.

In RBCs merozoites can also develop into gametocytes required for sexual reproduction of the parasite that will be completed inside the mosquito host. 10-1000 gametocytes can be picked up by a mosquito

host during a bloodmeal, where they fertilise to form zygotes in the mosquito gut that develop into ookinetes. If ookinetes survive the mosquito immune responses they form oocysts in the gut lining, where the parasite reproduces asexually: undergoing sexual recombination and meiosis, resulting in haploid cells that reproduce asexually in oocysts to form sporozoites. Sporozoites migrate to the salivary glands of the female mosquito and can be injected, together with its saliva, into another human host the next time the infected female mosquito takes a bloodmeal. The life cycle takes approximately two weeks to complete inside the human host (Lee et al. 2014), and 10-18 days in the mosquito host (CDC 2020).

The complexity of the parasite's life cycle allows for deployment of several immune evasion strategies by *Plasmodium*, to ensure its survival in the human host, posing a massive challenge to the development of a malaria vaccine (reviewed by Gomes et al. (2016)). *Plasmodium* has developed advanced motility mechanisms (including gliding motility that relies on sophisticated protein-ligand interactions) to transverse across various cell types and reach the liver of their vertebrate host. In the process, it forces immune phagocytic (Kupffer) cells into apoptosis, whilst the liver itself is an immuno-privileged organ protected from strong immune responses. In the liver, the parasites invade hepatocytes by exploiting their cholesterol uptake pathway, and hide within a parasitophorous vacuole that protects them from autophagic degradation. After the liver stage is complete, the merozoites migrate again in search of RBCs, enclosed in host-derived membranes known as merosomes, to further avoid recognition by phagocytes.Through intracellular survival in RBCs the parasites avoid interavtion with immune cells. RBCs in particular, do not express major histocompatibility complex (MHC) class I molecules on their surface, thereby escaping CD8+ T-cell detection. In addition, the parasite modulates infected erythrocytes to bind uninfected ones and form rosette clusters that further shield infected RBCs from immune recognition. Besides host cell remodelling, the parasite makes use of antigenic variation to achieve immune evasion: in particular, it expresses highly polymorphic, variable antigenic proteins on its surface, adapting to the host immune response by using both natural selection and random drift (Chang et al. 2013), to achieve distinct rounds of parasitaemia that are responsible for the cyclic, long-lasting and recurring nature of malaria.

1.1.2 Public health impact

Approximately 241 million people got infected and 627,000 people died from malaria in 2020, worldwide, with 95% of malaria cases and 96% of malaria deaths occurring in Africa (WHO 2021*a*). Nearly half of the world is at risk of being infected with the disease being endemic in 87 countries and territories (CDC 2021).

At higher risk of dying from malaria are individuals who have not yet developed immunity to the disease or are in some way immunocompromised (WHO 2021*a*, CDC 2021). This includes children under

the age of 5, pregnant women, people infected with HIV, and travellers who have not encountered the malaria parasite before.

The public health burden of malaria goes beyond increased mortality. Even if infection does not directly lead to death, chronic sub-clinical re-infections contribute to malnutrition, increase the severity of other infectious diseases, and may cause chronic health conditions, like anaemia and epilepsy (Snow et al. 2003). Cerebral malaria, which affects half a million children per year in Africa, kills 10-40% of patients, whilst 5-20% of survivors experience neurological sequelae, including behavioural disorders and cognitive impairment that debilitates those affected in carrying out executive and intellectual tasks (Sachs & Malaney 2002).

Even though there is a staggering difference in estimates of annual malaria deaths[1], it is important to note that the downward trend in annual deaths that has been observed in the last two decades has stalled or even been reversed (**Figure 1.3**). This is mostly attributed to the rise of insecticide resistance across the African continent (see section 1.2.2); whilst the Covid-19 pandemic is also thought to have hindered access to healthcare services, accurate malaria diagnosis, availability of preventative commodities, and it has reportedly stalled malaria prevention programs (Heuschen et al. 2021, Sherrard-Smith et al. 2020).

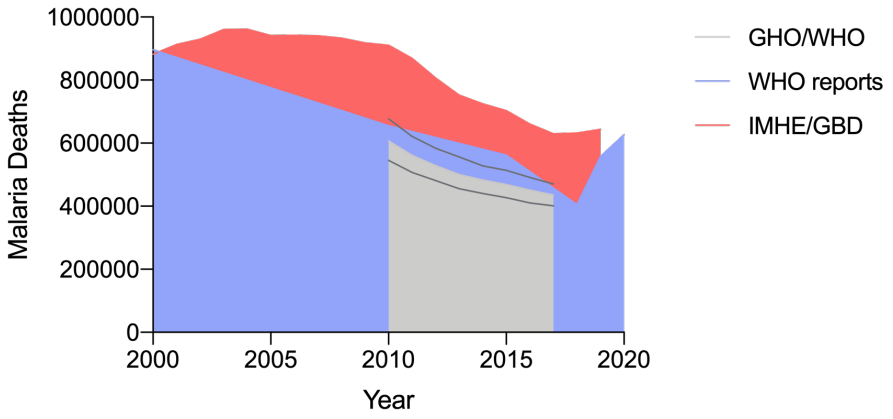


Figure 1.3: Estimated global malaria deaths per year. Data was taken from The Global Health Observatory (2019) of the World Health Organisation (WHO) (GHO/WHO, light grey, maximum and minimum data ranges are also plotted in dark grey), WHO (2021*b*) reports (WHO reports, blue) and the Global Disease Burden (2019) database (IMHE/GBD, pink); and plotted using GraphPad Prism 9.

[1]Estimated numbers of annual malaria deaths differ significantly depending on the organisation, as each organisation has access to different datasets and uses different criteria to record events.

1.1.3 Socioeconomic impact

Malaria disproportionately affects the poorest of all people. For very low income households the yearly total cost for malaria treatment amounted to ∼20% of the household's annual income (with country-specific variations) (Chima et al. 2003). Socioeconomic factors such as education, employment, income and household contribute to increased burden in the poorest countries and regions of sub-Saharan Africa (Degarege et al. 2019). For example, poorly constructed houses enable easy access to malaria vectors, and in densely populated areas the chance of community transmission increases. In addition, low income households cannot afford treatment regimes and control measures (if not provided for free). Even if these are available lack of education might lead to ineffective application of control measures, and decreased awareness with regards to when medical treatment needs to be sought out.

A vicious circle is formed as adults are susceptible to malaria re-infection in regions of high disease prevalence, which leads to temporary inability to work or study (Chima et al. 2003). Moreover, women and adolescents bear the socioeconomic burden of malaria unevenly, as on top of being more prone to severe disease due to pregnancy and to residual health defects such as anaemia, they also bear the responsibility of treating affected family members (Ricci 2012).

Overall, malaria has significantly slowed the economic growth of African countries, costing them an estimated $12 billion, annually (Gallup & Sachs 2001). In countries with highest prevalence (**Figure 1.1A**), malaria accounts for up to 40% of public health expenditures Uguru et al. (2009). It is estimated that if malaria transmission were to be significantly reduced or eliminated between 2013-2035 it would save African countries $208.6 billion (Purdy et al. 2013). This sum could instead be diverted towards other health emergencies, as well as education and business investment leading to unprecedented growth and development of African countries, as was the case when malaria got eradicated from subtropical countries like Greece, Portugal and Spain in the 1950s, leading to accelerated economic growth (Sachs & Malaney 2002).

1.2 Existing Interventions

1.2.1 Drugs, chemoprevention, vaccines

The most effective treatment available to halt severe *falciparum* malaria and prevent death is artemisin-based combination therapy (ACT), whereby a fast-acting artemisin-based compound is administered together with a compound from a different drug class. Despite the combination-based approaches of administering malaria treatments, artemisin-resistant strains of the malaria parasite have been detected in Africa, threatening current efforts to control and eliminate the disease (Nsanzabana 2019, Uwimana et al. 2020, 2021, Balikagala et al. 2021). This necessitates the discovery of novel antimalarials, and the vigilant enforcement of preventative strategies to reduce the number of cases.

Seasonal malaria chemoprevenetion (SMC), whereby a different class of antimalarial drugs are administered together once a month, has been shown to be effective in preventing disease in children, with serious side-effects being rare (Baba et al. 2020, Cairns et al. 2021). However, there has been evidence for the selection of resistance by the parasite (Baba et al. 2020), which can partially compromise the effectiveness of the strategy in areas of high malaria prevalence (Desai et al. 2016).

Vaccines have also shown some promise in preventing malaria disease. In 2021, the WHO approved a malaria vaccine for the first time. However the approved RTS,S vaccine showed only modest protection against severe disease ($\sim$30%), whilst requiring the delivery of multiple vaccine doses (3-4), administered at regimented intervals, which can be challenging when scaling up vaccination efforts (RTS S Clinical Trials Partnership 2012, 2015). According to modeling predictions the RTS,S vaccinations can prevent $\sim$5% of malaria deaths per 30 million vaccine doses administered (Hogan et al. 2020), and perhaps more if vaccinations are administered in conjunction with SMC (Chandramohan et al. 2021), however at the moment this is insufficient to eliminate the disease alone. More effective vaccines are currently under development at clinical (Datoo et al. 2021), or pre-clinical stages (Mwakingwe-Omari et al. 2021).

Although promising methods to prevent malaria exist or are currently in development, the rise of resistance to antimalarials, combined with the relativelty high cost and challenges involved in their delivery, necessitate the sustained use of vector control in order to eliminate the disease (Greenwood 2008).

1.2.2 Vector control

The overall reduction in malaria cases observed in the past two decades (**Figure 1.3**) is mainly attributed to the massive scale-up in the use of insecticides against the mosquito vectors of malaria (Kleinschmidt & Rowland 2021, Hemingway et al. 2016). Despite recent progress in deployment of preventative drugs or vaccines, insecticide-based methods are still the most effective in controlling malaria spread (Bhatt

et al. 2015). These are mainly applied in the form of indoor residual spraying (IRS) and more commonly insecticide-treated nets (ITNs) that are cheaper and easily distributed (Kleinschmidt & Rowland 2021). Combined, in 2015, IRS and ITNs averted approximately 500,000 malaria cases. By comparison, ACT averted slightly over 100,000 cases (Bhatt et al. 2015).

The most common class of insecticides used up until recently were pyrethroids, however due to evolved resistance to insecticides by the mosquito vectors, and its wide spread in Africa (Clarkson et al. 2021, Moyes et al. 2020), insecticide-based methods of control have turned to novel classes of insecticides (Hemingway et al. 2016).

Even with effective applications of existing vector control strategies there is still residual malaria transmission owing to several factors, including the behaviour of certain malaria vectors (Okumu et al. 2013). As such, existing vector control interventions have been deemed insufficient to eliminate malaria (Feachem et al. 2019, Gari & Lindtjørn 2018, Mendis et al. 2009). In addition to the need for novel insecticide classes and formulations and improved deployment of existing methods, there is need for novel vector control interventions (Nolan 2021).

Genetic control of the malaria vector could supplement existing interventions to suppress residual malaria transmission. Genetic control exploits the mating properties of the vector species to introduce genetic factors into the population that reduce its fitness, leading to a decrease in population size, or prevent disease transmission in some way.

1.3 Genetic Technologies for Vector Control

1.3.1 Self-limiting approaches

The sterile insect technique (SIT) was the first genetic control method to be developed for population control of the mosquito in the 1950s (Dyck et al. 2021, Black et al. 2011), and was also applied successfully in the 1970s suppressing the population of malaria vector *Anopheles albimanus* in Central America (Lofgren et al. 1974). SIT relies upon the mass sterilisation and release of males into the natural population.

Sterilisation can be achieved by use of mutagenic chemical agents or irradiation (Black et al. 2011), however it is important that sterilised males retain their ability to seek out and mate with wild females. It is also important that sterilised males are released at large enough quantities, overflooding wild males so that they dominate matings. The release ratio of sterilised-to-wild male required to have a significant population-wide sterility effects largely depends on the target insect, and can vary from 5:1 to 1000:1 (Dyck et al. 2021, Dunn & Follett 2017, Shelly & McInnis 2016). Moreover, to sustain long-term population suppression inundative releases are required.

A common pitfall of SIT is also that sterilisation can impact upon male competitiveness (Alphey & Andreasen 2002). Reduced competitiveness of sterile males has previously led to failure of SIT implementation against *Aedes* and *Culex* mosquito populations, despite adequate flooding ratios (Dame et al. 2009).

An alternative to SIT is release of insects carrying dominant lethals (RIDL), whereby male sterilisation is achieved through genetic engineering (Thomas et al. 2000). Adapting this technology further, female-specific and/or late-acting dominant lethality can be engineered to improve the sterility effect in dense populations, that face high resource competition early in development (so that modified individuals survive in the expense of wild-type early on in development but exhibit sex-specific lethality or sterility later on in development) (Tan et al. 2013, Phuc et al. 2007, Gong et al. 2005, Heinrich & Scott 2000). Nonetheless, even female-specific RIDL approaches have suffered from decreased male competitiveness, being only 5% as competitive as wild-type males, according to field trials performed in Brazil and the Cayman Islands (Harris et al. 2011, Carvalho et al. 2015).

Even if fully competitive mosquitoes, with bi-sex late-acting sterility were to be released (optimal scenario), the implementation strategy would require daily release of 20% of the wild-type female population for a prolonged period of time (Gentile et al. 2015). In extremely large and dense mosquito populations, such as that of *Anopheles gambiae*, which is predicted to comprise of $\sim$10 billion mosquitoes on average, with seasonal variations (Khatri et al. 2019), it may be impossible to apply SIT-like control, as it would be prohibitively expensive, whilst the connectivity of *Anopheles gambiae* populations is likely to lead to fre-

quent re-invasion of emptied niches, and rapid loss of any suppressive effect (Hemming-Schroeder et al. 2020).

An. arabiensis that breeds in temporary rain pools in arid regions and faces a massive population reduction in the dry season might be more amenable to this type of genetic control, if sterile insects are released in low-density populations during the peak of the dry season (Ndo et al. 2018). However, modelling predicts that a weekly release of 165,000 insects would be required to have an effect on *An. arabiensis* populations within 50 hectares of the release site (Kaiser et al. 2021). Note that this is the lowest estimate and it might already exceed current capabilities.

The limitations of SIT and RIDL lie in the fact that they are inherently self-limiting, meaning that their sterility effect disappears from the target population without sustained mass-release of modified insects. To tackle the malaria vector there is need for a cost-effective solution that can act over a meaningful time-frame. Engineered selfish genetic elements called gene drives, could provide a solution to this problem, as they have the potential to spread throughout entire insect populations, within just a few years, and without requiring mass-releases of the same scale as self-limited approaches, because of their self-sustaining nature.

1.3.2 Selfish genetic elements in nature

In nature, there are selfish genetic elements (SGEs) that, even though present at very low frequencies initially, have been found to spread to near-fixation, in a self-sustaining manner, by biasing their inheritance to achieve rates of inheritance greater than 0.5 as predicted by Mendelian genetics (this type of inheritance is referred to as super-Mendelian inheritance (Sinkins & Gould 2006)). Importantly, these SGEs can increase in frequency even if detrimental to the fitness of their host (Hurst & Werren 2001), as long as their inheritance rates outweigh their fitness cost (Sinkins & Gould 2006). It has long been theorised that SGEs can be re-purposed to spread useful traits such as parasite refractoriness in malaria vectors, or traits associated with a fitness cost that will lead to a decrease in vector population size (Sinkins & Gould 2006, Burt 2003, Ribeiro & Kidwell 1994). Engineered SGEs are commonly referred to as gene drive systems (GDSs), and there are several different classes of these that have been proposed or are currently under investigation (Figure 1.4).

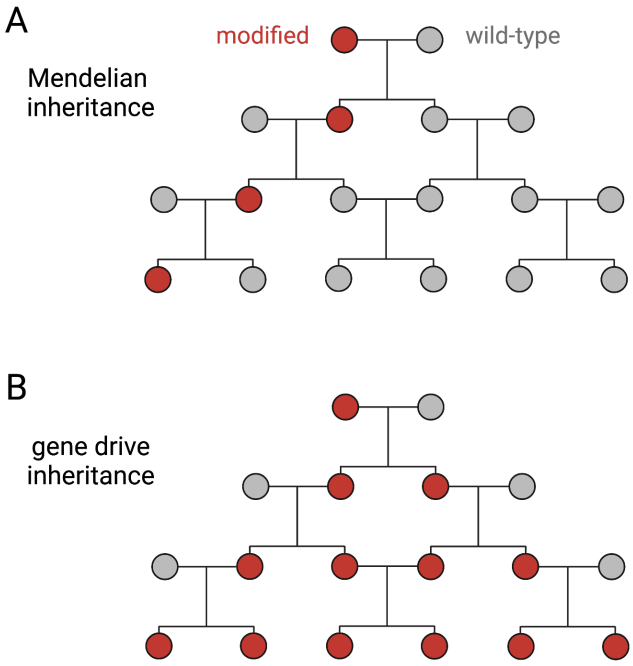


Figure 1.4: Mendelian vs. gene drive inheritance. (A) A modified individual will spread the modification to half of its offspring upon mating to a wild-type individual, under normal Mendelian inheritance. **(B)** A modified individual containing a modification linked to a gene drive, will spread the modification to almost all of its offspring, allowing the gene drive-associated trait to rise in frequency.

Autonomous replicators

Autonomous replicating elements such certain types of transposable elements (TEs) encode proteins that help them increase their copy number in the germline of carrier organisms, leading to increased inheritance rates in the carrier's offspring (Sinkins & Gould 2006). TEs are incredibly common in nature: they comprise approximately 50% of the maize genome, 45% of the human genome, and 15% of the fruitfly and malaria mosquito genome (Saleh et al. 2019, SanMiguel 1996, Biemont & Cizeron 1999, Holt et al. 2002).

A famous example of how a natural TE increased its frequency in a naive population is that of the *P* element in *Drosophila*. It is thought that horizontal gene transfer led to *Drosophila melanogaster* acquiring a single copy of a self-replicating *P* element from *D. willistoni* (Engels 1997). Within a few decades, *P* elements had spread to the entire population of wild *D. melanogaster*, the only populations escaping the *P* element invasion being those contained within science laboratories (Engels 1997).

Efforts have been made to adapt TEs in mosquitoes to drive desirable traits in populations (O'Brochta 2003), based on the *Drosophila P* element paradigm (Carareto et al. 1997). However, although they were useful as genetic engineering tools for the introduction of transgenes into the mosquito genome (Catteruccia et al. 2000, Nolan et al. 2002), their unpredictability, low rates of re-mobilisation in certain mosquito species, copy number variation, make them unsuitable and unreliable as GDSs (O'Brochta 2003, Sinkins & Gould 2006).

Gene converters

Gene conversion is not as common as retrotransposition in nature (Hurst & Werren 2001), however there is a class of selfish genes, homing endonuclease genes (HEGs), that can force unidirectional gene conversion pre-meiotically to favour their own inheritance. Burt (2003) was the first to note the potential of HEGs, and/or other driving endonucleases, in genetically modifying entire populations for vector control.

HEGs express an endonuclease that recognises and cleaves a site on the homologous chromosome (lacking the HEG) equivalent to the one within which the HEG is integrated, using a 14-40 bp site recognition sequence (Stoddard 2005). By exploiting the host's endogenous homology-directed repair (HDR) machinery, the HEG gets copied over into the site of the cut, increasing its frequency as it gets converted from heterozygosis to homozygosis (**Figure 1.5A**). This process is commonly referred to as 'homing' (Stoddard 2005).

Natural HEGs have been inserted into the mosquito genome and shown to be active (Windbichler et al. 2007), as well as able to function as GDSs, when provided with an artificial target site (Windbichler et al. 2011). However, to be useful as genetic control tools they would need to be engineered to recognise specific target sequences in the mosquito genome (Deredec et al. 2011). Though purposefully engineering homing endonucleases to shift their recognition sequence is possible (Rosen et al. 2006), it proved difficult and extremely time-consuming, with restrictions in the diversity of sequences that can be targeted depending on the nuclease (Chan et al. 2013, Werther et al. 2017).

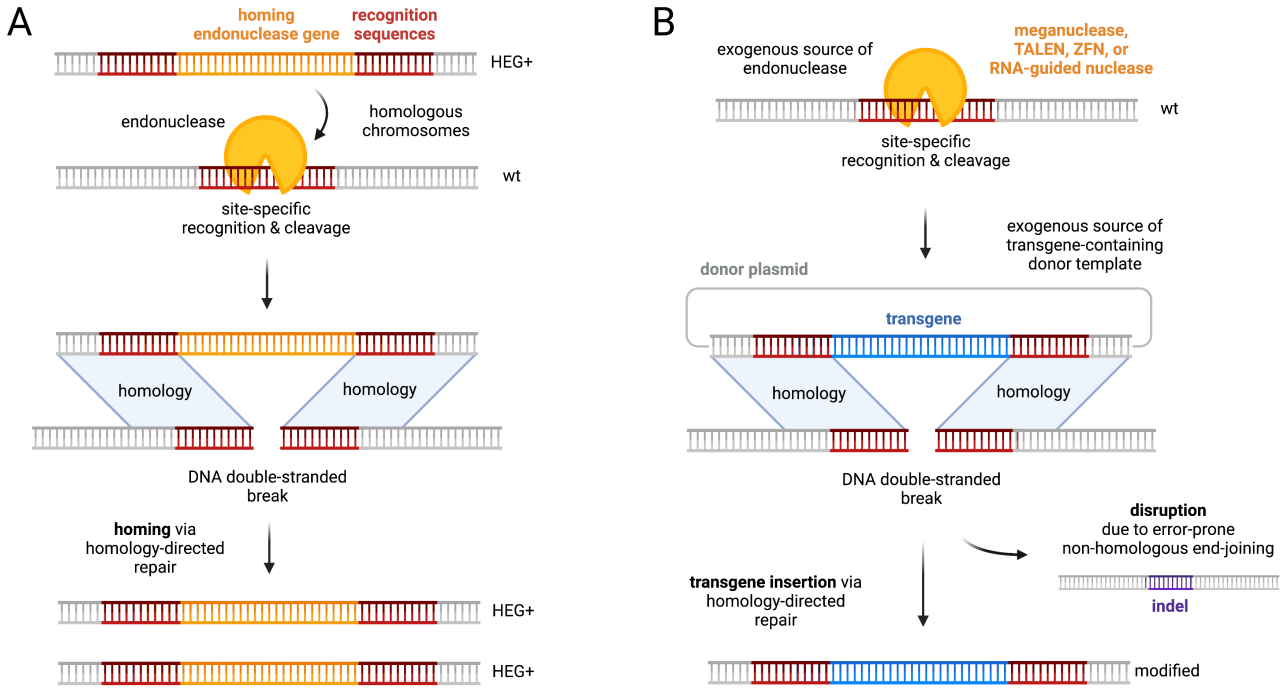


Figure 1.5: The homing reaction and how it can be exploited for transgene insertion. (A) A HEG (yellow) is embedded within the recognition sequences (red) of the endonuclease it encodes. When a HEG+ paired with a wild-type (wt) chromosome (heterozygosis) containing the HEG nuclease recognition sequence, the nuclease can cleave it producing a DSB. Upon repair of the DSB through HDR, based on homologous regions between the two chromosomes, the HEG+ chromosome is used as a template for HDR leading to incorporation of the HEG sequence at the site of the cut, in a process called homing. Homing disrupts the wt recognition sequence and protects the HEG from self-cleavage. After homing the HEG is found in homozygosis. **(B)** Endonucleases (meganucleases/TALENs/ZFNs/RNA-programmable nucleases) can be used for gene editing. An exogenous source of nuclease is provided either in plasmid form to be expressed endogenously or as a purified protein, to cleave its target. A donor plasmid containing the transgene of interest flanked by regions of homology to the upstream and downstream regions of the DSB can be used as a template for repair to incorporate the transgene at the DSB site through HDR. If instead of HDR, the non-homologous end-joining (NHEJ) pathway is used for repair of the DSB, insertion/deletions (indels) might get incorporated into the site of the cut, disrupting the target sequence.

Sex ratio distorters

To increase the inheritance rates of cytoplasmic genes, exclusively transmitted by female gametes, a range of endosymbiotic bacteria, Microsporidia and even organelles have been found to cause sex ratio distortion by inducing male sterility, parthenogenesis, feminisation or male-killing (Hurst & Werren 2001).

Sex ratio distortion can also be achieved by segregation distorters known as meiotic drive elements. These bias their inheritance by leading to preferential maturation of gametes carrying these elements, and if located on sex chromosomes they can lead to preferential inheritance of either the X or Y chromosome and subsequently bias the sex ratio of the progeny (Sinkins & Gould 2006, Hurst & Werren 2001). Gradually, the population is dominated by one sex that is unable to seek mates, and in this way the population size can be drastically reduced or even eliminated. Such Y-driving elements have been described in the *Aedes aegypti* and *Culex pipiens* mosquitoes, where they cause preferential inheritance of Y-bearing gametes by inducing pre-meiotic shredding of the X chromosome, leading to degeneration of most X-bearing gametes (Wood & Newton 1991). Male sex ratio distortion of up to 90% can be achieved in this way (Wood & Newton 1991).

Similarly, the *Physarum polycephalum*-derived I-*Ppo*I HEG was found to be able to recognise and cleave ribosomal DNA (rDNA) repeated motives conserved in eukaryotes (Lin & Vogt 1998), including rDNA repeats located on the X chromosome of the malaria mosquito (Windbichler et al. 2007). When meiotically expressed from an autosome, it was able to induce ∼90% male sex bias in *Anopheles gambiae* due to X-shredding through cleavage of the X-linked rDNA repeats (Windbichler et al. 2008). If I-*Ppo*I were instead expressed from the Y, it would function as a gene drive (Y-drive) and be an ideal candidate for self-sustaining vector control of the malaria mosquito. Due to X-shredding, most offspring carrying the Y-linked I-*Ppo*I would be male (XY), and by definition they would also carry the I-*Ppo*I, causing it to be inherited at super-Mendelian rates, whilst reducing the population of females.

However, the Y chromosome of the mosquito is largely uncharacterised and although site-specific engineering and expression from the Y chromosome was possible (Bernardini et al. 2014), it has not yet been possible to overcome the challenges posed by meiotic sex chromosome inactivation (MSCI) that inhibits any expression from the Y during meiosis (Taxiarchi et al. 2019), when the X-shredder should be active.

Post-segregation distorters

Some SGEs increase their inheritance rates by inhibiting survival of non-carrier individuals after fertilisation and zygote formation (Hurst & Werren 2001). An example of this is the *Medea* (maternal effect dominant embryonic arrest) element in the *Tribolium castaneum* beetle, which kills the offspring of females that do not inherit it through expression of a maternal toxin and a zygotic antidote (Beeman & Friesen 1999). Finding genes and regulatory elements that allow these properties to create an artificial toxin-antidote GDS is challenging (Hay et al. 2010), but using maternally deposited microRNAs (maternal toxin) that silence embryonically-required genes, then re-expressing a codon-scrambled version of the same genes that can no longer be targeted by the microRNAs (zygotic antidote) can provide a solution, as demonstrated in *Drosophila* (Chen et al. 2007, Akbari et al. 2013, Buchman et al. 2018), provided the gene and regulatory elements are sufficiently small in size.

Endosymbiotic bacteria can also increase their transmission through cytoplasmic incompatibility, which involves a SGE rescue mechanism, similar to toxin-antidote systems (Hurst & Werren 2001). A very well-documented example of this is the pattern of inheritance of the maternally-transmitted *Wolbachia* endosymbionts in arthropods (Stouthamer et al. 1999). Namely, modified sperm of *Wolbachia*-infected males cannot complete fertilisation with uninfected eggs, however a rescue function in infected eggs allows fertilisation and zygote formation (Sinkins & Gould 2006). Interestingly, *Wolbachia* strains in *Aedes* mosquitoes have also been found to block transmission of several viruses (Moreira et al. 2009), and decrease mosquito lifespan (McMeniman et al. 2009), which makes them particularly attractive tools for sustainable vector control. Overflooding a wild population with *Wolbachia*-infected males can also act as an SIT-like approach[2] to reduce the population of *Aedes* mosquitoes and reduce dengue transmission as demonstrated in field and semi-field trials (Hoffmann et al. 2011, Nguyen et al. 2015, Liu et al. 2022, The Project Wolbachia – Singapore Consortium & Ching 2021, Crawford et al. 2020).

Although *Wolbachia* can infect wild-type strains of various malaria vectors (Baldini et al. 2014, Jeffries et al. 2018), it has not yet been adapted for genetic control of the malaria mosquito.

Finally, it is worth pointing out that post-segregation distorters have much slower invasion dynamics compared to other types of SGEs, as they are usually removing wild-type genotypes from the population rather than doubling their own inheritance.

[2]This self-limiting method of vector population control using *Wolbachia* and other endosymbionts is often referred to as the incompatible insect technique (IIT).

1.4 Advances in Genome Engineering

1.4.1 A brief history of gene editing

The ability to introduce precise gene edits in the genome of virtually any organism, has revolutionised the study of biological processes (Kim 2016). Traditionally, geneticists relied on studying mutations that occurred spontaneously or introduced by chemical mutagens or radiation treatments (Randhawa & Sengar 2021). Later technologies utilised TEs that allowed the introduction of transgenes in semi-random genomic locations (see 1.3.2). However, the above technologies lacked desired specificity. In the 1980-1990s, endonucleases with large recognition sequences, like the ones encoded by HEGs (meganucleases), were first used to introduce gene edits in precise genomic locations (Rudin & Haber 1988, Rouet et al. 1994).

Meganucleases can catalyse a DNA DSB on a specific genomic location, and by exploiting the endogenous repair machinery, scientist can either choose to knock-out a gene, exploiting error-prone NHEJ, or incorporate a transgene at the site of the cut, through HDR (**Figure 1.5B**). Nonetheless, these types of endonucleases were constrained in recognising very specific DNA sequences each, and were not easily reprogrammable to recognise any other sequence of choice (see 1.3.2) (Randhawa & Sengar 2021, Kim 2016).

The discovery of programmable endonucleases, such as zinc-finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs) provided a marked improvement in modifying specific DNA regions, but they were still reliant upon protein domain shuffling to define their target region that posed certain constraints on the combinations of nucleotides that could be targeted (Urnov et al. 2010, Joung & Sander 2012).

It was not until the development of programmable RNA-guided nucleases (Jinek et al. 2012, Gasiunas et al. 2012), that precise genome editing became easy, cheap, accessible, and adaptable to almost any sequence of interest (Mali, Esvelt & Church 2013, LaManna & Barrangou 2018, Walton et al. 2020).

1.4.2 CRISPR

From bacterial adaptive immunity to gene editing

In 2005, Mojica et al. reported that bacterial and archaean repeated sequences were interspersed by foreign genetic material derived from transmissible genetic elements, such as bacteriophages and conjugative plasmids. These were termed clustered regularly interspersed short palindromic repeats (CRISPR) and it was speculated that they form part of a bacterial adaptive immunity system (Mojica et al. 2005), which was later demonstrated experimentally, with the added knowledge that a protein known as CRISPR-associated protein 9 (Cas9) is also required to bring about CRISPR immunity (Barrangou et al. 2007).

The next key discoveries were that: (a) phage-derived spacer sequences from CRISPR arrays transcribe into small RNAs (CRISPR RNAs – crRNAs) to guide antiviral defense (Brouns et al. 2008); (b) CRISPR immunity destroys DNA complementary to the crRNA spacers (Marraffini & Sontheimer 2008) and responsible for this function is the Cas9 protein which cleaves target DNA at a precise position, three nucleotides upstream of a proto-spacer adjacent motif (PAM) (Garneau et al. 2010), typically consisting of 3-4 nucleotides (NGG in the case of Cas9); and (c) another small RNA with complementarity to crRNA, the tracrRNA (trans-activating CRISPR RNA) is required for Cas-mediated cleavage (Deltcheva et al. 2011).

Finally, it was shown that CRISPR systems are self-contained and transferable to non-native bacteria (Sapranauskas et al. 2011), and that Cas9 can be reprogrammed to target almost any target site of choice by changing the sequence of the crRNA (Gasiunas et al. 2012, Jinek et al. 2012), whilst the cr- and tracrRNAs can be fused together to a single guide RNA (gRNA) for ease (Jinek et al. 2012), making CRISPR a great tool for genome engineering (Figure 1.6). CRISPR, as a gene editing tool, was also adapted for use in eukaryotic cells, while it was demonstrated that it can be programmed to cleave multiple target sites at the same time, and also bring about HDR (Cong et al. 2013, Mali et al. 2013).

Genome engineering using CRISPR

In its most common form as a genome editing tool, CRISPR is comprised by a Cas9 endonuclease and a guide RNA (gRNA) with an invariable motif that binds the Cas9 and a variable ∼20 bp sequence that directs it to a specific genomic location, which it can cleave. By changing these ∼20 bp one can direct the CRISPR nuclease to any desirable sequence to catalyse a DSB, provided it is followed by a ∼3 bp PAM in the target genome. CRISPR can be used to generate end-joining mutations for gene knock-out through NHEJ repair of a cleaved target site, or for site-specific integration of a transgene provided by donor template through HDR (**Figure 1.5B**).

Since then, the CRISPR toolbox has been expanded with newly discovered Cas9 orthologs, and other

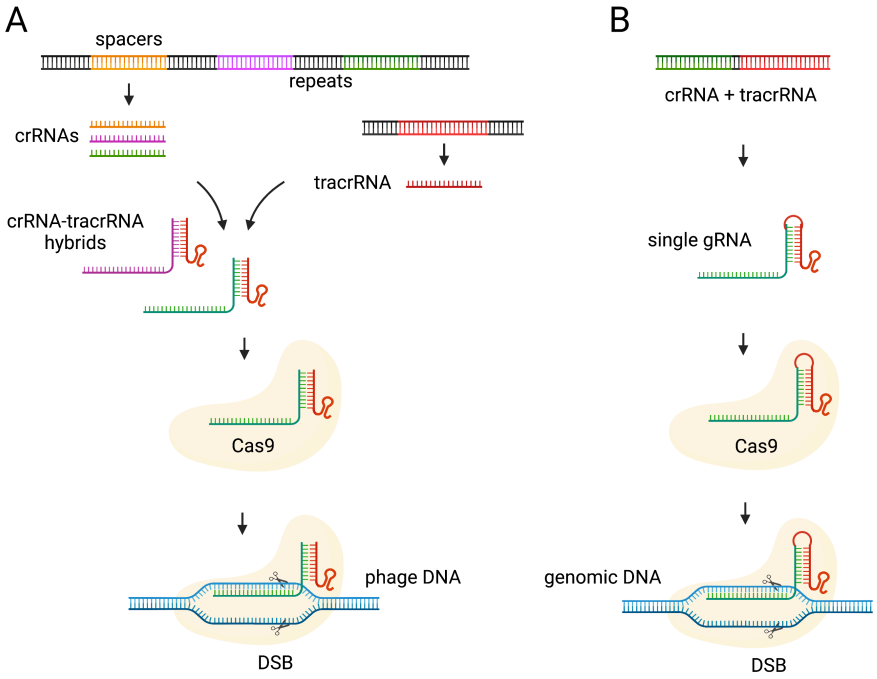


Figure 1.6: Adapting the naturally-occuring CRISPR system for gene editing. (A) CRISPR systems in bacteria act as adaptive immunity machines. DNA fragments of foreign invaders (spacers) are acquired and incorporated in between repeated regions, to form a CRISPR array. A whole CRISPR array transcript is cleaved by Cas proteins that recognise the repeat regions (not shown) to form mature CRISPR-RNAs (crRNAs). These complex with trans-activating crRNA (tracrRNA) to form hybrids able to bind Cas9 and direct it to catalyse a DSB on crRNA-complementary DNA (e.g. phage DNA) when the foreign invader is re-encountered. **(B)** In engineered CRISPR systems crRNA and tracrRNA are fused into a single 20 nt long guide RNA (gRNA) that can form a complex with Cas9 to induce site-specific DSB onto a genomic DNA (gDNA) region of interest. Inspired by Sander & Joung (2014).

nucleases with different properties to suit a variety of applications, with ongoing efforts to keep adding to the toolbox by mining the genomes of bacterial and arcaheal CRISPR systems Barrangou & Doudna (2016). CRISPR has further been adapted for transcriptional regulation and epigenome editing. By exploiting a dead Cas9 (dCas9) which binds indefinitely to its target, but lacks the ability to catalyse cleavage, DNA accessibility to RNA polymerases can be blocked, leading to transcriptional silencing. This process is known as CRISPR interference (CRISPR interference) (Gilbert et al. 2014). By contrast, using a dCas9 to recruit transcriptional activators can enhance expression of nearby genes in a process termed as CRISPR activation (CRISPRa) (Gilbert et al. 2014). Moreover, a dCas9 can be fused to domains that directly write, read or erase chromatin marks to alter epigenetic signatures, such as histone acetylation and methylation (Nakamura et al. 2021).

Moreover, Cas9 variants have been developed as single or dual Cas9 nickases with several applications (Barrangou & Doudna 2016). Importantly, Cas nickases (nCas) have allowed important functions of CRISPR such as base and prime editing, which allow the incorporation of base pair changes or other small indels, withour reliance on HDR (Anzalone et al. 2020). Base editing can induce transition point

mutations through an nCas-deaminase fusion, whilst prime editing can introduce any point mutation or small indel by employing an nCas-reverse transcriptase fusion and a prime editing gRNA (pegRNA) that functions as a template for repair (Anzalone et al. 2020). Additionally, fusion of nCas to a transposase or recombinase can facilitate targeted transgene insertion, circumvening the HDR pathway (Anzalone et al. 2020).

Applications

Overall, the CRISPR gene editing technology and its variations have opened up new horizons for genetics research. CRISPR provides a precise and easy-to-use molecular mechanism for editing cells, tissues and whole organisms, with a multitude of applications in experimental and applied systems (Barrangou & Doudna 2016), such as:

- Genome-wide screens to identify the function of essential genes and causative agents of disease; but also to investigate the function and importance of non-coding DNA regions.
- Basic biological research in small and large model organisms.
- Cell and gene therapies, aiming to correct disease-causing mutations.
- Anti-microbial and anti-viral applications, based on mimicking the native function of CRISPR.
- Improving the nutritional content and agronomic traits of crops.
- The development of improved diagnostic platforms.
- Biological control of insect vectors of disease and pest species (see 1.5.1).

1.5 CRISPR-based Gene Drives

1.5.1 Surge of new genetic technologies for vector control

Though HEGs are very efficient at biasing their inheritance, they are difficult to reprogram to recognise any genomic site of interest (see section 1.3.2). In principle, any programmable nuclease, such as ZFNs and TALENs, could be designed to function as a HEG. However, attempts to build GDSs with ZFNs and TALENs suffered from instability due to internal recombination of the repetitive elements present in the genes encoding them (Simoni et al. 2014).

Following the discovery of CRISPR, there was a surge in the development of novel genetic technologies for vector control. Due to the versatility of CRISPR, Cas9 was adapted as a self-sustaining driving endonuclease, akin to a HEG nuclease, by inserting it together with a gRNA within their target locus (which can be simply altered by using a different gRNA and integration site) (Gantz & Bier 2015, Gantz et al. 2015, Hammond et al. 2016, Kyrou et al. 2018), to ensure its autonomous spread. Allelic drives also rely on autonomous CRISPR-induced homing for spread (Guichard et al. 2019, Kaduskar et al. 2022).

Based on the same principles, self-limiting CRISPR-based gene drive systems were developed, or are currently under development, to achieve localised spread. These either spread non-autonomously by splitting the Cas9 and gRNA into different genomic locations (e.g. split drives (Li et al. 2020, Kandul et al. 2020, López Del Amo et al. 2020), integral drives (Hoermann et al. 2021), and tethered homing drives (Metzloff et al. 2021)); or contain added features that impede their spread to achieve localised population control (e.g. male-drive female sterile (MDFS) systems (Hammond et al. 2016) and Y-linked or male-linked editors (Burt & Deredec 2018)).

Systems for sex ratio distortion relying on CRISPR targeting rDNA repeats rather than I-*Ppo*I have also been developed (Fasulo et al. 2020, Meccariello et al. 2021, Galizi et al. 2016). Moreover, CRISPR has facilitated the engineering of self-sustaining toxin-antidote and cleave-and-rescue gene drive systems (Champer, Lee, Yang, Liu, Clark & Messer 2020, Oberhofer et al. 2019), as well as SIT approaches by being used to engineer simultaneous female-specific lethality and male-specific sterility, in a strategy known as precision-guided SIT (Kandul et al. 2019).

The present thesis will primarily focus on self-sustaining CRISPR-based homing drives that spread autonomously by emulating HEGs (Figure 1.5), however lessons learned will be applicable to most of the aforementioned technologies.

1.5.2 Population replacement vs. population suppression

Most CRISPR-based GDSs described in section 1.5.1, can be designed for both population replacement (Gantz et al. 2015, Oberhofer et al. 2019, Champer, Lee, Yang, Liu, Clark & Messer 2020, Hoermann et al. 2021, Adolfi et al. 2020, Pham et al. 2019), and population suppression approaches to vector control of the malaria mosquito (Hammond et al. 2016, Kyrou et al. 2018, Simoni et al. 2020, Fuchs et al. 2021) (Figure 1.7).

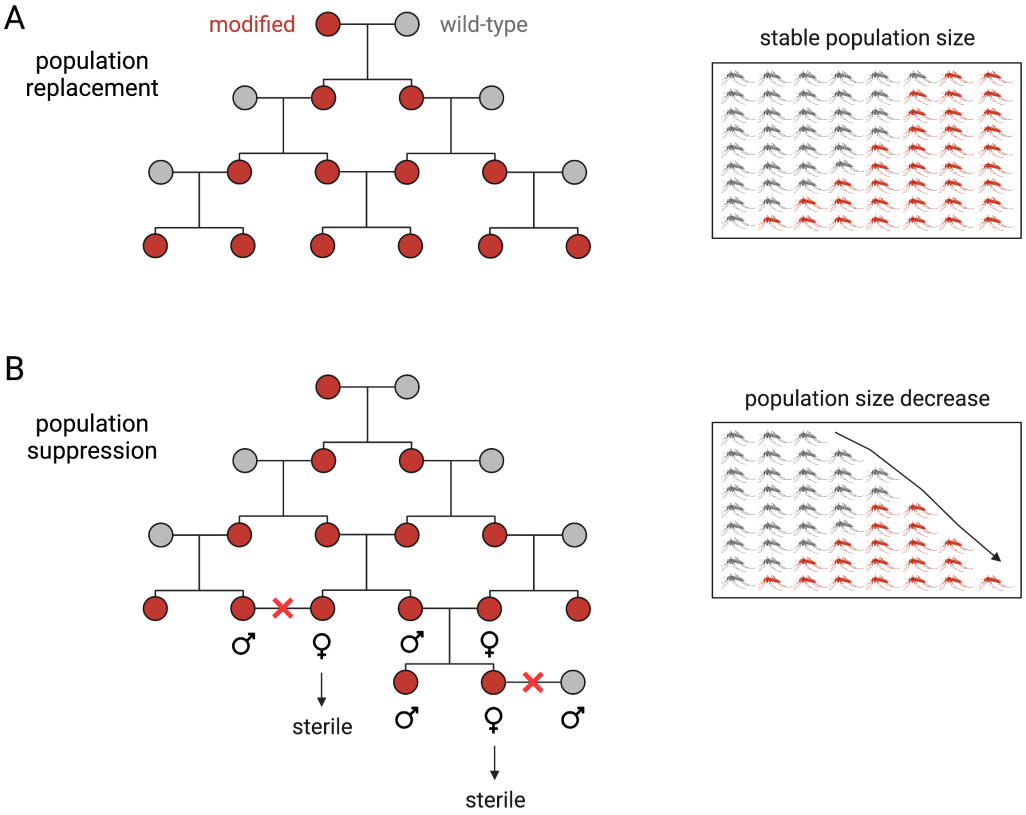


Figure 1.7: Population replacement vs. population suppression as strategies for vector control. (A) Population replacement strategies aim at genetically modifying a native population to carry anti-pathogen properties, or other desirable traits, using gene drive to spread the genetic modifications at super-Mendelian ratios. **(B)** Population suppression gene drive strategies aim at genetically modifying native populations to carry a fitness cost, eventually aimed at population collapse, to reduce disease transmission. Spreading a trait that carries a fitness cost is possible by targeting a haplo-sufficient gene that it only detrimental to only one of the two sexes, for example in females. This means that every individual will be able to spread the gene drive besides females that inherit the gene drive from both parents.

Population replacement

Population replacement is aimed at limiting the vectorial capacity of a target species by spreading genes in the population to that effect (effector genes). Most effectors considered in *An. gambiae* consist of exogenous anti-pathogen genes that can inhibit parasite growth at various stages of development inside the mosquito host (Figure 1.2, left) (Marshall et al. 2019, Dong et al. 2022). Alternatively, the knock-out or knock-down of host genes essential to parasite development is also being explored (Dong et al. 2022). Combinatorial approaches recruiting multiple anti-pathogen effectors simultaneously are also under development to delay the onset of parasite resistance (Marshall et al. 2019, Dong et al. 2020).

In its most simple configuration, one or several anti-pathogen genes will be carried by a homing GDS as in Gantz et al. (2015) (Figure 1.8A). However, this strategy is susceptible to loss of the effector gene over time, which can impede successful implementation of the strategy in the field (Beaghton et al. 2017, Burt 2003). Effector-less GDSs will most likely have a selective advantage over the original designs, due to higher geneticx stability compared to the original constructs, or if the effector genes are associated with a fitness cost in the mosquito (James 2005, Burt 2003). To ensure efficient homing, small effector genes are generally preferred (Gantz et al. 2015), expressed in a spatiotemporally restricted manner under gut-specific or blood meal-inducable promoters, rather than ubiquitously, to limit potential fitness costs (Hoermann et al. 2021, Gantz et al. 2015).

Split integral drives have been proposed to overcome the problem of putative loss of the effector gene (Burt 2003, Nash et al. 2019, Hoermann et al. 2021). Under this configuration the Cas9 (with a gRNA responsible for its homing) and the effector (with a gRNA responsible for its homing) are integrated separately. Under this configuration Cas9 drives autonomously, to enable the entire population to express Cas9, whilst the effector construct drives non-autonomously, depending on Cas9 presence. Additionally, a native promoter is used to drive tissue-specific expression of both constructs (germline-specific and gut-specific, respectively) being integrated in native or artificial intronic regions of a native gene, to further reduce construct loss (Hoermann et al. 2021, Nash et al. 2019). Note that, so far, only an integral effector has been experimentally demonstrated (Hoermann et al. 2021). An integral Cas9 might constitute too large of a construct and cause intron disruption, making this configuration non-viable.

To ensure successful population invasion, it is also important to limit the creation and selection of target-site resistance, that could reverse Cas9 spread (Beaghton et al. 2017) (see chapter 1.6). To achieve this, population replacement drives should be designed to confer the minimum possible fitness cost, e.g. by being integrated in a neutral locus showing minimal functional constraint (Burt 2003) (Figure 1.8A). However, the latter poses challenges since neutral loci contain a high amount of sequence diversity. Divergent sequences would fail to be recognised by the gRNA and escape cleavage and homing, but if the

locus is indeed neutral they should not fall under positive selection. Nonetheless, sequence variation could significantly delay the spread of the drive, especially considering how variable non-functionally constrained loci of the mosquito genome are. In fact, *Anopheles gambiae* is considered to be one of the most genetically diverse eukaryotic species, with one variant allele, every ∼2 bases of the portion of the genome that was accessible to SNP-calling (61%) (Miles et al. 2017, Clarkson et al. 2020). Unfortunately, well-conserved sequences arise due to functional constraint, meaning that their disruption by a gene drive would likely incur a fitness cost. In this case, alleles that block homing would come under positive selection, replacing the gene drive. Several solutions have been proposed to mitigate against this and will be explored in more detail in subsection 1.6.2 of section 1.6.

The biggest hurdle that population replacement strategies will need to overcome is parasite resistance to the anti-parasitic effector(s) (Burt 2003). Nonetheless, population replacement strategies aiming to increase the susceptibility of the mosquito population to insecticides, rather than reduce their vectorial capacity have also been proposed (Guichard et al. 2019, Kaduskar et al. 2022). These rely on an autonomously homing CRISPR-based drive (like in Figure 1.8) that is found in synteny with an allele causing insecticide susceptibility. At the same time they express an additional gRNA that targets the allele on the same gene causing insecticide resistance. Thus, whenever the resistant allele is encountered by the gene drive, it gets cleaved by CRISPR and reverted back to the susceptible allele through HDR. It was recently demonstrated in the fruitfly that, in this way, a population resistant to insecticides can be efficiently replaced by a susceptible one (Kaduskar et al. 2022).

Population suppression

Population suppression approaches aim to modify the target population to reduce its reproductive capacity and by consequence reduce the number of disease-transmitting mosquitoes (Burt 2003). (Figure 1.7B). Choosing a target gene is crucial for ensuring gene drive spread despite the fitness cost incurred by its disruption.

Ideally the target gene should be:

1. Essential for the survival or reproduction of the population.

2. Haplosufficient, meaning that one functional copy of the gene is sufficient for function - this allows gene drive carriers with with only a single copy of the drive to be fit and able to propagate it in the population.

3. Sex-specific, affecting only one of the two sexes so that the modification can spread at super-Mendelian rates (Figure 1.7B).

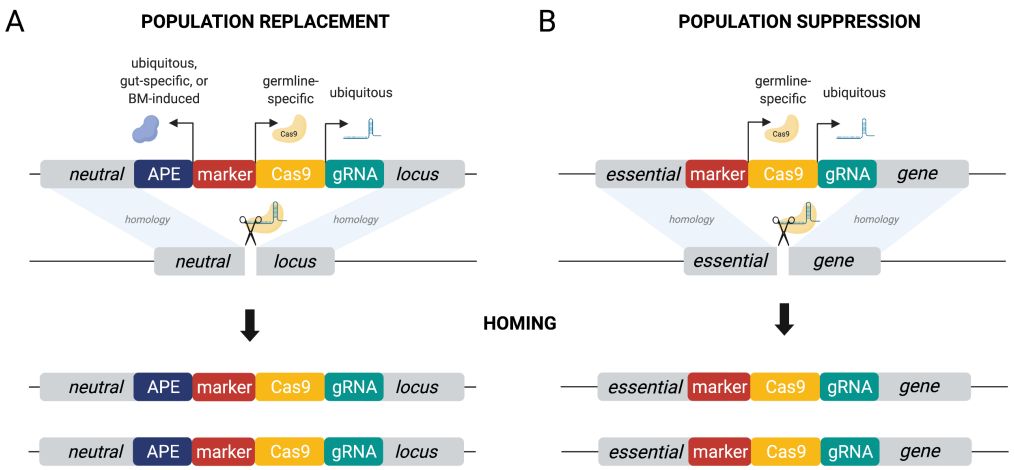


Figure 1.8: Homing gene drive designs for vector population replacement vs. population suppression. **(A)** The population replacement gene drive is integrated within a neutral locus and comprises a Cas9 constitutively expressed in the germline, a ubiquitously-expressed gRNA complementary to the region within which the gene drive is integrated, a selectable marker, and an anti-parasitic effector (APE) that can be ubiquitously or constitutively expressed, for example in the gut, post-bloodmeal, in the salivary glands, etc. **(B)** The population suppression gene drive is integrated in such way as to interrupt an essential gene and comprises a Cas9 constitutively expressed in the germline, a ubiquitously-expressed gRNA complementary to the region within which the gene drive is integrated and a selectable marker.

4. Required for function only in the soma, whilst the gene drive catalyses homing pre-meiotically in the germline (Deredec et al. 2011) (Figure 1.9).[3]

Genes that control aspects of mosquito demography can be targeted, including genes that influence fecundity and sex-specific embryo, larval or adult survival (Deredec et al. 2011). However, targeting late-acting genes required for adult survival is advantageous in two ways:

(A) It would not influence density-dependent factors that impact upon mosquito survival - such as resource competition, in contrast to early-acting genes (including those controlling female fertility and early survival) that would relax density-dependent factors, which are at play earlier during the mosquito life cycle (Deredec et al. 2011). This would lead to greater spread dynamics and an overall reduced number of female mosquitoes surviving long enough as adults to be able to transmit malaria.

(B) If the targeted gene allowed the mosquitoes to survive long enough to produce offspring (>5 days), but not long enough to transmit malaria (>10 days), then this strategy would impose a lower fitness cost compared to targeting early acting genes, and improve the dynamics of spread. A lower fitness cost would also reduce the selection pressure for resistance (see section 1.6) (Deredec et al. 2011).

There are attractive gene candidates in *Aedes aegypti* (with *Anopheles* orthologs) that cause a flightless phenotype in females, which should reduce adult survival in the wild, and could fulfill some if not all of

[3]Note that in theory other combinations of these requirements could also allow the gene drive to spread in super-Mendelian rates. For example, a gene required pre-meiotically for fertility could be targeted (early germline), provided that homing occurs during meiosis (late germline).

the above criteria (O'leary & Adelman 2020, Fu et al. 2010).

Even better, one could target genes that bring about a sterility effect not immediately but in their offspring (i.e. by causing maternal effect bisexual sterility - 'grandchildless' phenotype) (Burt 2003). This should reduce the fitness cost associated with the drive further and by consequence reduce the evolutionary pressure for selection of resistance. Though several genes whose knockout brings about a grandchildless effect in fruitflies exist (Boswell & Mahowald 1985, Niki 1984), orthologs with the same function have not been identified in mosquitoes.

Early-acting genes, whose knockout bring about female sterility, have been more straightforward to identify in mosquitoes. Hammond et al. (2016) identified three *An. gambiae* genes essential for female fertility, that could be good gene targets for a population suppression strategy, as they were found to be haplosufficient, and required for function in somatic tissues. These were: AGAP005958 (ortholog of *Drosophila yellow-g*), AGAP007280 (ortholog of *Drosophila nudel*), and AGAP011377 (Hammond et al. 2016). CRISPR-based gene drives were built to disrupt each of those genes as in Figure 1.8B. To contain gene drive activity in the germline, Cas9 was expressed under the control of the *vasa2* promoter and 3'untranslated region (UTR) (Papathanos et al. 2009). However, Cas9 leakiness into off-target somatic tissues (non-germline) introduced NHEJ mutations at the target site that caused partial or full sterility in females heterozygous for each GDS (Hammond et al. 2016, Hammond et al. 2017, Hammond et al. 2021a). Specifically, females heterozygous for the AGAP005958 GDS (vasa-5958) were completely sterile, whilst females heterozygous for the AGAP011377 (vasa-11377) and AGAP007280 GDSs (vasa-7280) maintained partial, albeit very low, fertility. Overall, only the vasa-7280 GDS met the minimum homing rate requirements for efficient spread, against the fitness cost of drive heterozygote females (Deredec et al. 2011) (Figure 1.9). Indeed, the vasa-7280 drive was able to spread for four consecutive generations in population invasion experiments, demonstrating for the first time that synthetic GDSs can invade and suppress populations (Hammond et al. 2016). However, spread of this drive was thwarted by the selection of target site resistance (Hammond et al. 2017) (see section 1.6 and subsection 1.6.1.4).

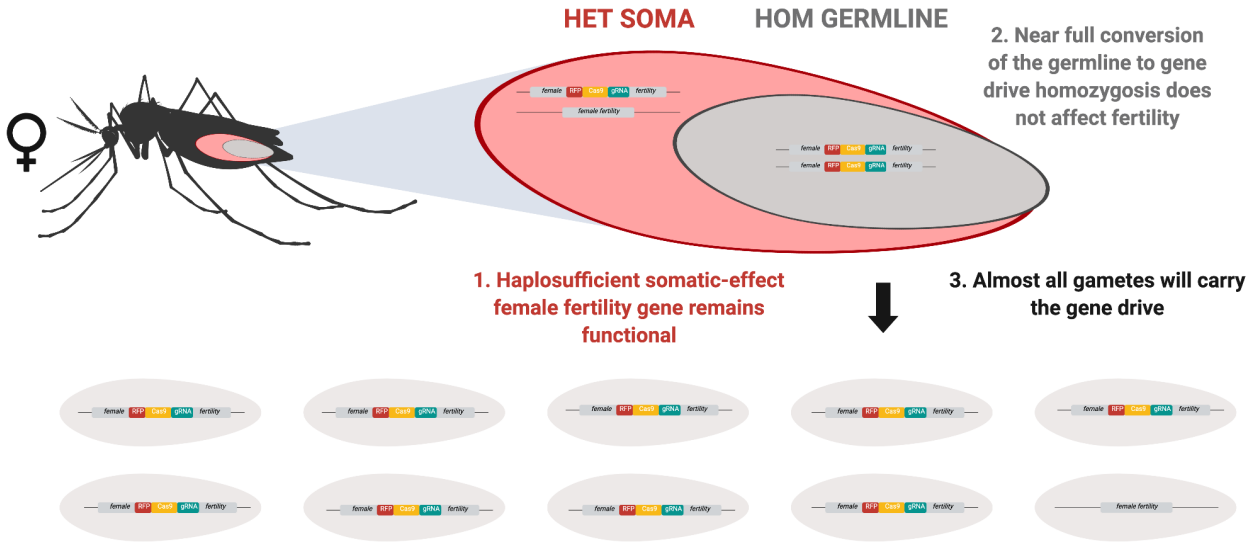


Figure 1.9: Ensuring fitness of female gene drive heterozygotes is crucial for gene drive spread. By targeting a haplosufficient female fertility gene that is somatically required for function, homing in the germline causing full knock-out of the targeted gene should not impact upon fertility. The gene drive allele is multicoloured, whilst the wild-type allele is grey. HOM = homozygous for the gene drive, HET = heterozygous for the gene drive.

1.6 Resistance to Gene Drive

1.6.1 Types of resistance

Resistance to gene drive refers to any genetic change (or a set of genetic changes) in the mosquito, that render the drive non-functional. For population replacement strategies resistance, even if generated in the mosquito, it is unlikely to get selected for, unless the GDS confers a fitness cost (which should be avoided for success of the strategy, see section 1.5.2). However, as with any vector and pest control strategy the use of population suppression gene drives will create an evolutionary pressure for the selection of resistance (Deredec et al. 2011).

For example, novel genetic elements might evolve, that shutdown the expression of the GDS, or block its activity at the protein level, such as short interfering RNAs, or anti-Cas proteins, respectively. Additionally, behavioural changes might evolve to limit the function of the drive, such as mating avoidance of gene drive carriers. However, it is hard to imagine such changes within the short evolutionary timeframe that gene drives act.

Additionally, the gene drive element might lose its ability to function due to the accumulation of mutations that disrupt (e.g. the Cas9), or lead to loss of one of its essential components, for example by error-prone/incomplete HDR repair (e.g. loss of the gRNA component). In this case, functional gene drives will out-compete non-functional ones lacking drive. The latter might also get selected out of the population if they confer a fitness cost, as is the case with suppression drives.

Moreover, cryptic population structuring (this does not refer to geographical barriers) might prevent invasion by a gene drive due to sub-populations of the same species not inter-breeding freely (Tennessen et al. 2021). However, even very low levels of admixture between distinct sub-populations should be enough to allow gene drive spread, due to the self-sustaining nature of homing gene drives; and perhaps this is something that can be achieved in a laboratory setting if needed, by introgressing the drive into different genetic backgrounds.

Since CRISPR-based gene drives rely on specific recognition and cleavage of a target site, based on sequence complementarity to a gRNA (Figure 1.8), the most parsimonious path to resistance is the evolution of a target site that can no longer be recognised by the gRNA and cleaved by the nuclease. This is commonly referred to as 'target site resistance' and will be the focus of the present thesis. All homing based gene drives are susceptible to this type of resistance since they are reliant upon sequence-specific recognition.

Target site resistance in the context of population replacement drives is only problematic if it the drive is associated with a fitness cost and thus the resistant allele has a selective advantage over the drive allele, otherwise they lack drive and are unlikely to increase in frequency.

In the case of population suppression drives, we distinguish two types of target site resistance: **R1** alleles that block homing and restore partial or full function of the essential target gene, and **R2** alleles that can block homing, but still disrupt the function of the targeted gene. **R1** alleles are certain to come under strong selection, as previously demonstrated, and reverse gene drive spread (Hammond et al. 2017, Hammond et al. 2021a), as long as they are generated in appreciable frequencies that prevent their stochastic loss from the population.

Conversely, **R2** alleles will continue to disrupt the target site and as such will not offer a direct evolutionary advantage over the gene drive, especially if created at low frequencies, and are therefore unlikely to get selected (Beaghton et al. 2019). However, they can still affect population-wide dynamics of drive if readily generated (Beaghton et al. 2019). Under this scenario the gene drive and R2 alleles can co-exist at an equilibirum, preventing complete population elimination (Beaghton et al. 2019). This is because they prevent the drive from spreading each time they are found in heterozygosity with the drive, and protect wild-type alleles when found in heterozygosity to them. Thus, they slow the spread of the drive, and help maintain essential alleles required for population survival. This effect is exacerbated if the gene drive itself confers extra fitness costs (for example reduced fertility in female heterozygotes when a haplosufficient female fertility gene is targeted).

In this thesis **R3** alleles are also introduced to represent those that though functional (like **R1** alleles) can only partially block gene drive homing (in contrast to **R1** alleles that fully block gene drive homing).

Resistant alleles might get generated at the target site *de novo* due to spontaneous introduction of mutations during meiosis. In particular, spontaneous mutation rates by base substitution were estimated at $1 * 10^{-9}$ in *An. coluzzii* (Rashid et al. 2022).

Resistant alleles, and particularly those that carry no fitness cost (R1) might already be established in certain natural populations, prior to implementation. Pre-existing R1 alleles would completely prevent gene drive invasion in populations in which they are widespread, and could get further selected considering they would offer a fitness advantage. This is why it is important to target sites that show low amounts of sequence variation in nature (see section 1.6.2).

Importantly, target site mutations might also get generated by gene drive activity itself, through error-prone EJ repair of the cleaved target site (Figure 1.10). These are several order of magnitude more frequent than *de novo* mutagenesis events, arising in approximately half of the chromosomes that did not get repaired by HDR ($\sim 1 * 10^{-2}$) (Hammond et al. 2016).

Generation of resistant alleles, particularly R1, should be reduced to the minimum to ensure success of a gene drive (Burt 2003).

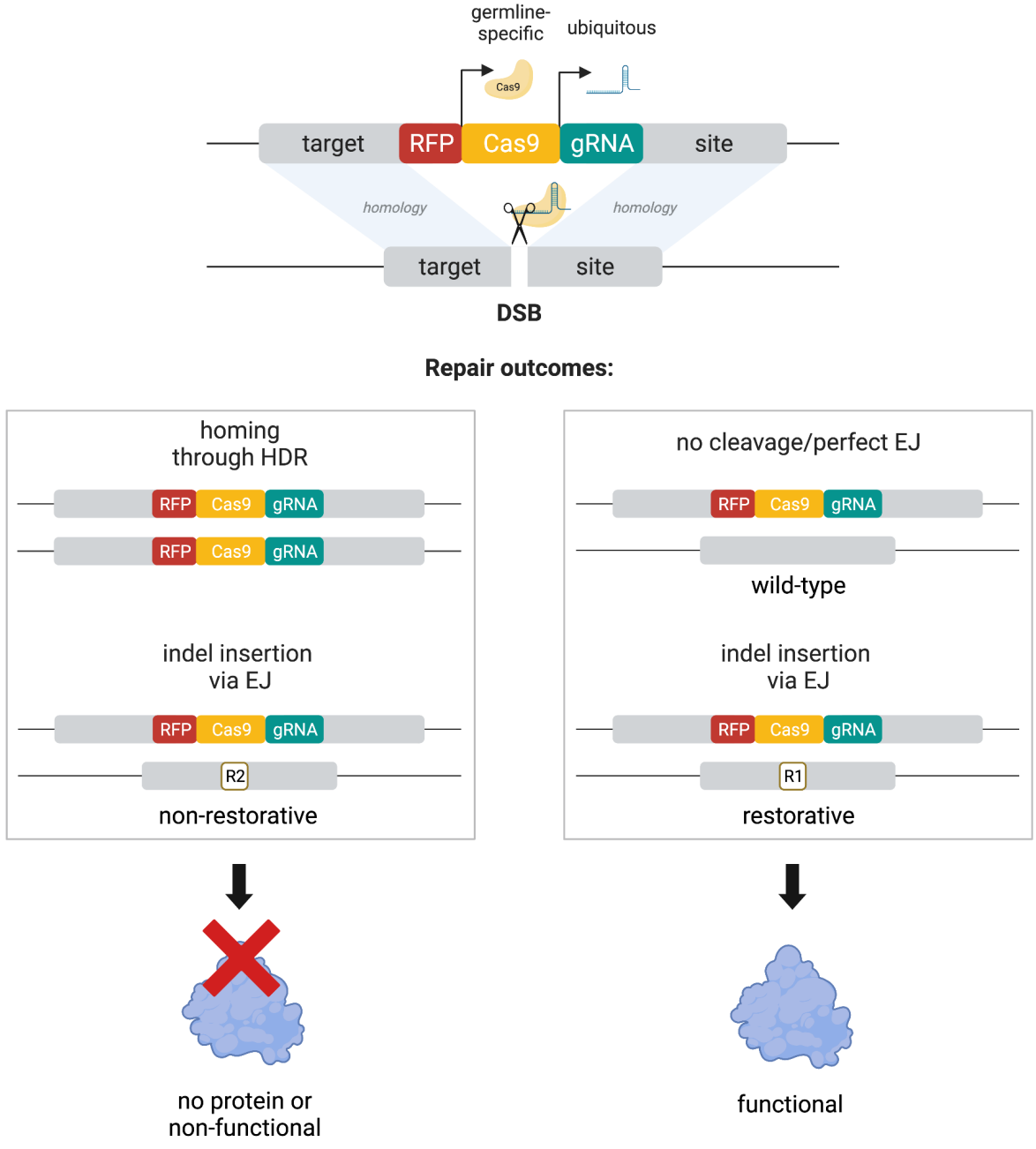


Figure 1.10: The creation of resistant mutations through gene drive activity. After nuclease-mediated DSB of the gene drive target site, there are two main avenues for repair of the broken chromosome: HDR which leads to homing of the gene drive into the site of the cut or error-prone EJ repair that can lead to the incorporation of a putative cut-resistant mutation at the site of the cut that might (R1) or might not restore (R2) the function of the targeted gene.

1.6.2 Strategies to mitigate resistance

1. Choice of target gene

As mentioned in section 1.5.2, the type of target gene chosen (e.g. influencing survival vs. fertility, acting immediately or in the following generation, acting in one or both sexes, etc.) can influence the strength of selection for resistance, depending on the fitness cost it incurs by being disrupted by the GDS (fitness cost: bisex>unisex, lethality>fertility>adult survival>fertility of progeny) (Burt 2003, Deredec et al. 2011, Bier 2021). However, the choice of target gene is, so far, somewhat limited in *An. gambiae*, since only a handful of genes have been identified and characterised as putative gene drive targets (Hammond et al. 2016, Kyrou et al. 2018, Krzywinska et al. 2021).

2. Restricting spatiotemporal Cas9 expression to the germline

The rate at which different resistant alleles are created largely depends upon the rates of error-prone EJ, which occurs in the small fraction of cleaved chromosomes that do not repair by HDR. Several studies have shown that EJ mutations are primarily introduced due leaky Cas9 activity in embryonic tissues, where EJ repair is preferred over HDR, due to deposition of the nuclease from one of the parents into the early embryo (Hammond et al. 2017, Champer et al. 2017, Hammond et al. 2021a) (Hammond et al. 2017, Champer et al. 2017). Specifically, the *vasa2* regulatory elements that were previously employed for Cas9 expression in a gene drive targeting 7280 (Hammond et al. 2016), were found to induce strong maternal deposition of the nuclease, leading to high levels of EJ mutations, many of which remained functional (**R1**) and completely reversed gene drive spread (Hammond et al. 2017, Hammond et al. 2021a). Spatiotemporally restricting Cas9 expression to the germline, where HDR is dominant, by expressing it under the regulatory elements of *nanos* (*nos*), or *zero population growth* (*zpg*), was shown to delay the onset of resistant mutations in *Anopheles gambiae* (Hammond et al. 2021a). This is also supported by low rates of EJ creation by *nos::Cas9* in the fruitfly (Champer et al. 2018).

3. Targeting highly conserved sites

Though restricting nuclease expression to the germline delayed the onset of resistance, it did not completely inhibit the creation of resistant mutations, causing gene drive spread to be reversed nonetheless (Hammond et al. 2021a). This is because EJ mutations can also be introduced in the germline, during meiosis, in the rare fraction of unhomed chromosomes (i.e. those where homing by HDR did not proceed) (Hammond et al. 2021a). Avoiding the generation of functional R1 resistance in its entirety is critical for successful gene drive spread.

One way this can be achieved is by targeting functionally constrained sites that cannot tolerate se-

quence diversity. Kranjc et al. (2021) have queried the *An. gambiae* genome for sites that are highly conserved in nature to facilitate rational gene drive design. This would reduce the chance of encountering a pre-existing resistant allele in the natural population, whilst EJ mutants generated by the gene drive would most likely consist of non-functional R2 alleles, over functional R1.

Indeed, a gene drive targeting a highly conserved site on the *doublesex* gene, reached fixation and achieved complete elimination of caged laboratory populations within a year of its release (Kyrou et al. 2018, Hammond et al. 2021b) (see section 1.7).

4. Using alternative nucleases

Another way to limit the generation of both R1 and R2 alleles at the target site is to use alternative nucleases to shift DNA repair away from EJ and towards HDR, or to avoid CRISPR-induced cuts and DNA repair at the Cas9-recognition site altogether (Hammond & Galizi 2017).

Employing Cas9 nickases (nCas9) that produce a single-stranded break could presumable offer advantages. Repair of DNA nicks (i.e. single stranded DNA breaks) is very efficient at repairing into the wild-type DNA sequence, and when two nickases are used together to produce two distinct but closely located nicks it was hypothesised that they could trigger alternative repair pathways, in favour of HDR, rather than EJ (Del Amo et al. 2022). However, when tested in *Drosophila melanogaster* the coordinated action of nCas9s was not able to exceed the rates of HDR produced by Cas9, and it still resulted in unwanted EJ outcomes (Del Amo et al. 2022).

Using a Cpf1 endonuclease was also proposed (Hammond & Galizi 2017). The properties of Cpf1 offer many advantages that could be exploited in a gene drive design. Firstly, it produces a staggered DSB, in contrast to Cas9's blunt-ended DSB, which causes the formation of large deletions upon EJ repair (more likely to constitute R2 alleles), over small indels (more likely to constitute R1 alleles) (Zaidi et al. 2017). It is also more amenable to multiplexing compared to Cas9 (see section 6 below), and shows the least off-target activity, which could have advantages in terms of getting regulatory approval for its use (Zaidi et al. 2017).

Finally, FokI[4]-dead Cas9[5] (FokI-dCas9) fusions have been proposed as ways to mitigate target site resistance, since their cleavage site is located outside of their recognition sequence (Hammond & Galizi 2017). Thus, introduced variation due to gene drive activity should not block future target site recognition and homing, and will therefore not constitute R1 or R2 mutations.

Nonetheless, alternative nucleases have not been tested in the mosquito, and there are concerns that increasing the gene drive size by including two nCas9s or a dCas-FokI fusion in the gene drive construct

[4]FokI is a type IIS restriction endonuclease.

[5]Dead Cas9 (dCas9) is a mutant form of Cas9 that lacks Cas9's endonuclease function but maintains the ability to bind gRNA and recognise specific sequences complementary to it.

might reduce homing efficiency.

5. Biasing EJ mutation outcomes towards non-functional alleles

The discovery that micro-homologies at the target site can influence EJ repair outcomes in favour of R1 resistance (Hammond et al. 2017), also raised the possibility of intentionally biasing mutation outcomes to favour non-functional R2 alleles over R1, by selecting sites that favour out-of-frame over in-frame microhomology-mediated end-joining (MMEJ) mutations (Kranjc 2022).

6. Targeting multiple sites simultaneously

Finally, a combinatorial approach can be adopted, whereby multiple sites are targeted simultaneously to reduce the likelihood that resistance at one of the sites could reverse gene drive spread (Marshall et al. 2017, Champer et al. 2018). This can be done by employing a gene drive that, in addition to a Cas9, encodes an array of distinct gRNAs ('multiplexed gRNAs') that target multiple non-overlapping sites on the same gene, or sites present on different genes. Alternatively, distinct gene drives containing a single gRNA each can be consecutively released. However, multiplexing gRNAs into a single gene drive construct that targets closely located sites offers the additional advantage of actively removing resistant mutations generated at one site, by homing using another target that remains cleavable (Figure 1.11). Indeed, a homing suppression gene drive with multiplexed gRNAs was able to avoid R1 resistance in *D. melanogaster* (Yang et al. 2022), and the *Aedes aegypti* mosquito (Anderson et al. 2022). However, multiplexing has never been tested before in the malaria mosquito, *Anopheles gambiae*.

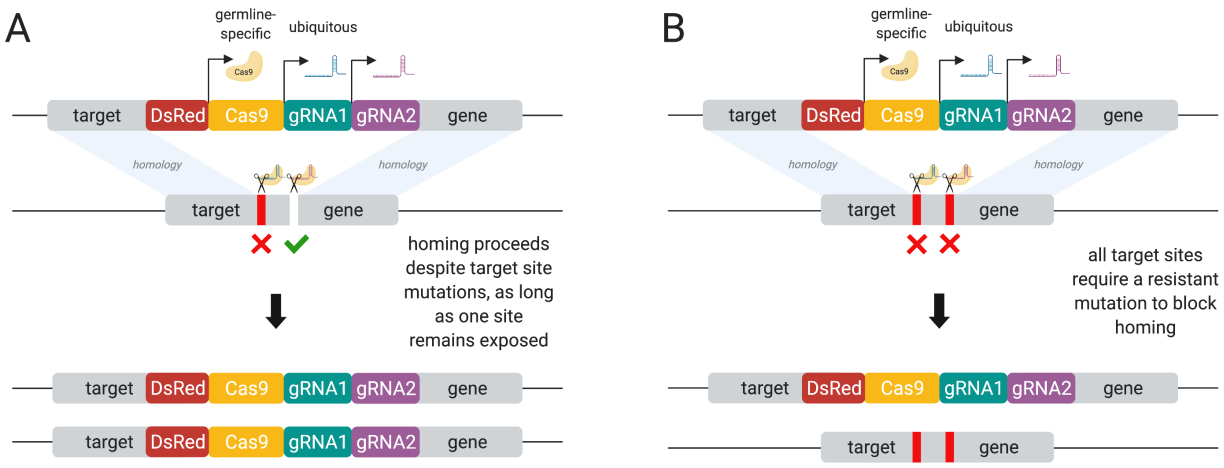


Figure 1.11: Multiplexed gene drives mitigate resistance by expressing multiple gRNAs that target many sites simultaneously. (A) Resistant mutations (shown in red) are removed from the target locus as long as one of the targeted sites remains cleavable, to permit homing. **(B)** To block gene drive homing all target sites would need to carry a resistant mutation (shown in red).

1.7 Targeting the *doublesex* gene

1.7.1 The *doublesex* gene

Doublesex (DSX) is a transcription factor that acts as the terminal 'double-switch' of the sex determination cascade in insects, responsible for the development of sexually dimorphic physical characteristics in both males and females (Figure 1.12) (Price et al. 2015). It belongs to the Doublesex/Mab-3 Related Transcription factor (DMRT) family of zinc-finger proteins, and is alternatively spliced into two distinct sex-specific isoforms (Figure 1.13) (Scali et al. 2005). Despite the initial signals for sexual differentiation being largely variable amongst animal phyla, the end-point of the somatic sexual differentiation cascade is fundamentally conserved in metazoans. Specifically, the DMRT family is present in most animal genomes and has likely pre-eumetazoan origins, whilst the DSX protein itself was found to be conserved across 30 orders of hexapods. Noteably, other more distantly related arthropods (hexapod outgroups), also express *dsx* homologs that are contained in two distinct gene copies, rather than being alternatively spliced (Price et al. 2015).

DSX transcription factors retain a universal domain conserved across the *dmrt* family of genes (DM domain) that consists of a DNA-binding domain (DBD domain) and an oligomerisation domain (OD1 domain). They also contain a second oligomerisation domain specific to DSX (OD2 domain) with a common N-terminus in both sexes, and a sex-specific C-terminus resulting from alternative splicing of *dsx* transcripts in males and females (Verhulst & Van de zande 2015) (Figure 1.14). A mutation that blocked oligomerisation of the DSX female isoform produced female flies that exhibit an intersex phenotype in *D. melanogaster*, whilst males remained unaffected (Erdman et al. 1996). Therefore, oligomerisation is essential for DSX function. Similarly, in the silk moth, *Bombyx mori*, expression of the male DSX isoform in females led to an intersex phenotype, whilst males were unaffected by ectopic expression of the female isoform (Suzuki et al. 2005). In the broad-horned beetle, *Gnatocerus cornutus* knock-down of the female specific isoforms, also led to an intersex phenotype in females while males were unaffected (Gotoh et al. 2016).

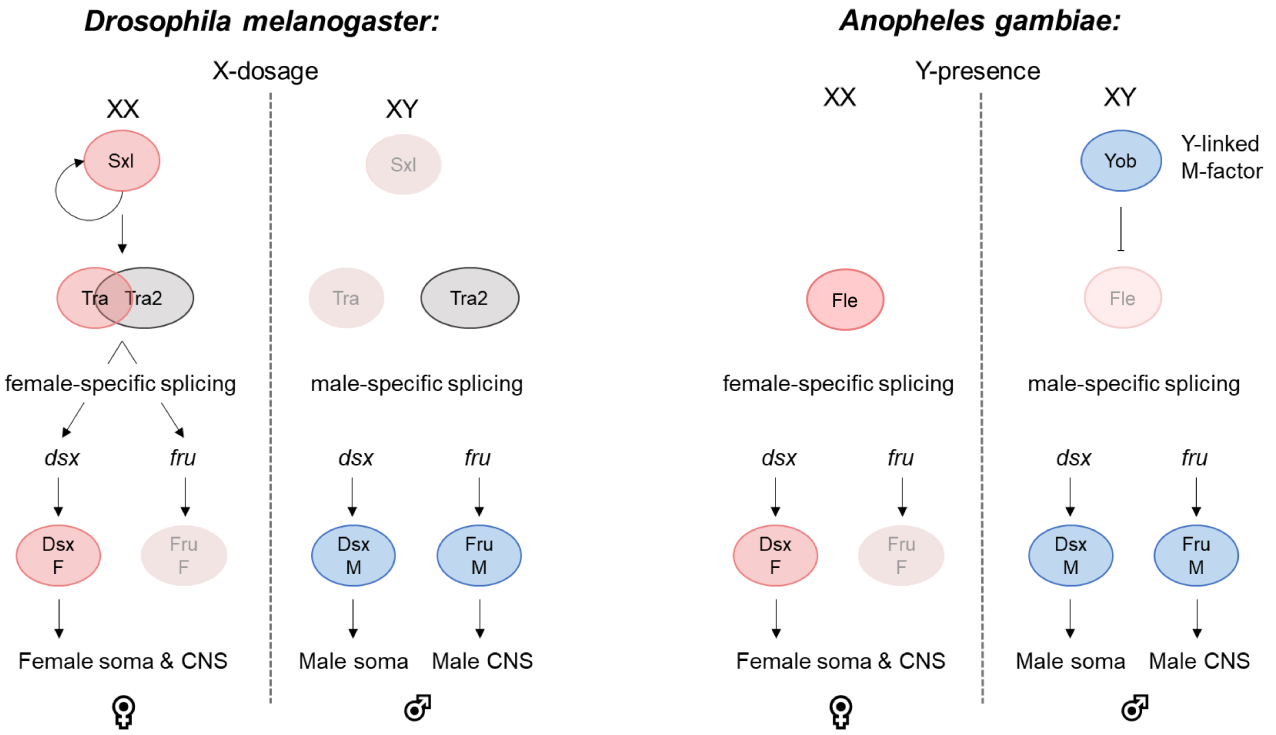


Figure 1.12: Overview of the sex determination molecular cascade in *D. melanogaster* and *An. gambiae*. In *Drosophila* a double dose of X-linked elements is the primary signal that initiates the female-specific splicing of *sex lethal* (*sxl*) which is maintained through an autoregulatory positive feedback loop (Erickson & Quintero 2007). The SXL splice factor controls female-specific splicing of *transformer* (*tra*). TRA in turn complexes with TRA2, that is present in both sexes, to control the splicing of *doublesex* and *fruitless* (*fru*) (Billeter et al. 2006). Although the female-specific isoform of *fru* is thought to be non-functional the *dsx* female-specific isoform contributes to the development of female morphological and behavioural features. In males, the absence of funcitonal SXL and TRA splicing factors causes *dsx* and *fru* to be spliced into male-specific isoforms controlling the development of the male soma and central nervous system (CNS), respectively (Billeter et al. 2006). In mosquitoes, a Y-linked maleness factor, Yob, is though to be the primary signal involved in sex determination, initiating a splicing cascade culminating in the sex-specific splicing of *dsx* and *fru* (Krzywinska et al. 2016). Though homologs of TRA2 also exist in the mosquito, they were not found to control the alternative splicing of sex determinants *dsx* and *fru* (Krzywinska et al. 2021). Instead, females express another factor termed *femaleless* (*fle*) that is more closely related to the male-determinant *Nix* in *Aedes*, rather than *tra2*. Fle controls the alternative splicing of female *dsx* and *fru*, whilst it is thought to be suppressed by Yob, in males, to allowing male-specific splicing of *dsx* and *fru* to proceed. F = female, M = male.

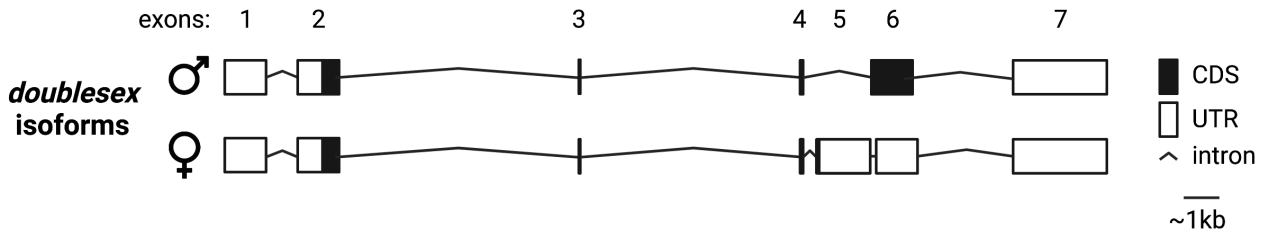


Figure 1.13: The male and female-specific *dsx* isoforms in *An. gambiae*. The female isoform contains an extra exon (exon 5) and lacks the exon 6 CDS which instead serves as a UTR. CDS = coding sequence, UTR = untranslated region.

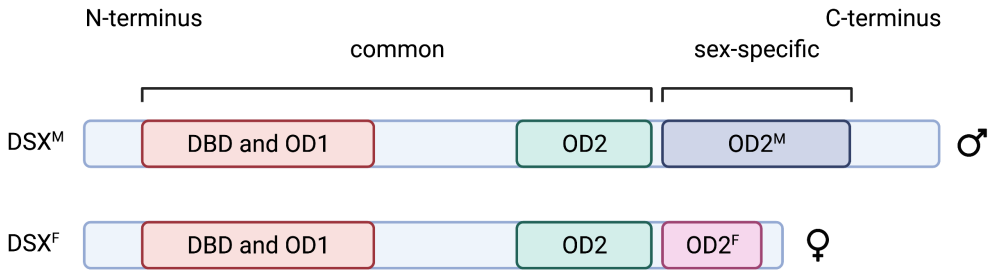


Figure 1.14: The protein domains of the male and female-specific DSX isoforms in *An. gambiae*. DBD = DNA-binding domain, OD1 = oligomerisation domain 1, OD2 = oligomerisation domain 2., F = female isoform, M = male isoform. Figure not to scale.

1.7.2 Sequence conservation at *dsx*

In *An. gambiae* the *dsx* gene (AGAP004050) consists of 7 exons and 6 introns (Figure 1.13). The main difference between its male- and female-specific isoform is that the male isoform skips exon 5 and contains an additional C-terminal domain, which is only present as a 3' UTR in females (Figure 1.13) (Scali et al. 2005). Interestingly, the entirety of the female-specific intron 4 shows a great amount of intra- and inter-specific sequence conservation (Kyrou 2021). It has been previously shown that intronic sequences flanking alternatively spliced exons tend to show high levels of conservation between distantly related species, despite the average level of conservation between intronic sequences being low, in general (Sorek & Ast 2003). It is speculated that the DNA sequence of the female-specific intron 4-exon 5 junction is unable to tolerate most nucleotide variation due to sequence-specific requirements for binding of a splicing factor that regulates alternative splicing, presumably femaleless (FLE) (refer to Figure 1.12) (Krzywinska et al. 2021).

It is worth noting that in *D. melanogaster*, the female-specific exon of *dsx* is preceded by a non-canonical 3' acceptor splice site to allow exon-skipping in TRA-lacking males (refer to Figure 1.12) (Lynch & Maniatis 1995). In contrast, the *An. gambiae* gene has retained the canonical CAAG acceptor splice site, which seems to be conserved prior exon 5 in all mosquito species examined (Figure 1.15). Perhaps there is a male-specific factor that recognises sequences surrounding the CAAG motif and masks that site to promote exon-skipping.

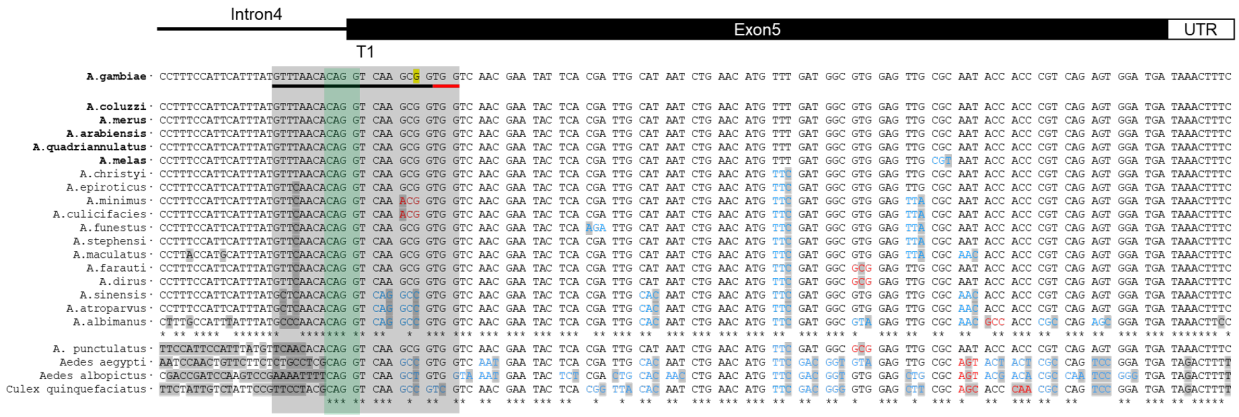


Figure 1.15: Sequence alignment of the exon 5 CDS of *dsx* and its flanking regions. Species names are shown to the left. Species in bold belong to the *An. gambiae* species complex (sensu lato). Nucletodies that differ compared to the *An. gambiae* sensu stricto sequence on the top of the graph are shaded in dark grey. Nucleotides are shown in light blue or red, depending on whether a variation causes a synonymous or non-synonymous amino acid change in the exon 5 CDS. The most distantly related species are separated (last four). Asterisks denote the nucleotide positions that remained unchanged in all species examined. The location where the Ag(QFS)1 gRNA (T1) binds is underlined and shaded in light grey. The PAM is underlined in red. The canonical 3' acceptor splice cite is shaded in green. A SNP that was previously identified in natural populations of *An. gambiae* is highlighted in yellow (Miles et al. 2017, Kyrou et al. 2018).

1.7.3 A population suppression gene drive targeting *doublesex* in *An. gambiae*

Kyrou et al. (2018) disrupted the 5' intron-exon boundary of the female-specific exon of *dsx* in *An. gambiae* by introducing a GFP cassette at that site, preventing the formation of a functional female transcript (*dsx*F-). The *dsx*F- disruption was found to be recessive, with heterozygous females following a normal sexual development, whilst females homozygous for the disruption developed into intersex individuals (Figure 1.16). Intersex mosquitoes, despite having a female (XX) genotype, showed underdeveloped male morphological features, including semi-plumose antennae and unrotated claspers (Figure 1.16). Anomalies in the proboscis did not allow intersex mosquitoes to blood-feed (required for egg production), whilst internally intersex mosquitoes lacked ovaries and a spermatheca and instead possessed male accessory glands (Kyrou 2021). As such, intersex individuals were completely sterile (Kyrou et al. 2018). Combined with the fact that males were unaffected by the *dsx*F knockout (Kyrou et al. 2018), and that the gene is somatically required for sexual development and fertility in females, the disrupted site fulfilled the requirements of a gene drive target site (see section 1.5.2).

In addition, what makes exon 5 of *dsx* an improved target compared to the previously studied 7280 (*nudel*) target gene (Hammond et al. 2016), is that it is extremely well-conserved across anopheline species (Neafsey et al. 2015) (Figure 1.15), implying higher functional constraint and lower likelihood for target site resistance to arise (Kyrou 2021).

Kyrou et al. (2018) developed a population suppression gene drive in *An. gambiae* that was designed to

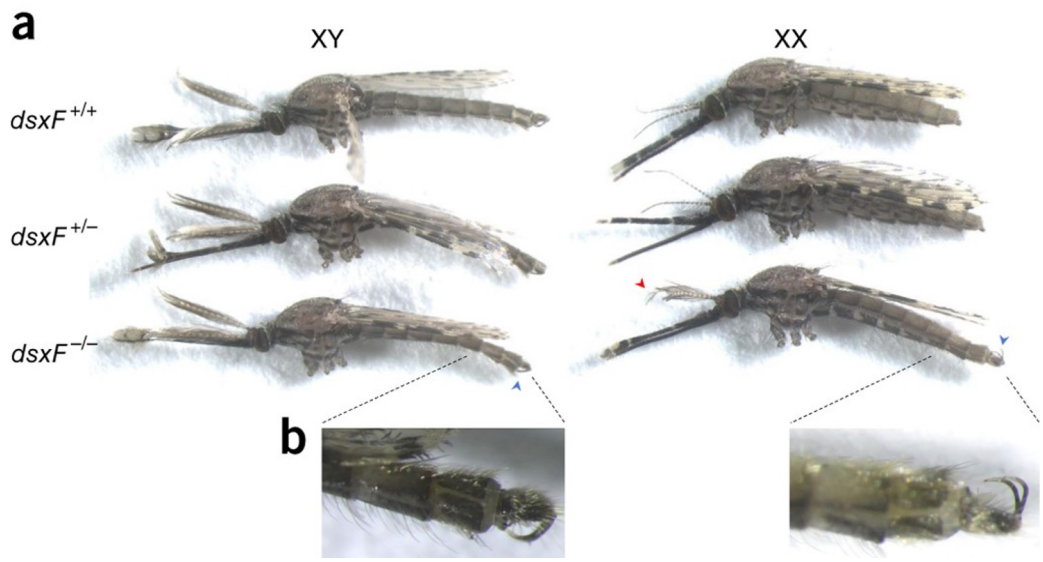


Figure 1.16: Morphological appearance of males and females heterozygous (dsxF$^{+/-}$) or homozygous (dsxF$^{-/-}$) for the exon 5 null allele. (A) Only XX females homozygous (dsxF$^{-/-}$) for the null allele showed an intersex anatomy characterised by the presence of underdeveloped male characteristics such as semi-plumose antennae (red arrowhead) and unrotated claspers (blue arrowheads). This group also showed anomalies in the proboscis and accordingly could not blood-feed and produce offsrping. Individuals of all other genotypes developed normally as males (XY) or females (XX). Representative samples of each genotype are shown. **(B)** Magnification of the external genitalia. All dsxF$^{-/-}$ females carried claspers that were dorsally rotated rather than in the normal ventral position. This figure from Kyrou et al. (2018) was adapted with permission from Kyros Kyrou.

mitigate resistance both by employing the germline-restricted *zpg* promoter to express Cas9 and by targeting a highly conserved site (T1), located at the female-specific intron 4-exon 5 boundary of *dsx* (Figure 1.15). This gene drive strain was named Ag(QFS)1 and was shown to spread to fixation within a year of release into laboratory-contained mosquito populations [6], leading to complete population elimination in both small and large-scale cage trials (Kyrou et al. 2018, Hammond et al. 2021b). Pooled amplicon sequencing revealed no evidence of resistance in either setting: despite in-frame mutations being generated, they did not seem to be tolerated at *dsx*. This was the first demonstration that a population suppression gene drive can spread to fixation, without selection of resistance, and induce complete population elimination. Attempts to target other highly conserved loci were not as fruitful, and ultimately led to the evolution of resistance (Carrami et al. 2018, Fuchs et al. 2021). The only other suppression gene drive that has been shown to achieve complete population elimination, since, was also targeting the same locus on *dsx* (Simoni et al. 2020). Modelling predicts that these gene drives could cause a stable ~95% population suppression 4 years post-release in a regional scale, in the absence of R1 alleles (North et al. 2020). Such a strong suppression of the malaria vector, together with the continued use of other malaria prevention measures (described in 1.2), might be enough to cause, at least regional, malaria elimination.

[6] QFS stands for driving female sterility.

Chapter 2

Aims & Objectives

Gene drive designed for population suppression of the malaria mosquito that are targeted to DSX have the potential to eliminate malaria across substantial regions of sub-Saharan Africa (Kyrou et al. 2018, North et al. 2020), should resistance be effectively contained. It is therefore critically important to predict, pre-empt, and mitigate resistance before a first release into the wild (Burt 2003). Though no R1 resistance was detected against the Ag(QFS)1 gene drive in both small and large-scale cage trials (Kyrou et al. 2018, Hammond et al. 2021b), from these experiments we cannot exclude the possibility that this may occur if gene drive mosquitoes were to be released in sub-Saharan Africa. The challenge cannot be underestimated as the dramatic change in scale, from laboratory populations of 10^3 to wild populations in excess of 10^9 (Khatri et al. 2019), provides unprecedented opportunities for resistance to occur, due to both natural variation and gene drive activity (Figure 2.1). During the course of this PhD I aimed to: (a) assess the potential of Ag(QFS)1 to spread through natural populations of the malaria mosquito in sub-Saharan Africa, by carrying out a deep investigation of resistance at the Ag(QFS)1 target site, and (b) develop novel gene drives designed to counteract naturally-occuring and evolved resistance.

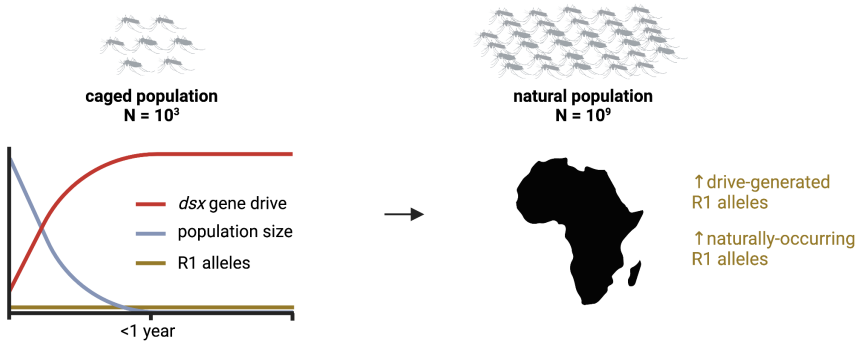


Figure 2.1: The chance that resistant mutations will get selected against the gene drive is higher in the natural environment. No R1 resistant alleles were detected against the Ag(QFS)1 gene drive, in laboratory populations. However, the natural population is several orders of magnitude larger and so there is a higher chance that resistant mutations will be encountered, drive-generated or pre-existing in nature.

2.0.1 Aim 1: To evaluate the landscape of natural variation at putative gene drive target sites on *dsx*

Target site resistance can be pre-existing in the natural population of malaria mosquitoes (see 1.6.1). Despite the high conservation of the female-specific exon (exon 5) of *dsx* (Figure 1.15), the *Anopheles gambiae* 1000 genomes project (Ag1000G) had previously shown that 2.9% amongst 765 wild-caught *An. gambiae* s.l. mosquitoes contained a G→A SNP, within the Ag(QFS)1 gene drive target site (Miles et al. 2017, Kyrou et al. 2018) (Figure 1.15). Since then, more than 2,000 new genomes have been added to the Ag1000G database. Thus, I aimed to survey the latest mosquito population genomics data (The Anopheles gambiae 1000 Genomes Consortium 2021), for the presence of natural variants at the Ag(QFS)1 target site, extending this search to include novel putative gene drive target sites on exon 5.

2.0.2 Aim 2: To evaluate the landscape of drive-induced variation at the Ag(QFS)1 target site

Target site resistance can also be drive-induced (see 1.6.1 and Figure 1.10). Hammond et al. (2018) showed that drive-resistant mutations (i.e. those that block gene drive activity) can either disrupt gene function altogether (R2 alleles) or maintain gene function (R1 alleles) and increase their frequency in the population, whilst displacing gene drive alleles (Hammond et al. 2017, Hammond et al. 2021a). Arising as a result of MMEJ, the most common resistant alleles generated at the 7280 gene drive target site unfortunately consisted of in-frame, rather than out-of-frame mutations. In general, in-frame mutations have a higher chance of being functional as they cause amino acid insertions, deletions, or substitutions, at the point of the mutation without disrupting the rest of the coding sequence. Conversely out-of-frame mutations, are largely non functional, as they cause the shifting of the gene's open reading frame, causing the translation of a nonsense gene product (note that in some rare cases even frameshift mutations are tolerated (Wang et al. 2022)). In the Hammond et al. (2018) study in-frame mutations varied a lot in size: from 3, to 6, or even up to 12 bp in length, leading to 1, 2 or 4 amino acid deletions respectively, that did not overtly disrupt 7280 function, presumably because they did not consist of catalytic residues in the 7280 protein.

In general, gene drive studies have shown that nucleotide deletions are the most common EJ repair outcomes, followed by insertions, whilst SNPs are very rare (Hammond et al. 2018, Hammond et al. 2021a).

For the *dsx* gene drive, Ag(QFS)1, it is estimated that every time a gene drive biases its inheritance there is a 0.5% chance that the drive-induced cut will repair by error-prone EJ, causing the generation of putatively resistant alleles (Hammond et al. 2016, Hammond et al. 2021a, Kyrou et al. 2018) (Figure

1.10). This means that approximately 1 in 200 progeny of gene drive carriers will carry an EJ mutant. So far only non-restorative R2 mutations have been detected at the Ag(QFS)1 target site (T1). A cut-resistant in-frame 3 bp deletion, leading to a single amino acid deletion, that was detected is thought to be non-functional since it did not reverse gene drive spread in laboratory population experiments, though generated at an appreciable frequency (Kyrou et al. 2018, Hammond et al. 2021b). This result is in stark contrast to the results of the previous gene drive targeting 7280, and confirms the assumption that the *dsx* target site is more functionally constrained than the 7280 one. Nonetheless, it is unlikely that any single site will be completely 'resistant-proof', as is the case for any suppressive technology, including insecticides and antibiotics (Hammond & Galizi 2017).

The rate at which functional resistance could occur against Ag(QFS)1, would dictate how long the gene drive may remain effective post-release. Thus, measuring this rate is instrumental for the design of an efficient gene drive release strategy that, combined with ongoing efforts to control the population of the malaria mosquito and minimise infections, could lead to disease eradication. Importantly, complete population elimination is not essential to eradicate malaria (Burt & Deredec 2018). Driving the vector population to a low enough number to completely curb malaria transmission is possible, upon which re-population of the mosquito vector due to the emergence of R1 resistance, would not necessitate a resurgence in disease transmission as: (a) populations of *An. gambiae* may fail to fully recover since their niche might have already been occupied by other insects, (b) local elimination of the parasite would prevent disease transmission assuming no or controlled imported cases, and (c) follow-ups using traditional control measures (e.g. mass insecticide treatment) could be targeted to regions where malaria is detected, or where the population of *An. gambiae* reaches a critical threshold that puts the human population at high risk of disease re-introduction.

Assessing rare repair outcomes that were not detected in populations of 600 mosquitoes (Kyrou et al. 2018, Hammond et al. 2021b), would be extremely labour-intensive, and potentially impossible within the confines of a standard molecular biology lab at the scale at which it may need to be done if informative for wide-scale application. This is primarily due to the strain's extremely high transmission rates (averaging at about 98%), which only leaves a small fraction of individuals containing a putative resistant EJ mutation (Kyrou et al. 2018).

Thus, my second aim was to develop and perform a high-throughput mutagenesis screen to purposefully generate a high number of EJ mutations at the T1 target site, determine at what frequency each mutation gets generated, and assess their functionality, to understand if they fall under the R1 or R2 category. This will avoid having to screen thousands of gene drive mosquito progeny, in order to discover a handful of mutations.

2.0.3 Aim 3: To deliberately reverse engineer putatively resistant alleles in the mosquito genome and assess their functionality, as well as their cleavability by the gene drive

Kyrou et al. (2018) showed that, *in vitro*, the natural G→A SNP at T1 is cleavable by a *Streptococcus pyogenes* Cas9 (SpCas9) when complexed with the gRNA (T1) employed by the Ag(QFS)1 gene drive. Nonetheless, *in vitro* cleavage can be more promiscuous and does not always recapitulate *in vivo* cleavage results (Garrood et al. 2021). Moreover, aims 1 and 2 were expected to yield novel putative R1 alleles. Thus, my third aim was to engineer the top R1 candidates discovered under Aims 1 and 2 into laboratory mosquito strains, to: (a) confirm their functionality and (b) determine if they are cleavable by the gene drive *in vivo*.

2.0.4 Aim 4: To generate and test multiplexed gene drive strains designed to mitigate resistance

Pre-empting for the eventuality of resistance at any one given site, the final aim of my PhD was to generate and test gene drive strains that target multiple sites on *dsx* exon 5 simultaneously, as a way to mitigate potential resistance. Multiplexing gRNAs that target more than 2 sites at the same time has previously been proposed as a method to counteract resistance, and has been applied successfully in *D. melanogaster*, but never in mosquitoes (see paragraph 5 of section 1.6.2).

Chapter 3

Methods

3.1 Assessing natural variation at three QFS gene drive target sites (T1, T2 and T3)

3.1.1 MalariaGEN data analysis

The Ag1000 phase 3 data release was accessed from Google Collaboratory, following instructions on the MalariaGEN website. The dataset includes genome-wide SNP calls from whole-genome sequencing of 2,784 wild-caught mosquitoes collected from 19 countries (The Anopheles gambiae 1000 Genomes Consortium 2021).

The Ag1000 phase 3 data was installed as such:

```
!pip install -q numpy dask[array] zarr gcsfs fsspec
!pip install -q malariagen-data
import malariagen_data
ag3 = malariagen_data.Ag3("gs://vo_agam_release/")
df_sample_sets = ag3.sample_sets()
df_samples = ag3.sample_metadata(sample_sets="v3_wild")
```

The positions corresponding to the three putative gene drive target sites (T1, T2, T3) were located on the contig 2R genotype array as such:

```
pos, ref, alt = ag3.snp_sites(contig="2R")
pos = allel.SortedIndex(pos)
loc_target_site = pos.locate_range(<insert genomic coordinate start>,
<insert genomic coordinate end>)
print(loc_target_site)
```

Each of the sites were defined using the genotype array coordinates, were each row is a position on the target site, and each column a mosquito sample:

```
import allel
gt = ag3.snp_genotypes("2R")
gt = allel.GenotypeDaskArray(gt)
target_site = gt[<insert contig array coordinate start>:<insert contig
array coordinate end>].compute()
```

The genotype coordinates for each SNP within a target site were obtained. This information included the position of the SNP and information on the mosquito sample within which it was identified (sample ID, country, species):

```
gens_coord = {}
bp_coord = target_site[:,0,0].size
sampl_coord = target_site[0,:,0].size
for row in range(bp_coord):
 for col in range(sampl_coord):
   check_coord = (target_site[row,col,:]==[1,0])
   if check_coord.any() == False:
     if row not in gens_coord.keys():
       print(row,col)
       print(df_samples.country[col])
       print(df_samples.species[col])
```

For example, in the below output, a SNP is present in position 4 of the queried sequence, in sample no. 452 which was an *Anopheles gambiae* mosquito collected from the Democratic Republic of Congo.

```
4 452
Democratic Republic of Congo
gambiae
```

The target site positions, reference alleles and alternate (SNP) alleles per each position were defined.

```
pos_target_site = pos[loc_target_site]
ref_target_site = ref[loc_target_site].compute()
alt_target_site = alt[loc_target_site].compute()
```

Finally, the nature of each SNP, namely the base change it corresponds to and whether it was found in homozygosis or heterozygosis was deciphered.

```
import numpy as np
gens = {}
bp = target_site[:,0,0].size
sampl = target_site[0,:,0].size
for row in range(bp):
 for col in range(sampl):
   check = (t1[row,col,:]==[1,0])
   if check.any() == False:
     if row not in gens.keys():
       print(row)
       print(ref_target_site[row])
       print(alt_target_site[row])
       print(pos_target_site[row])
       gens[row] = []
     gens[row].append(list(t1[row, col,:]))

false_bps = list(gens.keys())
for i in false_bps:
 print(gens[i])
 print(len(gens[i]))
```

For example, in the below output, 4 is the location of the SNP in the target site sequence; C (0) is the reference base; A, T, and G (1, 2, 3) are the alternate bases; 48714641 is the position of the SNP in the 2R contig. Therefore [2,2] corresponds to a homozygous C→T SNP, and [0,2] corresponds to a heterozygous C→T SNP, whilst 64 is the total number of mosquitoes with said SNP. Note that these are sense strand reads, and doublesex is found in the antisense strand so C→T would translate to G→A.

```
4
b'C'
[b'A' b'T' b'G']
48714641
[[2, 2], [0, 2], [0, 2], [2, 2], [0, 2], [0, 2], [0, 2], [2, 2],
[0, 2], [0, 2], [0, 2], [2, 2], [0, 2], [0, 2], [2, 2], [0, 2],
[0, 2], [0, 2], [0, 2], [0, 2], [0, 2], [0, 2], [0, 2], [0, 2],
[0, 2], [0, 2], [0, 2], [0, 2], [2, 2], [0, 2], [0, 2], [0, 2],
[0, 2], [0, 2], [0, 2], [0, 2], [0, 2], [0, 2], [0, 2], [0, 2],
[0, 2], [0, 2], [0, 2], [0, 2], [0, 2], [0, 2], [0, 2], [0, 2],
[0, 2], [0, 2], [0, 2], [0, 2], [0, 2], [0, 2], [0, 2], [0, 2],
[0, 2], [0, 2], [0, 2], [0, 2], [0, 2], [0, 2], [0, 2], [0, 2]]
64
```

3.2 Mosquito rearing and microinjections

3.2.1 Mosquito rearing

The wild-type mosquito strain used in all experiments is the *Anopheles* G3 strain, which is known to be a mix of major malaria vectors *An. gambiae* and *An. coluzzii*. G3 and transgenic mosquitoes were reared in cubicles with temperature and humidity control of 26±2ºC and 65±10%, respectively, and 12 hours daylight cycling, inside an arthropod level 2 containment (ACL2) insectary facility. Larvae were maintained in trays with 500 ml 1% tonic salt water, at a larval density of 200 per tray and fed on NishiKoi food pellets. Adults were fed on a 10% glucose solution and females were blood-fed on cow blood, approximately 5 days after being allowed to mate to males, using Hemotek membrane feeders. Females are allowed to lay eggs on water 2-3 days PBM.

3.2.2 Embryo microinjections and strain establishment

Mosquito embryo microinjections were performed on freshly laid embryos or as previously described by Fuchs et al. (2013), Kyrou et al. (2018). All embryo microinjections for strains described in the present thesis (besides a tiny fraction performed by myself) were performed by Andrew Hammond or Louise Marston, for whom I aligned the eggs. For random genomic integrations, a given plasmid containing 5' and 3' piggyBac repeats flanking the region to be integrated, was co-injected into G3 embryos together with a helper plasmid expressing a piggyBac transposase under the *vas2* promoter (Volohonsky et al. 2015). Site-specific integrations were either performed using CRISPR targeting or RMCE. For CRISPR-assisted integrations the region to be integrated was assembled in a plasmid flanked by homology arms complementary to 2kb upstream and downstream of the chosen integration site. This was co-injected together with a CRISPR plasmid expressing Cas9 under a germline-specific promoter (usually *vas2* or *zpg*) and one or multiple gRNAs directing the Cas9 to cleave the site(s) in which (or among which) the region of interest would be integrated. Gene drive constructs flanked by attB sites were co-injected into the offspring of docking line heterozygotes (containing a marker flanked by attP sites) together with a helper plasmid expressing the ΦC31 integrase to catalyse recombination between the attP and attB sites (Volohonsky et al. 2015), leading to RMCE as described in Hammond et al. (2016).

All plasmids were microinjected at a concentration of 300 ng/μl in phosphate-buffered saline solution (PBS). After 48 hours, hatched L1 larvae (G0) were screened under a fluorescence microscope for transient expression of fluorescent proteins expressed by the microinjected plasmids under the 3xP3 promoter[1]. Positive G0 larvae (transients) showing fluorescence in the tail were grown to adulthood and out-crossed to the G3 strain. Negative G0 larvae (non-transients) were crossed to each other. G1 progeny of both crosses were screened for a full fluorescent phenotype (see footnote) indicative of successful transformation. New strains were set up using the offsrping of a single G1 transformant.

[1]The 3xP3 promoter is an artificial promoter element that contains three tandem copies of the 'P3' sequence: an artificial ideal binding site for paired-type homeodomain dimers, fused upstream of a basal promoter element. It drives expression in drosophilid photoreceptors and is also functional in a number of other insect species, including the *An. gambiae* mosquito, where it drives expression in the optic lobe of the head and dorsal ganglia that run the length of the larval body.

3.3 Generation and screening for Cas9-induced variation to select for functional resistance

3.3.1 An assay designed to enrich for functional resistance (R1) at 7280

Male and female heterozygotes (F0) for a *zpg-7280* gene drive strain described in Hammond et al. 2021a were crossed to each other, allowed to lay *en masse*, and their progeny screened for drive heterozygosity (F1) on the basis of intermediate levels of DsRed expression using complex object parametric analyser and sorter (COPAS) as in Marois et al. (2012) (by comparison homozygotes show higher levels of DsRed fluorescence). A total of 246 COPAS-selected heterozygote females were crossed to an equal number of wild type males in four replicate cages (with 36, 70, 64 and 70 individuals in each replicate), allowed to lay, and their F2 progeny were collected at L1. Genomic DNA (gDNA) was then mass-extracted from entire samples of each replicate at F1 and F2 using the Wizard Genomic DNA purification kit (Promega) and used for pooled amplicon sequencing analysis as described in Hammond et al. 2021a. Mutant alleles were included in the analysis if present at over 1% frequency amongst non-reference (wild-type) alleles in any replicate of the F1 generation.

3.3.2 Generation of an autosomal linked editor (ALE) strain

To generate an ALE strain targeting the same site on *dsx* as Ag(QFS)1, G3 embryos were microinjected with the previously described p17410 plasmid Kyrou et al. (2018), together with a piggyBac transposase helper plasmid (see 3.2.2). The resulting strain was marked by *3xP3::DsRed* and expressed a *U6::gRNA* complementary to *dsx* T1 along with a *zpg::Cas9* that was insufficient in producing high levels of *dsx* mutagenesis.

3.3.3 High-throughput mutagenesis screen at *dsx*

To induce *dsx* T1 mutagenesis, an F1 cross of >100 heterozygous ALE dsRed+ males to >100 heterozygous *vasa2::Cas9* YFP+ females was performed 5 times, and females of each cross were blood-fed 5, 10 and 15 days after being crossed to produce offspring. 4 days PBM 4,000-12,000 L1 offspring of the F1 cross were screened, and their DsRed+YFP- subsection was selected using COPAS (Marois et al. 2012). 50 DsRed+YFP- males were crossed to 50 GFP+ null heterozygous females (*dsxF-*, described in Kyrou et al. (2018)) in an F2 cross. The F2 cross was performed 33 times to balance Cas9-induced mutations across the known null allele in their progeny, and examine whether they can restore *dsx* functions in females. Each F2 cross was blood-fed once, and 4 days PBM 2,000-6,000 L1 offspring were screened and their GFP+ subsection was selected using COPAS. GFP+ mosquitoes were reared to adulthood and 5 days post-

emergence they were offered a blood-meal. Subsequently, they were knocked-out using CO2 to examine their anatomy for signs of being intersex (identified through the presence of semi-plumose antennae and underdeveloped claspers as in Figure 1.16) and manually sorted into 4 different pools per parental cage: males, intersex, blood-fed (BF) females, non-BF females. A pool of 100 intersex mosquitoes was analysed from each parental cage by pooled amplicon sequencing. As controls, pools of wild-type mosquitoes were also sequenced, as well as the progeny of the ALE strain crossed to wild-type and/or the *dsxF-* strain. Also note that 8 F2 single crosses were also performed, and their progeny analysed in the same way using pooled amplicon sequencing (without visual inspection for phenotypic differences), to determine how many mutations does each single mutagenised male transmit to its offsrping.

Mosquitoes that developed normally as females were individually analysed by Sanger sequencing, wild-type controls were also sequenced. Individual females that were shown to contain a SNP by Sanger sequencing were also sent for pooled amplicon sequencing, using wild-type Sanger-sequenced samples as controls. Specifically, F3 GFP+ individuals were batch-collected from large cages containing >500 mosquitoes, and separated into three groups on a CO2 pad: males, anatomical females, and anatomical intersex, for long-term storage (>6-12 months due to Covid-19 interruptions). Prior to gDNA extraction anatomical females were individually separated, however a low level of cross-contamination of the samples from ruminant leg/wing pieces is possible.

3.3.4 Pooled amplicon sequencing

Pools of 100 adult mosquitoes at a time were subjected to gDNA extraction using the Promega Wizard Genomic DNA Purification kit, PCR amplification under non-saturating conditions as Hammond et al. 2021a and Kyrou et al. (2018) (i.e. a primer annealing temperature of 68ºC and allowing the PCR reaction to run for only 20 amplification cycles) using primers Illumina-AmpEZ-4050-F1 (`ACACTCTTTCCCTACACGACGCTCTTCCGATCTACTTATCGGCATCAGTTGCG`) and Illumina-AmpEZ-4050-R1 (`GACTGGAGTTCAGACGTGTGCTCTTCCGATCTGTGAATTCCGTCAGCCAGCA`). Samples were then sequenced using Illumina sequencing (AmpEZ service, Genewiz/Azenta Life Sciences) and results were analysed using CRISPResso2 (Clement et al. 2019) and on Python to correctly assign allele names/identities and set detection thresholds.

3.3.5 Sanger sequencing

Single mosquito samples were subjected to gDNA extraction using the Qiagen DNeasy Blood & Tissue kit, PCR amplification using primers dsx-intron4-F1 (`GTGAATTCCGTCAGCCAGCA`) and dsx-exon5-R4 (`AACTTATCGGCATCAGTTGCG`) using the FastCycling PCR kit with a primer annealing temperature of 58ºC and extension time of 10 seconds, and Sanger sequencing by Eurofins Genomics using primer dsx-exon5-R2 (`TGAATTCGTTTCACCAAACACAC`). Sequence alignments were performed on Benchling.

3.4 CRISPR-mediated cassette exchange (CriMCE) for the generation of SNP variant strains

3.4.1 Molecular cloning of CRISPR plasmids

We used Golden Gate cloning to insert a dual gRNA expression cassette into the p174 master vector from Kyrou et al. (2018), to generate CRISPR vectors p174102 and p17404 needed to catalyze genomic cleavage for the insertion of a placeholder cassette and the variant of interest, respectively. We first amplified a gRNAscaffold-U6terminator-U6promoter sequence, from plasmid p131 using primers containing *Bsa*I sites (capitals), and gRNA sequences (capitals): *Bsa*I-T1-U6-F (gagGGTCTCatgctGTTTAACACAGGTCAAGCGGgttttagagctagaaatagcaagt) and BsaI-T3-U6-R (gagGGTCTCaaaacCTCTGACGGGTGGTATTGCagcagagagcaactccatttcat), to add *dsx* targeting gRNAs onto p174 and BsaI-G1-U6-F (gagGGTCTCatgctGGTTAATTCGAGCTCGCCCGgttttagagctagaaatagcaagt) and BsaI-G2-U6-R (gagGGTCTCcaaaacCAACTAGAATGCAGTGAAACagcagagagcaactccatttcat) to add placeholder targeting gRNAs. The PCR products were inserted into p174, through GoldenGate cloning, to create CRISPR vectors p174102 and p17404, containing a zpg::hCas9, a 3xP3::DsRed::SV40 marker and U6-expressed doublesex-targeting gRNAs (T1 and T3) or placeholder-targeting gRNAs (G1 and G3), respectively.

3.4.2 Molecular cloning of the placeholder donor plasmid

A 3xP3::GFP::SV40 marker cassette was amplified from plasmid pK101 from Kyrou et al. (2018) using primers SgsI-3xP3-F (GGCGCGCCCCACAATGGTTAATTCGAGC) and SgsISV40-R (GGCGCGCCAAGATACATTGATGAGTTTGGAC). Genomic DNA regions 1.8 kb upstream and downstream of the doublesex intron 4-exon 5 splice junction were amplified using primer pairs: 4050-KI-Gib1 (GCTCGAATTAACCATTGTGGACCGGTCTTGTGTTTAGCAGGCAGGGGA) with 4050-KI-Gib31 (TCCAAACTCATCAATGTATCTTGGCGCGCCATAAATGAATGGAAAGGTAAGGC), and 4050-KI-Gib32 (GAGCTCGAATTAACCATTGTGGGGCGCGCCGTATCTTTGTATGTGGGTGTGTG) with 4050-KI-Gib4 (TCCACCTCACCCATGGGACCCACGCGTGGTGCGGGTCACCGAGATGTTC), to make up the right and left homology arms, respectively, of the donor plasmid. To generate the placeholder donor plasmid pHolder-dsx the three PCR products were combined with a digested vector backbone containing a 3xP3::DsRed::SV40 marker cassette in a four-fragment Gibson assembly, so that the *dsx* homology arms flank the GFP placeholder cassette.

3.4.3 Molecular cloning of variant donor plasmids

An intermediate plasmid (pVar-dsx) was Gibson-assembled to contain the same vector backbone and homology arms as pHolder, and a sequence containing *Bsa*I cloning sites, flanking the region of interest

of an otherwise intact exon 5. This allowed the Golden Gate cloning of annealed oligos containing three different doublesex exon 5 variants: a G/A SNP (`GTTTAACACAGGTCAAGCAGTGGT`, chromosome 2, position 47,997,665), a C/T SNP (`GTTTAACACAGGTCAAGTGGTGGT`, chromosome 2, position 47,997,666) and a G/T SNP (`GTTTAACACAGGTCAATCGGTGG`, chromosome 2, position 47,997,667 (Figure 3.1A-C). The same plasmid, pVar-dsx, can be used to clone and study more variants at the same target site in the future.

3.4.4 Embryo microinjections

To generate the placeholder strain, wild-type embryos were microinjected with the p174102 CRISPR plasmid and pHolder donor plasmid (Figure 3.1D). In transformants, this caused the excision of the coding sequence (CDS) of the female-specific exon 5 of the doublesex gene and its replacement with a GFP marker cassette. All microinjection survivors (G0) were crossed to wild-type mosquitoes and positive transformants (G1) were identified through fluorescence microscopy, as GFP+. To generate the SNP variant strains, placeholder homozygote males were crossed to placeholder heterozygote females, distinguished using the COPAS fluorescence-based larval sorter (Marois et al. 2012).

Their progeny was microinjected with the p174104 CRISPR plasmid and each of the variant donor plasmids (pVar-dsxGA, pVar-dsxCT, and pVar-dsxGT) (Figure 3.1E). In successful transformants, this caused the CriMCE of the marked placeholder for the doublesex exon 5 variants. Injected survivors (G0) were distinguished from non-injected survivors (G0), as they exhibited red fluorescence in their posterior, due to successful injection of the p174104 CRISPR plasmid, containing a DsRed cassette in its backbone, which acted as a co-injection marker. All injected survivors (G0) were crossed to wild-type and females were deposited to lay eggs individually. A decreased inheritance of the marked placeholder (GFP+) in G1 progeny indicated CriMCE of the placeholder for the variant sequence.

3.4.5 Variant strain isolation

Progeny (G1) of microinjected individuals heterozygous for the placeholder strain that were GFP- (that could be either carry the SNP in heterozygosis or be wild-type) were crossed to heterozygotes of the placeholder strain that was utilised in this case as a balancer, in single crosses (Figure 3.2 left). After being allowed to lay offspring (G2), molecular genotyping was performed on G1 parents as described in section 3.3.5. Single crosses of the GFP+ progeny (G2) of variant heterozygotes were performed and upon allowing them to lay offsrping (G3), G2 parents were also genotyped (3.3.5). The GFP- (G3) offsrping of SNP heterozygotes would by definition be homozygous for the SNP of interest. These individuals were kept and inbred to secure the strain of interest.

Progeny (G1) of microinjected males homozygous for the placeholder strain that are GFP-, indicative of CriMCE, were crossed to each other en masse, and their progeny (G2) were subsequently also crossed to each other in single crosses (Figure 3.2 right). After G2 parents gave offspring, they were genotyped and the offsrping of homozygous single crosses was kept.

Specifically the former method was applied to isolate the C→T and G→T SNPs, whilst the latter method was applied to isolate the G→A SNP strain in homozygosis. However, note that it was only successful for isolation of G→T and G→A. For the C→T strain homozygous females never gave offsrping when single-crossed to homozygous males, despite being fully fertile otherwise, implying that there might be a cost in homozygous males that prevented them at least from mating in single crosses (where swarming behaviour is absent).

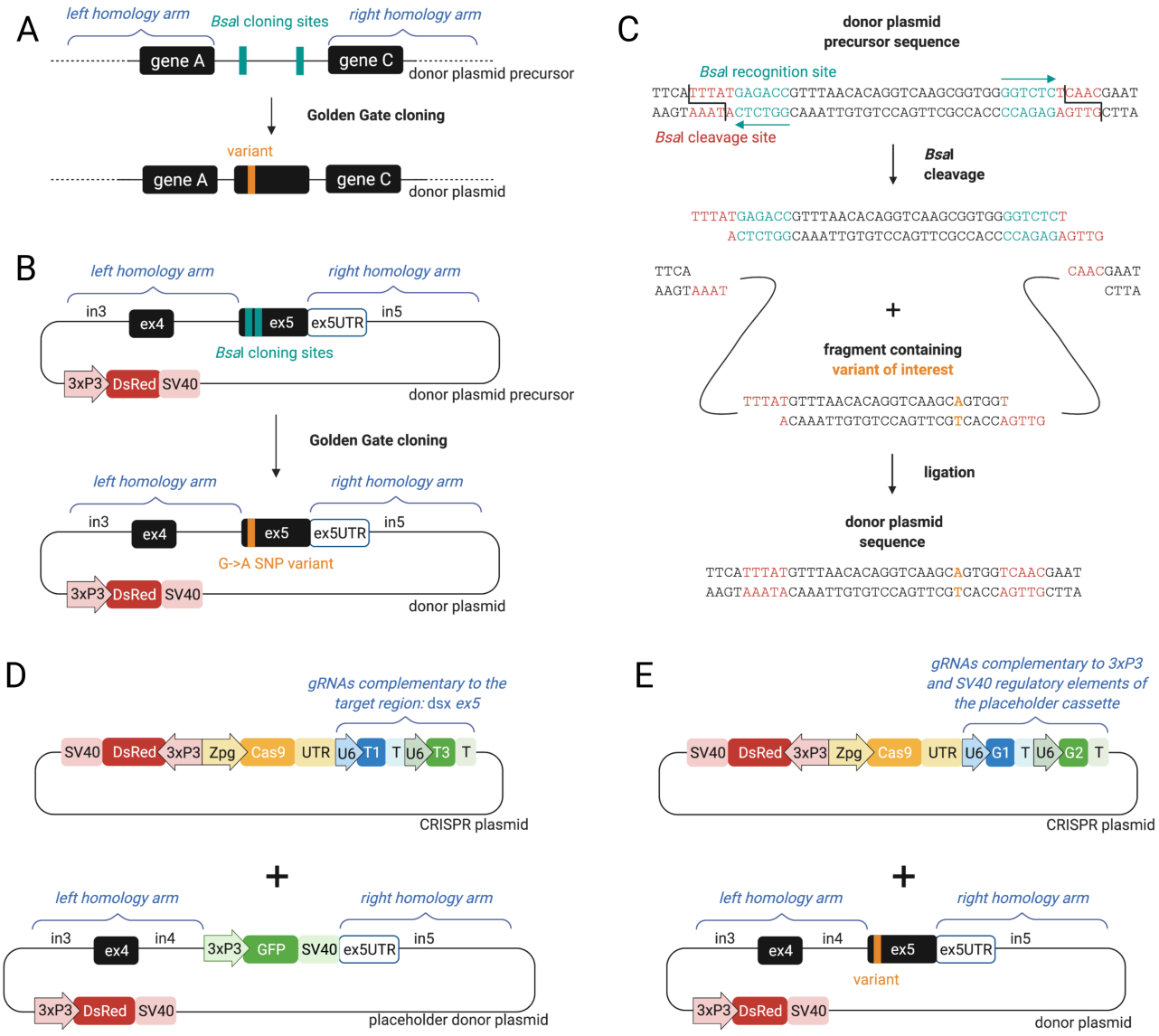


Figure 3.1: Generation of the variant donor plasmids through Golden Gate cloning and graphical representation of microinjected plasmids. (A) The donor plasmid precursor contains *Bsa*I sites (turquoise) flanked by homology arms complementary to the regions upstream and downstream of the target locus. Through Golden Gate cloning the variants of interest (orange) were inserted between the *Bsa*I cloning sites, removing them in the process leaving no molecular trace behind. **B-C** To create a variant donor plasmid (using the G→A SNP as an example), *Bsa*I recognition sites were introduced upstream and downstream of the 23 bp locus of interest on exon 5. *Bsa*I recognition and cleavage sites are distinct, therefore they were placed facing outwards in the donor precursor sequence, to ensure that upon cleavage they get lost, exposing staggered DNA ends on the plasmid precursor. A fragment containing the variant of interest and complementary staggered DNA ends was then ligated onto the plasmid precursor to make-up the final variant donor plasmid. **(D)** Co-injected plasmids used to generate the marked placeholder strain. **(E)** Co-injected plasmids used to generate the variant strains. **(E)** Co-injected plasmids used to generate the variant strains. **(D-E)** CRISPR plasmids, marked by DsRed (top), express Cas9 under the control of the germline-specific *zpg* regulatory elements, along with two gRNAs under the control of ubiquitous polIII promoters (U6): targeted to the *dsx* exon 5 (D, T1 and T3) or the placeholder cassette (E, G1 and G2). Donor plasmids for HDR (bottom), were designed to contain either the placeholder GFP cassette (D, green) or the *dsx* exon 5 bearing the variant of interest (E, orange), flanked by 1.8 kb homology arms complementary to the target region in *doublesex*. Note that donor plasmid backbones are marked by DsRed.

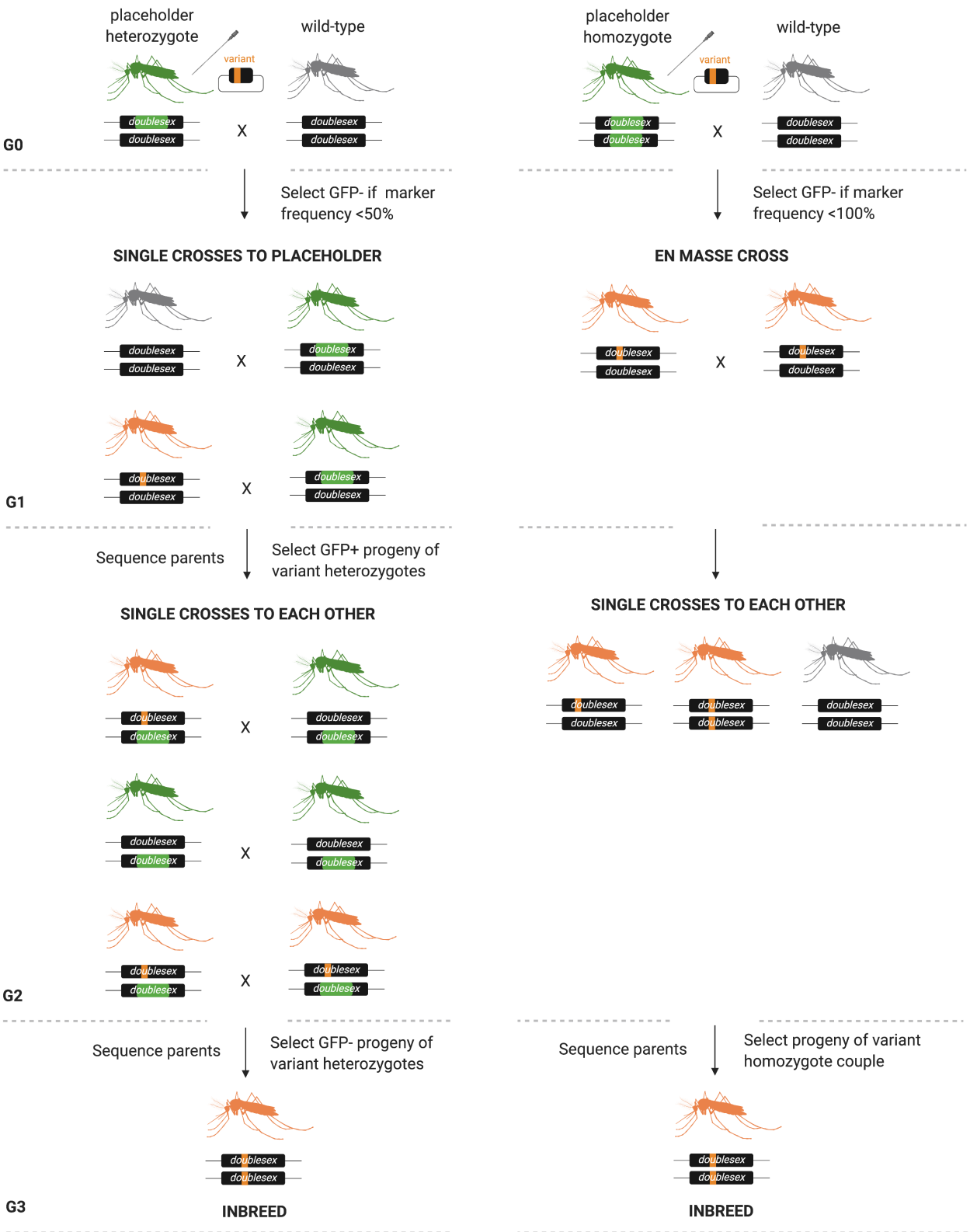


Figure 3.2: Isolation of SNP homozygous strains. A schematic of how homozygous SNP strains were obtained starting from injections of placeholder heterozygotes (left) or homozygotes (right). The method on the left was applied for isolation of the G→T SNP homozygous strain and the method on the right was applied for isolation of the G→A SNP homozygous strain.

3.5 Determining the functionality and gene drive cleavability of *dsx* variants

3.5.1 Assessing the fertility of SNP females

Female fertility of SNP homozygotes and heterozygotes was determined by double-blinded assays, whereby mixed populations of wild-type and SNP carrier females were crossed to wild-type males and allowed to lay eggs individually 2-3 days post-bloodmeal. Eggs and larvae were counted no more than 1 day after they were laid or hatched respectively. The genotype of each mother (homozygous or heterozygous for the SNP or wild-type) was determined through Sanger sequencing, as in section 3.3.5. Note that all females who gave zero eggs or larval offspring, including those that died post-bloodmeal, after being deposited to lay eggs, and those that were putatively unmated have been included in the analysis.

3.5.2 Assessing SNP cleavability by the gene drive

To determine the extent to which each SNP was cleavable by the gene drive, SNP carrier females were crossed to dsRed+ Ag(QFS)1 gene drive males and their trans-heterozygote Ag(QFS)1/SNP progeny, as well as Ag(QFS)1/wt, of both sexes, were crossed to wild-type. Females were allowed to lay eggs individually 2-3 days post-bloodmeal. The rate at which Ag(QFS)1 was inherited amongst their progeny was determined through fluorescence microscopy, since the Ag(QFS)1 trasngene is marked by DsRed.

3.5.3 Assessing the fertility of Ag(QFS)1/SNP females

To determine the fertility of Ag(QFS)1/SNP trans-heterozygotes, SNP carrier females were crossed to dsRed+ Ag(QFS)1 gene drive males and their trans-heterozygote Ag(QFS)1/SNP progeny, as well as Ag(QFS)1/wt, of both sexes, were crossed to wild-type. At the same time wild-type females were crossed to wild-type males as a control. Females were allowed to lay eggs individually 2-3 days post-bloodmeal. The egg and larval output of Ag(QFS)1/C→T, Ag(QFS)1/G→T and Ag(QFS)1/wt females was also recorded and compared to that of the wild-type control.

3.6 Resistance modelling

3.6.1 Simulations of gene drive spread

Discrete generation models of gene drive spread in the presence of various types of resistance were built by Dr Bhavin Khatri using R, after discussions with Prof Austin Burt, Dr Andrew Hammond and myself. Models were based on the single gene drive target model described in Khatri & Burt (2022), which consideres the rate of EJ mutations during gene drive activity, however this model only considers functional (R1) and non-functional resistant mutations (R2). It was therefore expanded to include functional partially resistant EJ outcomes (R3), by following the same logic (Figure 3.3).

Non-spatial Wright–Fisher stochastic simulations of drive were used with separate sexes throughout, but with coupling to population dynamics using density-dependent Beverton–Holt growth. The simulations are stochastic but were ran assuming a very large population size (10^{12}) to approximate a deterministic outcome. In these, there are five alleles that can occur: W (wild-type), D (drive), R1 (fully functional, complete resistance), R2 (non-functional, complete resistance) or R3 (fully functional, incomplete resistance - meaning that gene drive homing is still permitted to some extent (Figure 3.3). Fertility-weighted average homing rates were incorporated in the model to take into account the difference in both homing rate and fertility of the different parental classes of gene drive (e.g. M→M, M→F, F→F, and F→M). Specifically, the fertility-weighted homing rates used in the model were: 0.93 for D/W genotypes, calculated using the Ag(QFS)1 data in sections 4.4.3 and 4.4.4; and 0.42 for D/R3 genotypes, calculated using the Ag(QFS)1/R3 data in sections 4.3.3 and 4.3.3.

These alleles are assumed to have no effect on male fitness, while fitness effects in females are as follows: W alleles, as well as R1 and R3 were assumed to have zero fitness cost (fitness = 1), D and R2 alleles are deleterious in homozygosis (fitness = 0). In heterozygosis D alleles were also assumed to have a fitness cost that varied depending on whether they were coupled to W (cost of 20% as well as 80% relative to wild-type was modelled, i.e. 0.8 or 0.2 relative fertility to wild-type, respectively), R1 (no cost), R2 (no cost), or R3 alleles (intermediate cost of 26%, i.e. 0.74 relative fertility to wild-type).

We modelled homing gene drive spread in the presence of no resistance (N), non-functional resistance only (R2 only), partial resistance (R2+R3), and all types of resistance, including non-functional, full and partial functional resistance co-existing (R1+R2+R3). We used the rates of creation of R1 and R3 mutations that were calculated using the mutagenesis screen described in section 4.2, and combined them with rates of error-prone EJ in the mosquito germline determined previously by (Kyrou et al. 2018) and (Hammond et al. 2016) (see parameters table 3.1 for specific values).

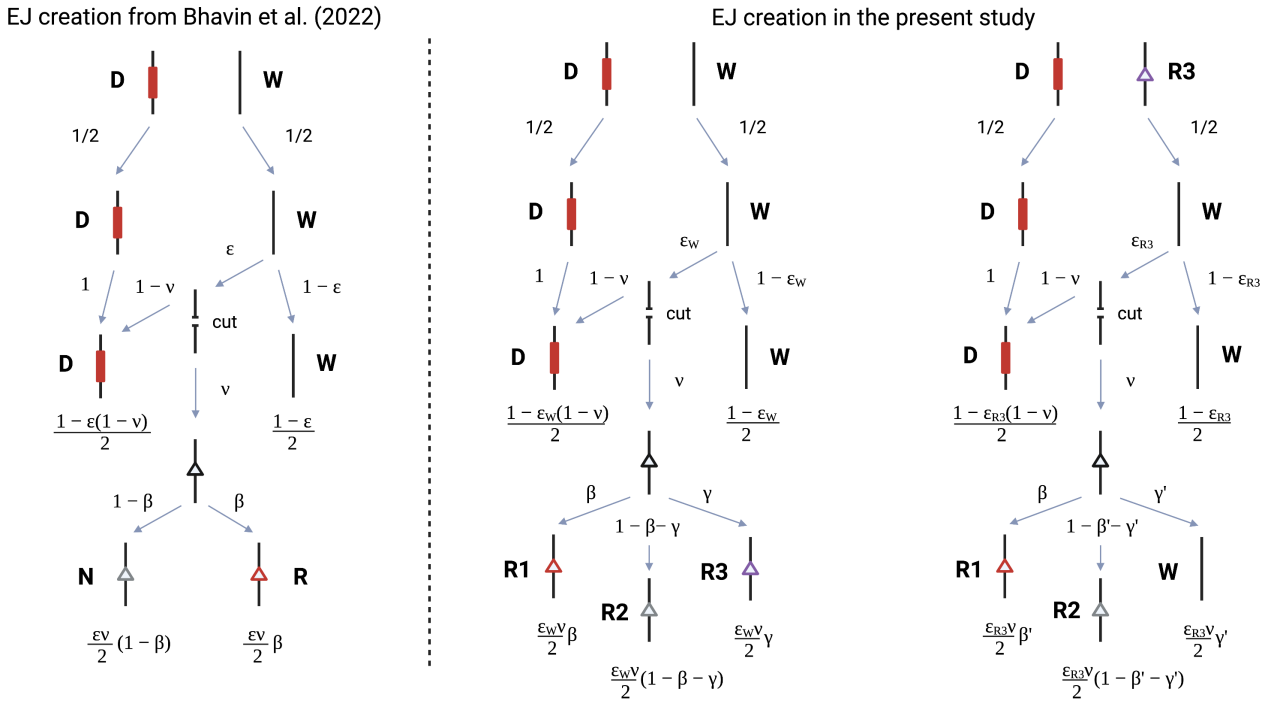


Figure 3.3: Diagram of resistant allele creation during gametogenesis, and how that is incorporated into the model. Abbreviations: EJ = end-joining, D = drive allele, W = wild-type allele, N = non-functional EJ mutation, R = functional resistant mutation, R1 = functional resistant mutation, R2 = non-functional resistant mutation, R3 = functional partially resistant mutation. Model parameters: ε = cleavage efficiency, ε_W = cleavage efficiency of wild-type, ε_R3 = cleavage efficiency of R3, β = fraction of R or R1, γ = fraction of R3, β' = fraction of R1, γ' = fraction of wild-type.

Table 3.1: Modelling parameter values.

Parameters	Description	Values
ε_W	Cleavage rate of WT that leads to non-WT allele	0.966
$1-\varepsilon_W$	Frequency of unmodified WT allele after exposure to nuclease	0.034
ν	Frequency of WT conversion to EJ mutant allele	0.035
$1-\nu$	Fraction of drive alleles amongst cleaved (that do not become WT)	0.965
β	Fraction of EJ mutations that become an R1 allele	0.001
$1-\beta-\gamma$	Fraction of EJ mutations that become an R2 allele	0.996
γ	Fraction of EJ mutations that become an R3 allele	0.003
-	Fraction of non-drive, cleaved WT that are converted to EJ	0.500
$\varepsilon_R 3$	Cleavage rate of R3	0.428
$1-\varepsilon_R 3$	Frequency of unmodified R3 allele after exposure to nuclease	0.572
ν'	Frequency of R3 conversion to EJ mutant allele	0.027
$1-\nu$'	Fraction of drive alleles amongst cleaved (that do not become R3)	0.973
β'	Fraction of EJ mutations that become an R1 allele	0.001
$1-\beta$'$-\gamma$'	Fraction of EJ mutations that become an R2 allele	0.999
γ'	Fraction of EJ mutations that become a WT allele	0.000
-	Fraction of non-drive, cleaved R3 that are converted to EJ	0.020

3.7 Testing of multiplexed gene drive systems

3.7.1 Molecular cloning of CRISPR plasmids with multiplexed gRNAs

To create the p174102 vector required for generation of the Ag(Dsx) docking line and Ag(QFS)2 gene drive strain, primers containing *Bsa*I sites (capitals) and gRNA sequences (capitals): BsaI-T1-U6-F (gagG GTCTCatgctGTTTAACACAGGTCAAGCGGgttttagagctagaaatagcaagt) and BsaI-T3-U6-R (gagGGTCT CaaaacCTCTGACGGGTGGTATTGCagcagagagcaactccatttcat), were used to amplify a gRNAscaffold-U6terminator-U6promoter sequence from plasmid p131. The PCR product was inserted into the p174 master vector (Kyrou et al. 2018), through Golden Gate cloning to create p174102, containing zpg::hCas9, a 3xP3::DsRed::SV40 marker and U6-expressed gRNAs targeting sites T1 and T3 on doublesex exon 5. To create the p174103 vector for generation of the Ag(QFS)3 strain, oligonucleotide sequences dsx-T2-F (TGCTGTCTGAACATGCTTTGATGGCG) and dsx-T2-R (AAACCGCCATCAAACATGTTCAGAC) were annealed and cloned into the p139 vector containing a gRNAscaffold-U6terminator-U6promoter-*Bsa*I spacer cloning site-gRNAscaffold-U6terminator-U6promoter sequence, through GoldenGate cloning. The resulting plasmid was amplified using primers *Bsa*I-T1-U6-F and *Bsa*I-T3-U6-R, like before, to enable GoldenGate cloning of the PCR product into the p174 plasmid to create the p174103 plasmid, containing *zpg::hCas9*, a *3xP3::DsRed::SV40* marker and U6-expressed gRNAs targeting sites T1, T2 and T3 on doublesex exon 5.

3.7.2 Molecular cloning of the donor plasmid required to facilitate RMCE of multiplexed gene drives

The pDLv3 donor vector, used for generation of the Ag(Dsx) strain, was created by Gibson assembly. A *3xP3::GFP::SV40* marker cassette flanked by attP sites was amplified using primers 4050-KI-Gib34 (ATGTT TAACACAGGTCAAGACCGGTACCCCAATCGTTCA) and 4050-KI-Gib35 (GCGGAAAGTTTATCATCCACTCACGCG TTCAGGATTATATCT). Genomic regions 1.8 kb upstream and downstream of the dsx intron4-exon5 splice junction were amplified using primer pairs 4050-KI-Gib1 (GCTCGAATTAACCATTGTGGACCGGTCTTGTGT TTAGCAGGCAGGGGA) with 4050-KI-Gib33 (TGAACGATTGGGGTACCGGTCTTGACCTGTGTTAAACATAAATG), and 4050-KI-Gib36 (AGATATAATCCTGAACGCGTGAGTGGATGATAAACTTTCCGCAC) with 4050-KI-Gib4 (T CCACCTCACCCATGGGACCCACGCGTGGTGCGGGTCACCGAGATGTTC), to make up the left and right homology arms, respectively, of the donor plasmid. The pK101 plasmid from (Kyrou et al. 2018) was digested with restriction enzymes *Mlu*I and *Bsh*TI to release its 5.4 kb backbone containing a *3xP3::DsRed::SV40* marker, which was combined with the three amplicons in a four-fragment Gibson assembly, so that the *dsx* homology arms were placed to each side the attP-flanked GFP cassette, to produce pDLv3.

3.7.3 Generation of the Ag(Dsx) docking line

To generate the Ag(Dsx) docking line to accommodate RMCE of the gene drive constructs, wild-type embryos of the G3 strain were microinjected with the p174102 CRISPR plasmid and the pDLv3 donor plasmid. In transformants, CRISPR elements expressed by p17402 caused excision of the region between the T1 and T3 cut sites on *dsx* exon 5, and its replacement with the attP-flanked GFP marker cassette from the pDLv3 plasmid, facilitated by HDR. All microinjection survivors (G0) were outcrossed to wild-type mosquitoes and positive transformants were identified through fluorescence microscopy as GFP+ and RFP-.

3.7.4 Generation of the Ag(QFS)2 and Ag(QFS)3 multiplexed gene drive strains

To generate the Ag(QFS)2 and Ag(QFS)3 multiplexed gene drive strains, heterozygotes of the Ag(Dsx) strain were crossed to each other, and their progeny were microinjected with the p174102 or p174103 CRISPR plasmids, respectively, together with a vasa-integrase helper plasmid (Volohonsky et al. 2015), to facilitate RMCE of the GFP cassette of Ag(Dsx) for each of the gene drive constructs. All microinjection survivors (G0) were outcrossed to wild-type mosquitoes and positive transformants were identified through fluorescence microscopy as dsRed+ and GFP-. Construct orientation was determined via PCR using primer pairs zpg-term-F1 (`GCTGTACTACATCTCGTGGACG`) with dsx-exon5-R2, which would only produce an amplicon if the gene drive construct is integrated the same orientation as the *doublesex* gene (fwd), and zpg-term-F1 with dsx-intron4-F1, which would only produce an amplicon if the gene drive construct is integrated in reverse (rev) with respect to *doublesex*.

3.7.5 Phenotype assessment of multiplexed gene drive strains

Gene drive homing of strains Ag(QFS)2 and Ag(QFS)3 was assessed by crossing heterozygous dsRed+ gene drive parents of both sexes that inherited the drive from either males or females to wild-type. Ag(QFS)2 male individuals that inherited the gene drive from either males or females, were also crossed to heterozygous females of a GFP+ *dsxF-* strain (Kyrou et al. 2018), where a GFP cassette interrupts target site T1 but leaves T3 exposed to cleavage. Trans-heterozygous GFP+ dsRed+ offspring of both sexes were crossed to wild-type to determine rates of gene drive inheritance in the offspring. After allowing females to lay eggs individually, 2-3 days post-bloodmeal, the rate at which each gene drive (DsRed+) was inherited amongst their progeny was determined through fluorescence microscopy.

The extent to which gRNA dissociation takes place in Ag(QFS)2 and Ag(QFS)3 strains was determined by performing gDNA extractions on single mosquito samples of long-standing lab colonies, over 3 consecutive generations (G1 N=27, G2 N=96 and G3 N=96) for Ag(QFS)2 (2 year old strain) and a single

generation (N=43) for Ag(QFS)3 (1 year old strain), and PCR amplification using primers: zpg-term-F1 and dsx-exon4-F6 (`GCACACCAGCGGATCGACGAAG`) with an expected amplicon of 1,341 bp and 1,588 bp for Ag(QFS)2 and Ag(QFS)3 respectively, without instability. An amplicon of 1,094 bp indicated loss of a single or two gRNAs, whilst an amplicon of 1,835 indicated duplication of a single gRNA. The zpg-term-F1 primer was used for Sanger-sequencing to determine the nature of each event.

The fertility of Ag(QFS)1, Ag(QFS)2 and wild-type females was also compared through fertility assays performed simultaneously to allow direct comparison. These looked at egg and larval output, mating ability and post-bloodmeal mortality. Females that inherited each drive from either males or females, as well as wild-type females, were crossed to wild-type males, and allowed to lay eggs individually 2-3 days post-bloodmeal. Their egg and larval output were counted no longer than 1 day post laying and hatching, respectively. The number of females that had died in any of the 3-4 days post-bloodmeal were recorded. Females were also interrogated for their mating ability by examining their spermathecae for sperm presence under an EVOS high resolution light microscope. Note that Kyrou et al. (2018) had previously performed phenotypic assays of Ag(QFS)1, where unmated females and females that did not bloodfeed were excluded from the analysis. In the present study all females that were examined and did not lay eggs were included in the analysis to estimate a more accurate measurement of female fertility, since blood-feeding and mating ability might be affected by the *dsxF* knockout.

Finally, the number of females showing a mosaic intersex phenoype amongst progeny of males of different gene drive strains that were integrated in forward (fwd) or reverse (rev) orientation with respect to the *dsx* gene, were counted. This included Ag(QFS)1 (fwd orientation), Ag(QFS)2 (rev orientation) and five independently created Ag(QFS)3 strains of either orientation.

3.8 Population invasion experiments in small cages

3.8.1 Ag(QFS)2 cage trials

Population invasion experiments were performed in duplicate with each G0 population consisting of 300 wild-type females, 150 wild-type males and 150 Ag(QFS)2 DsRed+ heterozygous males that inherited the gene drive paternally, as in (Kyrou et al. 2018). The initial caged release was performed using age-matched pupae. Mosquitoes were left to mate for 5 days post-emergence before being blood-fed on cow-blood. Two days later they were allowed to lay eggs. 650 eggs were randomly selected to seed the next generation, whilst the rest were photographed and counted using JMicroVision. Emerged larvae were screened for DsRed using fluorescence microscopy to determine drive presence in the population. After pupation the number of normally developed females was recorded to give an indication of the reproductive capacity of the cage, before allowing all pupae indiscriminately to seed the next generation. Note that sterile intersex mosquitoes are indistinguishable from males at the pupal stage.

3.9 Data analysis tools

3.9.1 Figures & data visualisation

Graphs were plotted on Graphpad Prism 9, and figures were designed on Biorender (full licence), Adobe Illustrator and PowerPoint.

3.9.2 Programming

Google Colaboratory was used to access and analyse the Ag1000 genomic datasets. CRISPResso2 was used to analyse pooled amplicon sequencing data (Clement et al. 2019). Visual Studio Code (Python) was used for handling and analysing sequencing data.

3.9.3 Statistics

All statistical comparisons were performed using Graphpad Prism 9.

3.9.4 Other

Eggs (from the cage trials) were counted using JMicroVision.

Chapter 4

Results

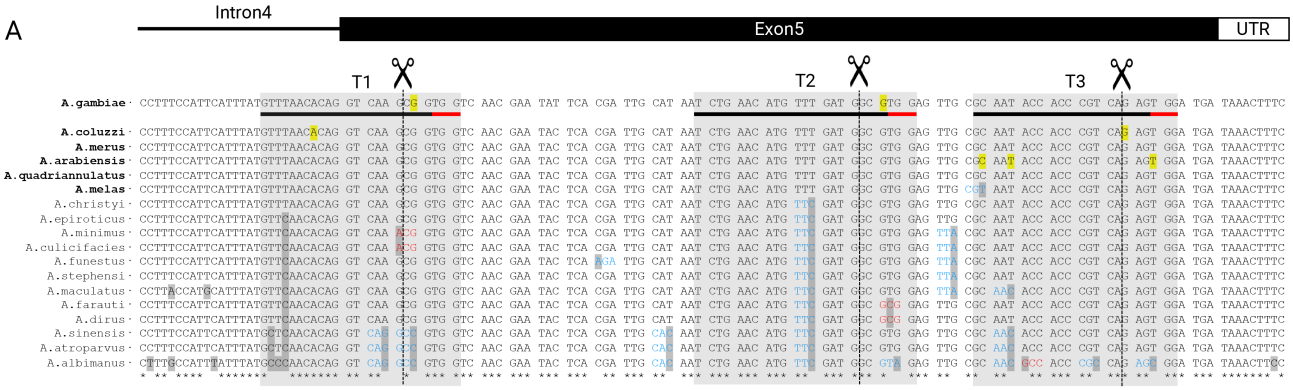
4.1 Evaluating the landscape of naturally-occurring resistance against gene drive at *dsx*

4.1.1 Background

The Ag(QFS)1 gene drive was successful in eliminating laboratory populations primarily due to the high functional constraint of the T1 target site, as indicated by the high sequence conservation of that site amongst anopheline species (Figure 1.15). As a consequence, none of the EJ mutations generated at the gene drive target site led to a functional gene product, which is required to constitute **R1** resistance (see section 1.6.1).

Despite this, a G→A SNP (2R:48714641) was found at T1 at a 2.9% frequency amongst 765 wild-caught *An. gambiae* s.l. mosquitoes (Miles et al. 2017, Kyrou et al. 2018) (Figure 1.15). Though not widespread in Africa, the same SNP was found at an appreciable 25% frequency in Angolan populations, and at lower frequencies in Gabon and Cameroon, raising the possibility that it might be functional and resistant to gene drive provided that it blocks Cas9 cleavage. If so, it would constitute an **R1** allele that is already well-distributed in certain natural populations, and it could prevent gene drive invasion in those populations in the future. If SNP carriers were to migrate beyond their geographical region they could also prevent or reverse gene drive spread in other regions, or repopulate niches in which the malaria mosquito has previously been suppressed due to gene drive activity. It is therefore critical to: (a) gain knowledge of all natural polymorphisms that exist on a given target site prior to release of a gene drive, (b) assess whether these are functional in the lab, and (c) determine if the gene drive can cleave them or not.

Since the last analysis of wild-caught mosquito genomes for the presence of variants at T1, the sample of available genomes on the Ag1000G database has increased reaching 2,784. Thus, I surveyed the latest mosquito population genomics data (The Anopheles gambiae 1000 Genomes Consortium 2021), for the presence of natural variants at the Ag(QFS)1 target site (T1), extending this search to include 2 novel gene drive target sites (T2 and T3) on exon 5 that were later used in a multiplexed gene drive strategy (for methods see 3.1.1) (Figure 4.1A). Specifically, T1 and T3 are the target sites of a novel multiplexed gene drive: Ag(QFS)2, whilst another multiplexed gene drive: Ag(QFS)3 targets T1, T2 and T3, both developed as part of this PhD (see section 4.4).



B

SNP*	Target site	Position	Amino acid change	*An. gambiae* (n)	*An. coluzzii* (n)	*An. arabiensis* (n)	Countries	Maximum Allelic Frequency	Total Allelic Frequency	HET	HOM
A→G	T1	8	N/A (intronic)	1	0	0	Burkina Faso	0.17%	0.02%	100%	0%
G→A	T1	19	silent (Ala)	28	36	0	Angola, DRC, Gabon, Cameroon	25.9%	1.20%	91%	9%
G→A	T2	20	non-silent (Val→Met)	1	0	0	Cameroon	0.11%	0.02%	100%	0%
C→T	T3	2	silent (Arg)	0	0	1	Cameroon	0.11%	0.02%	100%	0%
T→C	T3	5	silent (Asn)	0	0	1	Tanzania	0.17%	0.02%	100%	0%
G→A	T3	17	silent (Gln)	0	1	0	Mali	0.22%	0.02%	100%	0%
T→C	T3	21	silent (Ser)	0	0	2	Tanzania	0.33%	0.04%	100%	0%

*None of the above SNPs are found in synteny.

Figure 4.1: Natural inter- and intra-specific variation on the coding sequence of the female-specific exon of *dsx*. (A) To interpret the multiple sequence alignment refer to figure 1.15. Three QFS gene drive target sites have been highlighted in grey: T1, T2 and T3. Bases that may vary within the same sensu stricto (s.s.) species are highlighted in yellow and specific substitutions present at those variable sites are analysed in table B, below. **(B)** Natural SNP variants that were identified amongst 2,784 wild-caught mosquitoes of the *Anopheles gambiae*, *coluzzii* and *arabiensis* species. SNPs that led to no amino acid substitution are labelled as 'silent'. Maximum allelic frequency refers to the frequency of each SNP in the country in which it was most frequent. The percentage of mosquitoes contained each SNP in heterozygosis (HET) or homozygosis (HOM) is also shown.

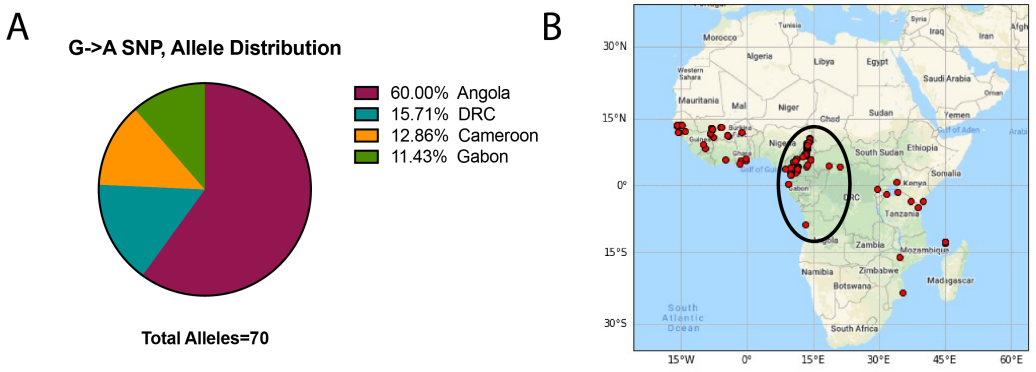


Figure 4.2: The geographical distribution of the G→A T1 SNP (2R:48714641). (A) The frequency of G→A alleles sampled from each country that it was present. **(B)** Map showing the sampling sites in which the G→A T1 SNP was detected (circled).

4.1.2 Natural variation within gene drive target sites on *dsx* exon 5

The Ag1000G data analysis yielded two variable nucleotides in T1: one A→G SNP (2R:48714652) on intron 4, present in a single *An. coluzzii* mosquito sampled from Burkina Faso; as well as the known G→A silent SNP (2R:48714641) (Figure 4.1). The G→A SNP was identified in both *An. gambiae* and *An. coluzzii* mosquito samples, in 28 and 36 mosquitoes respectively; in geographically proximal sites of

Central Africa, in Angola, the Democratic Republic of Congo (DRC) – which was previously unsampled, Gabon and Cameroon (Figure 4.1B, Figure 4.2). Moreover, the G→A SNP was present in homozygosis in 9% of all carrier mosquitoes, indicating that it might be functional (Figure 4.3).

Therefore, the SNP frequencies in the Angolan population were tested for adherence to the expected Hardy-Weinberg (H-W) equilibrium frequencies:

Hardy-Weinberg equation:
$p^2 + 2pq + q^2 = 1$,
where p^2 is the frequency of homozygotes,
$2pq$ is the frequency of heterozygotes,
and q^2 is the frequency of wild-type.

Given the G→A allelic frequency in Angola (25.9%), the expected frequencies of the different allelic classes are:

$p^2 = 0.067$ or 6.7% for homozygotes.
$2pq = 0.384$ or 38.4% for heterozygotes.
$q^2 = 0.549$ or 54.9% for wild-type.

These values do not deviate significantly (Chi-squared test, p = 0.9889) from the observed values of:

$p^2 = 0.074$ or 7.4% for homozygotes.
$2pq = 0.370$ or 37.0% for heterozygotes.
$q^2 = 0.556$ or 55.6% for wild-type.

This suggests absence of a selection pressure for its removal or enrichment in the population, implying that it is equivalent to wild-type in terms of functionality and thus a putative **R1** allele, if able to block gene drive activity.

At the T2 site, only a single SNP was detected. This was a G→A SNP (2R:48714592) proximal to the PAM that caused a Valine-to-Methionine conservative amino acid substitution (Figure 4.1), and was present in a single *An. gambiae* mosquito in Cameroon. It is unclear whether this mutation is functional or not, however it is unlikely that this is a sequencing error as low quality reads are filtered out (The Anopheles gambiae 1000 Genomes Consortium 2021).

At the T3 site, four distinct low frequency SNPs were detected. These are not expected to cause an

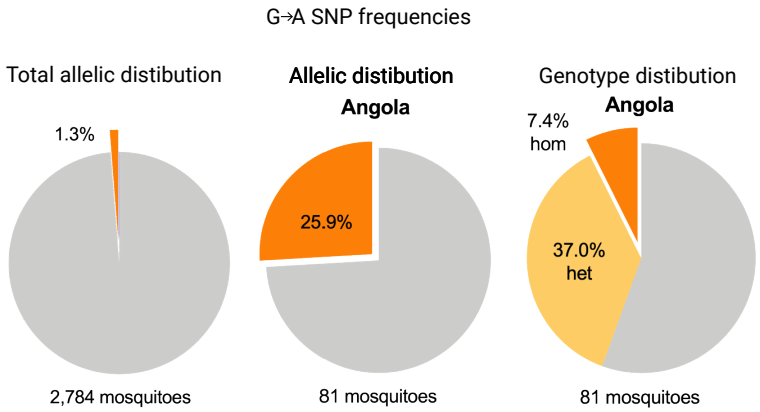


Figure 4.3: Allelic distribution of the G→A T1 SNP (2R:48714641) that was observed at Hardy-Weinberg equilibrium in Angola. Left: its total allelic distribution in Africa (orange = SNP alleles, grey = wild-type alleles). Middle: its allelic distribution in Angola (orange = SNP alleles, grey = wild-type alleles). Right: the percentage of individual mosquitoes that were heterozygote (het, yellow) or homozygote (hom, orange) for the SNP (grey = wild-type individuals).

amino acid substitution and are therfore more likely to be functional than the T2 SNP (Figure 4.1). The first two: a C→T SNP (2R:48714581) present in a single *An. arabiensis* mosquito sampled from Cameroon; and a T→C SNP (2R:48714578) present in a single *An. arabiensis* mosquito sampled from Tanzania, are present on the 3' end of the target site, away from the PAM region. The third SNP identified at T2 is a G→A PAM-proximal substitution (2R:48714566) present in a single *An. coluzzii* mosquito sampled from Mali. Since this substitution is located right at the cut site, it also raises the possibility that it may be generated by gene drive activity through EJ repair. Finally, a T→C SNP (2R:48714563) was identified at T3 in two *An. arabiensis* mosquitoes sampled from Tanzania.

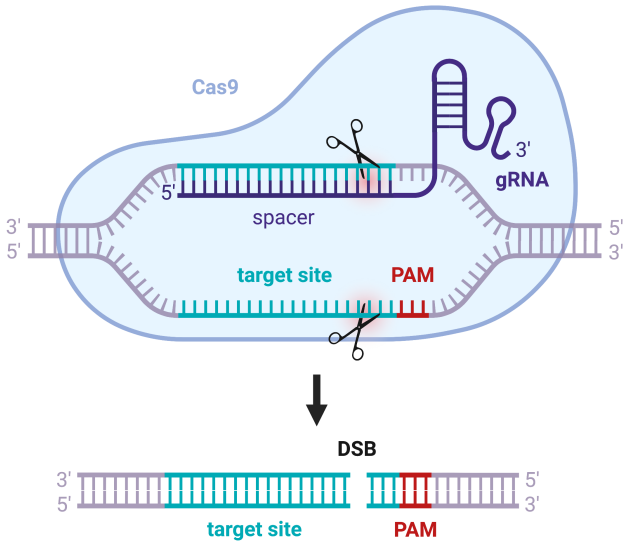
Importantly, none of the natural SNPs reported here were found in synteny. Multiple SNPs occurring on the same haplotype would be a lot more problematic as they could provide resistance to a multiplexed gene drive strategy.

4.1.3 Prediction of cleavage efficiency at variable target sites

Most of the natural variants discovered are assumed to be functional, especially those inducing silent amino acid substitutions (Figure 4.1). For these to constitute fully drive-resistant alleles (either **R1** or **R2**) they would also need to block gene drive activity.

The version of Cas9 employed by the Ag(QFS)1 gene drive[1] is occassionally able to cleave target sites that contain one or more base pair mismatches to the employed gRNA (Zhang et al. 2015). Mismatches are tolerated to various degrees depending on their position along the target site. In general, mismatches closer to the 5' end of the gRNA are better tolerated than those at the 3' end that are proximal to the

[1]Ag(QFS)1 contains a human codon optimised human Cas9 (hCas9) that is the same in its amino acid composition as *Streptococcus pyogenes* Cas9 (SpCas9).



[h!]

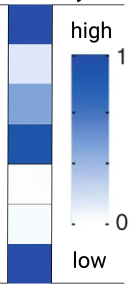
Figure 4.4: Schematic of an RNA-guided Cas9 endonuclease catalysing a DSB. The Cas9 nuclease (light blue) is recruited to genomic DNA (gDNA) by a gRNA (purple) containing a ∼20 bp spacer sequence complementary to a gDNA target site. The target sequence needs to be followed by a short (typically 3-4 bp) PAM for the nuclease to be able to catalyse a DSB. Cas9 typically requires an 'NGG' PAM sequence and cleaves 3 bp upstream the NGG motif, as shown in the schematic.

PAM (Zhang et al. 2015) (Figure 4.4), however within the PAM-proximal region the degree of mismatch tolerance can vary depending on the identity of each particular mismatch (Hsu et al. 2013). Hsu et al. (2013) tested the SpCas9 cleavage efficiency for over 400 mismatched gRNA/target sequences in human cells, and scored the extent of cleavage at each target site position, depending on the nature of base-pair-mismatch. Their findings were used to predict cleavage efficiency by canonical[2] gene drive gRNA/Cas9 complexes (Figure 4.5). Based on Hsu et al. (2013)'s data: PAM-distal SNPs are expected to be cleavable by the canonical gRNA/Cas9 complex, with the exception of one of T→C SNP discovered at position 15 of the T3 target site (Figure 4.5). Conversely, PAM-proximal SNPs show low to intermediate likelihood of being cleaved with the exception of another T→C SNP at T3 that falls on the variable base (N) of the NGG PAM and is therefore predicted to be fully cleaved by a canonical gRNA/Cas9 complex (Figure 4.5).

Since the G→A SNP at T1 was the most abundant variant in the natural population, is expected to be functional, and shows low modification efficiency (Figure 4.5), it was prioritised for testing, despite previously having been cleaved *in vitro* (Kyrou et al. 2018).

[2]Canonical gRNAs would be those that are fully complementary to the wild-type target site.

SNP	Target site	Position away from PAM	Mismatched RNA(r):DNA(d) pair	Modification efficiency
A→G	T1	13	rU:dG	
G→A	T1	2	rC:dA	
G→A	T2	1	rC:dA	
C→T	T3	18	rG:dT	
T→C	T3	15	rA:dC	
G→A	T3	3	rC:dA	
T→C	T3	N/A*	rA:dC	



*located within the variable position (N) of the NGG proto-spacer adjacent motif (PAM).

Figure 4.5: Variant cleavage prediction based on Hsu et al. (2013). Mean cleavage levels are shown for the different natural variants detected. Cleavage depends on both the distance away from the PAM – typically variation at PAM-distal sites is better tolerated, and the nature of the gRNA-target DNA mismatch. Darker blue signifies a greater likelihood of cleavage according to the data of Hsu et al. (2013).

4.2 Evaluating the landscape of drive-induced resistance against the Ag(QFS)1 gene drive

4.2.1 Background

After evaluating the landscape of naturally-occurring resistance against three putative gene drive target sites at *dsx*, I set out to evaluate the landscape of drive-induced resistance against Ag(QFS)1. This is extremely difficult to do in the context of a homing gene drive since most of the drive-induced cuts are repaired by HDR in the germline leading to homing (Figure 1.10), and only a very small fraction of progeny end up carrying putatively resistant EJ mutations.

An assay designed to enrich for functional resistance (R1) at 7280

To enrich for functional resistance (R1) at the 7280 site, we devised a novel assay using automated screening and artificial selection to detect functional (R1) EJ mutations (Hammond et al. 2018). F0 drive heterozygotes (GD/WT) were crossed to each other and the rare fraction of their heterozygote F1 offsrping was selected using COPAS, a high throughout larval sorter that can select larvae based on fluorescence. These individuals (GD/ND) had inherited a single copy of the gene drive (GD) from one parent coupled to a drive-lacking (unhomed) allele that was previously exposed to nuclease activity (ND) and might contain an EJ mutation (R1 or R2) or be unmodified (WT) (Figure 4.6). To enrich for R1 alleles, F1 GD/ND females were crossed to wild-type males. F1 GD/R2 females are predicted to be sterile due to carrying a homozygous disruption at 7280, and therefore only F1 GD/WT and GD/R1 females would give F2 offspring. Thus, by sequencing the F2 offsring as well as their F1 parents it is possible to distinguish R2 from R1 alleles: any EJ mutation detected in the F2 would constitute an R1 allele, whilst mutations that were only present in the F1 but filtered out from the F2 would constitute R2 alleles (Figure 4.6).

This assay is useful in assessing the resistance potential of gene drives prior to cage trial testing, and it was efficient in determining the rates of creation of the various EJ mutations at the 7280 target site. However, when replicated using the *dsx* Ag(QFS)1 gene drive, it was difficult to get a sufficiently high number of F1 individuals to assess, due to the extremely low rates of meiotic EJ of this gene drive (Kyrou et al. 2018). To illustrate, fewer than 200 EJ events (1.75%) would be expected amongst 10,000 progeny of drive heterozygotes. In addition, due to the high functional constraint of the *dsx* target site, very few (if any) of these mutations would be expected to be functional. Due to this, in the context of Ag(QFS)1, it is even more important to examine a high number of distinct mutations.

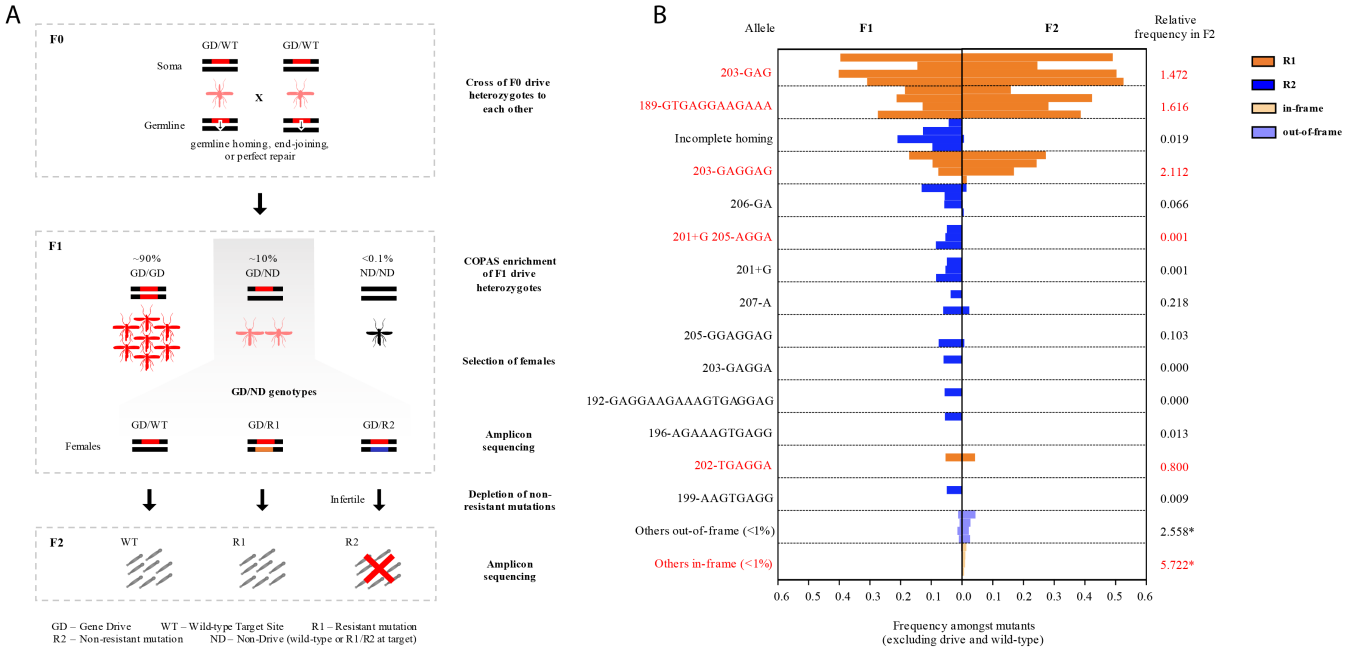


Figure 4.6: An assay to enrich for functional drive-resistant mutations at the 7280 target site. (A) A minimum of 100 males and 100 females heterozygous for the zpg-7280 gene drive developed by Hammond et al. 2021a were crossed to each other and allowed to lay eggs *en masse*. Their F1 progeny that contained a putatively mutated copy of 7280 (unmodified WT, R1 or R2) coupled to the null gene drive allele (GD) were selected on the basis of intermediate levels of DsRed fluorescence that marked the GD allele, using a COPAS larval sorter. F1 females were then crossed to wild-type males, and after giving F2 progeny, both the F1 and F2 were sequenced to reveal the type and frequency of mutations they carried. **(B)** Mutations detected in the F1 but not in the F2 are presumably non-functional R2, whilst those that are enriched in the F2 are functional R1 (see key). The relative change in frequency of each mutation exceeding 1% frequency amongst mutated allele is shown on the right of the graph. The change in frequency in mutations found below 1% is also shown (*).

Testing an autosome-linked editor (ALE)

We hypothesised that repeated cleavage of the T1 target site of Ag(QFS)1 would eventually force the generation of EJ mutations resistant to further cleavage. To test this we generated an ALE strain comprising an out-of-locus[3] *zpg::Cas9* and *U6::gRNA(T1)*, using piggyBac transposition. However, only 1/24 individuals exposed to the ALE carried an EJ mutation. To exclude that this is due to inactivity of the ALE we assessed whether the out-of-locus *zpg*-expressed nuclease can instead cause homing of an in-locus GFP cassette at the *dsx* target site through HDR (Figure 4.7B). Indeed, the ALE was able to induce homing of a GFP cassette at the T1 target site. However, it was not as efficient in doing so as the in-locus Ag(QFS)1 construct (Figure 4.8), as it was only able to produce GFP transmission rates averaging at 79.1% (52.4-97.6%) in males, and 91.7% (70.5-100.0%) in females, compared to the 94.1% (49.6-100.0%, p-value $<$ 0.0001) and 99.3% (83.3-100.0%, p-value $<$ 0.0001) respective transmission rates of the Ag(QFS)1 gene drive.

[3]Integrated at a different genomic locus than its target.

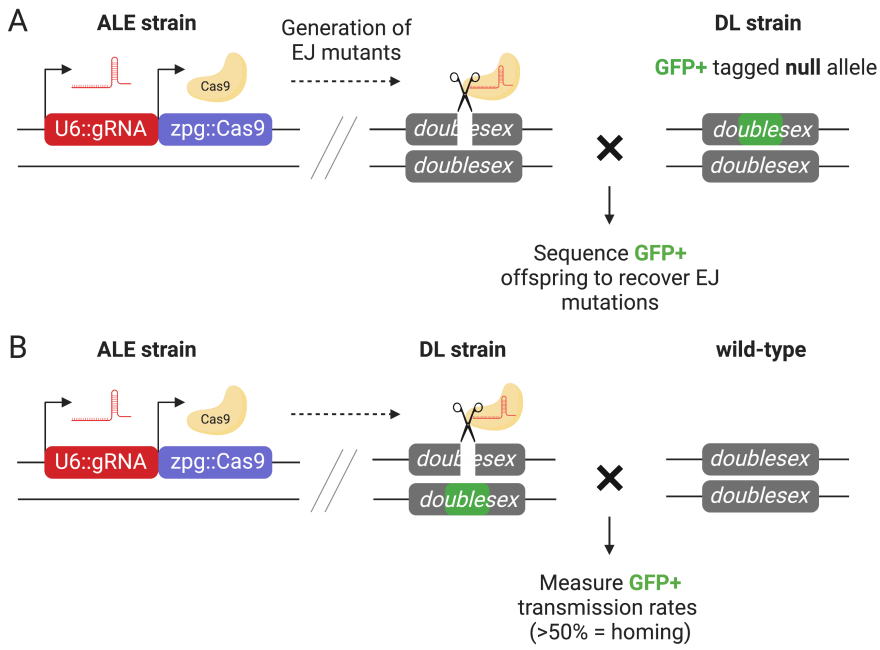


Figure 4.7: Testing an out-of-locus source of Cas9 for in-locus activity. (A) Testing an ALE strain for its ability to generate in-locus EJ mutations. An ALE strain containing a random piggyBac integration of a ubiquitously-expressed gRNA targeting *dsx* and a *zpg::Cas9*, marked by *3xP3::DsRed* was crossed to a docking line (DL) containing a GFP cassette interrupting the T1 target site of Ag(QFS)1. 24 GFP+ progeny of this cross were sequenced to determine whether the ALE induced any EJ mutations at the cut site. **(B)** Testing an ALE strain for its ability to induce in-locus HDR. Transheterozygotes for the ALE and DL constructs were crossed to wild-type to assess whether the ALE can induce homing of the GFP into T1 target site.

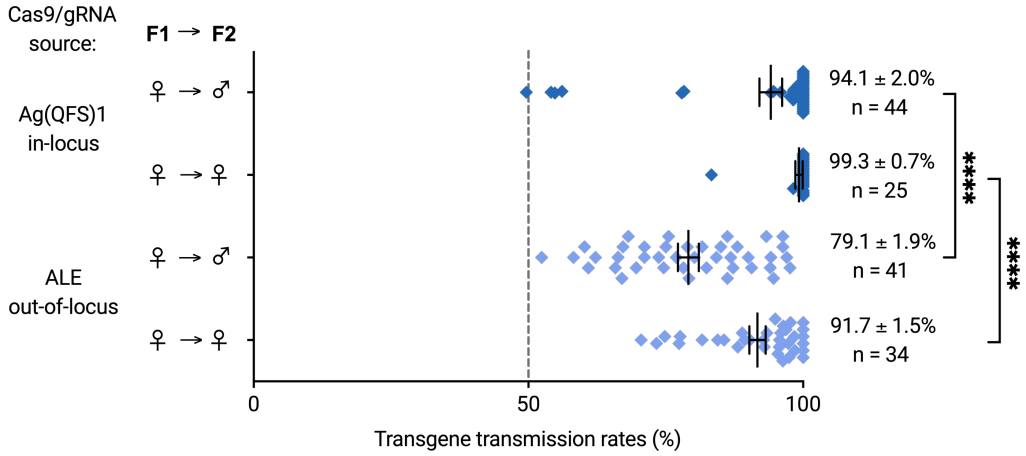


Figure 4.8: The ALE can induce homing of a GFP cassette at *dsx*. Ag(QFS)1 transgene transmission rates are shown in dark blue and GFP transmission rates are shown in light blue. Transmission rates in the progeny of female or male parents (F2) that inherited the Cas9-containing transgenes maternally (F1) are shown. In the latter case the Cas9 and gRNA were provided by an out-of-locus ALE, rather than the in-locus Ag(QFS)1 gene drive. Kruskal-Wallis statistical comparisons are shown to the right (* * * * = p-value < 0.0001).

4.2.2 A high-throughput mutagenesis screen revealed putatively restorative nucleotide substitutions at the Ag(QFS)1 gene drive target site on *dsx* exon 5.

Using a maternal source of embryonically deposited Cas9 can force the generation of EJ mutations.

First generation gene drives built using the *vasa2* promoter caused high levels of EJ due to strong maternal Cas9 deposition into the early embryo that is likely lacking the machinery to perform HDR. Specifically, over 70% of non-drive alleles in the offspring of *vasa2::Cas9* carriers contained an EJ mutation (Hammond et al. 2021a). To generate a high amount of EJ mutations we leveraged maternal deposition of *vasa2*-expressed Cas9 and devised a high-throughput screen to assess them (Figure 4.10A).

First, males of the ALE strain (DsRed+) (containing a gRNA against the T1 site on *dsx*) were crossed to females expressing *vas2::Cas9* (YFP+) (F1 cross), to generate mosaic individuals carrying EJ mutations (Figure 4.10A). Of those we inspected a sample of females that inherited the gRNA (DsRed+) and found that all (100%) exhibited full or partial intersex phenotype (n=117), characterised by the presence of partially or fully male genitalia in the pupal stage, irrespective of carrying the Cas9 transgene (YFP+) (Figure 4.9). These results are consistent with strong maternal deposition of Cas9 and subsequent EJ mutagenesis of the T1 site (Figure 4.9), rendering those individuals effectively homozygous for a *dsx*F null in sufficient part of their soma and germline to be observable as intersex. By contrast, the gRNA showed only modest levels of paternal deposition as only 7% of females that did not inherit the gRNA transgene exhibited an intersex phenotype (n=84) (Figure 4.9).

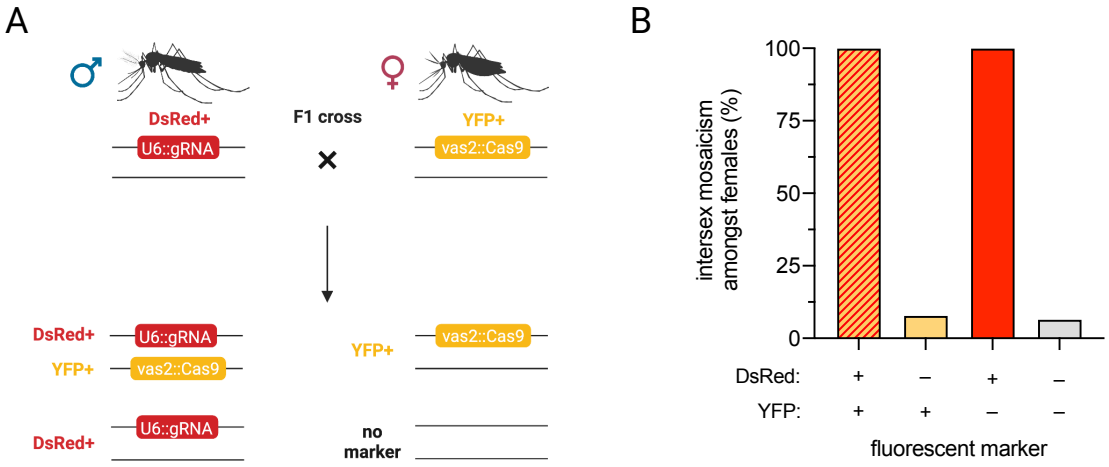


Figure 4.9: Intersex mosaicism in the female offsrping of vasa-Cas9 females and ALE males. (A) *vas2::Cas9* YFP+ females were crossed to ALE DsRed+ males. **(B)** Their female offspring were examined for intersex mosaicism at the pupal and adult stage, depending on their genotype. Abbreviations: YFP = yellow fluorescent protein.

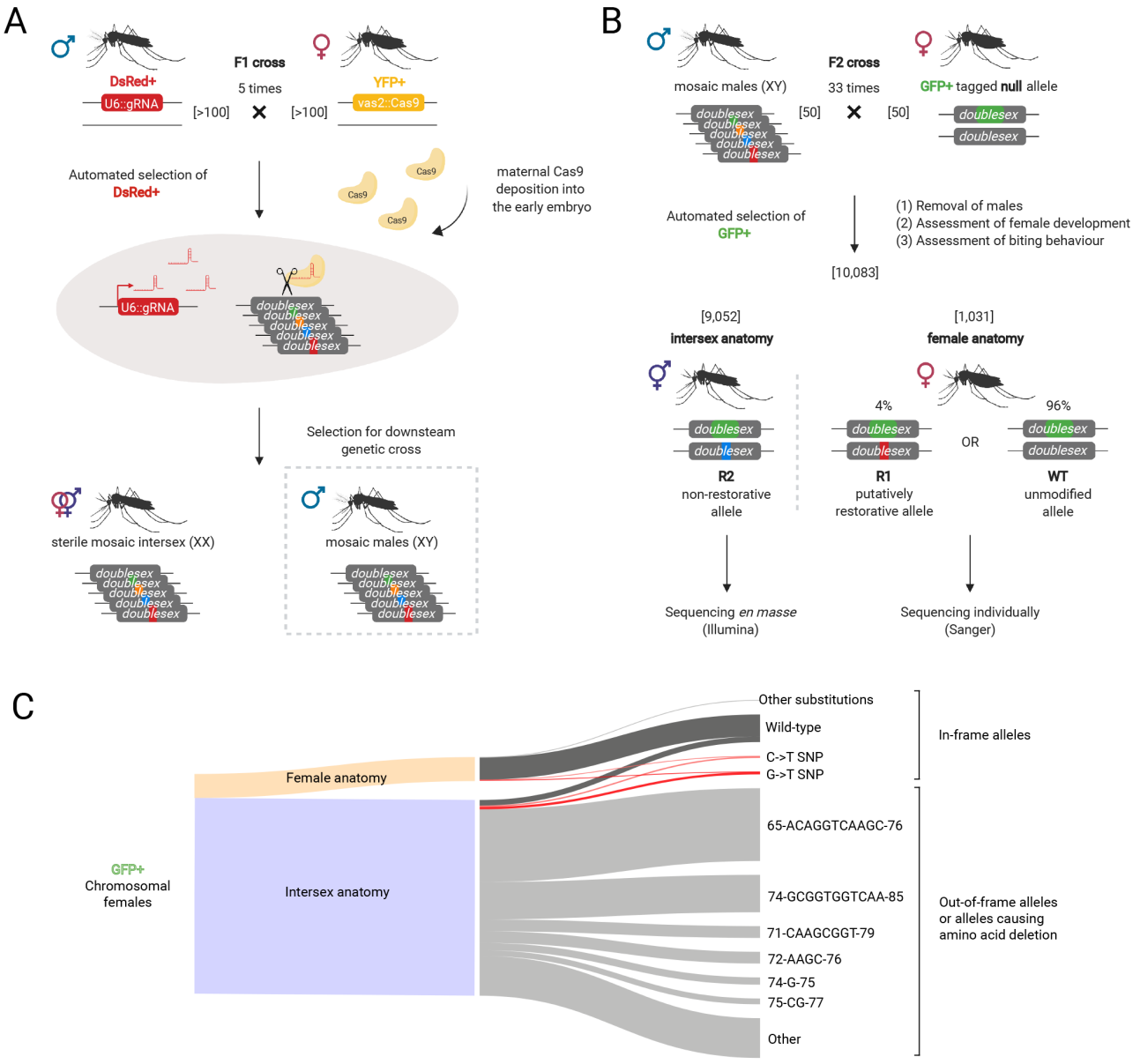


Figure 4.10: A high-throughput mutagenesis screen used to generate and assess Cas9-induced EJ mutations. (A) A minimum of 100 males of the ALE strain (DsRed+) expressing a ubiquitous gRNA targeting the T1 site on *dsx* exon 5, were crossed to a minimum of 100 females expressing *vas2::Cas9* (YFP+). This F1 cross was repeated 5 times. *Vas2*-expressed Cas9 gets maternally deposited into the embryo. Combined with a source of gRNA (DsRed+) it can induce germline and somatic mutations, even in individuals that did not inherit the *vas2::Cas9* (YFP+) transgene. DsRed+YFP- offspring were selected using the COPAS larval sorter. Due to the high mutational load females developed as mosaic intersex individuals that were sterile. **(B)** Mosaic males that are unaffected by *dsx* mutagenesis were selected and crossed to "docking line" (DL) females containing a female-specific null copy of *dsx* (*dsxF-*), due to T1 site disruption by a GFP marker (F2 cross). The GFP+ fraction of the offspring was selected, to balance mutations inherited from males across the *dsxF-* allele. To determine whether mutations can restore *dsx* function we assessed female development and blood-feeding behaviour: females that developed as intersex must contain R2, non-restorative mutations, and those that developed normally, and could blood-feed must contain putatively restorative, R1 mutations, or a wild-type unmodified allele. We analysed anatomically intersex mosquitoes *en masse* through pooled amplicon Illumina sequencing and anatomical females individually by Sanger sequencing to identify the mutated alleles. **(C)** The Sankey diagram shows the relative proportion of GFP+ females that showed a female vs. intersex anatomy, and the relative proportion of each of the most common mutations discovered in them. The figure was generated using Biorender and Sankeymatic.

Deposited Cas9 induced EJ mutations early on during embryonic development.

To characterise the extent of mosaicism, we examined the nature and frequency of mutations generated in the progeny of 8 individual males, using Sanger sequencing, and found that mosaics typically carry 1-4 EJ alleles (2.75 on average) with almost no unmodified targets (Figure 4.11). Additionally, the most common mutation differ between different progeny batches. This would indicate that mutations created as a result of Cas9 deposition occur early during embryonic development.

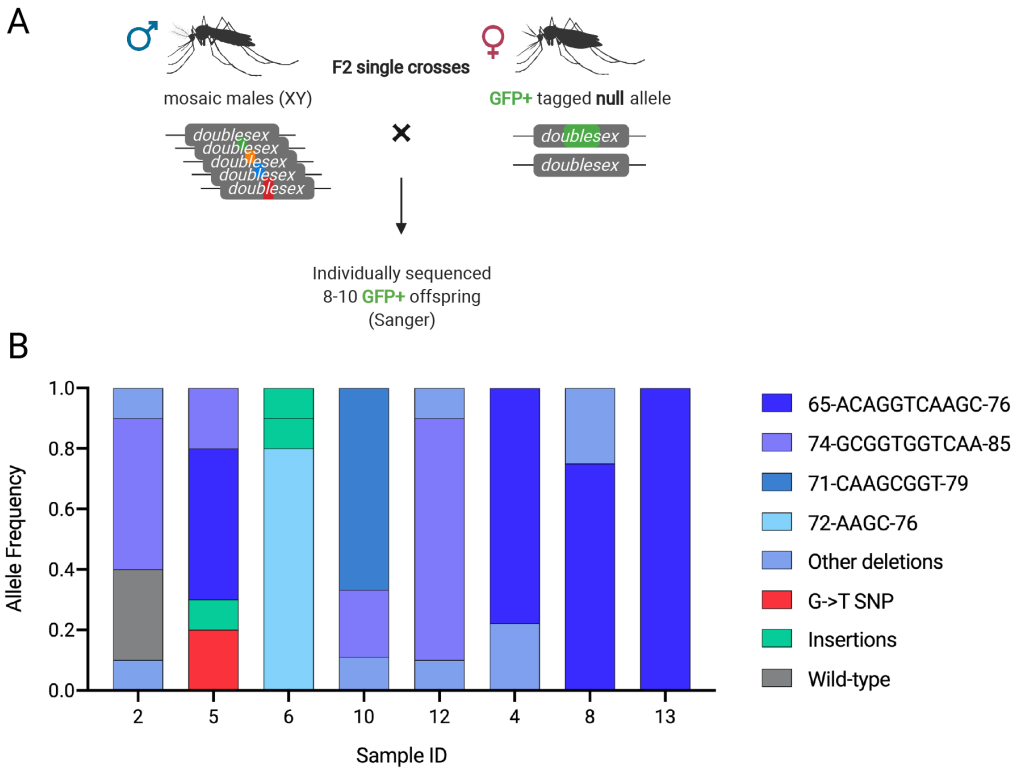


Figure 4.11: The number and types of mutations inherited in the progeny of mosaic males during the mutagenesis screen. (A) Schematic of the single crosses that were set up to determine the number of distinct mutations that each mosaic male parent contributed to its offsrping. **(B)** The portion of distinct mutations inherited amongst offspring of single F2 crosses (each denoted with a distinct sample ID).

Mating behaviour of mosaic males might have introduced biases in favour of certain mutations.

To mitigate against stochastic bias in the resistant allele creation rates, caused by early germline events, both F1 and F2 crosses were performed in separate batches (Figure 4.10A-B). To determine if mutant alleles could encode a functional *dsx* copy, F2 males were crossed to females carrying a GFP-marked *dsxF*-null allele (Kyrou et al. 2018) (Figure 4.10B). We examined 10,083 chromosomal females that carried the marked null mutation in the F3 balanced across individual EJ or unmodified alleles (Figure 4.10B, Table 4.1). Of these, 89.8% developed as intersex, confirming they must contain an R2 mutation, whereas the remaining 10.2% developed with normal female anatomy and must therefore contain an unmodified wild-type or R1 allele (Figure 4.10B, Table 4.1).

Table 4.1: The number of mutant females assessed as part of the mutagenesis screen performed at the T1 gene drive target site.

GFP+ chromosomal females collected	N	%
Total	10,083	100.0
Female anatomy	1,031	10.2
Intersex anatomy	9,052	89.8

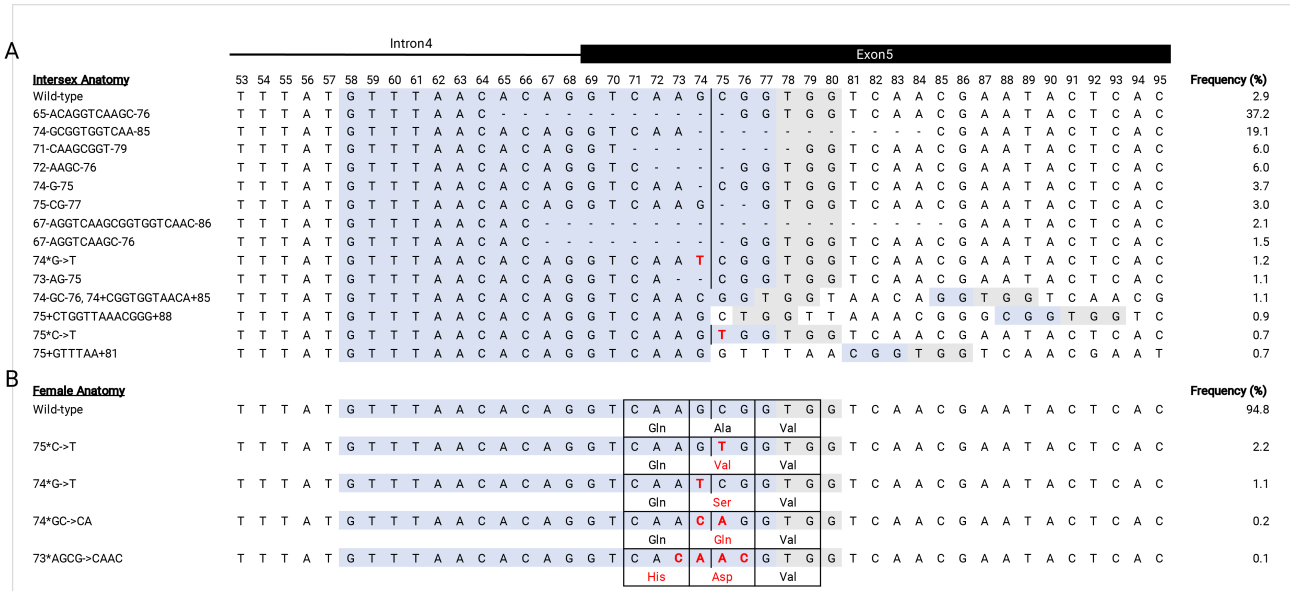


Figure 4.12: Sequence alignment of the alleles discovered through the mutagenesis screen at the T1 target site. Deletions are denoted by dashes, and SNPs are depicted in red. The T1 gRNA binding site is highlighted in blue and the PAM in grey. The frequency of each allele amongst intersex (A) and anatomical females (B), respectively is shown to the right. **(A)** Alleles discovered in anatomically intersex (sterile) GFP+ females. **(B)** Alleles discovered in GFP+ females that developed normally, as well as the amino acids encoded by codons carrying SNPs.

Targeted pooled amplicon sequencing confirmed that the majority of R2 mutations were out-of-frame deletions, the most common of which had been observed previously (Kyrou et al. 2018) (Figure 4.12A).

Moreover, we observed that there was a strong bias in the most common mutation that each parental F2 cross contributed to the next generation (Figure 4.14). Specifically, on average, in the offsrping of each cage, the most common mutation made up 45.2±17.7% of all reads. In clutches where this was most prominent (e.g. X1, R1, U1), it might suggest that a small number of males dominated most matings. In clutches showing a variety of evenly distributed mutations (e.g. F1, G20-1, C20-1) it is more likely that matings were also more evenly distributed (Figure 4.14).

It is worth noting, that according to this setup, mutations causing reduced male competitiveness would be selected out of the population and not be prominently detected in the offspring of F2 mosaics.

Nucleotide substitutions may be tolerated at the gene drive target site.

Sanger sequencing of 852 anatomical females, revealed that 96.4% carried a wild-type allele paired to the GFP+ null allele, whereas the remaining 3.6% carried putative R1 resistance, including a C→T substitution (2.2%) (2R:48714642), a G→T substitution (1.1%) (2R:48714643), and two complex alleles (0.4%) (Figure 4.12B, Table 4.2). All four mutations caused a change to the amino acid sequence, and the two most common SNPs generated semi-conservative (Ala→Val) and conservative (Ala→Ser) amino acid substitutions respectively (Figure 4.12B).

Table 4.2: The portion of mutant females carrying putatively restorative mutations at the T1 gene drive target site.

Sanger-sequenced anatomical females	Blood-fed (BF) or not (non-BF)	N	%
Total		852	100.0
Wild-type	All	808	94.8
	BF	629	
	non-BF	107	
	n/a	72	
C→T SNP	All	19	2.2
	BF	14	
	non-BF	5	
G→T SNP	All	9	1.1
	BF	9	
	non-BF	0	
Others modified	All	3	0.4
	BF	2	
	non-BF	1	
Mixed modified*	All	13	1.5
	BF	10	
	non-BF	3	

*Mixed modified reads likely result from mosaicism due to residual *zpg*-expressed Cas9 activity from the ALE transgene.

Wild-type and putatively restorative mutations were also detected at low frequency (2.9% wild-type, 1.2% G→T and 0.7% C→T) amongst the intersex fraction of GFP+ chromosomal females (Figure 4.10C, Figure 4.12A). In certain clutches they were more prominent than others (Figure). These could be present in mosaics generated by residual *zpg::Cas9* activity from the gRNA-expressing transgene (Figure 4.7A). This hypothesis is supported by the fact that in sequencing the offsrping of the control crosses of the ALE strain to the GFP-marked null, in the absence of a *vas2*-expressed nuclease, 37-39.9% mutagenesis was detected solely due to *zpg::Cas9* activity (Figure).

To confirm that anatomical females that were revealed to contain a SNP by Sanger sequencing were not mosaic with a majority WT genotype, amplicons from a sample of single anatomical females (that yielded high enough quantity and quality of gDNA to allow this) were Illumina-sequenced individually (Figure 4.13). Indeed these females contained a single genotype in their majority, raising the possibility that the G→T and C→T SNPs can produce a functional female isoform of *dsx*.

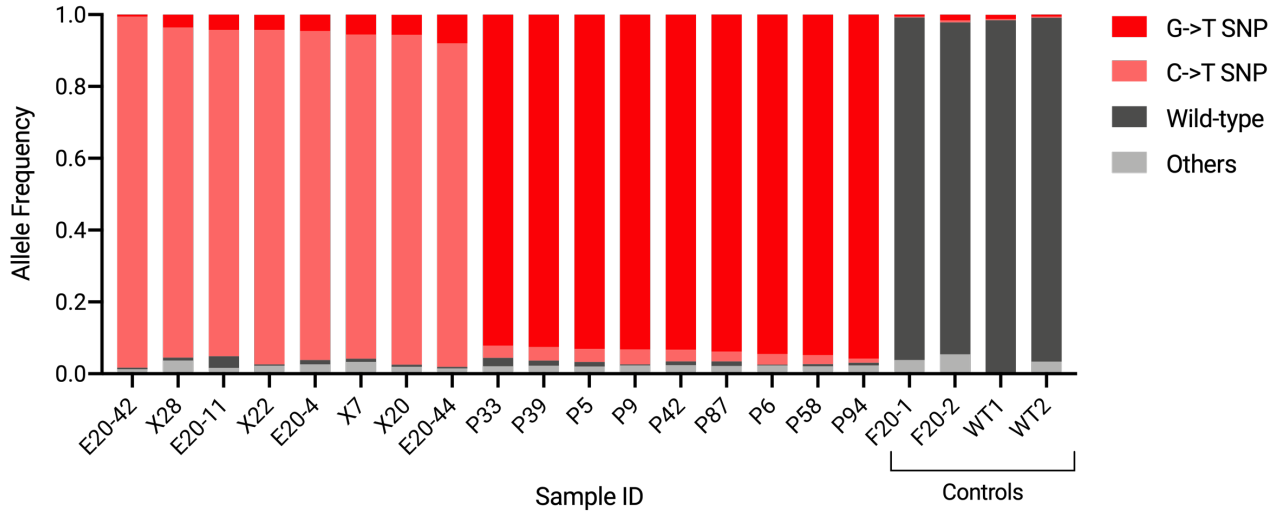


Figure 4.13: Analysis of the genotype of anatomical females carrying a SNP paired to a null *dsx*F allele. Pooled amplicon sequencing of single F3 GFP+ anatomical females, from the mutagenesis screen in 4.10) was performed. Samples analysed were previously shown to carry a single allele paired to the *dsx*F null mutation, through Sanger sequencing: the C→T SNP (majority pink), the G→T SNP (majority red) or a WT allele (majority dark grey).

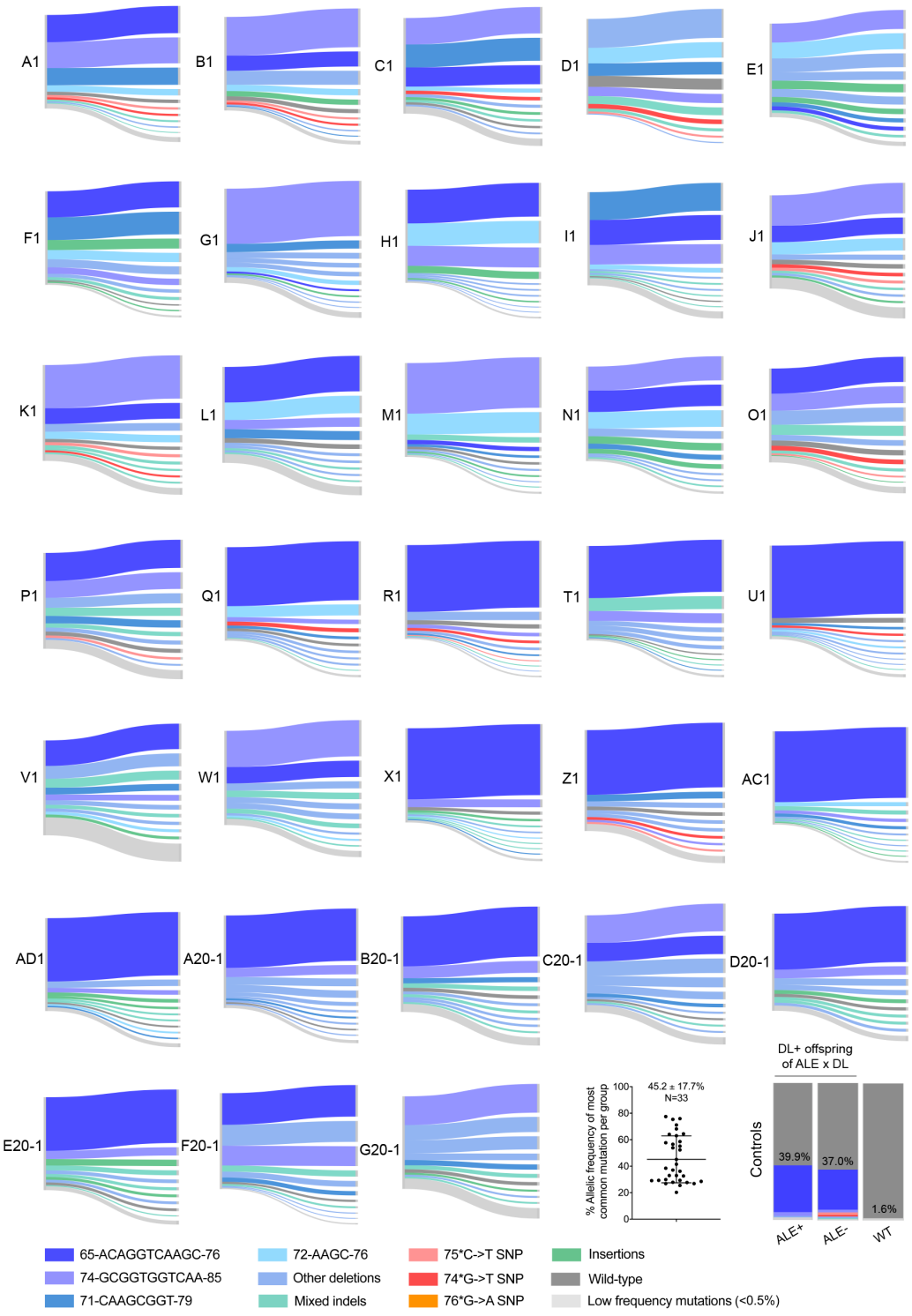


Figure 4.14: The types of Cas9-induced EJ mutations recovered in the intersex offspring of distinct crosses of mosaic males to *dsxF-* females. A minimum of 50 ALE+ mosaic males containing a multitude of Cas9-induced EJ mutations were crossed to a minimum of *dsxF-* females (cross 2 in Figure 4.10B). This cross was performed in replicate cages 33 times (cages A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, P, Q, R, T, U, V, W, X, Z, AC, AD, A20, B20, C20, D20, E20, F20, G20). For each cross, 100 GFP+ intersex offspring were analysed through pooled amplicon sequencing. The panels show the relative portion of each mutation recovered in intersex individuals. The graph shows the allelic frequency of the most common mutation present in each intersex offsrping pool is also shown. As controls we also sequenced 100 wild-types, as well as the GFP+ offsrping of a cross of 50 ALE males to 50 *dsxF-* DL females, i.e. in the absence of the maternally depostied *vas2*-expressed Cas9, shown in the last column panel. Both ALE+ (DsRed+) and ALE- (DsRed-) offsrping contained a similar amount of background mutagenesis. The figure was generated using Sankeymatic and Graphpad Prism.

Estimation of the rate of resistant allele creation as a result of gene drive activity.

Assuming a fertility-weighted average of 93% homing for Ag(QFS)1, and EJ mutations being present in half the non-drive progeny of heterozygotes (according to the data of Kyrou et al. (2018)), 519,771 progeny of heterozygotes would need to be examined to identify the 9,096 confirmed EJ mutants described here. Therefore, screening for resistance in this way is a 100× improvement, compared to the screening of non-drive offspring in the ranges of 100 that was done previously (Hammond et al. 2017, Kyrou et al. 2018, Hammond et al. 2021b).

It remains to be seen whether differences in timing of expression under the *zpg* promoter alters the landscape of resistance, perhaps affecting how resistant alleles cluster in the progeny of heterozygotes. In our screen, each F2 founder contributed an average of 2.75 distinct mutations in their offspring (Figure 4.11).

Among the 9096 individuals carrying confirmed EJ alleles, only 31 ($3.4*10^{-3}$) were found to contain putative R1 alleles. Accounting for the few anatomical females that were not sequenced (which may have contained WT or putative R1 alleles), we estimate the true frequency of putative R1s to be $4.1*10^{-3}$ amongst all EJ alleles.

Assuming relative allele frequencies are not influenced by the source of Cas9 (i.e. *vasa* vs *zpg*-controlled expression), and a fertility-weighted average of 93% homing[4], we estimate that putative R1 alleles are created at a frequency of $7.2*10^{-5}$, on average, in the progeny of Ag(QFS)1 heterozygotes. Note that the likelihood of resistant allele creation will flactuate depending on whether the gene drive carrier is male or female, and whether they had inherited the drive allele paternally or maternally (Kyrou et al. 2018, Hammond et al. 2021a).

[4]Gene drive homing rates vary depending on parental class (M→M, M→F, F→M, F→F), however not all classess contribute equally to the offsrping pool of the next generation because their fertility also varies. To account for this a fertility-weighted average of homing was calculated.

4.3 Assessing the potential for resistance against the *dsx* gene drive

4.3.1 Background

We selected three variants of the T1 gene drive target site for closer examination (in the interest of time): the G→A SNP that is prevalent in Angola (see section 4.1), and the two most common Cas9-induced variants (the C→T and G→T SNPs) that were selected for being able to restore the *dsx* phenotype in females through resistance screening (see section 4.2). First, we wanted to verify that the variants enable normal female sexual development and reproductive potential, as predicted by the analyses in 4.1 and 4.2. Second, we wanted to test whether they can block gene drive activity. Considering their proximity to the PAM, we assumed that the variants are unikely to be cleaved efficiently *in vivo* sequence (Hsu et al. 2013) (Figure 4.5). To this end we set out to engineer the three variants in the genome of laboratory mosquitoes (*An. gambiae* G3 strain).

It is crucial that these are engineered without a molecular transformation marker or other gene editing debris (such as ruminant *loxP* or other attachment sites used by recombinase enzymes), to ensure that any observed phenotype is solely caused by the variant of interest, and not due to an additional exogenous sequence.

4.3.2 CriMCE: A novel method to engineer precise marker-less variants in the genome of insects.

In insects, independent of the chosen technology, engineering small marker-less edits remains inefficient, with transformation rates rarely exceeding 5% (Bosch et al. 2020, Grigoraki et al. 2021, Kistler et al. 2015). Moreover, isolating transformants that lack a molecular marker is very inefficient relying upon large numbers of single crosses and molecular identification of variants.

Inspired by RMCE that can use attachment sites flanking a transgene to introduce another transgene in its place, we devised a novel method, which we termed CriMCE, that relies on the exchange of a marker cassette ('placeholder') for the variant of interest, using CRISPR instead of a recombinase enzyme to improve efficiency and exclude ruminant attachment sites (Morianou et al. 2022). Using CriMCE, visual detection of a marker-less edit is possible, through the loss of a marked placeholder, such as a fluorescent protein (Figure 4.15).

First, we generated a placeholder strain by inserting a GFP cassette in place of the entire female-specific exon (exon 5) of *dsx* via CRISPR-mediated HDR (Figure 4.16A). This strain ('referred to as DLv7' in the lab) was isolated based on GFP fluorescence, and displayed an intersex phenotype in homozygous females, consistent with the null mutation of Kyrou et al. (2018).

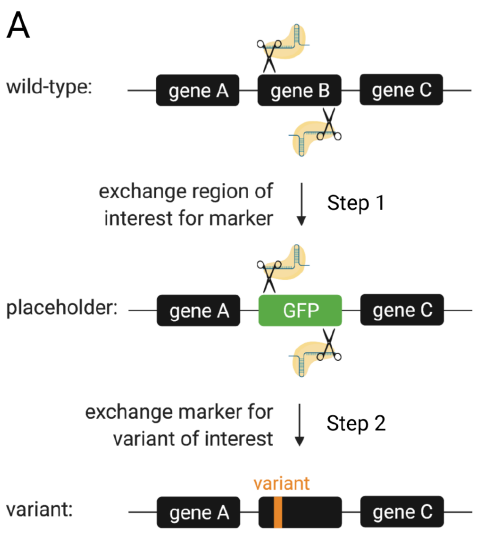


Figure 4.15: CRISPR-mediated cassette exchange (CriMCE) is a two-step method for engineering the detection and isolation of marker-less edits via CRISPR-mediated homology-directed repair. Step 1: To generate a marked placeholder strain, the region of interest (gene B) is replaced by a marker (GFP, green). Step 2: The marker is replaced by the native sequence containing the variant of interest (orange), through CriMCE, to obtain a marker-less strain carrying the variant.

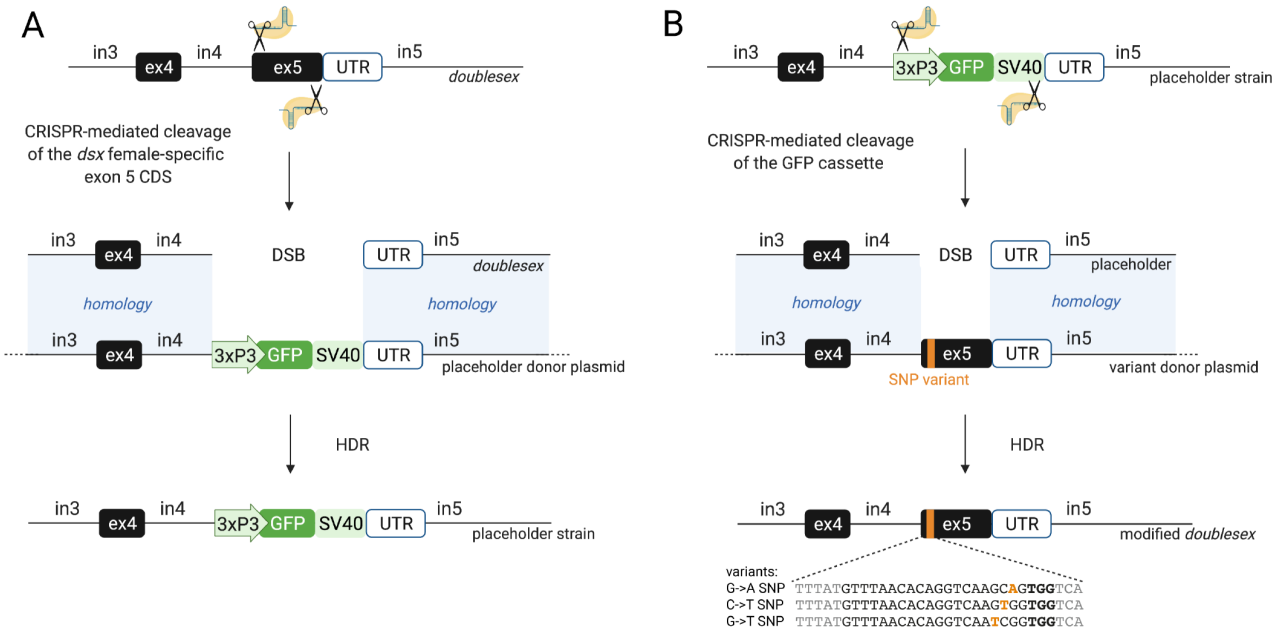


Figure 4.16: CriMCE relies upon the generation of a marked placeholder strain, and the subsequent exchange of the placeholder for the variant of interest through CRISPR-mediated HDR. (A) To generate the marked placeholder strain, the entirety of the exon 5 coding sequence (CDS) was removed via two CRISPR-mediated double-stranded breaks (DSBs) and replaced with a 3xP3::GFP::SV40 marker cassette (green) from a donor plasmid that served as a template for HDR. **(B)** To generate a strain carrying the variant of choice (G→A, C→T or G→T SNPs at exon 5) the marker cassette was removed via two CRISPR-mediated cleavages and exchanged for the exon 5 CDS containing the variant of interest (orange) from a donor plasmid, through HDR.

We then performed CRISPR-mediated exchange of the placeholder for the marker-less SNPs of interest (G→A, C→T or G→T) (Figure 4.16B), by injecting both placeholder homozygotes and heterozygotes

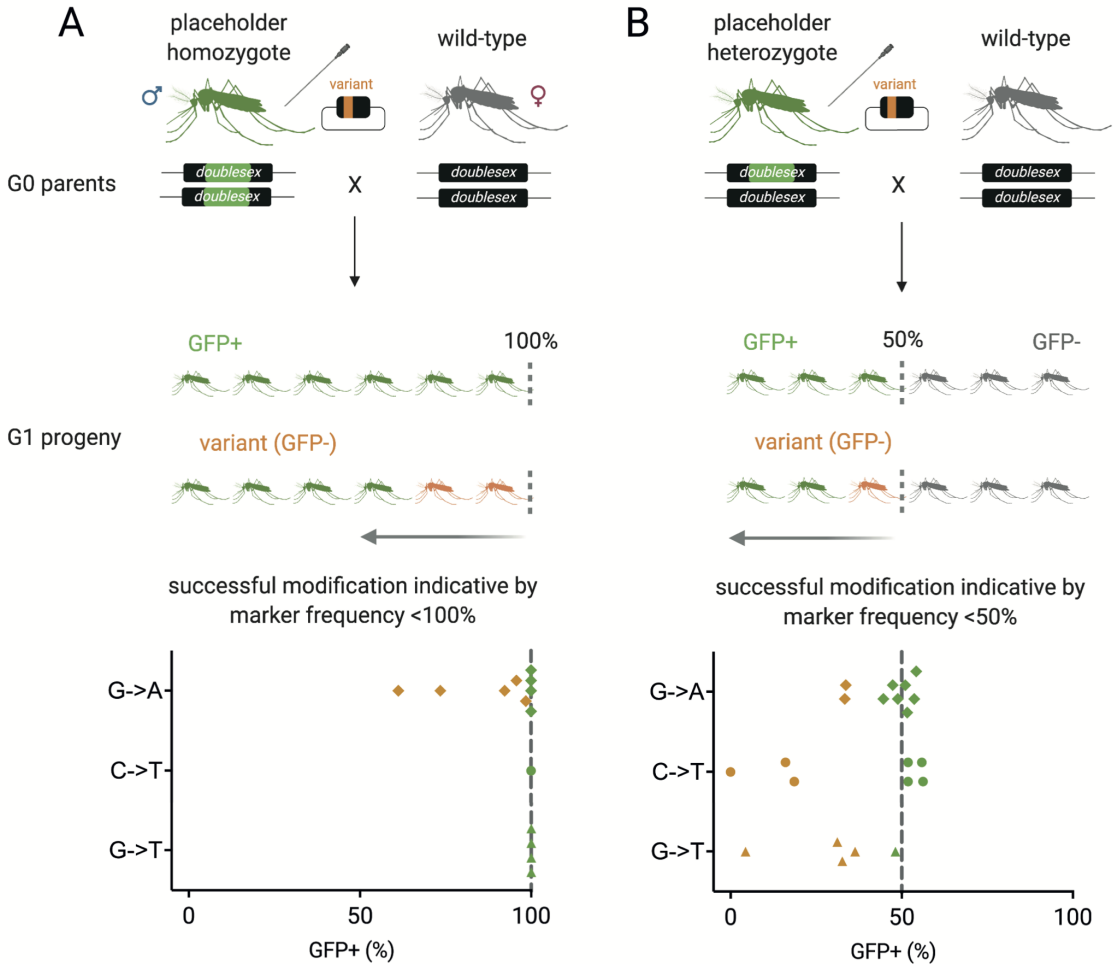


Figure 4.17: Visual detection of marker-less edits. The introduction of a marker-less variant using CriMCE is evidenced by reduced rates of marker inheritance in the progeny of microinjected individuals of the placeholder strain. Marked placeholder male homozygotes (A) and heterozygotes of both sexes (B), were microinjected with a CRISPR helper plasmid and a variant donor plasmid to facilitate CriMCE of the placeholder for one of the variants of interest (G→A, C→T, G→T). G0 parent injected mosquitoes (green) were individually crossed to wild-type (grey) and their G1 progeny screened for GFP fluorescence. Successful introduction of each marker-less variant via CriMCE, was evidenced by a marker frequency of less than 100% in the progeny of placeholder homozygotes, and a marker frequency of less than 50% in the progeny of placeholder heterozygotes (orange). Lack of modification was evidenced by a marker frequency equal to 100% in the progeny of placeholder homozygotes and a marker frequency normally distributed around 50% in the progeny of placeholder heterozygotes (green).

with a CRISPR plasmid (expressing RFP, Cas9 and gRNAs targeted to the placeholder), and a template for repair encoding the variant of interest (See methods section for detail 3.4). To maximise the recovery of editing events, we selected only the fraction of injected mosquitoes that showed transient RFP fluorescence and mated these to wild-type.

CriMCE-induced editing was evidenced by loss of GFP (<100% GFP inheritance) among the offspring of placeholder homozygotes, or by significant deviation below the Mendelian expectation of 50% GFP inheritance among the offspring of placeholder heterozygotes (Figure 4.17). We saw rates of precise editing up to 39% for the G→A SNP (evidenced by 61% GFP inheritance in the offspring of placeholder

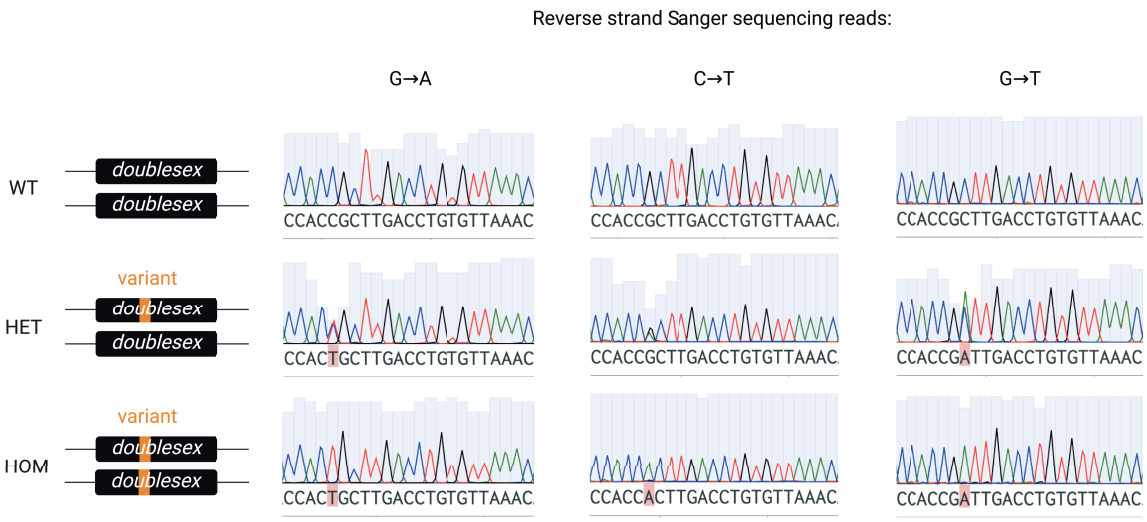


Figure 4.18: Molecular validation of successful CriMCE-induced genetic modification through Sanger sequencing. Sanger sequencing chromatographs from single GFP negative mosquitoes. Top: WT, example of an unedited individual. Middle: example of a heterozygous edited individual carrying the SNP variant of interest (G→A, C→T or G→T), evident through a double peak in the chromatograph. Bottom: example of a homozygous edited individual carrying the SNP variant of interest in homozygosis (G→A, C→T or G→T), evident through a single modified peak in the chromatograph. Note that reverse strand sequencing chromatographs are shown.

homozygotes) (Figure 4.17A), up to 100% for the C→T SNP, and up to 92% for the G→T SNP variant (evidenced by 0% and 4% GFP inheritance in the offspring of placeholder heterozygotes, respectively) (Figure 4.17B). Incorporation of the SNPs of interest was confirmed by Sanger sequencing (Figure 4.18). Notably, we did not detect any end-joining (EJ) events (N=55). Owing to the high rates of editing by CriMCE, G1 transformants that showed low levels of GFP inheritance can be immediately crossed to the placeholder strain that will act as a balancer, for rapid characterisation of each marker-less edit (Figure 3.2).

In two G1 clutches with altered GFP inheritance we also detected variant donor plasmid integration, evidenced by RFP at 2% and 18% amongst GFP negatives (with a median of 0% taken across all modified clutches). These were excluded from the analysis when calculated the efficiency of the programmed edit.

To compare our method to previously developed strategies employing HDR and prime editing, three measures of transformation efficiency were calculated: the percentage of G0 founders that gave G1 transformants, the G1 transformant to G0 injected survivor ratio, and the G1 transformant percentage out of all G1 screened (Figure 4.19). If the G1 transformant to G0 injected survivor ratio is high, then a high number of transformants can be obtained from a smaller number of injected survivors; whilst having a high percentage of G1 transformants out of total G1 screened, implies a reduced requirement for screening, whether this is done visually, like in the present study (less laborious), or by PCR and sequencing analysis, like in previous studies (more laborious). As a reference, the efficiency of locus-specific marked

transgene insertion through RMCE and HDR are also shown (Figure 4.19).

In total, visible editing was detected in the progeny of 7/18 (38.9%) G0 micro-injected individuals with the G→A construct, 3/8 (37.5%) G0 micro-injected individuals with the C→T construct, and 4/9 (44.4%) G0 micro-injected individuals with the G→T construct (Figures 4.17 and 4.19). These rates are amongst the highest reported amongst all methods for targeted gene editing (Figure 4.19).

CriMCE offered a marked improvement in transformation efficiency when compared to other approaches employed to introduce marker-less edits (Figure 4.20). Specifically, CriMCE shows a mean G1 transformant to G0 injected survivor ratio of 5.76 (±2.37 s.d.), compared to 0.14 (±0.10 s.d.) for direct HDR (Welch's t-test p=0.031) and 1.06 (±0.83 s.d.) for prime editing; and a mean G1 transformant per G1 screened percentage of 10.5% (±6.0% s.d.), compared to 1.0% (±0.5% s.d.) for direct HDR and 1.4% (±1.1% s.d.) for prime editing (Welch's t-test p=0.058) (Figure 4.20).

Transgenesis method and study		Organism	Eggs injected N	G0 Injected survivors N	G1 transformants N	Total G1 screened N	G0 Founders N (%)	G1 transformant to G0 injected survivor ratio	G1 transformants per G1 screened %
Introduction of precise marker-less edit									
CriMCE present study	G->A	*Anopheles gambiae*	380	18* (59)	111*+	1716	7/18 (38.9)	6.17	6.47
	C->T	*Anopheles gambiae*	1025	21	166*+	953	3/8 (37.5)	7.90	17.42
	G->T	*Anopheles gambiae*	963	23	74*+	97	4/9 (44.4)	3.22	7.62
HDR	Kistler et al. (2015)	*Aedes aegypti*	636	61**	4*+	620	N/A	0.07	0.65
	Grigoraki et al. (2021)	*Anopheles gambiae*	338	19	4*+	290	1/19 (5.0)	0.21	1.38
Prime Editing Bosch et al. (2021)***	Plasmid pegRNA	*Drosophila melanogaster*	50	18	3*+	1767	1/18 (5.6)	0.17	0.17
	Plasmid pegRNA+ sgRNA	*Drosophila melanogaster*	50	15	28*+	1594	6/15 (40.0)	1.20	1.76
	Synthetic pegRNA	*Drosophila melanogaster*	50	11	20*+	866	4/9 (44.4)	1.82	2.31
Introduction of marked transgene									
RMCE Hammond et al. (2016)	7280	*Anopheles gambiae*	540	56**	15*	4000	N/A	N/A	0.38
	11377	*Anopheles gambiae*	500	21**	4*	2990	N/A	N/A	0.13
	5958	*Anopheles gambiae*	400	49**	2*	4000	N/A	N/A	0.05
HDR	Gratz et al. (2014)	*Drosophila melanogaster*	N/A	50	599*	7657	9/50 (18.0)	11.98	7.82
	Gantz et al. (2015)††	*Anopheles stephensi*	680	251**	2*	25,712	N/A	0.01	0.01
	Hammond et al. (2016) 7280	*Anopheles gambiae*	350	48	278*	1536	9/48 (18.8)	5.79	18.10
	Hammond et al. (2016) 5958	*Anopheles gambiae*	760	26	51*	3184	3/26 (11.5)	1.96	1.60
	Adolfi et al. (2020)	*Anopheles stephensi*	504	184**	96*	25,293	N/A	0.52	0.38
	Ang et al. (2022) *190-perfect*	*Aedes aegypti*	N/A	271**	350	9,774	13/13 (100.0)º	1.29	3.6
	Ang et al. (2022) *64+234-perfect*	*Aedes aegypti*	N/A	355**	207	22,158	8/17 (47.1)º	0.58	0.93

Figure 4.19: Comparison of CriMCE to different transgenesis methods for the introduction of small precise marker-less edits or marked transgenes. Efficiency of each method is measured through the G1 transformant to G0 injected survivor ratio and the % of G1 transformants isolated from screened G1 progeny, as well as the number of G0 founders.

Notes on Figure 4.19: *Only 18 out of 59 G0 injected survivors were kept and crossed to obtain G1 transgenics, due to Covid-19 restrictions in April 2020.

*In most studies G0 injected survivors are not being distinguished from non-injected survivors through transient expression of a fluorescent marker. The Kistler et al. (2015), Gantz et al. (2015), Hammond et al. (2016), Adolfi et al. (2020), Ang et al. (2022) studies did not use such a method to distinguish injected survivors, or used all injected survivors (whether or not they showed signs of injection) to obtain transgenics.

**Showing the set of injections with greater success for each method of prime editing: (a) using pegRNA expressed from a plasmid to provide cleavage and a template for repair, (b) using plasmid pegRNA together with an sgRNA to provide cleavage, (c) injecting a synthetic pegRNA straight away.

***** Identified visually.

****** Identified through sequencing.

†The number of transformants is equal to the number of individuals lacking a fluorescent marker in the progeny of placeholder homozygotes. The number of transformant in the progeny of placeholder heterozygotes it was estimated using this formula: (Total G1)/2 - GFP+ - RFP+.

††Note that the transgene integrated by HDR in the Gantz et al. (2015) study was significantly larger in size compared to all other studies, which could have reduced efficiency of integration.

ºThe number of G0 founder pools that gave G1 transformants.

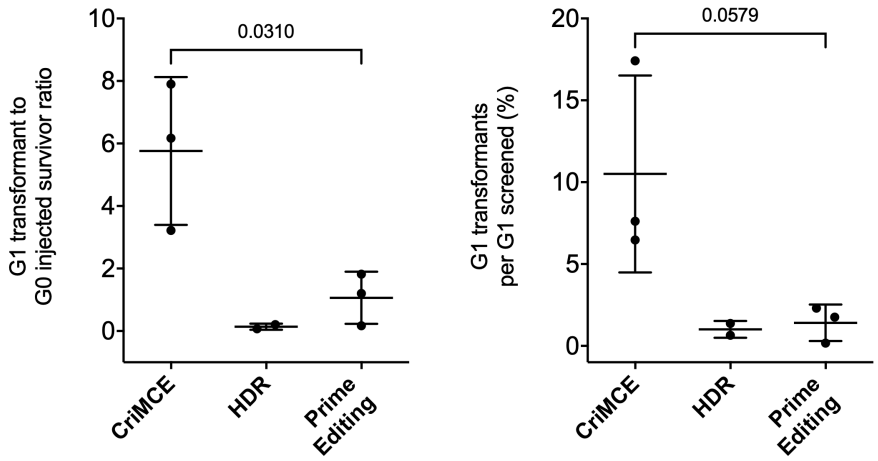


Figure 4.20: Comparison of CriMCE to different transgenesis methods for the introduction of small precise marker-less edits. Welch's t-test p-values of statistical comparisons between CriMCE and prime editing are shown on top of each graph. HDR could not be statistically compared due to its small sample size.

4.3.3 Natural and artificially selected variants of the T1 site are functional and provide different levels of resistance to gene drive

Fecundity of variant-bearing females.

After successfully engineering the SNP variants in the genome and confirming their presence though rounds of single crosses and Sanger sequencing (Figure 4.18), I set out to assess their fecundity. SNP heterozygotes were crossed to each other to generate a pool of SNP heterozygotes, SNP homozygotes, as well as wild-type females. These were then crossed to wild-type males and allowed to lay individually to measure their egg and larval output, as well as egg hatching rates (Figure 4.21, 4.22 and 4.23). This phenotype assessment was carried out in a blind manner and SNP genotypes were determined by Sanger sequencing after all data had been collected. Simultaneous rearing of all genotypes, allowed for an accurate statistical comparison between SNP carriers and the wild-type control.

As expected, we did not find a significant reduction in the fertility of SNP-bearing females. The egg and larval output of females carrying each variant, either in homozygosis or heterozygosis, did not vary significantly from that of their respective wild-type control (Figures 4.21, 4.22, 4.23). This suggests that the SNPs restore the function of *dsx*, at least with respect to female fecundity.

Note that hatching rates were significantly reduced in offspring of G→A homozygotes (p=0.0463) and C→T heterozygotes (p=0.0445). However, this did not impact the overall larval outputs, and could be caused by slight variability in rearing conditions and/or fitness/competitiveness of wild-type males, despite efforts to maintain similar rearing as much as possible [5].

[5]Note that depending on the week that fertility assays were carried out, the fertility of WT females was gravely different. For example, the WT control of the G→A crosses is almost 50% less fertile than that of the G→T crosses. This is because there is great variability between weeks in the fitness of wild-type males maintained by the lab rota, as well as the humidity and temperature of the rearing cubicles, which affects swarming and mating behaviour, showing the importance of contemporaneous controls.

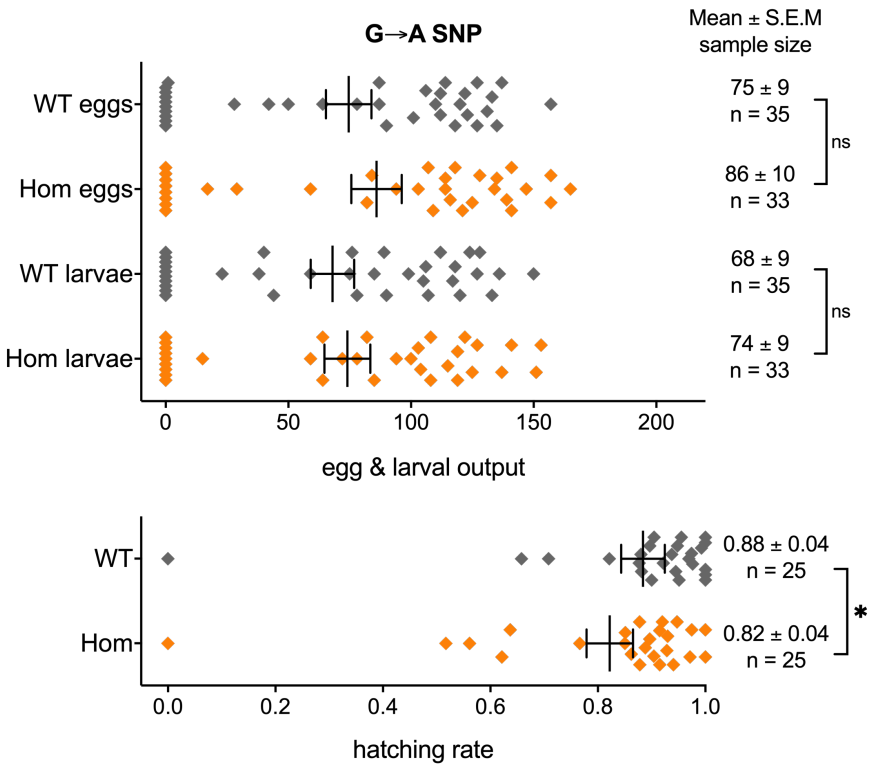


Figure 4.21: Fecundity of females carrying the natural G→A variant at T1. The datasets did not pass the D'Agostino-Pearson normality test and therefore each experimental dataset was compared to the wild-type control using the Mann-Whitney non-parametric test. Abbreviations: WT = wild-type, Hom = homozygous for the G→A variant, ns = not significant (WT vs Hom egg output p-value = 0.3051, WT vs Hom larval output p-value = 0.6459), ∗ = significant with p-value < 0.05 (WT vs Hom hatching rate p-value = 0.0463).

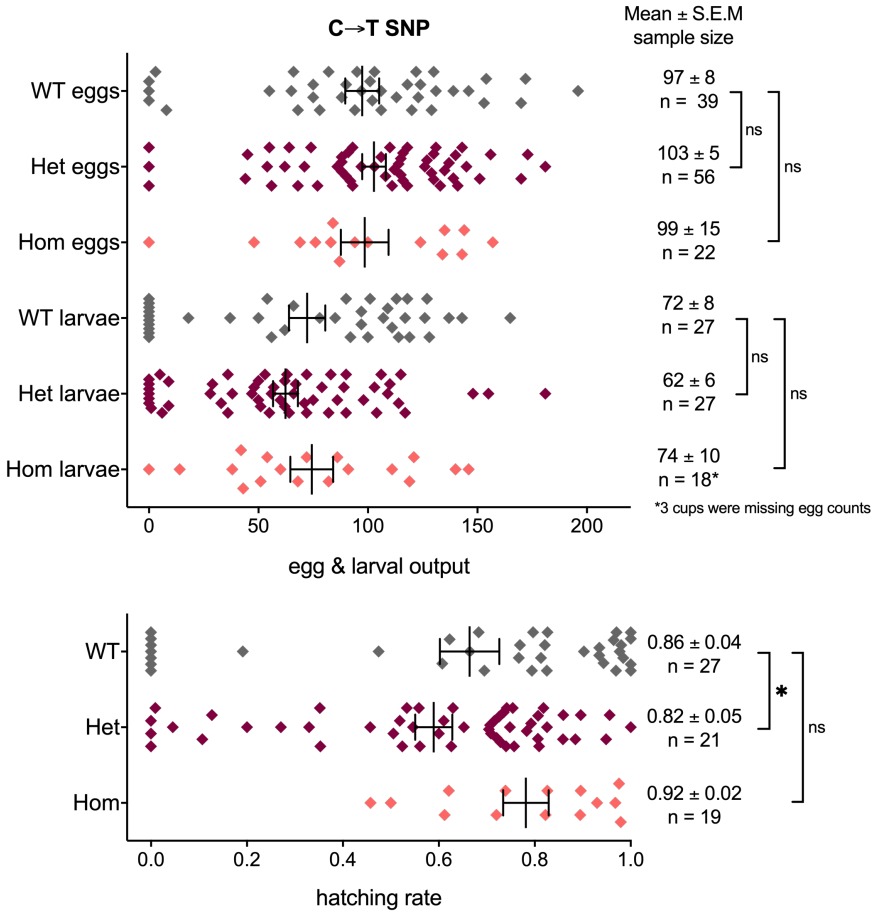


Figure 4.22: Fecundity of females carrying the Cas9-induced and artificially selected C→T variant at T1. Only egg output datasets passed the D'Agostino-Pearson normality test and were analysed using ordinary ANOVA, allowing for multiple comparisons to the wild-type control. Larval output and hatching rate datasets were compared to the appropriate wild-type control using the Kruskall-Wallis non-parametric test. Abbreviations: WT = wild-type, Het = heterozygous for the C→T variant, Hom = homozygous for the C→T variant, ns = not significant (WT vs Het egg output p-value = 0.7869, WT vs Hom egg output p-value = 0.9946, WT vs Het larval output p-value = 0.4326, WT vs Hom larval output p-value > 0.9999, WT vs Hom hatching rate p-value > 0.9999), ∗ = significant with p-value < 0.05 (WT vs Hom hatching rate p-value = 0.0445).

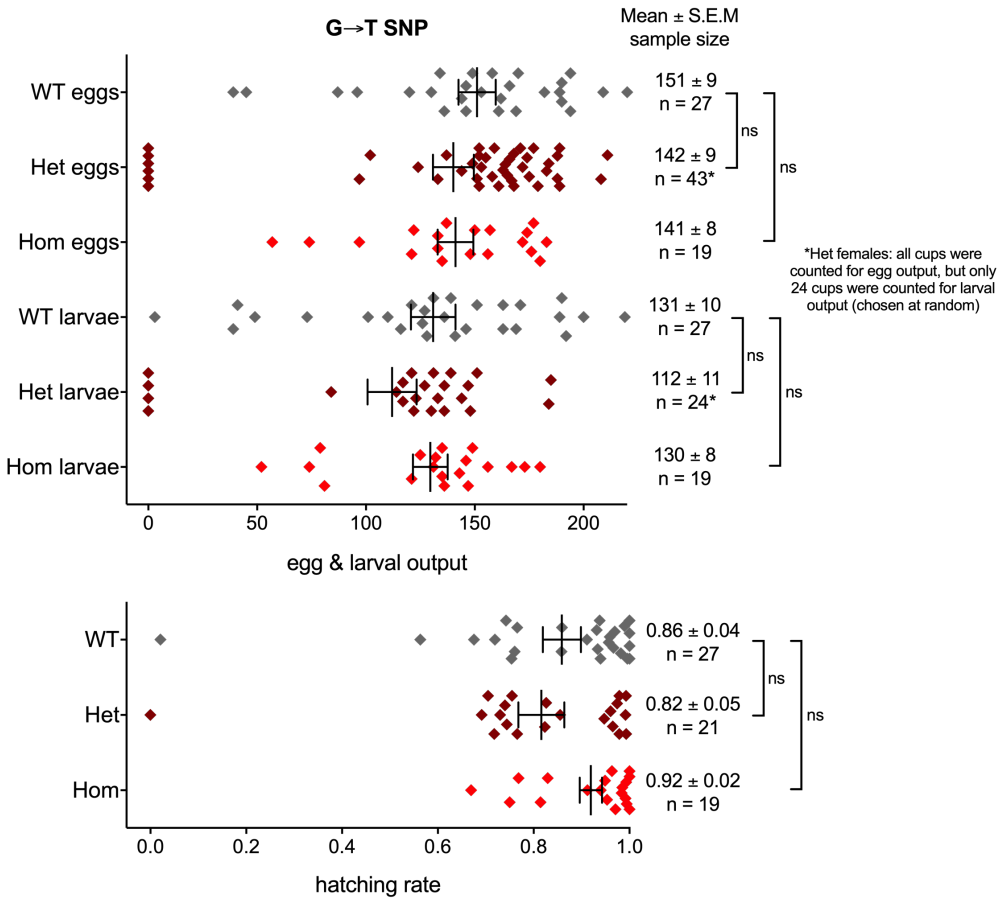


Figure 4.23: Fecundity of females carrying the Cas9-induced and artificially selected G→T variant at T1. The datasets did not pass the D'Agostino-Pearson normality test and therefore each experimental dataset was compared to the wild-type control using the Kruskall-Wallis non-parametric test, allowing for multiple comparisons. Abbreviations: WT = wild-type, Het = heterozygous for the G→T variant, Hom = homozygous for the G→T variant, ns = not significant (WT vs Het egg output p-value $>$ 0.9999, WT vs Hom egg output p-value $>$ 0.9999, WT vs Het larval output p-value = 0.4874, WT vs Hom larval output p-value $>$ 0.9999, WT vs Het hatching rate p-value = 0.4863, WT vs Hom hatching rate p-value = 0.6497).

Gene drive transmission in the presence of variants.

To check whether each of the variants could block gene drive transmission, trans-heterozygous Ag(QFS)1/SNP individuals of both sexes, were crossed to wild-type. Gene drive transmission was measured in their offspring and compared to an Ag(QFS)1/wild-type control (Figure 4.24). An average transmission rate above the Mendelian expectation of 50% would indicate that the gene drive can still cleave the variable target site and home into it despite a gRNA/target base mismatch, whilst transmission rates normally distributed around 50%, i.e. the expected Mendelian rate of inheritance of any non-driving genetic element, would indicate full resistance to gene drive activity (no homing).

In agreement with in vitro cleavage assays (Kyrou et al. 2018), the natural G→A variant did not provide resistance to gene drive homing, allowing for similar gene drive transmission rates as the wild-type allele (Figure 4.24, 100.0% from females, and 90.2-100.0% from males, p-value>0.9999). In contrast, in the presence of the Cas9-induced variants (C→T and G→T) the gene drive produced significantly lower transmission rates (Figure 4.24, p-value<0.0001). Specifically, the C→T SNP still allowed some homing to take place evidenced by biased inheritance in the range of 63.2-100% in females and 55.3-76.3% in males, that is translated to a 19.5% and 32.0% reduction in gene drive transmission in females and males, respectively (Figure 4.24). The strongest peturbation of homing was observed for the G→T SNP, which completely blocked gene drive transmission, resulting in the gene drive being inherited in Mendelian rates, averaging at 52.9% (range 41.7-65.2%) for females and 50.5% (range 35.0-65.0%) for males (Figure 4.24).

The present data demonstrate that mutations which only partially block gene drive activity, whilst restoring gene function, are possible, like the C→T SNP. We termed this type of resistance **R3** since it has not been described before. Conversely, the G→T SNP was fully resistant to gene drive activity, whilst restoring gene function, thereby behaving like a typical R1 mutation.

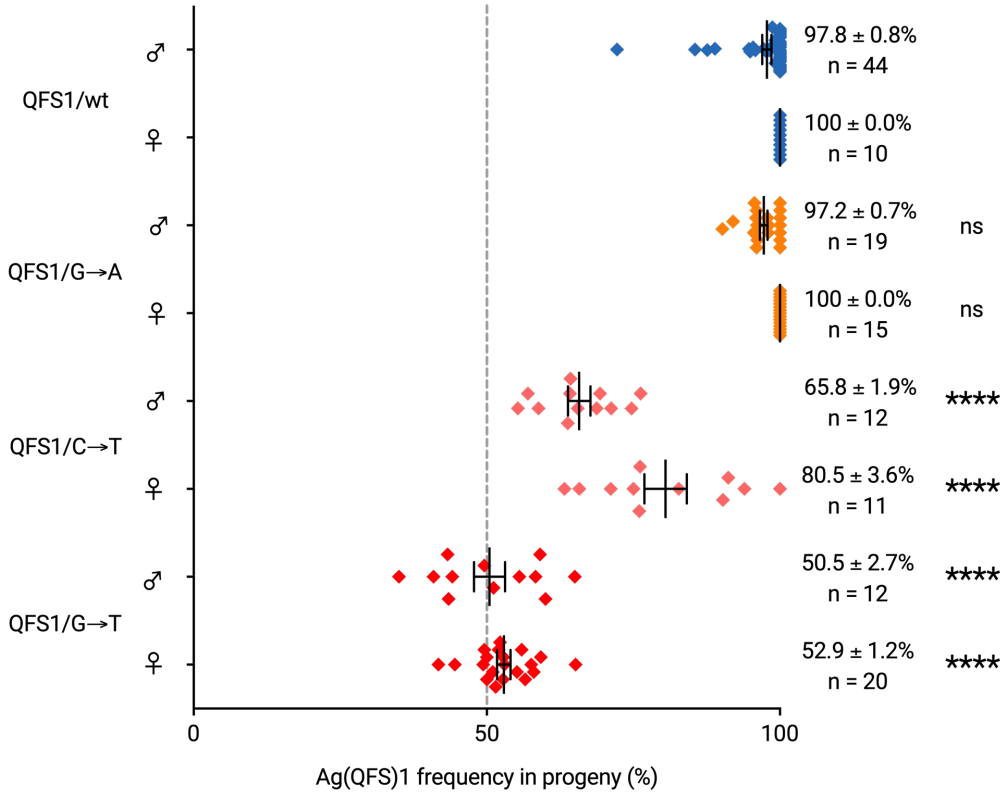


Figure 4.24: Each SNP provided a different level of resistance to gene drive activity. Ag(QFS)1 frequency in the progeny of heterozygous (Het) Ag(QFS)1 males or females crossed to wild-type. The Ag(QFS)1 allele in Het parents has been inherited paternally and is paired to a wild-type (QFS1/wt), G→A SNP (QFS1/G→A), C→T SNP (QFS1/C→T) or G→T SNP allele (QFS1/G→T). Mean Ag(QFS)1 transmission rates and the standard error around the mean (S.E.M.) are shown to the right of the graph, together with the sample size (n) and Kruskall-Wallis statistical comparison to the QFS1/wt control (ns: non-significant with p-value > 0.9999, $**$: significant with p-value < 0.0001). Note that part of the Ag(QFS)1/wild-type transmission rate dataset was taken from Kyrou et al. (2018).

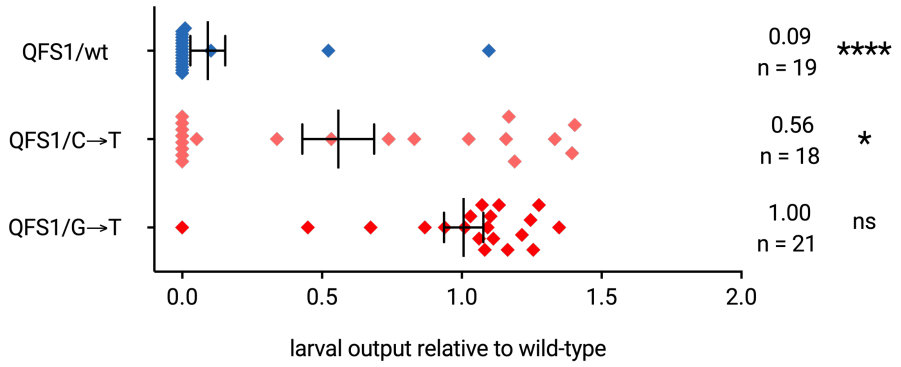


Figure 4.25: Target site variants restore the fertility of heterozygous gene drive carrier females. The relative larval output of Ag(QFS)1 Het females, where Ag(QFS)1 is paired to a wild-type (QFS1/wt), C→T SNP (QFS1/C→T) or G→T SNP allele (QFS1/G→T), compared to the average larval output of wild-type females. Mean relative larval output is shown to the right of each graph, together with the sample size (n) and non-parametric statistical comparisons to each wild-type control (Mann-Whitney: * * ** : significant, with p-value < 0.0001 for QFS1/wt, ns: non-significant, with p-value = 0.0593 for QFS1/C→T and p-value = 0.7324 for QFS1/G→T; Kolmogorov-Smirnov: **** : significant, with p-value < 0.0001 for QFS1/wt, * : significant, with p-value = 0.0289 for QFS1/C→T and ns: non-significant, with p-value = 0.1570).

Fecundity of variant/gene drive trans-heterozygous females.

Consistent with its ability to completely block homing, the R1 allele (G→T) fully rescued the fitness cost seen in gene drive heterozygous females (Ag(QFS)1/wt), showing larval output equivalent to wild-type (Figure 4.25). Despite R3 alleles being susceptible to cleavage, they also rescued female fertility by 55%, presumably due to reduced cleavage of R3 in somatic tissues where DSX has an important function (Figure 4.25). This is consistent with the observed 39% reduction of cleavage in germline tissues (Figure 4.24). Both of these characteristics will most likley impact the spread of gene drive dynamics in a complex manner since reduced transmission rates will negatively impact upon spread, whilst increased fertility of drive-carrying females will increase spread of the drive and impact dynamics in a positive manner.

Modelling partial resistance.

In collaboration with Dr Bhavin Khatri we developed a model to predict the effects of partial resistance on gene drive spread, in the context of a natural population, based on the previous model of resistance generation developed by Khatri & Burt (2022).

Gene drive spread was modelled in the presence of no resistance (N), non-functional resistance only (R2 only), partial resistance (R2+R3), and all types of resistance, including non-functional, full and partial functional resistance co-existing (R1+R2+R3), assuming a low (20%) or high (80%) fitness cost of the drive in female heterozygotes (4.26). We used the rates of creation of R1 and R3 mutations that were calculated using the mutagenesis screen described in section 4.2, and combined them with rates of error-prone EJ

in the mosquito germline determined previously by (Kyrou et al. 2018) and (Hammond et al. 2016) (see section 3.6).

In agreement with Beaghton et al. (2019) we found that R2 resistance can prevent complete population suppression if the gene drive has a high cost, despite the fact that R2 alleles are non-functional. Conversely, in the presence of both R2 and R3 resistance, the fitness cost of the gene drive does not affect the extent of population suppression. The presence of R3 alleles prevents complete population modification, but long-term population suppression is still sustained, at around 40% of the original population size. This might vary depending on the functionality and cleavage susceptibility of the R3 allele in question. Finally, R1 resistance completely reverses gene drive spread, as previously experimentally demonstrated (Hammond et al. 2017, Hammond et al. 2021a).

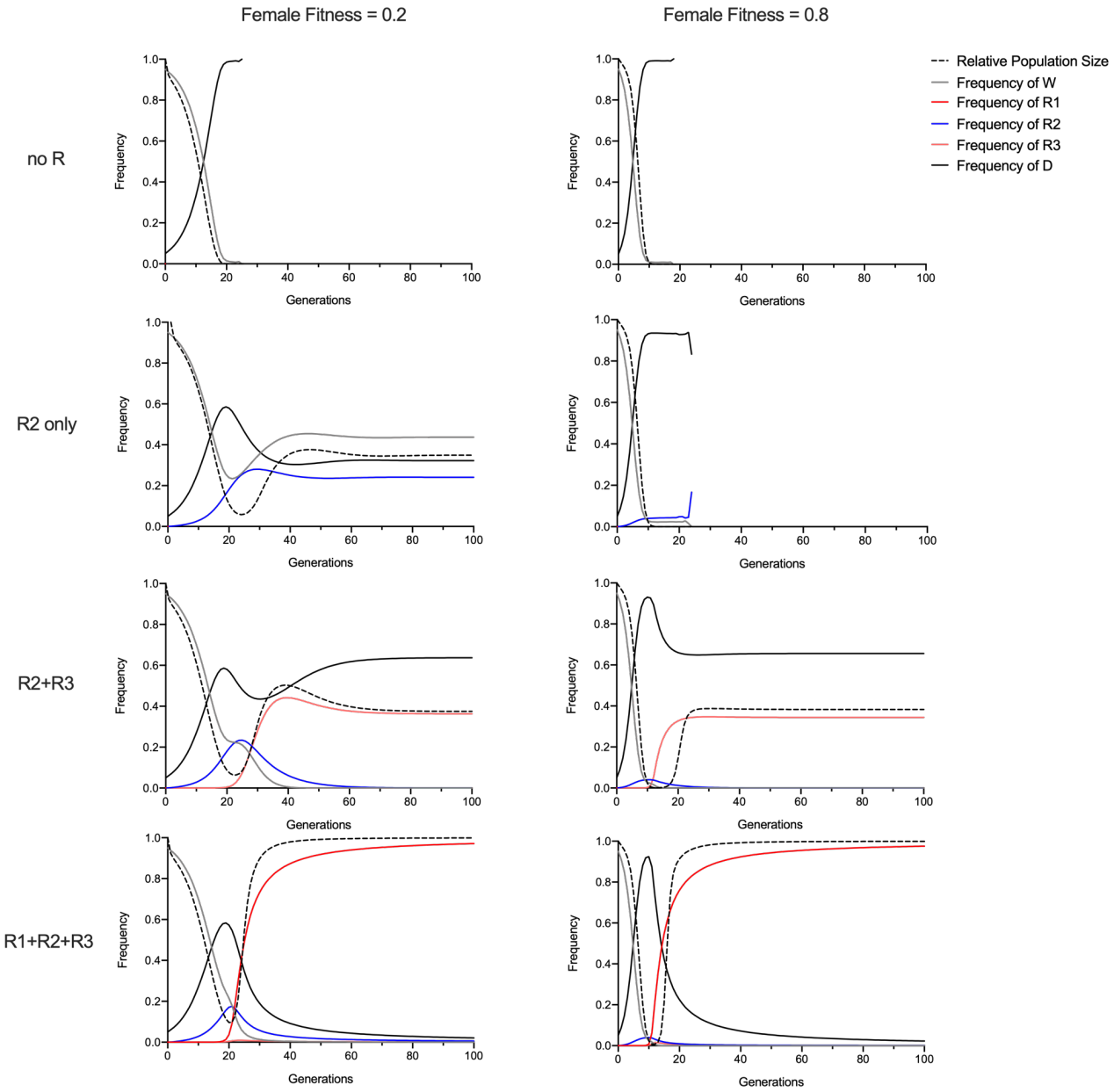


Figure 4.26: Deterministic modelling of gene drive spread in the presence of no, R1, R2 and R3 resistance. Abbreviations: W = wild-type, R1 = functional full resistance, R2 = non-functional resistance, R3 = functional partial resistance, D = gene drive.

4.4 Mitigating resistance using a multiplexed gene drive

4.4.1 Background

As determined in section 4.2, it is possible to select for functional resistance against the Ag(QFS)1 gene drive, as a result of Cas9 activity at its target site. This is despite the *dsx* target site being amongst the most conserved in the whole of the mosquito genome (Kranjc et al. 2021). If these resistant alleles were to be generated in natural populations, they could lead to major disruption of gene drive deployment for malaria control.

To overcome resistance a combinatorial approach can be adopted, whereby multiple non-overlapping sites can be targeted simultaneously (see Figure 1.11). Under this scenario, if resistance were to occur at a single site, provided that the alternative target site is still cleavable, the gene drive could still catalyse homing, and in the process remove resistance at the first target site. Resistant mutations would have to exist simultaneously in order to block gene drive homing, whilst being able to combine to give off a functional gene product.

Multiple sites can be targeted together by expressing several gRNAs from a single gene drive construct. There are multiple ways to achieve this (McCarty et al. 2020). The simplest is to include multiple gRNA expression cassettes in the same gene drive construct, using the same or different pol III promoters. Alternatively, the gRNAs can be expressed under the same promoter: interspersed by tRNAs (Yang et al. 2022), or self-cleaving ribozymes (Xu et al. 2017). However, the latter approaches have never been successfully tested in the malaria mosquito. Therefore, we decided to multiplex using repeated U6-gRNA cassettes.

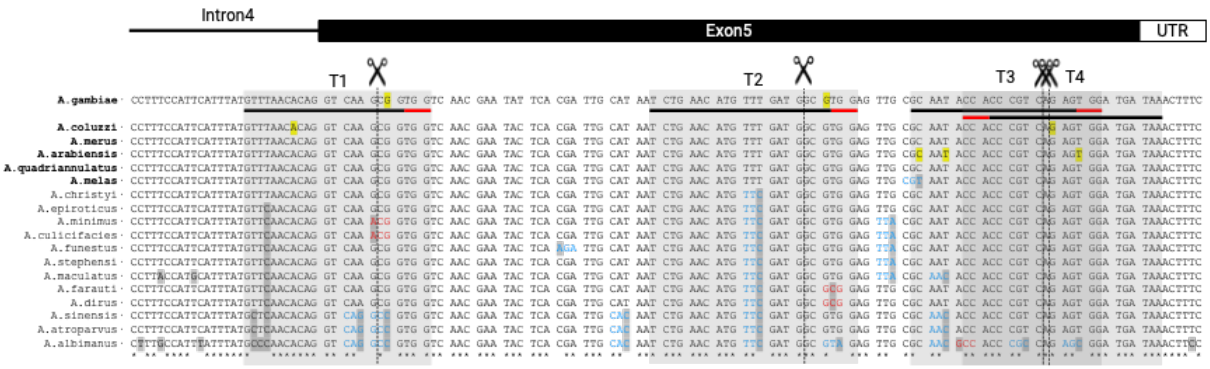


Figure 4.27: Four gene drive target sites spanning the female-specific exon 5 of *dsx*. See Figure 4.1 for a more detailed description.

4.4.2 Generation of multiplexed gene drive strains

In addition to the T1 site targeted by Ag(QFS)1 we identified three further targets on the female-specific exon 5 of *dsx* that showed high amounts of sequence conservation (Figure 4.27). I cloned multiple gRNA cassettes, whereby gRNAs were initially multiplexed in couples (T1+T2, T1+T3, T2+T4), and later cloned a triple gRNA-expressing construct (T1+T2+T3).

Before the generation of a multiplexed gene drive strain, novel docking lines had to be obtained to allow insertion of the gene drive construct in the region between the outer-most gene drive cut sites. Three new donor vectors were cloned to allow the creation of novel docking lines: DLv2, DLv3 and DLv4, that respectively could accommodate gene drives expressing T1+T2, T1+T3, or T2+T4. DLv3 was the only docking line that could also accommodate the T1+T2+T3 multiplexed gRNA construct, and so priority was given to obtaining this particular strain.

We performed wild-type (G3 strain) embryo micro-injections of a CRISPR helper vector, expressing Cas9 and gRNAs T1+T3, together with a GFP-marked HDR donor vector, so that the region between the gRNA target sites is excised and replaced by a GFP cassette, flanked with attP sites to allow downsream RMCE(Figure 4.28). As expected, the DLv3 transgenic strain replicated the recessive sterile/intersex phenotype of the docking line created by Kyrou et al. (2018), as both strains disrupt the female-specific exon of *dsx* (exon 5).

Heterozygous DLv3 transgenic individuals were crossed to each other and their offspring were micro-injected to obtain multiplexed gene drive strains through RMCE (Figures 4.28). The strains obtained simultaneously targeted the sites T1 and T3 or T1, T2 and T3 and were named Ag(QFS)2 and Ag(QFS)3, respectively (Figure 4.29).

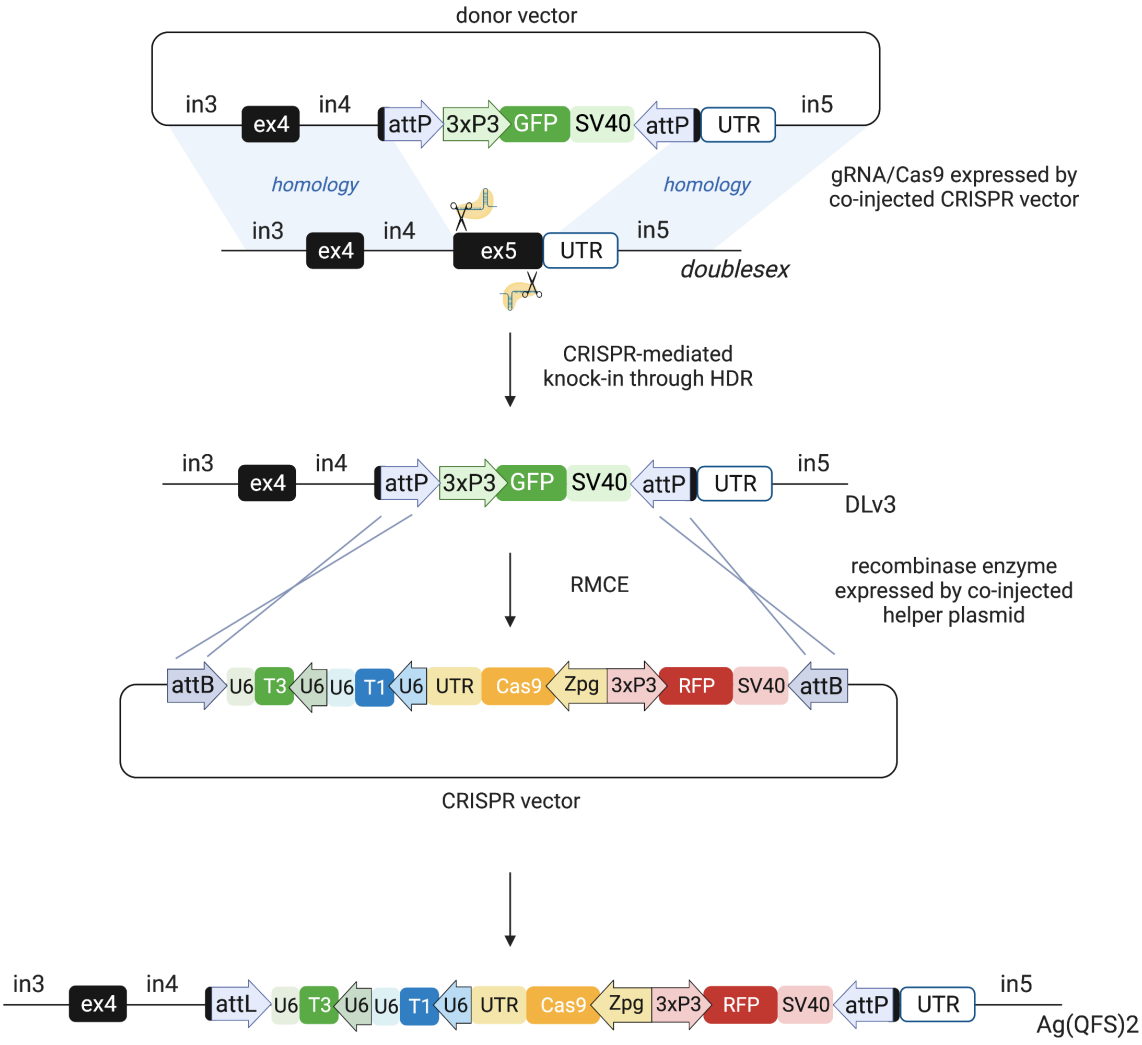


Figure 4.28: Pipeline to obtain a multiplexed gene drive strain. First a docking line was obtained through CRISPR targeting. The docking line (DLv3) contains an attP-flanked GFP cassette in place of a 76 bp deleted region between the T1 and T3 cut sites. The attP sites then allowed the RMCE of the gene drive construct that is flanked by attB sites, provided that a helper plasmid expressing a recombinase enzyme is provided. Upon RMCE the resulting gene drive strain, here Ag(QFS)2 is shown, contains a 3xP3-DsRed marker, a zpg-Cas9, and two U6-gRNA cassettes expressing gRNAs against the T1 and T3 sites, and is flanked by ruminant attL sites.

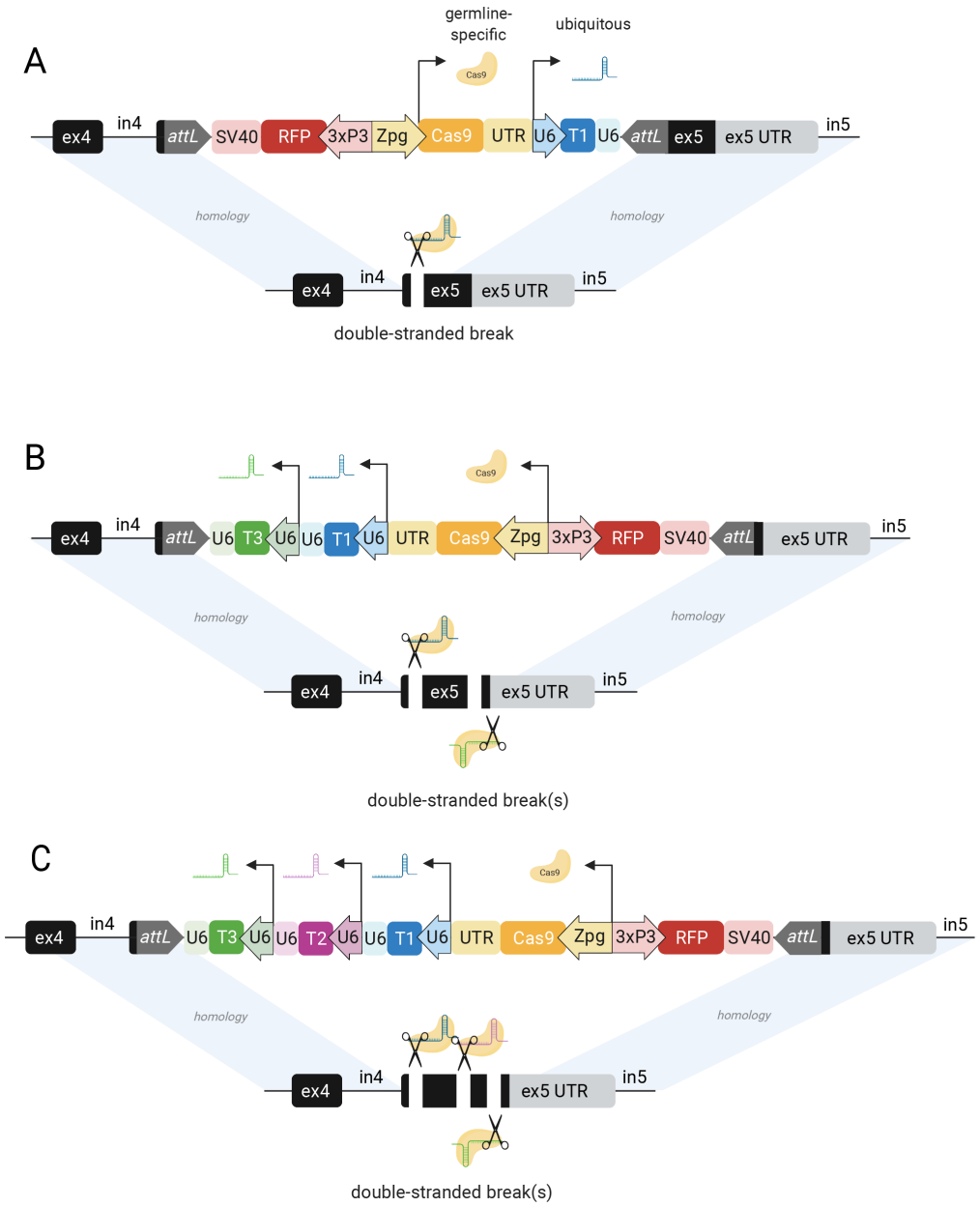


Figure 4.29: Schematics of the gene drive constructs targeting the *dsx* gene. (A) Ag(QFS)1, gene drive construct integrated in the same orientation as the *doublesex* gene. **(B)** Ag(QFS)2, gene drive construct integrated in the reverse orientation with respect to *doublesex*. **(C)** Ag(QFS)3, gene drive construct integrated in the reverse orientation with respect to *doublesex*. Components: attL = RMCE ruminant attachment sites, SV40 = viral terminator, RFP = DsRed fluorescent marker, 3xP3 = neuronal promoter, zpg = *zero population growth promoter* and UTR, Cas9 = human codon optimised *Streptococcus pyogenes* Cas9 (SpCas9), U6 = pol III promoter and terminator, T1/2/3 = gRNAs containing spacers complementary to sites T1/2/3 respectively and the same scaffold employed by Kyrou et al. (2018).

4.4.3 Multiplexed gene drives show equivalent or improved transmission rates

Transmission rates of multiplexed gene drives

The inheritance rates of each of the novel gene drive constructs: Ag(QFS)2 and Ag(QFS)3 were measured and compared to the transmission rates of Ag(QFS)1 (Figure 4.30A). Briefly, inheritance rates were determined from the fraction of RFP+ progeny of male (M) and female (F) gene drive heterozygotes that had inherited the gene drive either paternally (M→M, M→F) or maternally (F→M, F→F), when mated to wild-type.

Overall, multiplexed gene drives were inherited in similar or slightly greater ratios than Ag(QFS)1. Specifically, Ag(QFS)2 showed significantly higher rates of inheritance in the progeny of males ($p=0.0004$ for M→M and $p=0.0067$ for F→M) (Figure 4.30B). Ag(QFS)3 also showed increased transmission rates in the progeny of males that inherited the gene drive maternally ($p=0.0057$) (Figure 4.30C). These increases in transmission frequency might be due to the increased number of gRNAs offering a high likelihood of cutting and homing initiation at one of the targeted sites. In all other cases, the multiplexed gene drives were inherited in equivalent rates as Ag(QFS)1, when homing into a wild-type chromosome (Figure 4.30).

Resistant mutations do not affect transmission of multiplexed gene drives

To assess whether Ag(QFS)2 can bias its inheritance by solely using the novel T3 site to initiate homing, it was crossed to strains that either contained a GFP cassette or an R1 allele blocking the T1 site. Transheterozygote offspring were then crossed to wild-type. Indeed, Ag(QFS)2 was still able to produce high transmission rates, by targeting T3 alone (Figure 4.30B). This was despite the requirement for DNA resection of 1.4 kb in the case of the GFP cassette and 86 bp in the case of the R1 allele, demonstrating that different types of resistance at T1 can be effectively mitigated by multiplexing.

Tandemly repeated gRNA cassettes reduce construct stability

Construct stability was also tested in gene drive carriers that inherited the gene drive paternally, both for Ag(QFS)2 and Ag(QFS)3. We found that having multiple U6-gRNA expression cassettes, repeated in tandem (Figure 4.29), induced construct instability that was exacerbated with increasing number of gRNAs. Specifically, we observed a constant 1% (1/96) rate of T1 gRNA loss over 3 generations in Ag(QFS)2, and a 5% (2/43) T1+T2 gRNA loss and 2% (1/43) rate of gRNA duplication in Ag(QFS)3.

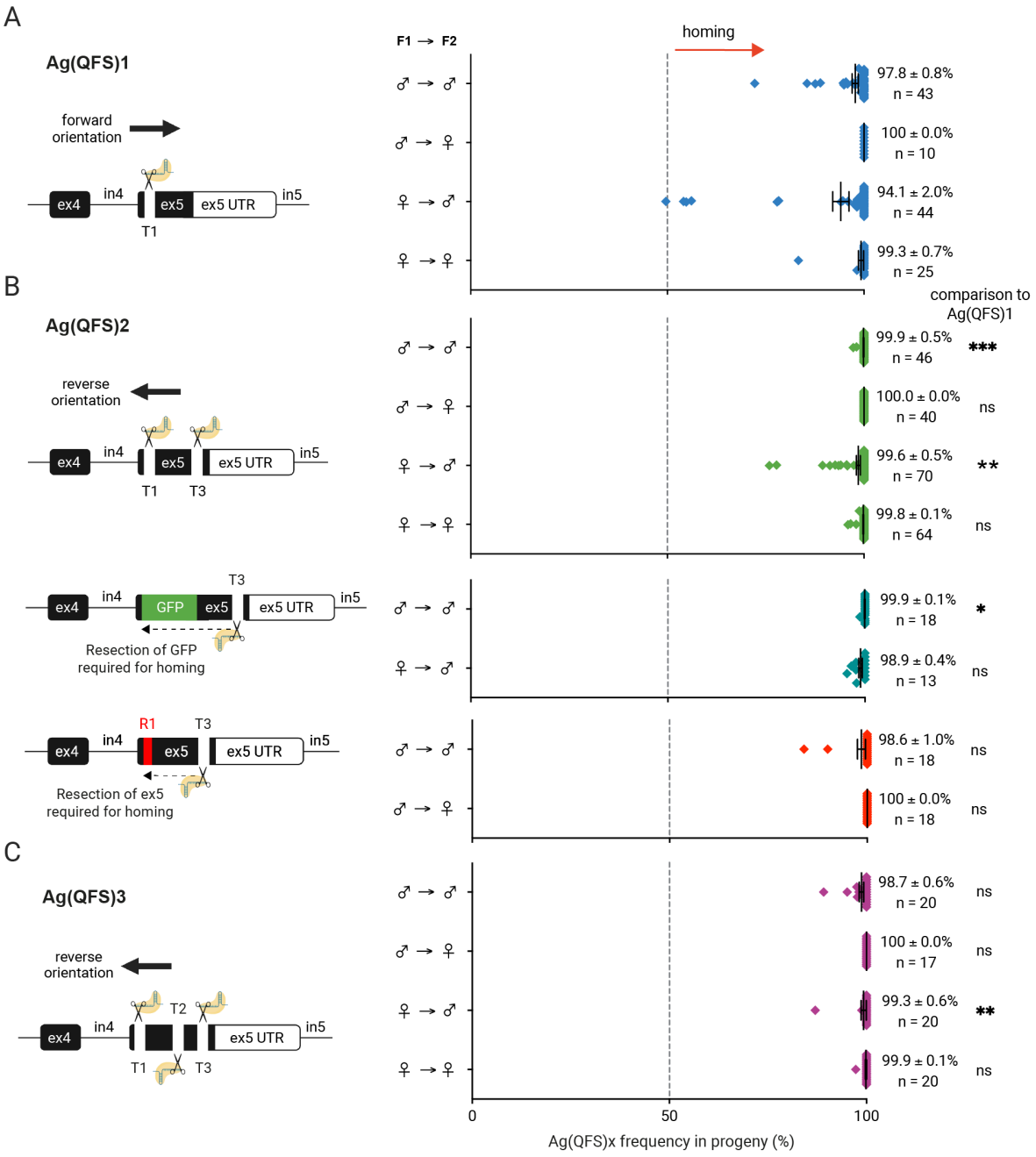


Figure 4.30: Transmission rates of multiplexed gene drives Ag(QFS)2 and Ag(QFS)3 in comparison to Ag(QFS)1. (A-C) Schematics to the left show the number and location of cut sites at each of the three putative targets (T1, T2 and T3) on dsx exon 5, depending on the gene drive (Ag(QFS)1, Ag(QFS)2 or Ag(QFS)3) and target region (wt, *dsxF*-, or R1), as well as the orientation of each gene drive construct with respect to the *dsx* gene, to correspond to the graphs on the right. Graphs show the Ag(QFS)1 frequency in the progeny of gene drive heterozygotes crossed to wild-type: (A) QFS1/wt (blue), (B) QFS2/wt (green), QFS2/*dsxF*-(GFP+) (teal), or QFS2/R1 (red), and (C) QFS3/wt (violet). The sex of the parents and grandparents of the progeny scored is shown to the left. For example: M→F means that the frequency of Ag(QFS)x inheritance in the progeny of heterozygous gene drive females that inherited the drive from males is shown. Mean Ag(QFS)x transmission rates and the standard error around the mean (S.E.M.) are shown to the right of the graph, together with the sample size (n) and Kruskall-Wallis statistical comparison to the QFS1/wt control (blue): (B) for QFS2/wt, $***$: significant with p-value = 0.0004, $**$: significant with p-value = 0.0067, ns: non-significant with p-value $>$ 0.9999; for QFS2/dsxF-, $*$: significant with p-value = 0.0206, ns: non-significant with p-value $>$ 0.9999; for QFS2/R1 ns: non-significant with p-value = 0.1803 and p-value $>$ 0.9999; (C) for QFS3/wt, ns: non-significant with p-value $>$ 0.9999, $**$: significant with p-value = 0.0057.

4.4.4 The fitness of multiplexed gene drive carriers is improved

Fecundity of multiplexed gene drive carriers

Next, the fecundity of multiplexed gene drive carriers was measured and compared to wild-type (Figure 4.31). Heterozygous gene drive carriers of both sexes that inherited the gene drive paternally or maternally were crossed to wild-type and the progeny of each female were counted as eggs and larvae hatched (Figure 4.31A-B, D-E). The hatching rate of the eggs was also calculated (Figure 4.31C, F).

Ag(QFS)2 females that inherited the drive paternally (M→F) showed reduced egg and larval output compared to wild-type, however this difference was not statistically significant (Figure 4.31A-B). The offspring of female gene drive carriers (both M→F and F→F) also showed reduced hatching rates compared to wild-type (Figure 4.31C).

Though male gene drive carriers should in theory be as fit as wild-type, as demonstrated by Kyrou et al. (2018), Ag(QFS)2 males that inherited the drive paternally (M→M) were significantly more fertile in terms of both egg and larval output (Figure 4.31A-B), but showed equivalent hatching rates to wild-type (Figure 4.31C). This might be due to unintended differences in the maintenance of the wild-type and Ag(QFS)2 strain, that could have led to the development of fitter Ag(QFS)2 males.

Ag(QFS)3 showed greater fitness than the wild-type control (Figure 4.31D-F). This was determined to be due to poor maintenance conditions of previous generations of wild-type mosquitoes by the laboratory rota, as we have discovered that fitness effects endure for several generations. As a result, it is not appropriate to draw solid conclusions from these datasets. Nevertheless, there is an overall trend implying reduced fertility of females that inherited the gene drive paternally, as they produced significantly less larval offspring compared to all other gene drive parental groups (p-values: M→F vs. M→M = 0.0280, M→F vs. F→M = 0.0317, M→F vs. F→F = 0.0424), in agreement with paternal effects suggested by Kyrou et al. (2018), Simoni et al. (2020).

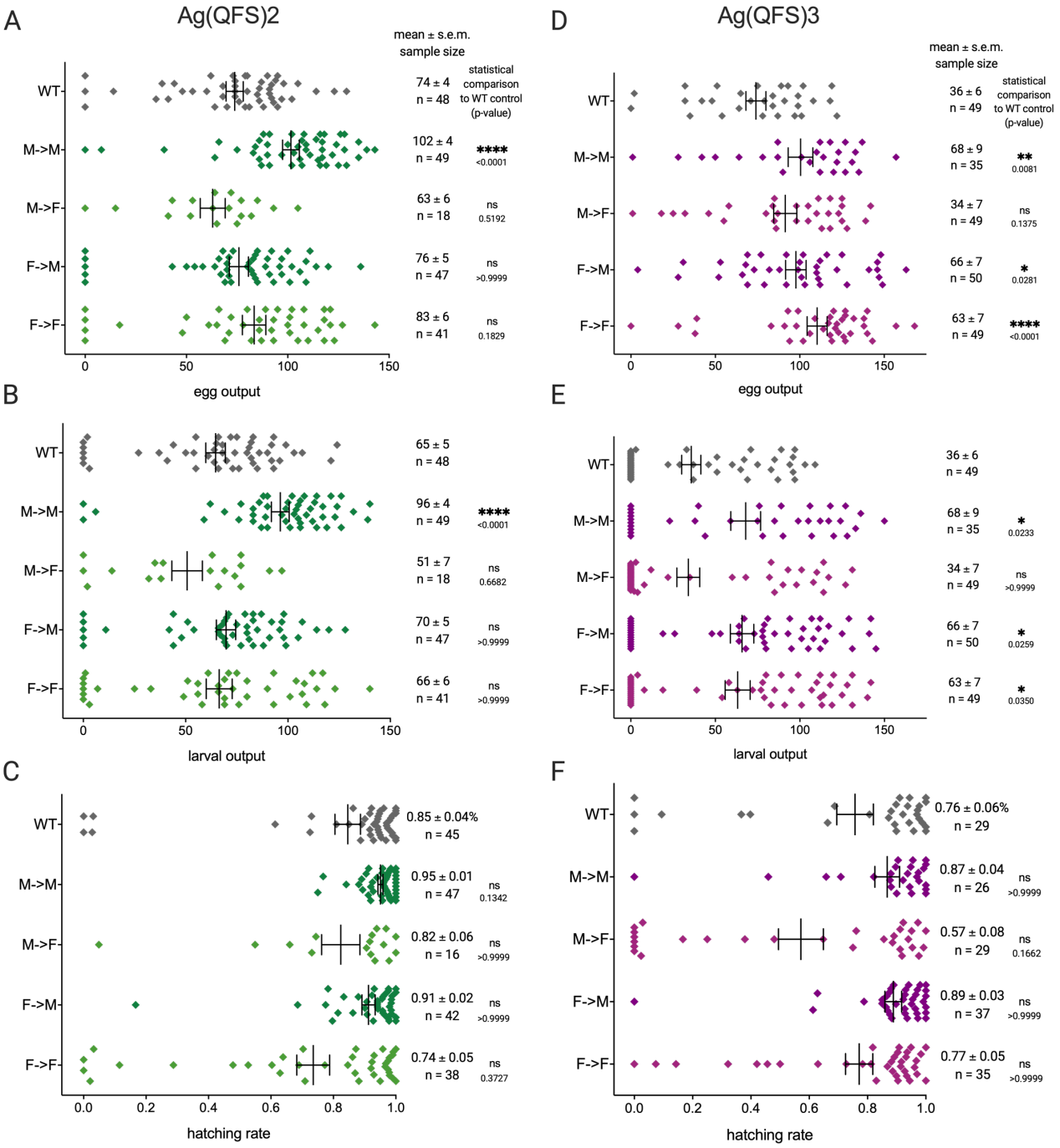


Figure 4.31: Fitness of Ag(QFS)2 (A-C) and Ag(QFS)3 (D-F) multiplexed gene drive carriers compared to wild-type. Male (M) and female (F) heterozygous gene drive carriers that inherited the drive paternally (M→M, M→F) or maternally (F→M, F→F) were crossed to wild-type, and the number of eggs (A, D) and larvae (B, E) produced per female parent were scored. The hatching rate of the eggs is also shown (C, F). For example: M→F means that the progeny of drive females that inherited the gene drive paternally were scored. Mean values, the standard error around the mean (S.E.M.), together with the sample size (n) and p-values derived from Kruskall-Wallis statistical comparisons to the wild-type control are shown to the right of each graph.

Ag(QFS)2 shows improved fitness compared to Ag(QFS)1

To be able to directly compare the fitness of Ag(QFS)1 and Ag(QFS)2 females, another phenotype assay was performed: Ag(QFS)1 and Ag(QFS)2 females that inherited the gene drive paternally (M→F) or maternally (F→F) were crossed to wild-type males and were assessed in terms of: egg and larval output, egg hatching rates, mating status, and PBM mortality, and compared to a simultaneous wild-type control.

Ag(QFS)1 females showed consistently lower fecundity than wild-type, in terms of egg and larval ouput, as well as egg hatching (Figure 4.32). Specifically, for Ag(QFS)1 females that inherited the drive paternally (M→F), egg output was 31% that of WT (significant, $p<0.0001$), larval output was 1% (significant, $p<0.0001$) that of WT, and hatching rates were 24% (significant, $p<0.0001$) that of WT. For Ag(QFS)1 females that inherited the drive maternally (F→F), egg output was 73% that of WT (non-significant, $p=0.3331$), larval output was 46% (significant, $p=0.0003$) that of WT, and hatching rates were 57% (significant, p=0001) that of WT.

Though Ag(QFS)2 females showed slightly reduced fecundity compared to wild-type, the differences were not statistically significant (n.s.)(Figure 4.32). Specifically their egg output was 87% for M→F ($p>0.9999$) and 89% for F→F ($p>0.9999$); their larval output 74% for M→F ($p=0.1846$) and 84% for F→F ($p>0.9999$); and their hatching rates 90% ($p=0.3751$) for M→F and 99% for F→F ($p>0.9999$), compared to WT.

Ag(QFS)1 and Ag(QFS)2 females were also directly compared to each other, to find that Ag(QFS)2 females were significantly fitter than Ag(QFS)1, which they were in most cases (Figure 4.32).

We also examined whether there are any differences in the fertility of gene drive females that inherited the gene drive paternally (M→F) vs. maternally (F→F). We found that Ag(QFS)1 females that inherited the gene drive paternally (M→F) showed reduced fecundity in all measures, though this difference was only significant in egg output ($p=0.0446$) (Figure 4.32). This paternal effect was largely reduced in Ag(QFS)2.

After noticing that Ag(QFS)1 females were dying post-bloodmeal, the rates of PBM mortality were also measured and compared between Ag(QFS)1, Ag(QFS)2 and WT (Figure 4.33). Ag(QFS)1 females showed higher PBM mortality than Ag(QFS)2 females, and in both cases females that inherited the gene drive paternally showed increased rates of mortality compared to those that inherited it maternally. Specifically, only 5% of wild-type females died PBM, whereas 14% of Ag(QFS)2 F→F, 21% of Ag(QFS)2 M→F, 40% of Ag(QFS)1 F→F, and 54% of Ag(QFS)1 M→F died PBM.

Intersex mosaicism also appeared strongest in Ag(QFS)1 females that inherited the gene drive from males (F→M) which in some cases developed single claspers (Figure 4.34A). Visibly mosaic intersex individuals that were dissected were always unmated, and in some but not in all cases they carried unfertilised

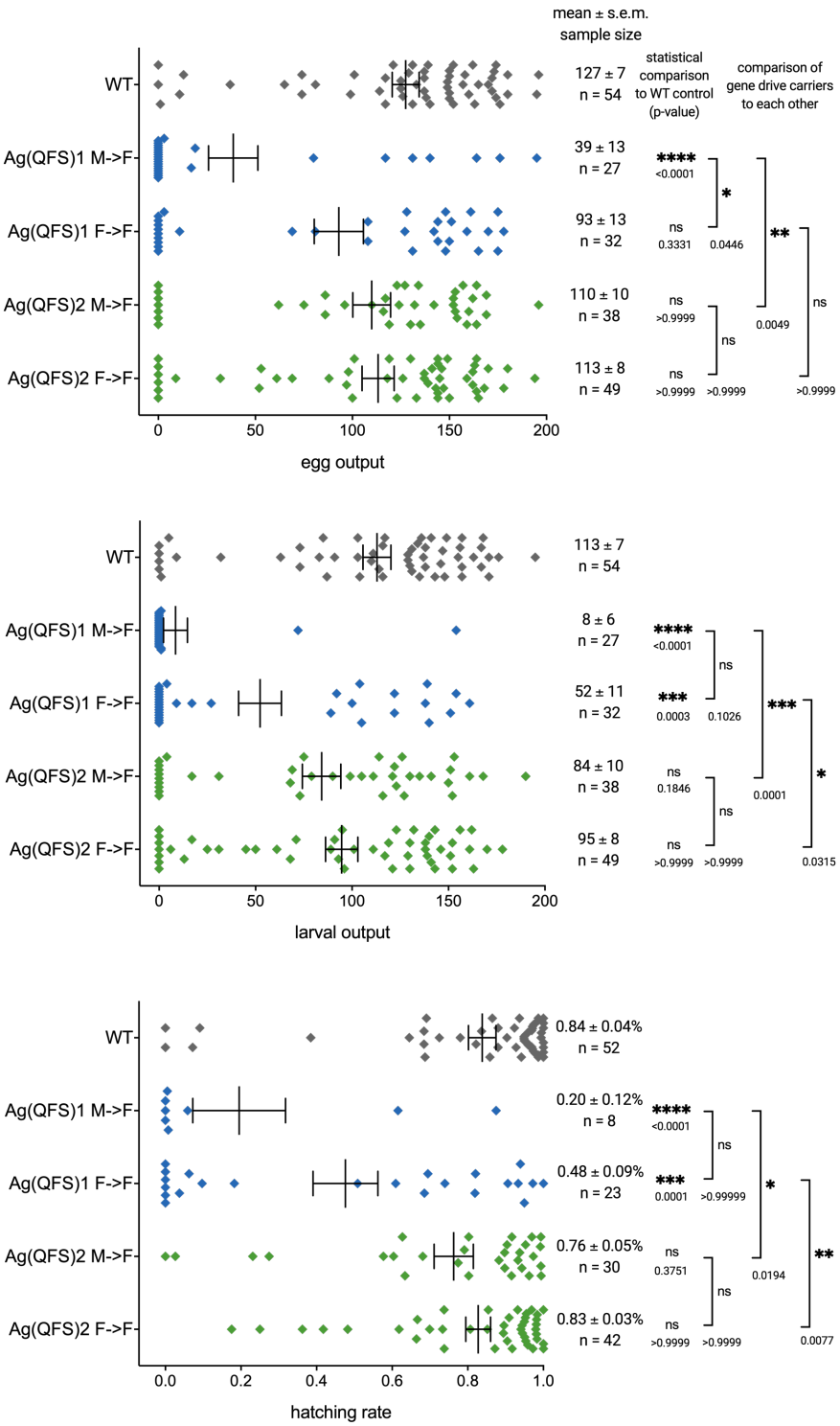


Figure 4.32: Fitness of Ag(QFS)1 vs Ag(QFS)2 females compared to wild-type. Female (F) heterozygous gene drive carriers that inherited the drive paternally (M→F) or maternally (F→F) were crossed to wild-type (WT), and the number of eggs (top) and larvae (middle) produced per female parent were scored. The hatching rate of the eggs is also shown (bottom). Mean values, the standard error around the mean (S.E.M.), together with the sample size (n) and p-values derived from Kruskall-Wallis statistical comparisons to wild-type control are shown to the right of each graph, and pairwise comparisons between specific groups as indicated by the brackets are also shown. Note that both mated and unmated females have been included in the analysis. However, dead females were excluded.

eggs (Figure 4.34A).

Taken together, the fitness of Ag(QFS)1 females is substantially lower than that of Ag(QFS)2 females, with a strong paternal effect that is more pronounced in the case of Ag(QFS)1. This is likely due to increased somatic Cas9 activity due to paternal deposition of the nuclease into the early embryo, causing a mosaic *dsx* knock-out intersex phenotype. Leakiness of the nuclease into somatic tissues of the adults might also be acting on top of this, increasing the levels of intersex mosaicism. Nevertheless, it was unclear what caused increased fecundity and viability of Ag(QFS)2 over Ag(QFS)1.

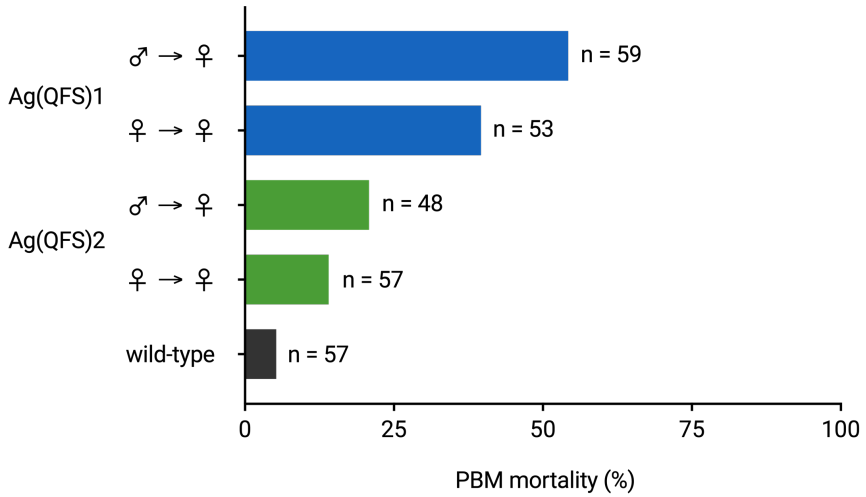


Figure 4.33: PBM mortality of Ag(QFS)1 vs Ag(QFS)2 females. Female (F) heterozygous gene drive carriers that inherited the drive paternally (M→F) or maternally (F→F) were crossed to wild-type (WT), and given a blood-meal 5 days after being allowed to mate. 2 days later they were allowed to lay eggs and 3-4 days later they were scored for mortality, and compared to a wild-type control. Note that the vast majority of females that died PBM did not lay eggs.

Genomic orientation impacts upon fitness of female gene drive carriers

The two main differences between the two Ag(QFS)1 and Ag(QFS)2 constructs is the fact that the latter cleaves two sites, instead of one site on the *dsx* exon 5 and that they are integrated in the opposite orientation in the genome[6]. Ag(QFS)1 is integrated to interrupt the female-specific exon 5, in the same orientation as the *dsx* gene, whereas Ag(QFS)2 is integrated in the same location, but in reverse with respect to the *dsx* gene[7].

We were able to test whether the underlying cause of the fitness disparity between the strains was a product of gene drive orientation or another function of multiplexed gene drives, as we obtained 5 independent integrations of the same Ag(QFS)3 gene drive construct, in both orientations: 3 in the reverse

[6]Note that the orientation here refers to directionality of the gene drive construct's Cas9 expression cassette.

[7]Technically there is a third difference between the two constructs, which is the fact that in the case of Ag(QFS)2 the region between the target sites T1 and T3 are also missing in the gene drive-carrying chromosome.

and 2 in the forward orientation with respect to *dsx*. It was therefore possible to test whether orientation of the gene drive can impact upon rates of intersex mosaicism. Intersex mosaicism can be visually detected in the pupal stage, but not as easily in the adult stage, as only extreme cases of mosaicism show externally, through development of claspers (Figure 4.34A). Female pupae that are mosaic develop abnormal male-like genitalia (Figure 4.34B), which is an intermediate phenotype between the fully female and fully knock-out intersex anatomy (full intersex are indistinguishable from males as pupae). The female offspring of male gene drive carriers containing constructs of Ag(QFS)1, Ag(QFS)2 and Ag(QFS)3, integrated in both orientations to interrupt the *dsx* female-specific exon 5, were thus scored for intersex mosaicism by inspecting their genitalia at the pupal stage. Gene drives integrated in the same/forward (fwd) orientation with respect to *dsx*, including Ag(QFS)1 and two Ag(QFS)3 strains, were approximately 10 times more likely to develop mosaic intersex genitalia (p-value = 0.0104), compared to those integrated in the reverse (rev) orientation, including Ag(QFS)2 and three Ag(QFS)3 strains (Figure 4.34C). These results signify that genomic orientation of the gene drive construct, and specifically its Cas9 expression cassette, plays an important role in off-target mutagenesis of somatic tissues, causing intersex mosaicism and associated sterility.

4.4.5 Ag(QFS)2 cage trials

To validate that Ag(QFS)2 can efficiently invade laboratory populations two replicate cage trials were performed. Replicating the experimental setup of Kyrou et al. (2018), Ag(QFS)2 gene drive males were released into a mixed wild-type population at a 25% frequency (12.5% initial allelic frequency). In both replicate populations the Ag(QFS)2 gene drive reached fixation (i.e. 100% gene drive frequency in the population) after 6-7 generations (Figure 4.35), as opposed to 7-11 generations for Ag(QFS)1 (Kyrou et al. 2018). Phenotypic females were completely eliminated by generations 7-8, and since all remaining intersex females were sterile the populations crashed (Figure 3.8). Note that when Ag(QFS)2 cages reached 100% gene drive frequency (i.e. in the generation prior to complete population elimination), they had an increased egg output, by 46% (5069 eggs, cage 1, generation 7 and 4313 eggs, cage 2, generation 6) compared to Ag(QFS)1 (2645 eggs, cage trial 1, generation 11 and 1668 eggs, cage trial 2, generation 7) (4.35). This is potentially owing to the higher fertility of Ag(QFS)2 female heterozygous gene drive carriers (4.32).

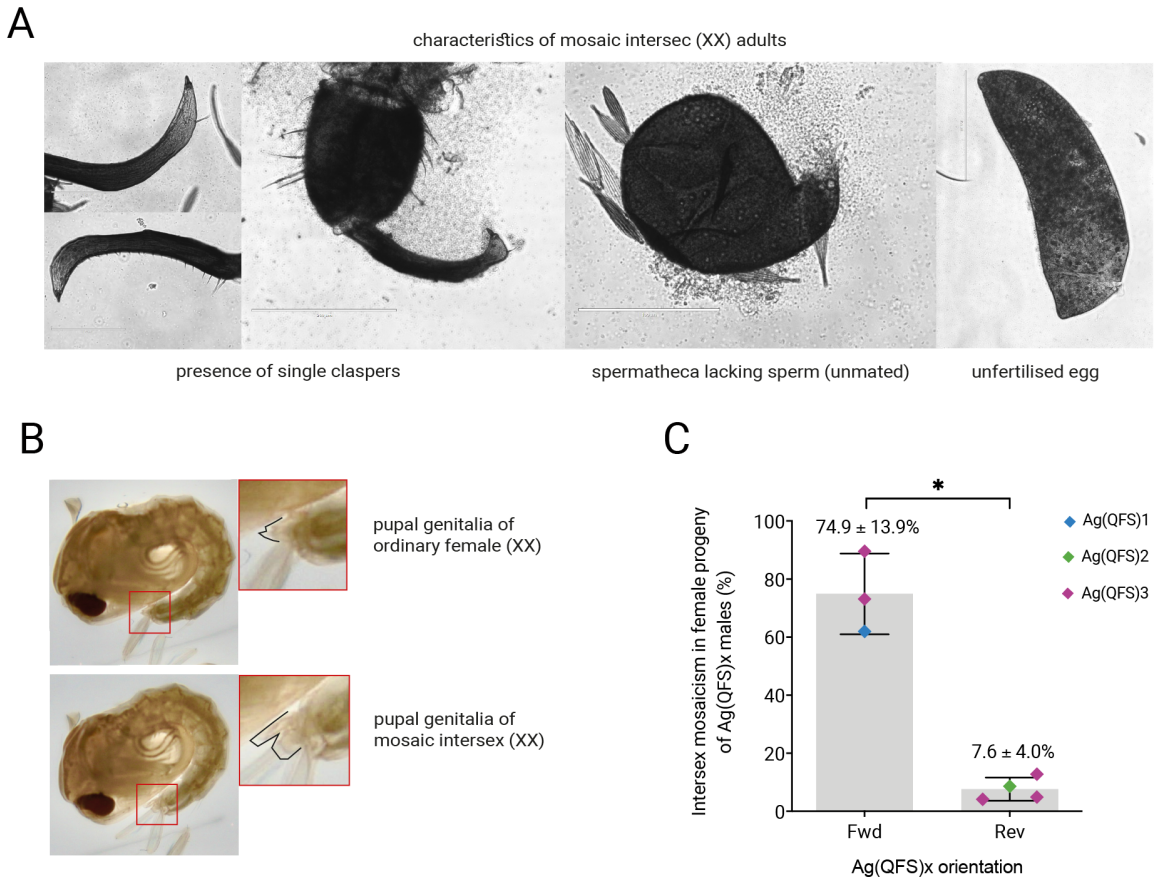


Figure 4.34: Intersex mosaicism in female gene drive carriers. (A) Phenotypic characteristics of intersex mosaic Ag(QFS)1 adult females that inherited the gene drive from males. **(B)** Example of ordinary versus mosaic intersex female genitalia. **(C)** Percentage intersex mosaics in the female offspring of male gene drive carriers that harboured the gene drive in the forward (fwd) or reverse (rev) orientation with respect to *dsx*. Means and standard deviations are shown above the graph. Error bars indicate the amount of standard deviation. A t-test, with Welch's correction to account for unequal standard deviation in the two datasets, was performed (p-value = 0.0104) to compare between the rates of mosaicism in females harbouring a gene drive in the fwd vs. rev orientation.

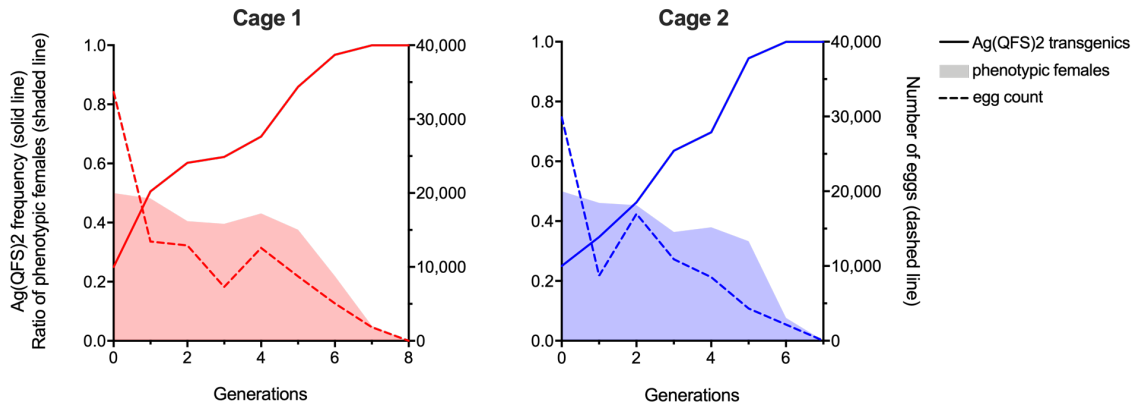


Figure 4.35: The Ag(QFS)2 gene drive achieved vomplete population elimination in two duplicate cages. Ag(QFS)2 heterozygous males that inherited the gene drive paternally, were released in a wild-type population at a 25% frequency (12.5% allelic frequency) in two replicate cages, and the presence of Ag(QFS)2 (solid lines) was tracked over time and for each distinct generation, until the caged populations were eliminated at generation 8 and 7 for Cage 1 and 2, respectively. The proportion of phenotypic females (as opposed to mosaic intersex, fully intersex and males) (shaded regions), as well as the egg output (dashed line) per generation was also recorded. 650 eggs were always kept and used to seed each subsequent generation.

Chapter 5

Discussion

5.1 Resistance to gene drive

Gene drives have incredible potential to control insect vectors of disease. So far, the greatest scientific obstacle in implementing gene drives to the field is the possibility that resistance will evolve against them, causing gene drive spread and its associated impact in limiting disease transmission to be reversed.

CRISPR-based gene drives that rely on homing risk resistance from naturally occurring or drive-induced sequence variants at the target site (outlined in section 1.6.1, Figure 1.10). The selection pressure for resistance is inversely correlated with the cost of the drive, and thus, resistance is a greater obstacle for suppression drives compared to those aimed at population replacement. Nonetheless, target site resistance has been quick to arise against replacement drives as well, as it is very hard to target a genomic locus that is non-variable and neutral at the same time.

A method to overcome this was proposed by Esvelt et al. (2014) who suggested targeting a highly conserved gene, but including a recoded version of it (i.e. containing a nucleotide sequence that encodes the same amino acid sequence as the native gene, whilst it is no longer recognised by the gRNA(s) of the gene drive) in the replacement drive construct to minimise the fitness cost imposed by disrupting a genetically encoded gene. The feasibility of this strategy was demonstrated by Adolfi et al. (2020), whereby invasion and genetic replacement of caged populations proceeded even in the presence of functional target site resistance (i.e. **R1** alleles). Nonetheless, recoding a gene is not a simple endeavour. The most highly conserved genes like *doublesex* are also conserved in the nucleotide level, and so recoding them might disrupt mRNA processing reliant upon a specific nucleotide sequence. Recapitulating the expression of the native gene, to ensure correct function, can also be challenging as re-inserting it out of context in a different genomic location, containing a different set of regulatory sequences, might alter its expression (Chen & Zhang 2016).

Irrespective of the gene drive approach, several measures must be taken to avoid the rapid creation and selection of target site resistance. So far, most efforts have been concentrated on restricting the spatiotemporal expression of the Cas9 nuclease to minimise the introduction of EJ mutations at the target site (Hammond et al. 2021a, Carballar-Lejarazú et al. 2020, Kyrou et al. 2018), and targeting highly conserved sites to minimise R1 mutation outcomes (Kyrou et al. 2018, Fuchs et al. 2021, Adolfi et al. 2020, Carrami et al. 2018) (see section 1.6.1). Even so, most population suppression gene drives have forced the generation of resistant alleles, even at sites with presumably high functional constraint (Carrami et al. 2018, Fuchs et al. 2021). The only population suppression gene drive that has been able to spread to fixation and eliminate caged populations, without selection of resistance against it, was the Ag(QFS)1 gene drive (Kyrou et al. 2018, Hammond et al. 2021b), and a variation thereof, in which the homing drive at *dsx* was combined with an X-shredding element (Simoni et al. 2020).

5.2 The resistance landscape to gene drives targeting the *doublesex* gene

5.2.1 Detection and evaluation of natural and Cas9-induced variants

Despite Ag(QFS)1's success in overcoming target site resistance by targeting a highly conserved site on *dsx*, all testing was performed in caged laboratory populations of ∼600 mosquitoes (Kyrou et al. 2018, Hammond et al. 2021b), and it remained unclear whether the gene drive would be as successful in natural populations that are several orders of magnitude larger (Figure 2.1) (Khatri et al. 2019). The first two aims of this thesis aimed to address this directly (see 2.0.1, and 2.0.2). First, by querying existing population genomics data for the presence of natural variants that could be resistant to gene drive (see section 4.1), and second, by designing a high throughput assay to generate a high number of Cas9-induced mutations at the gene drive target site, and individually assess their potential to encode a functional copy of DSX (see section 4.2).

The population analysis performed as part of the first aim of this PhD (section 2.0.2) revealed several low frequency SNPs along the CDS of the exon 5 of *dsx* (Figure 4.1) (see section 4.1). Amongst them, a single G→A variant, on the Ag(QFS)1 target site, was found at a high frequency in Angola (∼26%) at Hardy-Weinberg equilibrium (see 4.1.2). This implies it is under no selection pressure in that region and so it is presumed to be functional[1].

It is unclear whether any of the other discovered SNPs are functional or not. Amongst them there was 1 intronic SNP, 4 silent SNPs, and 1 SNP that causes a valine-to-methionine amino acid substitution, in a codon position that varies in three other anopheline species (Figure 4.1). However, it remains to be seen whether they are fully compatible with normal development, and if they can prevent gene drive cleavage.

Moreover, additional variants are highly likely given the relatively small dataset of the existing analysis, limited to ∼3,000 wild-caught mosquitoes[2], with bias towards certain collection sites, rather than uniform sampling across geographically distinct sites in Africa. For example, sampling from Nigeria, where most malaria cases occur is completely absent, whilst sampling from geographically unique regions within Central African countries (Angola, DRC, Zambia) is limited. The fact that the G→A natural variant was only found in Angola, the DRC, Gabon and Cameroon in proximal geographic locations, whilst absent from all other datasets, shows the need for sampling populations that exist in ecologically distinct niches. This can be facilitated by next-generation sequencing (NGS) methods to perform whole genome sequencing (WGS) of pooled samples or targeted short-read Illumina sequencing of the regions of interest (Davey et al. 2011).

[1] Note that being at high frequency only in a single region of Africa may imply that the variant carries no cost only in certain geographies.

[2] Though more genomes are to be made publicly available on the MalariaGEN website next year, making a total of ∼10,000 malaria vector genomes.

Moving forwards, to assess the resistance landscape at pre-defined gene drive target sites, two types of resistant screens were designed. First, an assay was developed to enrich for functional resistant mutations, that makes use of the gene drive strain of interest (Figure 4.6). This assay relies upon crossing of heterozygous gene drive carriers to each other, and automated screening to select the portion of their offspring that contain non-drive chromosomes previously exposed to nuclease activity, and that might therefore contain an EJ mutant, balanced across a copy of the gene drive (exposed/drive - e/d) that knocks out the target gene. By assessing the phenotype of e/d females, one can therefore determine which mutations are functional. For female fertility genes this process can be further automated, for example, by allowing all sorted e/d females the chance to mate and give offspring. In this case only mutations that are functional and not cleavable by the gene drive (d) will pass to the next generation, whilst non-functional EJ mutations will render e/d females sterile and get filtered out. Therefore, by sequencing both e/d mothers and their offspring it is possible to distinguish between functional (R1) and non-functional (R2) resistant mutations. This was demonstrated using the 7280 gene drive developed by (Hammond et al. 2021a), whilst Fuchs et al. (2021) also used a similar approach to assess an adult lethality gene drive targeting a highly conserved site on the AGAP029113 gene.

The described pipeline is relatively simple and can yield efficient results (Figure 4.6). It can be applied to gene drive strains to assess their efficacy, prior to more labour-intensive and time-consuming testing, such as cage trials. Additionally, because the resistance screening process is largely automated it can allow the screening of much larger populations (tens of thousands), in contrast to cage trial experiments that have so far been limited to sstudying populations of ∼600 mosquitoes.

Nonetheless, this assay is reliant upon the existence of gene drive strains in the first place, and cannot be used to assess putative target sites prior to strain development (Figure 4.6). Moreover, it relies upon full or at least good levels of fertility of gene drive heterozygotes which has not been the case for the majority of gene drives aimed at population suppression (Hammond et al. 2016). This can be overcome by adopting a strategy like the one described by Fuchs et al. (2021), where males heterozygous for the drive are crossed to females carrying one copy of a knock-out allele of a haplosufficient gene that should not bear a fitness cost. However, this requires the utilisation of an extra strain that might not be available, and in this way information about the cleavability of EJ variants is lost.

Finally, this pipeline is difficult to adapt for gene drive showing extremely high rates of homing like Ag(QFS)1, as it relies on the filtering of the relatively small fraction of individuals containing "unhomed" alleles (Figure 4.6). The smaller their portion the less EJ mutants can be investigated. Indeed, attempting this with Ag(QFS)1 was not fruitful, due to the gene drive's near perfect homing, leading to the production of a very low amount of non-drive carriers per generation for the assay to be useful (Kyrou et al. 2018).

In order to assess the Cas9-induced resistance landscape to Ag(QFS)1 we designed another high-

throughput assay by exploiting the high rates of EJ caused by embryonic activity of maternally deposited nuclease (Figure 4.10). By forcing the generation of Cas9-induced EJ mutations using the *vasa2* promoter (Papathanos et al. 2009, Hammond et al. 2016, Hammond et al. 2021a), and adapting the use of COPAS for high-throughput screening (Marois et al. 2012), the screening of over 50,000 edited individuals was made possible, and isolation of over 10,000 females containing a putatively resistant allele balanced across the female-specific *dsx* null mutation (*dsxF-*) from Kyrou et al. (2018). This allowed the characterisation of 9,096 mutated alleles, bypassing the need to interrogate more than half a million insects that would otherwise be required to assess the same amount of mutants, in the context of a gene drive, like Ag(QFS)1 (Figure 5.1).

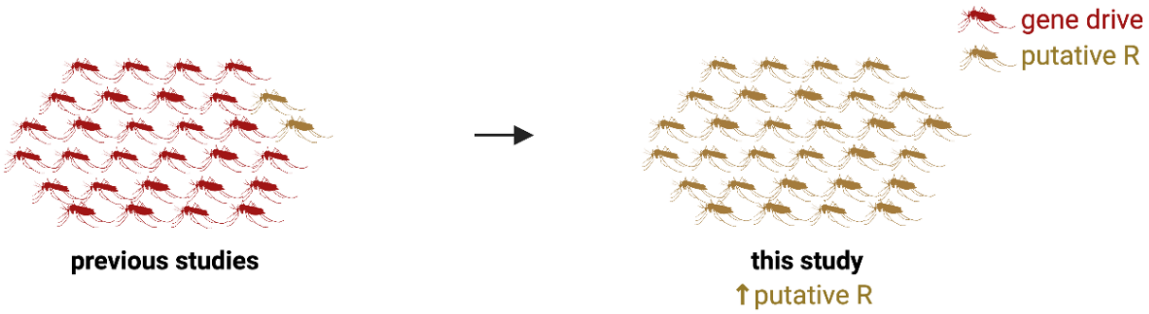


Figure 5.1: The number of putatively resistant alleles characterised in the present study greatly exceeds the number of alleles investigated in previous studies that tracked resistant alleles in the context of single generation gene drive crosses, or even multi-generational cage trials.

The vast majority of EJ alleles generated at the Ag(QFS)1 target site consisted of non-functional R2 mutations. This was the case even for in-frame indels, in contrast to what was previously observed when targeting the less functionally constrained site on the 7280 gene (Hammond et al. 2017), that could readily tolerate in-frame deletions of 3-6 bp, and even larger 12 bp deletions (Figure 4.6).

In general, in-frame mutations were rarer at the T1 site of *dsx*, compared to the target site at 7280, probably as a result of the nature of the micro-homologies surrounding the target site that were not in multiples of 3, in contrast to those at 7280. This raises the possibility of predicting MMEJ repair outcomes *in silico* for improved gen edrive target site selection (Kranjc 2022).

The only in-frame mutations that showed signs of functionality at *dsx* were rare nucleotide substitutions that collectively arose at a frequency of $3*10^{-3}$-$4*10^{-3}$ (Figure 4.12). Females containing the nucleotide substitutions described in Figure 4.12, developed normal external anatomy and were able to blood-feed (Table 4.2). Surprisingly, the substitutions were not silent (Figure 4.12). This might indicate that conservation of specific nucleotides, might be more important than conservation on the amino acid level, at the intron-exon boundary of the female-specific exon of *dsx*, potentially due to binding of splicing factors that modulate correct expression of the female-specific *dsx* variant. Even so, their absence from natural populations may indicate a cryptic fitness cost that could not be measured in our experiments.

What remains unclear is whether the ∼9,000 EJ alleles interrogated represent 9,000 distinct mutagenesis events, or not. Indeed we found that mutations are likely generated early on in embryogenesis, since each mutated male parent only contributed 2.75 distinct mutations to its offspring, on average, whilst the most dominant mutations tended to differ between males. and that most of the offspring therefore contain mutations that arose from the same mutagenesis event (Figure 4.11). This estimate is a minimum because the most common mutations, favoured by MMEJ repair outcomes, might have also arisen independently. Even so, if we assume that all 50 mutated males, from each of the 33 cages of the resistance assay, contributed 2.75 distinct mutations to the next generation, then we would expect that overall by performing this assay we examined ∼4,538 (=50∗2.75∗33) distinct mutagenesis events. Previous studies have looked at the generation of resistant alleles, in single generation gene drive crosses, or multi-generational cage trials, but never at this scale (Kyrou et al. 2018, Oberhofer et al. 2018, Terradas et al. 2021, Hammond et al. 2021a, Hammond et al. 2021b, Yang et al. 2022).

Moreover, we found that the mating behaviour of mutated males might have introduced biases in the types of alleles that were detected in their offsrping (Figure 4.14). If a similar assay were to be repeated in the future, it would benefit by including a higher number of replicate parental cages, but with a lower number of males per cage (e.g. 5-10), giving each male a better chance to mate, and avoiding several highly competitive males to dominate all matings. Another surprising finding is that the two most common SNPs discovered in anatomical females (C→T and G→T), as well as wild-type alleles, were also present in the sample of intersex individuals (Figures 4.10 and 4.14). This is likely due to mosaicism resulting from residual levels of *zpg*-Cas9 activitity from the ALE transgene in unmodified individuals, meaning that their presence in intersex does not imply the mutants are non-functional. This is further supported by the fact that only alleles that the only alleles that mutated anatomically female individuals carried were nucleotide substitutions (i.e. no indels). We further excluded the possibility that they are present in mosaicism with wild-type alleles by performing pooled amplicon sequencing of single mutant females (Figure 4.13), fortifying the hypothesis that these alleles may be functional.

Nonetheless, to minimise background noise in the future, gRNA-only strains should be crossed to *vas2*-Cas9 strains. The method could be further expanded by using multiple gRNAs spanning a single locus to see whether a combination of variants could restore function; or used in conjunction with a system like EvolvR or base editing that enables targeted mutagenesis within a tunable window, rather than placing focus around a single cut site (Halperin et al. 2018, Marr & Potter 2021), to screen for variants with complex phenotypes, such as insecticide resistance. Different promoters to *vas2* could also be considered. Ideally, one would choose a Cas9 promoter that causes high rates of EJ in the male germline, like the *beta2-tubulin* promoter. However, when this was attempted, the *beta2*-Cas9 gave insufficient levels of editing (40%) (from data I obtained not presented in this thesis). Finally, the assay could be further automated

by expressing the gRNA strain from the Y chromosome, meaning that sexing could also be automated by using the COPAS larval sorter (Figure 5.2). Though obtaining Y-linked (or male-linked) sexing strains strains is feasible (Bernardini et al. 2014, Lutrat et al. 2022), it is not straight-forward, because the Y chromosome is poorly (if at all) assembled in most species, and undergoes MSCI that shuts down gene expression in cells that undergo meiosis (Taxiarchi et al. 2019). Indeed when a Y-linked ALE was obtained in the lab it was not functional, and could not produce mutations even when crossed to the *vasa*-Cas9 strain, and despite visible expression of a DsRed reporter (Ignacio Tolosana, personal communication).

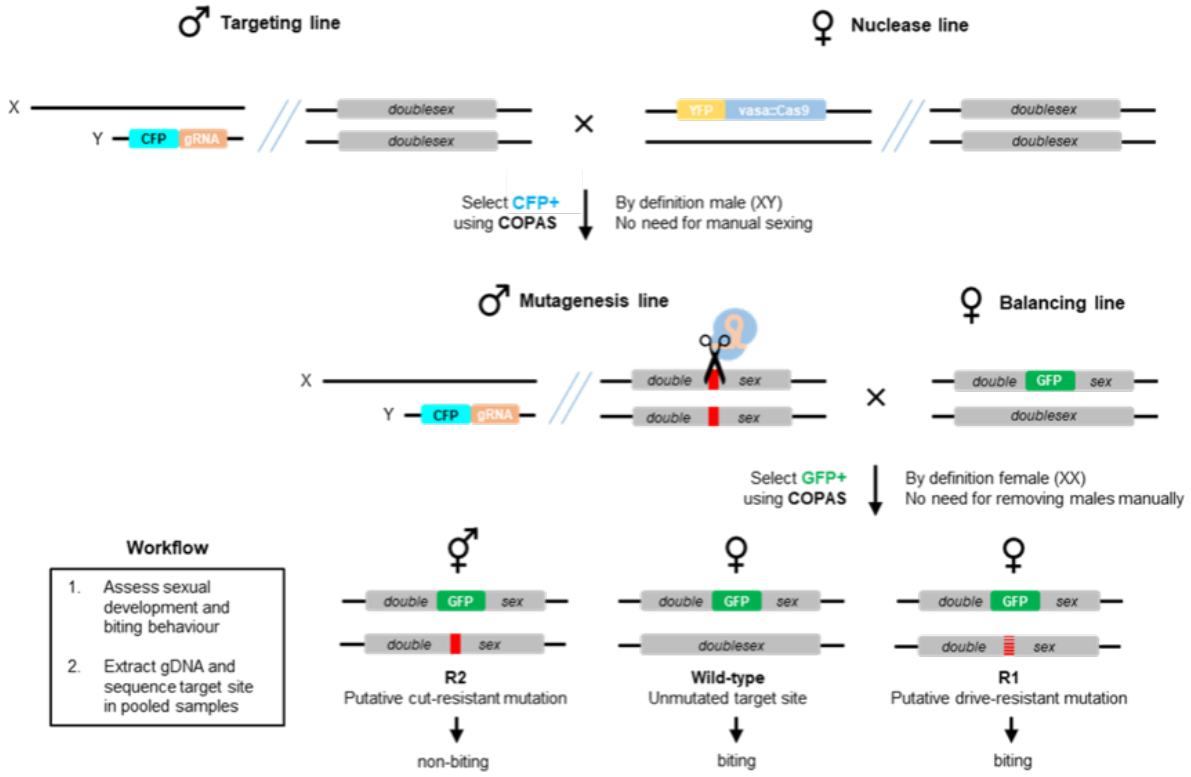


Figure 5.2: Increasing the throughput of the resistance assay by linking a gRNA-expressing transgene to the Y chromosome. By inserting an RFP-marked ubiquitously expressed gRNA on the Y chromosome, and use the fluorescence-based larval sorter COPAS, we can omit all requirements for manual sexing of mosquitoes and increase the throughput of the assay.

5.2.2 Engineering marker-less variants into the mosquito genome for testing

After the discovery of naturally-occurring and Cas9-induced variants that may be resistant to the Ag(QFS)1 gene drive, the next aim of this PhD was to engineer those in the mosquito genome to verify (a) their presumed functionality, and (b) whether they can block gene drive activity *in vivo*.

Simple and complex edits can be engineered in the genome of most organisms with precision using CRISPR by presenting a modified DNA template for HDR at a Cas9-induced break. Recently developed methods such as base editing and prime editing are ideal for the introduction of SNPs and other small edits (see 1.4.2), and they alleviate reliance on HDR that can be inefficient in many organisms. However, they have not been widely tested in insects and first attempts in *Drosophila* suggested that prime editing is no more efficient than HDR (Bosch et al. 2020), whilst base editing is effective but inherently imprecise (Marr & Potter 2021).

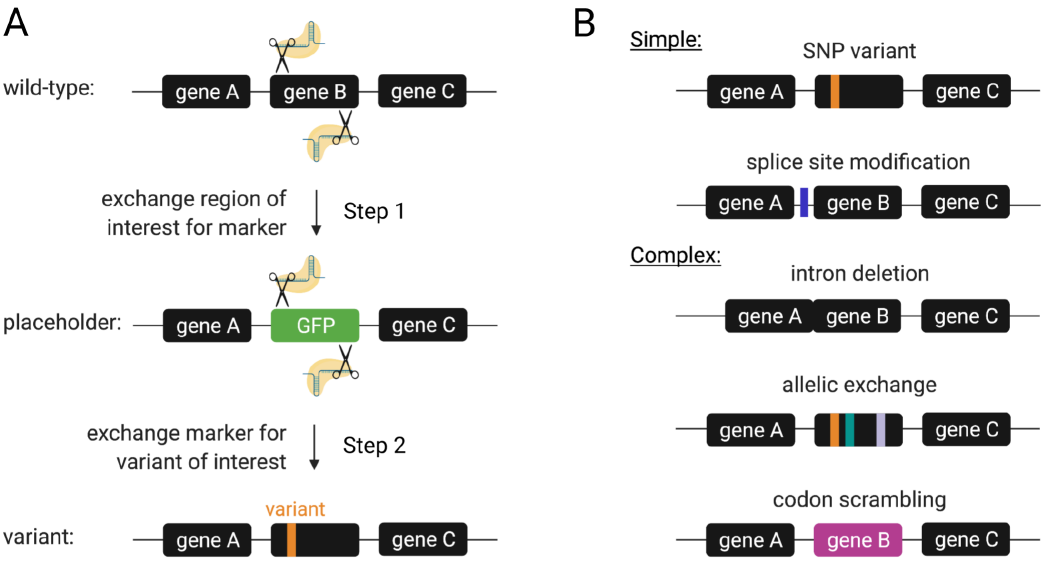
In insects, independent of the chosen technology, engineering small marker-less edits remains inefficient, with transformation rates rarely exceeding 5% (Bosch et al. 2020, Grigoraki et al. 2021, Kistler et al. 2015). Moreover, isolating transformants that lack a molecular marker is very inefficient, relying upon large numbers of single crosses and molecular identification of variants.

We developed a new strategy, that we termed CriMCE, for the purpose of engineering *dsx* variants into the mosquito genome, without the presence of a visual marker or any other molecular debris that could interfere with their function (Morianou et al. 2022).

CriMCE is based upon CRISPR-mediated cassette exchange of a marked placeholder for the variant of interest, to allow visual detection of the insertion and facilitate the engineering and isolation of marker-less edits (Figures 4.15 and 4.16). It therefore relies on HDR, unlike other methods for marked cassette exchange or removal, like recombinase-mediated cassette exchange (RMCE) and Cre-Lox recombination. This allows for comparatively high efficiency (when compared to RMCE) (Table 4.19), and uniquely traceless editing such that any phenotypic change can be attributed to the intended edit rather than ruminant attachment sites (Figure 5.3); whilst it is inherently precise, in contrast to base editing (Marr & Potter 2021).

The most common natural variant at the T1 site (G→A SNP) and the two most common Cas9-induced variants (C→T and G→T SNPs), were prioritised for testing. Using CriMCE we were able to engineer them into the mosquito genome, completely marker-less, after the first round of injections. The efficiency of the method was proven to be far greater than other methods for marker-less editing (Figure 4.20 and Table 4.19) (Kistler et al. 2015, Grigoraki et al. 2021, Bosch et al. 2020).

Surprisingly, rates of HDR-induced editing are high when marked mutations are introduced in *D. melanogaster* and *An. gambiae* (Table 4.19) (Gratz et al. 2014, Hammond et al. 2016), however drop sub-



C

	Visual identification	Scarless strain	Type of modification	Transgenesis steps	Generations to homozygosis (N)	Rearing effort	Generations sequenced
CriMCE	✓	✓	Simple & complex	2	3	medium	1-2
RMCE	✓	✗	Simple & Complex	2	3	medium	1-2
HDR	✗	✓	Simple & Complex	1	4	very high	3-4
base/prime editing	✗	✓	Simple	1	4	very high	3-4

Figure 5.3: CRISPR-mediated cassette exchange (CriMCE) is a two-step method for engineering, detection and isolation of marker-less edits. (A) Step 1: To generate a marked placeholder strain, the region of interest (gene B) is replaced by a marker (GFP, green). Step 2: The marker is replaced by the native sequence containing the variant of interest (orange), through CRISPR-mediated cassette exchange (CriMCE), to obtain a marker-less strain carrying the variant. **(B)** Examples of the types of simple and complex genetic modifications that can be obtained using CriMCE. **(C)** Comparison of CriMCE to other methods used to make precise genomic edits, including recombinase mediated cassette exchange (RMCE), direct homology-directed repair (HDR) of a wild-type sequence and base or prime editing.

stantially when SNPs are directly inserted into a wild-type genomic locus (Table 4.19) (Kistler et al. 2015, Grigoraki et al. 2021). Using CriMCE we achieved high rates of HDR editing consistent with those for marked transgene insertion (Table 4.19). In both cases, repair templates differ significantly from their target regions: transgenes introduced via HDR do not resemble their genomic target, while in the present study the wild-type target is replaced by a placeholder, which serves to differentiate it from the desired edit (Figure 4.16). Conversely, when direct HDR is used to induce small marker-less edits the repair template is almost identical to that of the wild-type target. It is unclear why sequence dissimilarity between the exogenous repair template and its target should boost the efficiency of editing, but perhaps it functions to shift repair away from using the unmodified homologous chromosome as a template.

By targeting CRISPR to non-coding regions of the placeholder, undesirable EJ events are also filtered out, as they are unlikely to affect marker expression, further boosting the efficiency of CriMCE, focusing molecular characterisation and rearing efforts. Indeed, no EJ mutations were observed in the fraction of marker-less individuals that were sequenced, indicating that absence of the marker was a reliable signal for precise cassette exchange. Expressing Cas9 under the *zpg* promoter (Hammond et al. 2021a), rather than the *vas2* promoter like in Hammond et al. (2016) might have also boosted HDR efficiency.

To further boost CriMCE editing efficiency, non-plasmid-based repair templates, such as single-stranded oligo-deoxyribonucleotides (ssODNs), could also be employed in certain organisms (Levi et al. 2020), though it is not clear whether this is beneficial in mosquitoes (Ang et al. 2022).

The CriMCE method can also mitigate against the risk of using previously untested and potentially inefficient gRNAs/pegRNAs that would otherwise expend undue effort on genetic crosses and molecular genotyping. Generating a marked placeholder prior to precise editing ensures that rare transgenesis using novel gRNA/pegRNAs is easily identifiable by a fluorescent marker. Previously tested guides can then be used to target the placeholder, inducing CriMCE. The present study validated the use of two gRNAs that target a universal placeholder which is designed to function across insect species.

Nonetheless, an extra transgenesis step is required to perform CriMCE, in contrast to other strategies for marker-less edit insertion, like prime editing and direct HDR (Figure 5.3C). However, this is unlikely to be a limiting factor due to its benefits that relate to increasing transformation efficiency, the fact that a single placeholder strain at a locus of interest can be used for insertion multiple distinct marker-less edits for testing, whilst it can also serve as a balancer strain to aid in the isolation of the desired edit (Figure 3.2).

CriMCE is particularly powerful for experiments aimed at introducing a range of modifications to a single locus of interest, as a single placeholder strain can be exchanged for any number of variants (Figure 5.3B). Indeed, Kaduskar et al. (2022) employ a similar approach, based upon exchange of a marked allele for engineering of *kdr* pyrethroid resistance mutations in *Drosophila* and could be further extended to

incorporate newly discovered insecticide resistant SNPs (Clarkson et al. 2021).

Moreover, CriMCE allows for complex mutations that are not possible using prime editing since the entire region ablated by the placeholder can be replaced with a region bearing any number of desired edits. This strategy, shown in Figure 5.3B, as 'allelic exchange' could allow multiple linked SNPs to be introduced across a wide genetic locus. This would be useful in assessing how a combination of resistant variants interact with each other to produce complex insecticide resistance phenotypes (Samantsidis et al. 2020). Other complex edits are also possible such as the introduction, modification or deletion of introns and splice sites, or complete codon scrambling by which a coding sequence is modified without affecting the encoded amino acid sequence (Figure 5.3B). The latter strategy could serve to engineer synthetic alleles that are resistant to gene drive elements as a mechanism for gene drive recall (Vella et al. 2017). Finally, integrating the placeholder cassette within intronic or neutral regions, could permit editing of haploinsufficient genes (Figure 5.4).

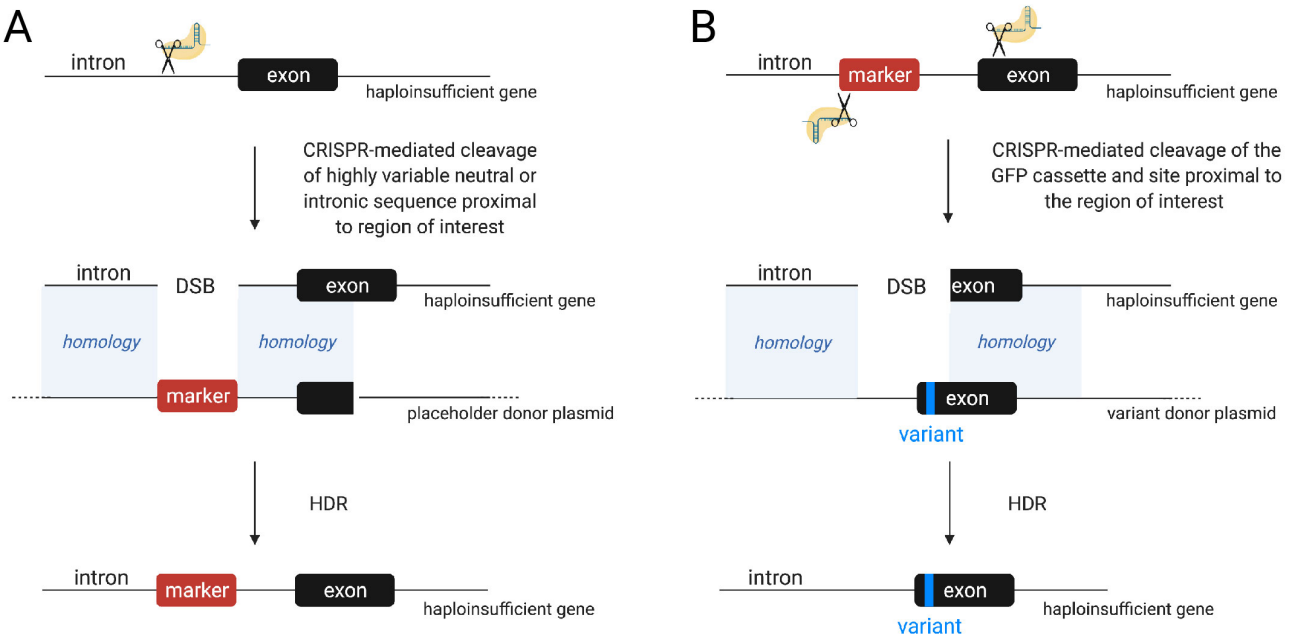


Figure 5.4: A strategy for introducing precise marker-less edits into haploinsufficient genes using CriMCE. CriMCE can be adapted to modify haplo-insufficient genes by introducing the marked placeholder into a neutral locus, like an intron, proximal to a given target site on a haploinsufficient exon. **(A)** To generate a marked placeholder strain, a highly variable intronic region, proximal to the haplo-insufficient exon, can be cleaved using CRISPR, and a marker cassette (red) is introduced from a donor plasmid, through HDR. **(B)** To generate a strain carrying the variant of choice on the exon the marker cassette can be removed via two CRISPR-mediated cleavages: one at the marker cassette and one near the site of interest; and exchanged for an intact sequence containing the variant of choice from a donor plasmid, through HDR.

5.2.3 Testing of putative drive-resistant variants

After successfully engineering the three putatively resistant SNPs (G→A, natural; C→T and G→T, artificially evolved) into the mosquito genome using CriMCE (Figure 4.18), rigorous testing was performed to assess their functionality (Figures 4.21, 4.22 and 4.23), as well as the extent to which they block gene drive activity (Figure 4.5). As predicted, all tested SNPs were found to be functional in homozygosis, leading to the development of fertile anatomical females. However, each SNP conferred different levels of resistance to homing ranging from no resistance (G→A), to partial (C→T, termed R3) and full resistance (G→T, typical R1).

The same finding was also reflected in the fertility of QFS1/SNP transheterozygous females (Figure 4.25). Namely, QFS1/R1 females were fully fertile, since somatic cleavage of the exposed allele, which is the cause of infertility, was blocked. Conversely, QFS1/R3 females showed intermediate levels of fertility (i.e. higher fertility than QFS1/wt heterozygotes, but lower fertility than QFS1/R1), since the R3 alleles could still be cleaved in off-target somatic tissues to some extent, causing sterility associated with intersex mosaicism in certain females.

This is the first time that partial/incomplete resistance to gene drives (R3) has been experimentally demonstrated. The model built in collaboration with Dr Bhavin Khatri and Prof Austin Burt, suggests that R3 alleles could come under strong selection (Figure 3.6). Though they would not completely prevent gene drive spread, they would reduce the levels of population suppression, and therefore need to be considered by field implementation models and strategies. Nonetheless, relaxed target site specificity can be somewhat beneficial since drive is not completely reversed by presence of R3 alleles like it would for presence of R1 (Hammond et al. 2017, Hammond et al. 2021a).

Despite intermediate Cas9 fidelity, no off-target effects were readily detected for *zpg*-expressed Cas9, partly owing to the uniqueness of the target site (since the most closely related sequence differed by 4 mismatches) and also due to confined expression of the nuclease under the *zpg* germline promoter (Garrood et al. 2021). This fortuitous balance between limiting off-target effects and allowing on-target cleavage of non-canonical sequences can benefit spread of self-sustaining gene drives.

However, Cas9 flexibility can question the extent to which targeting private alleles would be possible to confine spread of gene drives to specific populations (Willis & Burt 2021), considering that non-canonical on-target cleavage can occur, even if the mismatches are proximal to the target site PAM (Figure 4.5). Depending on the nature of the private and non-private alleles, and the variation position on the target site, it might be possible to predict their susceptibility to cleavage (Hsu et al. 2013).

Both private allele and gene drive strategies relying on target side re-coding (such as population replacement gene drives (Adolfi et al. 2020), or threshold-dependent tethered homing drives (Metzloff

et al. 2021), should target sufficiently divergent sequences that block undesirable Cas9 cleavage, preferably showing variation in the PAM region itself, or resort to using a high fidelity (HF) Cas9 (Zhang et al. 2021), to minimise unwanted cleavage of natural or the re-coded target site variants.

5.2.4 Resistance to gene drive in natural populations

The detection of the R1 and R3 resistant alleles at the Ag(QFS)1 target site in an experimental setting, might imply that these alleles could also be generated in natural populations, upon field release. As the model predicts, this would impair the potential spread and persistence of the gene drive, leading to partial (R3) or complete (R1) loss of the drive over time (Figure 4.26). Through our resistance screen we estimated a rate of R1 and R3 creation of $1*10^{-3}$ and $3*10^{-3}$ amongst all EJ mutations, respectively (Figure 3.1). If we assume that in the fraction of alleles that get repaired by EJ (0.015) in the context of gene drive the creation rate of R1s and R3s is the same, then we would expect that they would get generated in approximately 1 in 67,000 and 1 in 22,000 gene drive offspring, respectively.

This might mean that small populations with effective population sizes in the range of 200-2,000 mosquitoes (Wiltshire et al. 2018), could get suppressed by Ag(QFS)1, but the gene drive might struggle with invading larger contiguous populations, upwards of 2 million mosquitoes (O'Loughlin et al. 2014, Khatri et al. 2019), where resistant mutations might arise before the gene drive has had a chance to eliminate the population; and it is unclear what would happen in the context of medium-sized populations, like the one in Bana, Burkina Faso (Yao et al. 2022). A spatial model of resistance creation in the context of natural populations would help elucidate this. Moreover, to ensure the success of field implementations, strategies to mitigate resistance need to be put in place prior to a first gene drive release in the wild.

5.2.5 Mitigating resistance using a multiplexed gene drive strategy

There are several ways to mitigate resistance by multiplexed gene drives. One approach would be to use resistance screening to deduce which are the most likely resistant alleles to be generated at a given target site, and use several gRNAs to target those variants in addition to the wild-type target (provided they do not alter the PAM sequence). Another way to mitigate resistance would be to target multiple non-overlapping sites on the same wider target locus (as was done during the course of this PhD). Finally, one could target multiple sites each located on a distinct highly conserved target gene. These strategies could also be combined with one another.

Moreover, each chosen target site could be targeted simultaneously by expressing all gRNAs from the same gene drive construct (which is what is usually referred to as multiplexing) or by developing a series of follow-up gene drives, to be released sequentially into a natural population, when resistance to the previous gene drive excedes a certain threshold. Resistance testing in the laboratory can further inform this decision-making.

In the present study, we chose to mitigate resistance generated at one target site, by targeting multiple other sites at the same time by multiplexing gRNAs expressed from a single gene drive construct (Figure 1.11) (Burt 2003, Oberhofer et al. 2018, Yang et al. 2022). To this end, two gene drive strains were generated: Ag(QFS)2 and Ag(QFS)3, that simultaneously target 2 and 3 sites on the CDS of *dsx* exon 5, respectively (though more focus was placed on testing Ag(QFS)2). In this case, cut-resistant mutations need to be simultaneously present at all target sites, and combining to produce a functional protein, to constitute a drive-resistant allele.

Excitingly, when tested, Ag(QFS)2 was able to catalyse efficient homing despite the presence of a known R1, or GFP-marked null (equivalent to R2) allele, in one of the targeted sites (Figure 4.4.3), removing the resistant mutations in the process. This is the first time that resistance removal by a multiplexed gene drive is being experimentally demonstrated.

Multiplexing with more than 2-3 gRNAs has previously been suggested to reduce drive conversion efficiency, primarily due to saturation of Cas9 activity by multiple gRNAs, and the imperfect alignment of homology arms around the cut site (Champer, Oh, Liu, Wen, Clark, Messer & Champer 2020, Oberhofer et al. 2018). In contrast to this, both Ag(QFS)2 and Ag(QFS)3 showed extremely high rates of transmission, that were comparable to or higher than Ag(QFS)1 in some cases, even when we artificially created a 1,4 kb requirement for DNA resection before homing can be permitted (Figure 4.4.3). Having multiple gRNAs active at the same time, might increase the chance of successful cleavage, and thus lead to higher rates of gene drive homing (Champer et al. 2018).

We also did not observe any partial homing events, in contrast to previous studies (Oberhofer et al.

2018, Hammond et al. 2016), which might be owing to the reverse orientation of the gRNAs in the gene drive, compared to the equivalent, homologous targets on the target gene. However, multiplexing using tandemly repeated U6-gRNA cassettes led to some construct instability, owing to the repetition of identical U6 promoters and gRNA backbones. In particular, rare gRNA loss and duplication events were observed, which were more likely to occur with an increasing number of gRNAs.

Construct instability can be mitigated by reducing the repeating elements in the gRNA expression cassettes, reducing the chance of internal recombination events. This can be achieved by using a single U6 promoter to express a series of gRNAs that are linked together by self-cleaving ribozymes (Marshall et al. 2017, Xu et al. 2017), by utilising distinct gRNA backbones (though this can affect editing efficiency) (Wolabu et al. 2020), or by employing different U6 promoters from the same species (Anderson et al. 2020). Spacing the gRNA cassettes further apart (e.g. by integrating the fluorescence marker between them) might also reduce rates of recombination.

Surprisingly, heterozygous Ag(QFS)2 females showed ∼100% improved fertility, overall, compared to Ag(QFS)1 (Figure 4.32). The presence of multiple gRNAs was not expected to increase female fecundity, however the gene drives also differed in their orientation. Namely, Ag(QFS)1 was integrated in the same orientation as the *dsx* gene (fwd), whilst Ag(QFS)2 was integrated in the opposite orientation (rev) (Figure 4.29). Indeed, when the same Ag(QFS)3 construct was integrated in the same orientation as *dsx* females that inherited the gene drive paternally were over 10 times more likely to develop as mosaic intersex (sterile), than as normal females.

Directionality of promoter and enhancer elements is known to be largely variable (Ibrahim et al. 2018), and their activity can be sensitive to the orientation of the genes they regulate (Hozumi et al. 2013). Cas9 expression might therefore be affected by *dsx*-associated enhancers in a directional manner causing its ectopic expression in somatic tissues, leading to a mosaic phenotype for the *dsxF* knock-out, which ultimately renders females sterile.

Ectopic Cas9 activity can be caused by a leaky germline-specific promoter that drives its expression in somatic tissues, as well as in the germline, or by parental deposition of the Cas9 into the early embryo, where it can act on somatic tissue progenitors[3]. For example, Cas9 expressed by the *vas2* promoter is known to be maternally deposited (Papathanos et al. 2009, Hammond et al. 2016, Hammond et al. 2021a). In contrast, *zpg*-expressed Cas9 from the 7280 locus was not found to be parentally deposited, but it did presumably cause some somatic mutagenesis in females that reduced their fecundity, independent from whether they had inherited the gene drive maternally or paternally (Hammond et al. 2021a). *Zpg*-Cas9 expression from the *dsx* locus however seems to differ. In the context of Ag(QFS)1, females that inherit the gene drive paternally show significantly reduced fertility compared to females that inherit the

[3]Note that the gRNA is ubiquitously expressed at all tissues by a pol III (U6) promoter.

drive maternally, indicating that the Cas9 might be paternally deposited (as protein or mRNA) (Figure 4.32). At the same time, there is still some somatic leakiness in females that inherited the drive maternally, causing them to have somewhat reduced fertility. In the context of Ag(QFS)2, all females show increased fertility, and the paternal effects on fertility are a lot less pronounced, possibly owing to the orientation of the drive.

Thus, female fertility is not only affected by the promoter element driving Cas9 expression but also its position in the genome, and its orientation within that position. Moreover, this thesis provides evidence that positional effects can also alter the ability of Cas9 to induce homing, as when the Ag(QFS)1 construct was integrated out-of-locus and we measured its ability to induce homing of a GFP cassette in-locus, it did not bias GFP as strongly as expected from the bias of gene drive inheritance in-locus (Figures 4.8 and 4.4.3). Strong positional effects have also been observed by other groups (personal communication).

In general, note that here, Ag(QFS)1 females are reported to be less fertile than in Kyrou et al. (2018). Building upon knowledge gained since then, this thesis aimed at a more comprehensive assessment of fitness by including non-blood-fed females, as well as females that died post-blood-meal, in the analysis (that were previously excluded), resulting in the apparent decrease in Ag(QFS)1 female fitness (Figure 4.32).

This allowed the investigation of post-blood-meal mortality of gene drive carriers, on top of previously assessed phenotypes relating to female fitness (such as egg output, larval output, and egg hatching rate) that was previously not considered as a phenotype of the *dsxF* knock-out (Hammond et al. 2016, Kyrou et al. 2018). Mosquitoes exhibit strong sexual dimorphism with relation to their blood-feeding behaviour, since females require a blood-meal to produce eggs, whilst males do not feed on blood. Since *dsx* regulates the expression of sexually dimorphic traits, it might also regulate the expression of female-specific blood digestion enzymes, that when absent, due to a full or mosaic *dsxF* knock-out, lead to lethality upon blood ingestion.

Finally, this thesis demonstrated that Ag(QFS)2 can induce successful population elimination of caged laboratory populations, achieving faster spread than previously recorded for Ag(QFS)1 (Kyrou et al. 2018), presumably owing to the increased fertility of the heterozygous female gene drive carriers, and its ability to actively remove R2 mutations, that can slow down gene drive invasion.

Chapter 6

Conclusions

6.0.1 Established pipeline for the discovery of resistant mutations

A novel pipeline has been established for the creation, detection and assessment of drive-resistant mutations. This high-throughput resistance screen enabled the characterisation of over 9,000 individual female mosquitoes carrying EJ mutations at dsx, bypassing the need to screen more than a half a million gene drive-carrying insects that would be required to detect the same number of functional resistant alleles (R1), as reported here. This methodology can be applied to evaluate gene drives prior to field testing for their propensity to generate resistant mutations, and guide the implementation strategy accordingly to minimise the potential for resistance.

6.0.2 Established a method to engineer marker-less edits

A new method to engineer and isolate precise marker-less edits, termed CriMCE, was also established. Compared to other methods it benefits from being able to introduce uniquely traceless, and potentially complex edits, whilst it is 5-41x more efficient than other strategies based on standalone HDR or prime editing. Using this method it is possible to link small genetic changes with a biologically relevant outcome across a range of insect species, and it may have particular applications in the study of resistance to insecticides and gene drives. In the future, this method could also allow the engineering of alleles resistant to multiplexed (or simple) gene drives, for the purpose of gene drive recall.

6.0.3 *In vivo* testing revealed that functional target site SNPs can show various levels of drive-resistance

Here, putatively resistant SNPs were engineerd in the mosquito genome using CriMCE to verify their function *in vivo*, as well as assess the extent to which they can block gene drive activity, improving upon

previous testing of resistant alleles that was performed *in vitro* (Kyrou et al. 2018, Carballar-Lejarazú et al. 2020). This is an important step forwards since *in vitro* cell-free assays are less stringent, and often do not always translate to high or even detectable levels of cleavage *in vivo* (Garrood et al. 2021). Moreover, they do not take into account that *in vivo* cleavage can vary between the sexes, or that it might be partial, as was shown here (Figure 4.5). In the future, the development of reliable *in vitro* cleavage assays in mosquito cell lines might be able to bridge the gap between *in silico* or cell-free *in vitro* assays and *in vivo* experiments, surpassing the difficulty of generating and maintaining novel mosquito strains, and producing data that accurately reflect CRISPR activity *in vivo* (Krzywinska et al. 2022).

Moreover, the discovery that natural SNP variants can be partially or fully cleaved *in vivo*, despite being PAM-proximal, will have important implications for field testing of gene drive, allowing for better invasion dynamics in natural populations than previously thought.

6.0.4 Multiplexed gene drives can mitigate resistance

Testing of multiplexed gene drives revealed that construct orientation and position in the genome can affect fertility of females, as well as homing rates, presumably due to differences in levels and location of Cas9 expression (e.g. in non-germline tissues). An increasing number of gRNAs present in the construct increased homing rates, but reduced construct stability due to the repetition of identical elements present in each gRNA cassette. Finally, multiplexed gene drives can mitigate and actively remove R1 and R2 alleles, as expected; as well as reach fixation and eliminate caged laboratory populations in 6-7 and 7-8 generations respectively.

6.0.5 Concluding statement

The present thesis has demonstrated that targeting a highly conserved target site is not enough to completely mitigate resistance to gene drive, and additional strategies need to be adopted on top of that to ensure successful population suppression upon field implementation. Here, we have shown how targeting multiple sites simultaneously can remove resistant alleles, and recommend that a multiplexed gene drive strategy is adopted going forwards, to pre-empt the eventuality of resistance in natural populations. We also recommend the continued exploration of alternative nucleases as tools to reduce the number of resistant alleles being created by gene drive activity in the first place.

Bibliography

Adolfi, A., Gantz, V. M., Jasinskiene, N., Lee, H. F., Hwang, K., Terradas, G., Bulger, E. A., Ramaiah, A., Bennett, J. B., Emerson, J. J., Marshall, J. M., Bier, E. & James, A. A. (2020), 'Efficient population modification gene-drive rescue system in the malaria mosquito Anopheles stephensi', *Nature Communications 2020 11:1* **11**(1), 1–13.
URL: *https://www.nature.com/articles/s41467-020-19426-0*

Akbari, O. S., Matzen, K. D., Marshall, J. M., Huang, H., Ward, C. M. & Hay, B. A. (2013), 'A synthetic gene drive system for local, reversible modification and suppression of insect populations', *Current biology : CB* **23**(8), 671–677.
URL: *https://pubmed.ncbi.nlm.nih.gov/23541732/*

Alphey, L. & Andreasen, M. (2002), 'Dominant lethality and insect population control', *Molecular and Biochemical Parasitology* **121**(2), 173–178.
URL: *https://doi.org/10.1016/S0166-6851(02)00040-3*

Anderson, M. A., 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. & Alphey, L. (2022), 'A multiplexed, confinable CRISPR/Cas9 gene drive propagates in caged Aedes aegypti populations', *bioRxiv* p. 2022.08.12.503466.
URL: *https://www.biorxiv.org/content/10.1101/2022.08.12.503466v1*

Anderson, M. A., Purcell, J., Verkuijl, S. A., Norman, V. C., Leftwich, P. T., Harvey-Samuel, T. & Alphey, L. S. (2020), 'Expanding the CRISPR Toolbox in Culicine Mosquitoes: In Vitro Validation of Pol III Promoters', *ACS Synthetic Biology* **9**(3), 678–681.
URL: *https://pubs.acs.org/doi/full/10.1021/acssynbio.9b00436*

Ang, J. X. D., Nevard, K., Ireland, R., Purusothaman, D. K., Verkuijl, S. A., Shackleford, L., Gonzalez, E., Anderson, M. A. & Alphey, L. (2022), 'Considerations for homology-based DNA repair in mosquitoes: Impact of sequence heterology and donor template source', *PLOS Genetics* **18**(2), e1010060.
URL: *https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010060*

Anzalone, A. V., Koblan, L. W. & Liu, D. R. (2020), 'Genome editing with CRISPR–Cas nucleases, base editors, transposases and prime editors', *Nature Biotechnology 2020 38:7* **38**(7), 824–844.
URL: *https://www.nature.com/articles/s41587-020-0561-9*

Baba, E., Hamade, P., Kivumbi, H., Marasciulo, M., Maxwell, K., Moroso, D., Roca-Feltrer, A., Sanogo, A., Stenstrom Johansson, J., Tibenderana, J., Abdoulaye, R., Coulibaly, P., Hubbard, E., Jah, H., Kaman Lama, E., Razafindralambo, L., Van Hulle, S., Jagoe, G., Tchouatieu, A.-M., Collins, D., Gilmartin, C., Tetteh, G., Djibo, Y., Ndiaye, F., Kalleh, M., Kandeh, B., Audu, B., Ntadom, G., Kiba, A., Savodogo, Y., Boulotigam, K., Ali Sougoudi, D., Guilavogui, T., Keita, M., Kone, D., Jackou, H., Ouba, I., Ouedraogo, E., Alassana Messan, H., Jah, F., Janneh Kaira, M., Sire Sano, M., Chérif Traore, M., Ngarnaye, N., Yinusa Cassandra Elagbaje, A., Halleux, C., Merle, C., Iessa, N., Pal, S., Sefiani, H., Souleymani, R., Laminou, I., Doumagoum, D., Kesseley, H., Coldiron, M., Grais, R., Kana, M., Bosco Ouedraogo, J., Zongo, I., Eloike, T., Johnbull Ogboi, S., Achan, J., Bojang, K., Ceesay, S., Dicko, A., Djimde, A., Sagara, I., Diallo, A., Louis NdDiaye, J., Marcel Loua, K., Beshir, K., Cairns, M., Fernandez, Y., Lal, S., Mansukhani, R., Muwanguzi, J., Scott, S., Snell, P., Sutherland, C., Tuta, R., Milligan, P. & Partnership, A.-s. (2020), 'Effectiveness of seasonal malaria chemoprevention at scale in west and central Africa: an observational study'.
URL: *https://www.access-smc.org/*

Baldini, F., Segata, N., Pompon, J., Marcenac, P., Robert Shaw, W., Dabiré, R. K., Diabaté, A., Levashina, E. A. & Catteruccia, F. (2014), 'Evidence of natural Wolbachia infections in field populations of Anopheles gambiae', *Nature Communications 2014 5:1* **5**(1), 1–7.
URL: *https://www.nature.com/articles/ncomms4985*

Balikagala, B., Fukuda, N., Ikeda, M., Katuro, O. T., Tachibana, S.-I., Yamauchi, M., Opio, W., Emoto, S., Anywar, D. A., Kimura, E., Palacpac, N. M., Odongo-Aginya, E. I., Ogwang, M., Horii, T. & Mita, T. (2021), 'Evidence of Artemisinin-Resistant Malaria in Africa', *The New England journal of medicine* **385**(13), 1163–1171.
URL: *https://pubmed.ncbi.nlm.nih.gov/34551228/ https://pubmed.ncbi.nlm.nih.gov/34551228/?dopt=Abstract*

Barrangou, R. & Doudna, J. A. (2016), 'Applications of CRISPR technologies in research and beyond', *Nature Biotechnology 2016 34:9* **34**(9), 933–941.
URL: *https://www.nature.com/articles/nbt.3659*

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D. A. & Horvath, P. (2007), 'CRISPR provides acquired resistance against viruses in prokaryotes', *Science (New York, N.Y.)*

315(5819), 1709–1712.
URL: *https://pubmed.ncbi.nlm.nih.gov/17379808/*

Beaghton, A., Hammond, A., Nolan, T., Crisanti, A., Godfray, H. C. J. & Burt, A. (2017), 'Requirements for driving antipathogen effector genes into populations of disease vectors by homing', *Genetics* **205**(4), 1587–1596.
URL: *https://academic.oup.com/genetics/article/205/4/1587/6066441*

Beaghton, A. K., Hammond, A., Nolan, T., Crisanti, A. & Burt, A. (2019), 'Gene drive for population genetic control: Non-functional resistance and parental effects', *Proceedings of the Royal Society B: Biological Sciences* .
URL: *https://doi.org/10.1098/rspb.2019.1586*

Beeman, R. W. & Friesen, K. S. (1999), 'Properties and natural occurrence of maternal-effect selfish genes ('Medea' factors) in the red flour beetle, Tribolium castaneum', *Heredity* **82**, 529–534.
URL: *https://doi.org/10.1038/sj.hdy.6885150*

Bernardini, F., Galizi, R., Menichelli, M., Papathanos, P. A., Dritsou, V., Marois, E., Crisanti, A. & Windbichler, N. (2014), 'Site-specific genetic engineering of the Anopheles gambiae y chromosome', *Proceedings of the National Academy of Sciences of the United States of America* .
URL: *https://doi.org/10.1073/pnas.1404996111*

Bhatt, S., Weiss, D. J., Cameron, E., Bisanzio, D., Mappin, B., Dalrymple, U., Battle, K. E., Moyes, C. L., Henry, A., Eckhoff, P. A., Wenger, E. A., Briët, O., Penny, M. A., Smith, T. A., Bennett, A., Yukich, J., Eisele, T. P., Griffin, J. T., Fergus, C. A., Lynch, M., Lindgren, F., Cohen, J. M., Murray, C. L., Smith, D. L., Hay, S. I., Cibulskis, R. E. & Gething, P. W. (2015), 'The effect of malaria control on Plasmodium falciparum in Africa between 2000 and 2015', *Nature 2015 526:7572* **526**(7572), 207–211.
URL: *https://www.nature.com/articles/nature15535*

Biemont, C. & Cizeron, G. (1999), 'Distribution of transposable elements in Drosophila species', *Genetica* **105**, 43–62.
URL: *https://doi.org/10.1023/A:1003718520490*

Bier, E. (2021), 'Gene drives gaining speed', *Nature Reviews Genetics 2021 23:1* **23**(1), 5–22.
URL: *https://www.nature.com/articles/s41576-021-00386-0*

Billeter, J. C., Rideout, E. J., Dornan, A. J. & Goodwin, S. F. (2006), 'Control of Male Sexual Behavior in Drosophila by the Sex Determination Pathway', *Current Biology* **16**(17).
URL: *https://doi.org/10.1016/j.cub.2006.08.025*

Black, W. C., Alphey, L. & James, A. A. (2011), 'Why RIDL is not SIT', *Trends in Parasitology* **27**(8), 362–370.
URL: *https://doi.org/10.1016/j.pt.2011.04.004*

Bosch, J. A., Birchak, G. & Perrimon, N. (2020), 'Precise genome engineering in Drosophila using prime editing', *Proceedings of the National Academy of Sciences of the United States of America* **118**(1).
URL: *https://www.pnas.org/content/118/1/e2021996118 https://www.pnas.org/content/118/1/e2021996118.abstract*

Boswell, R. E. & Mahowald, A. P. (1985), 'tudor, a gene required for assembly of the germ plasm in Drosophila melanogaster', *Cell* **43**(1), 97–104.
URL: *https://doi.org/10.1016/0092-8674(85)90015-7*

Brouns, S. J., Jore, M. M., Lundgren, M., Westra, E. R., Slijkhuis, R. J., Snijders, A. P., Dickman, M. J., Makarova, K. S., Koonin, E. V. & Van Der Oost, J. (2008), 'Small CRISPR RNAs guide antiviral defense in prokaryotes', *Science (New York, N.Y.)* **321**(5891), 960–964.
URL: *https://pubmed.ncbi.nlm.nih.gov/18703739/*

Buchman, A., Marshall, J. M., Ostrovski, D., Yang, T. & Akbari, O. S. (2018), 'Synthetically engineered Medea gene drive system in the worldwide crop pest Drosophila suzukii', *Proceedings of the National Academy of Sciences of the United States of America* .
URL: *https://doi.org/10.1073/pnas.1713139115*

Burt, A. (2003), 'Site-specific selfish genes as tools for the control and genetic engineering of natural populations', *Proceedings of the Royal Society B: Biological Sciences* .
URL: *https://doi.org/10.1098/rspb.2002.2319*

Burt, A. & Deredec, A. (2018), 'Self-limiting population genetic control with sex-linked genome editors', *Proceedings of the Royal Society B: Biological Sciences* .
URL: *https://doi.org/10.1098/rspb.2018.0776*

Cairns, M., Ceesay, S. J., Sagara, I., Zongo, I., Kessely, H., Gamougam, K., Diallo, A., Ogboi, J. S., Moroso, D., van Hulle, S., Eloike, T., Snell, P., Scott, S., Merle, C., Bojang, K., Ouedraogo, J. B., Dicko, A., Ndiaye, J. L. & Milligan, P. (2021), 'Effectiveness of seasonal malaria chemoprevention (SMC) treatments when SMC is implemented at scale: Case–control studies in 5 countries', *PLOS Medicine* **18**(9), e1003727.
URL: *https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1003727*

Carareto, C. M. A., Kim, W., Wojciechowski, M. F., O'grady, P., Prokchorova, A. V., Silva, J. C. & Kidwell, M. G. (1997), 'Testing transposable elements as genetic drive mechanisms using Drosophila P element constructs as a model system', *Genetica* **101**, 13–33.
URL: *https://doi.org/10.1023/A:1018339603370*

Carballar-Lejarazú, R., Ogaugwu, C., Tushar, T., Kelsey, A., Pham, T. B., Murphy, J., Schmidt, H., Lee, Y., Lanzaro, G. C. & 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 of the United States of America* **117**(37), 22805–22814.
URL: *https://www.pnas.org/content/117/37/22805 https://www.pnas.org/content/117/37/22805.abstract*

Carrami, E. M., Eckermann, K. N., Ahmed, H. M., Héctor Sánchez, C. M., Dippel, S., Marshall, J. M. & Wimmer, E. A. (2018), 'Consequences of resistance evolution in a Cas9-based sex conversion-suppression gene drive for insect pest management', *Proceedings of the National Academy of Sciences of the United States of America* **115**(24), 6189–6194.
URL: *https://www.pnas.org/content/115/24/6189 https://www.pnas.org/content/115/24/6189.abstract*

Carvalho, D. O., McKemey, A. R., Garziera, L., Lacroix, R., Donnelly, C. A., Alphey, L., Malavasi, A. & 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), e0003864.
URL: *https://journals.plos.org/plosntds/article?id=10.1371/journal.pntd.0003864*

Catteruccia, F., Nolan, T., Loukeris, T. G., Blass, C., Savakis, C., Kafatos, F. C. & Crisanti, A. (2000), 'Stable germline transformation of the malaria mosquito Anopheles stephensi', *Nature* **405**(6789), 959–962.
URL: *https://pubmed.ncbi.nlm.nih.gov/10879538/*

CDC (2020), 'Malaria - Biology'.
URL: *https://www.cdc.gov/malaria/about/biology/index.html*

CDC (2021), 'Malaria's Impact Worldwide'.
URL: *https://www.cdc.gov/malaria/malaria_worldwide/impact.html*

Champer, J., Lee, E., Yang, E., Liu, C., Clark, A. G. & Messer, P. W. (2020), 'A toxin-antidote CRISPR gene drive system for regional population modification', *Nature Communications* **11**(1082).
URL: *https://doi.org/10.1038/s41467-020-14960-3*

Champer, J., Liu, J., Oh, S. Y., Reeves, R., Luthra, A., Oakes, N., Clark, A. G. & Messer, P. W. (2018), 'Reducing resistance allele formation in CRISPR gene drive', *Proceedings of the National Academy of Sciences of the United States of America* **115**(21), 5522–5527.
URL: *https://doi.org/10.1073/pnas.1720354115*

Champer, J., Reeves, R., Oh, S. Y., Liu, C., Liu, J., Clark, A. G. & 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), e1006796.
URL: *https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1006796*

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

Chan, Y. S., Takeuchi, R., Jarjour, J., Huen, D. S., Stoddard, B. L. & Russell, S. (2013), 'The Design and In Vivo Evaluation of Engineered I-OnuI-Based Enzymes for HEG Gene Drive', *PLOS ONE* **8**(9), e74254.
URL: *https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0074254*

Chandramohan, D., Zongo, I., Sagara, I., Cairns, M., Yerbanga, R.-S., Diarra, M., Nikièma, F., Tapily, A., Sompougdou, F., Issiaka, D., Zoungrana, C., Sanogo, K., Haro, A., Kaya, M., Sienou, A.-A., Traore, S., Mahamar, A., Thera, I., Diarra, K., Dolo, A., Kuepfer, I., Snell, P., Milligan, P., Ockenhouse, C., Ofori-Anyinam, O., Tinto, H., Djimde, A., Ouédraogo, J.-B., Dicko, A. & Greenwood, B. (2021), 'Seasonal Malaria Vaccination with or without Seasonal Malaria Chemoprevention', *The New England journal of medicine* **385**(11), 1005–1017.
URL: *https://pubmed.ncbi.nlm.nih.gov/34432975/ https://pubmed.ncbi.nlm.nih.gov/34432975/?dopt=Abstract*

Chang, H. H., Moss, E. L., Park, D. J., Ndiaye, D., Mboup, S., Volkman, S. K., Sabeti, P. C., Wirth, D. F., Neafsey, D. E. & Hartl, D. L. (2013), 'Malaria life cycle intensifies both natural selection and random genetic drift', *Proceedings of the National Academy of Sciences of the United States of America* **110**(50), 20129–20134.
URL: *https://www.pnas.org/doi/abs/10.1073/pnas.1319857110*

Chen, C. H., Huang, H., Ward, C. M., Su, J. T., Schaeffer, L. V., Guo, M. & Hay, B. A. (2007), 'A synthetic maternal-effect selfish genetic element drives population replacement in Drosophila', *Science* **316**(5824), 597–600.
URL: *https://pubmed.ncbi.nlm.nih.gov/17395794/*

Chen, X. & Zhang, J. (2016), 'The genomic landscape of position effects on protein expression level and noise in yeast', *Cell systems* **2**(5), 347.
URL: */pmc/articles/PMC4882239/ /pmc/articles/PMC4882239/?report=abstract https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4882239/*

Chima, R. I., Goodman, C. A. & Mills, A. (2003), 'The economic impact of malaria in Africa: a critical

review of the evidence', *Health Policy* **63**(1), 17–36.
URL: *https://doi.org/10.1016/S0168-8510(02)00036-2*

Clarkson, C. S., Miles, A., Harding, N. J., Lucas, E. R., Battey, C. J., Amaya-Romero, J. E., Kern, A. D., Fontaine, M. C., Donnelly, M. J., Lawniczak, M. K., Kwiatkowski, D. P., Donnelly, M. J., Ayala, D., Besansky, N. J., Burt, A., Caputo, B., Torre, A. d., Fontaine, M. C., J. Godfray, H. C., Hahn, M. W., Kern, A. D., Lawniczak, M. K., Midega, J., O'Loughlin, S., Pinto, J., Riehle, M. M., Sharakhov, I., Schrider, D. R., Vernick, K. D., Weetman, D., Wilding, C. S. & White, B. J. (2020), 'Genome variation and population structure among 1142 mosquitoes of the African malaria vector species Anopheles gambiae and Anopheles coluzzii', *Genome Research* **30**(10), 1533–1546.
URL: *https://genome.cshlp.org/content/30/10/1533.full*

Clarkson, C. S., Miles, A., Harding, N. J., O'Reilly, A. O., Weetman, D., Kwiatkowski, D. & Donnelly, M. J. (2021), 'The genetic architecture of target-site resistance to pyrethroid insecticides in the African malaria vectors Anopheles gambiae and Anopheles coluzzii', *Molecular Ecology* **30**(21), 5303–5317.
URL: *https://onlinelibrary.wiley.com/doi/full/10.1111/mec.15845*

Clement, K., Rees, H., Canver, M. C., Gehrke, J. M., Farouni, R., Hsu, J. Y., Cole, M. A., Liu, D. R., Joung, J. K., Bauer, D. E. & Pinello, L. (2019), 'CRISPResso2 provides accurate and rapid genome editing sequence analysis', *Nature Biotechnology 2019 37:3* **37**(3), 224–226.
URL: *https://www.nature.com/articles/s41587-019-0032-3*

Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P. D., Wu, X., Jiang, W., Marraffini, L. A. & Zhang, F. (2013), 'Multiplex genome engineering using CRISPR/Cas systems', *Science (New York, N.Y.)* **339**(6121), 819–823.
URL: *https://pubmed.ncbi.nlm.nih.gov/23287718/*

Crawford, J. E., Clarke, D. W., Criswell, V., Desnoyer, M., Cornel, D., Deegan, B., Gong, K., Hopkins, K. C., Howell, P., Hyde, J. S., Livni, J., Behling, C., Benza, R., Chen, W., Dobson, K. L., Eldershaw, C., Greeley, D., Han, Y., Hughes, B., Kakani, E., Karbowski, J., Kitchell, A., Lee, E., Lin, T., Liu, J., Lozano, M., MacDonald, W., Mains, J. W., Metlitz, M., Mitchell, S. N., Moore, D., Ohm, J. R., Parkes, K., Porshnikoff, A., Robuck, C., Sheridan, M., Sobecki, R., Smith, P., Stevenson, J., Sullivan, J., Wasson, B., Weakley, A. M., Wilhelm, M., Won, J., Yasunaga, A., Chan, W. C., Holeman, J., Snoad, N., Upson, L., Zha, T., Dobson, S. L., Mulligan, F. S., Massaro, P. & White, B. J. (2020), 'Efficient production of male Wolbachia-infected Aedes aegypti mosquitoes enables large-scale suppression of wild populations', *Nature Biotechnology 2020 38:4* **38**(4), 482–492.
URL: *https://www.nature.com/articles/s41587-020-0471-x*

Dame, D. A., Benedict, M. Q., Robinson, A. S. & Knols, B. G. (2009), 'Historical applications of induced sterilisation in field populations of mosquitoes', *Malaria Journal* **8**(Suppl 2), S2.
URL: */pmc/articles/PMC2777324/*

Datoo, M. S., Magloire Natama, H., Somé, A., Traoré, O., Rouamba, T., Bellamy, D., Yameogo, P., Valia, D., Tegneri, M., Ouedraogo, F., Soma, R., Sawadogo, S., Sorgho, F., Derra, K., Rouamba, E., Orindi, B., Ramos-Lopez, F., Flaxman, A., Cappuccini, F., Kailath, R., Elias, S. C., Mukhopadhyay, E., Noe, A., Cairns, M., Lawrie, A., Roberts, R., Valéa, I., Sorgho, H., Williams, N., Glenn, G., Fries, L., Reimer, J., Ewer, K. J., Shaligram, U., Hill, A. V. S. & Tinto, H. (2021), 'High Efficacy of a Low Dose Candidate Malaria Vaccine, R21 in 1 Adjuvant Matrix-M™, with Seasonal Administration to Children in Burkina Faso', *The Lancet* **397**(10287), 1809–1818.
URL: *https://doi.org/10.1016/S0140-6736(21)00943-0*

Davey, J. W., Hohenlohe, P. A., Etter, P. D., Boone, J. Q., Catchen, J. M. & Blaxter, M. L. (2011), 'Genome-wide genetic marker discovery and genotyping using next-generation sequencing', *Nature Reviews Genetics* **12**(7), 499–510.
URL: *https://www.nature.com/articles/nrg3012*

Degarege, A., Fennie, K., Degarege, D., Chennupati, S. & 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), e0211205.
URL: *https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0211205*

Del Amo, V. L., Juste, S. S. & Gantz, V. M. (2022), 'A nickase Cas9 gene-drive system promotes super-Mendelian inheritance in Drosophila', *Cell Reports* **39**(8), 110843.
URL: *https://doi.org/10.1016/j.celrep.2022.110843*

Deltcheva, E., Chylinski, K., Sharma, C. M., Gonzales, K., Chao, Y., Pirzada, Z. A., Eckert, M. R., Vogel, J. & Charpentier, E. (2011), 'CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III', *Nature* **471**(7340), 602–607.
URL: *https://www.nature.com/articles/nature09886*

Deredec, A., Godfray, H. C. J. & Burt, A. (2011), 'Requirements for effective malaria control with homing endonuclease genes', *Proceedings of the National Academy of Sciences of the United States of America* **108**(43), E874–E880.
URL: *https://www.pnas.org/content/108/43/E874 https://www.pnas.org/content/108/43/E874.abstract*

Desai, M., Gutman, J., Taylor, S. M., Wiegand, R. E., Khairallah, C., Kayentao, K., Ouma, P., Coulibaly, S. O., Kalilani, L., Mace, K. E., Arinaitwe, E., Mathanga, D. P., Doumbo, O., Otieno, K., Edgar, D., Chaluluka, E., Kamuliwo, M., Ades, V., Skarbinski, J., Shi, Y. P., Magnussen, P., Meshnick, S. & Ter Kuile, F. O. (2016), 'Impact of Sulfadoxine-Pyrimethamine Resistance on Effectiveness of Intermittent Preventive Therapy for Malaria in Pregnancy at Clearing Infections and Preventing Low Birth Weight', *Clinical Infectious Diseases* **62**(3), 323–333.
URL: *https://academic.oup.com/cid/article/62/3/323/2462857*

Dong, S., Dong, Y., Simões, M. L. & Dimopoulos, G. (2022), 'Mosquito transgenesis for malaria control', *Trends in Parasitology* **38**(1), 54–66.
URL: *https://doi.org/10.1016/j.pt.2021.08.001*

Dong, Y., Simões, M. L. & Dimopoulos, G. (2020), 'Versatile transgenic multistage effector-gene combinations for Plasmodium falciparum suppression in Anopheles', *Science Advances* **6**(20).
URL: *https://www.science.org/doi/full/10.1126/sciadv.aay5898*

Dunn, D. W. & Follett, P. A. (2017), 'The Sterile Insect Technique (SIT) – an introduction', *Entomologia Experimentalis et Applicata* **164**(3), 151–154.
URL: *https://onlinelibrary.wiley.com/doi/full/10.1111/eea.12619*

Dyck, V. A., Hendrichs, J. & Robinson, A. S. (2021), *Sterile Insect Technique*, CRC Press.

Engels, W. R. (1997), 'Invasions of P Elements', *Genetics* **145**(1), 11.
URL: *https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1207769/*

Erdman, S. E., Chen, H. J. & Burtis, K. C. (1996), 'Functional and genetic characterization of the oligomerization and DNA binding properties of the Drosophila doublesex proteins', *Genetics* **144**(4), 1639–1652.
URL: *https://pubmed.ncbi.nlm.nih.gov/8978051/*

Erickson, J. W. & Quintero, J. J. (2007), 'Indirect effects of ploidy suggest X chromosome dose, not the X:A ratio, signals sex in Drosophila', *PLoS Biology* **5**(12), 2821–2830.
URL: *https://doi.org/10.1371/journal.pbio.0050332*

Esvelt, K. M., Smidler, A. L., Catteruccia, F. & Church, G. M. (2014), 'Concerning RNA-guided gene drives for the alteration of wild populations', *eLife* **3**, 1–21.
URL: *https://doi.org/10.7554/eLife.03401*

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

16(3), e1008647.
URL: *https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1008647*

Feachem, R. G., Chen, I., Akbari, O., Bertozzi-Villa, A., Bhatt, S., Binka, F., Boni, M. F., Buckee, C., Dieleman, J., Dondorp, A., Eapen, A., Sekhri Feachem, N., Filler, S., Gething, P., Gosling, R., Haakenstad, A., Harvard, K., Hatefi, A., Jamison, D., Jones, K. E., Karema, C., Kamwi, R. N., Lal, A., Larson, E., Lees, M., Lobo, N. F., Micah, A. E., Moonen, B., Newby, G., Ning, X., Pate, M., Quiñones, M., Roh, M., Rolfe, B., Shanks, D., Singh, B., Staley, K., Tulloch, J., Wegbreit, J., Woo, H. J. & Mpanju-Shumbusho, W. (2019), 'Malaria eradication within a generation: ambitious, achievable, and necessary', *The Lancet* **394**(10203), 1056–1112.
URL: *http://www.thelancet.com/article/S0140673619311390/fulltext http://www.thelancet.com/article/S0140673619311390/a https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(19)31139-0/abstract*

Fu, G., Lees, R. S., Nimmo, D., Aw, D., Jin, L., Gray, P., Berendonk, T. U., White-Cooper, H., Scaife, S., Phuc, H. K., Marinotti, O., Jasinskiene, N., James, A. A. & Alphey, L. (2010), 'Female-specific flightless phenotype for mosquito control', *Proceedings of the National Academy of Sciences of the United States of America* **107**(10), 4550–4554.
URL: *www.pnas.org/cgi/doi/10.1073/pnas.1000251107*

Fuchs, S., Garrood, W. T., Beber, A., Hammond, A., Galizi, R., Gribble, M., Morselli, G., Hui, T. Y. J., Willis, K., Kranjc, N., Burt, A., Crisanti, A. & Nolan, T. (2021), 'Resistance to a CRISPR-based gene drive at an evolutionarily conserved site is revealed by mimicking genotype fixation', *PLOS Genetics* **17**(10), e1009740.
URL: *https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1009740*

Fuchs, S., Nolan, T. & Crisanti, A. (2013), 'Mosquito transgenic technologies to reduce Plasmodium transmission', *Methods in molecular biology (Clifton, N.J.)* **923**, 601–622.
URL: *https://pubmed.ncbi.nlm.nih.gov/22990807/*

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

Gallup, J. L. & Sachs, J. D. (2001), 'The Economic Burden of Malaria'.
URL: *https://www.ncbi.nlm.nih.gov/books/NBK2624/*

Gantz, V. M. & Bier, E. (2015), 'The mutagenic chain reaction: A method for converting heterozygous to

homozygous mutations', *Science* **348**(6233), 442–444.
URL: *https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4687737/*

Gantz, V. M., Jasinskiene, N., Tatarenkova, O., Fazekas, A., Macias, V. M., Bier, E. & 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 of the United States of America* **112**(49), E6736–E6743.
URL: *www.pnas.org/cgi/doi/10.1073/pnas.1521077112*

Gari, T. & Lindtjørn, B. (2018), 'Reshaping the vector control strategy for malaria elimination in Ethiopia in the context of current evidence and new tools: Opportunities and challenges 11 Medical and Health Sciences 1108 Medical Microbiology 11 Medical and Health Sciences 1117 Public Health and Health Services', *Malaria Journal* **17**(1), 1–8.
URL: *https://malariajournal.biomedcentral.com/articles/10.1186/s12936-018-2607-8*

Garneau, J. E., Dupuis, M. E., Villion, M., Romero, D. A., Barrangou, R., Boyaval, P., Fremaux, C., Horvath, P., Magadan, A. H. & Moineau, S. (2010), 'The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA', *Nature* **468**(7320), 67–71.
URL: *https://www.nature.com/articles/nature09523*

Garrood, W. T., Kranjc, N., Petri, K., Kim, D. Y., Guo, J. A., Hammond, A. M., Morianou, I., Pattanayak, V., Joung, J. K., Crisanti, A. & Simoni, A. (2021), 'Analysis of off-target effects in CRISPR-based gene drives in the human malaria mosquito', *Proceedings of the National Academy of Sciences of the United States of America* **118**(22).
URL: *https://pubmed.ncbi.nlm.nih.gov/34050017/*

Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. (2012), 'Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria', *Proceedings of the National Academy of Sciences of the United States of America* **109**(39), E2579–E2586.
URL: *https://www.pnas.org/content/109/39/E2579 https://www.pnas.org/content/109/39/E2579.abstract*

Gentile, J. E., Rund, S. S. & Madey, G. R. (2015), 'Modelling sterile insect technique to control the population of Anopheles gambiae', *Malaria Journal* **14**(1), 1–12.
URL: *https://link.springer.com/articles/10.1186/s12936-015-0587-5 https://link.springer.com/article/10.1186/s12936-015-0587-5*

Gilbert, L. A., Horlbeck, M. A., Adamson, B., Villalta, J. E., Chen, Y., Whitehead, E. H., Guimaraes, C., Panning, B., Ploegh, H. L., Bassik, M. C., Qi, L. S., Kampmann, M. & Weissman, J. S. (2014), 'Genome-

Scale CRISPR-Mediated Control of Gene Repression and Activation', *Cell* **159**(3), 647.
URL: */pmc/articles/PMC4253859/*

Global Disease Burden, G. (2019), 'GBD Results Tool, GHDx'.
URL: *http://ghdx.healthdata.org/gbd-results-tool*

Gomes, P. S., Bhardwaj, J., Rivera-Correa, J., Freire-De-Lima, C. G. & Morrot, A. (2016), 'Immune Escape Strategies of Malaria Parasites', *Frontiers in Microbiology* **7**(OCT), 1617.
URL: *https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5066453/*

Gong, P., Epton, M. J., Fu, G., Scaife, S., Hiscox, A., Condon, K. C., Condon, G. C., Morrison, N. I., Kelly, D. W., Dafa'Alla, T., Coleman, P. G. & Alphey, L. (2005), 'A dominant lethal genetic system for autocidal control of the Mediterranean fruitfly', *Nature Biotechnology 2005 23:4* **23**(4), 453–456.
URL: *https://www.nature.com/articles/nbt1071*

Gotoh, H., Ishiguro, M., Nishikawa, H., Morita, S., Okada, K., Miyatake, T., Yaginuma, T. & Niimi, T. (2016), 'Molecular cloning and functional characterization of the sex-determination gene doublesex in the sexually dimorphic broad-horned beetle Gnatocerus cornutus (Coleoptera, Tenebrionidae)', *Scientific Reports 2016 6:1* **6**(1), 1–10.
URL: *https://www.nature.com/articles/srep29337*

Gratz, S. J., Ukken, F. P., Rubinstein, C. D., Thiede, G., Donohue, L. K., Cummings, A. M. & Oconnor-Giles, K. M. (2014), 'Highly specific and efficient CRISPR/Cas9-catalyzed homology-directed repair in Drosophila', *Genetics* **196**(4), 961–971.
URL: *https://academic.oup.com/genetics/article/196/4/961/5935648*

Greenwood, B. M. (2008), 'Control to elimination: implications for malaria research', *Trends in Parasitology* **24**(10), 449–454.
URL: *https://doi.org/10.1016/j.pt.2008.07.002*

Grigoraki, L., Cowlishaw, R., Nolan, T., Donnelly, M., Lycett, G. & Ransoni, H. (2021), 'CRISPR/Cas9 modified An. gambiae carrying kdr mutation L1014F functionally validate its contribution in insecticide resistance and combined effect with metabolic enzymes', *PLOS Genetics* **17**(7), e1009556.
URL: *https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1009556*

Guichard, A., Haque, T., Bobik, M., Xu, X. R. S., Klanseck, C., Kushwah, R. B. S., Berni, M., Kaduskar, B., Gantz, V. M. & Bier, E. (2019), 'Efficient allelic-drive in Drosophila', *Nature Communications 2019 10:1* **10**(1), 1–10.
URL: *https://www.nature.com/articles/s41467-019-09694-w*

Halperin, S. O., Tou, C. J., Wong, E. B., Modavi, C., Schaffer, D. V. & Dueber, J. E. (2018), 'CRISPR-guided DNA polymerases enable diversification of all nucleotides in a tunable window', *Nature* **560**(7717), 248–252.
URL: *https://pubmed.ncbi.nlm.nih.gov/30069054/*

Hammond, A., Galizi, R., Kyrou, K., Simoni, A., Siniscalchi, C., Katsanos, D., Gribble, M., Baker, D., Marois, E., Russell, S., Burt, A., Windbichler, N., Crisanti, A. & Nolan, T. (2016), 'A CRISPR-Cas9 gene drive system targeting female reproduction in the malaria mosquito vector Anopheles gambiae', *Nature Biotechnology* **34**, 78–83.
URL: *https://doi.org/10.1038/nbt.3439*

Hammond, A., Karlsson, X., Morianou, I., Kyrou, K., Beaghton, A., Gribble, M., Kranjc, N., Galizi, R., Burt, A., Crisanti, A. & Nolan, T. (2021), 'Regulating the expression of gene drives is key to increasing their invasive potential and the mitigation of resistance', *PLOS Genetics* **17**(1), e1009321.
URL: *https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1009321*

Hammond, A., Kyrou, K., Gribble, M., Karlsson, X., Morianou, I., Galizi, R., Beaghton, A., Crisanti, A. & Nolan, T. (2018), 'Improved CRISPR-based suppression gene drives mitigate resistance and impose a large reproductive load on laboratory-contained mosquito populations', *bioRxiv* p. 360339.
URL: *https://www.biorxiv.org/content/10.1101/360339v1*

Hammond, A. M. & Galizi, R. (2017), 'Gene drives to fight malaria: current state and future directions'.
URL: *https://pubmed.ncbi.nlm.nih.gov/29457956/*

Hammond, A. M., Kyrou, K., Bruttini, M., North, A., Galizi, R., Karlsson, X., Kranjc, N., Carpi, F. M., D'Aurizio, R., Crisanti, A. & Nolan, T. (2017), 'The creation and selection of mutations resistant to a gene drive over multiple generations in the malaria mosquito', *PLoS Genetics* **13**(10), e1007039.
URL: *https://doi.org/10.1371/journal.pgen.1007039*

Hammond, A., Pollegioni, P., Persampieri, T., North, A., Minuz, R., Trusso, A., Bucci, A., Kyrou, K., Morianou, I., Simoni, A., Nolan, T., Müller, R. & Crisanti, A. (2021), 'Gene-drive suppression of mosquito populations in large cages as a bridge between lab and field', *Nature Communications 2021 12:1* **12**(1), 1–9.
URL: *https://www.nature.com/articles/s41467-021-24790-6*

Harris, A. F., Nimmo, D., McKemey, A. R., Kelly, N., Scaife, S., Donnelly, C. A., Beech, C., Petrie, W. D. & Alphey, L. (2011), 'Field performance of engineered male mosquitoes', *Nature Biotechnology 2011 29:11* **29**(11), 1034–1037.
URL: *https://www.nature.com/articles/nbt.2019*

Hay, B. A., Chen, C. H., Ward, C. M., Huang, H., Su, J. T. & Guo, M. (2010), 'Engineering the genomes of wild insect populations: Challenges, and opportunities provided by synthetic Medea selfish genetic elements', *Journal of Insect Physiology* **56**(10), 1402–1413.

Heinrich, J. C. & Scott, M. J. (2000), 'A repressible female-specific lethal genetic system for making transgenic insect strains suitable for a sterile-release program', *Proc. Nat. Acad. Sci. (USA)* **97**(15), 8229–8232.

Hemingway, J., Ranson, H., Magill, A., Kolaczinski, J., Fornadel, C., Gimnig, J., Coetzee, M., Simard, F., Roch, D. K., Hinzoumbe, C. K., Pickett, J., Schellenberg, D., Gething, P., Hoppé, M. & Hamon, N. (2016), 'Averting a malaria disaster: will insecticide resistance derail malaria control?', *The Lancet* **387**(10029), 1785–1788.

Hemming-Schroeder, E., Zhong, D., Machani, M., Nguyen, H., Thong, S., Kahindi, S., Mbogo, C., Atieli, H., Githeko, A., Lehmann, T., Kazura, J. W. & Yan, G. (2020), 'Ecological drivers of genetic connectivity for African malaria vectors Anopheles gambiae and An. arabiensis', *Scientific Reports 2020 10:1* **10**(1), 1–12.
URL: *https://www.nature.com/articles/s41598-020-76248-2*

Heuschen, A. K., Lu, G., Razum, O., Abdul-Mumin, A., Sankoh, O., von Seidlein, L., D'Alessandro, U. & Müller, O. (2021), 'Public health-relevant consequences of the COVID-19 pandemic on malaria in sub-Saharan Africa: a scoping review', *Malaria Journal* **20**(1), 1–16.
URL: *https://malariajournal.biomedcentral.com/articles/10.1186/s12936-021-03872-2*

Hoermann, A., Tapanelli, S., Capriotti, P., Del Corsano, G., Masters, E. K., Habtewold, T., Christophides, G. K. & Windbichler, N. (2021), 'Converting endogenous genes of the malaria mosquito into simple non-autonomous gene drives for population replacement', *eLife* **10**.

Hoffmann, A. A., Montgomery, B. L., Popovici, J., Iturbe-Ormaetxe, I., Johnson, P. H., Muzzi, F., Greenfield, M., Durkan, M., Leong, Y. S., Dong, Y., Cook, H., Axford, J., Callahan, A. G., Kenny, N., Omodei, C., McGraw, E. A., Ryan, P. A., Ritchie, S. A., Turelli, M. & O'Neill, S. L. (2011), 'Successful establishment of Wolbachia in Aedes populations to suppress dengue transmission', *Nature 2011 476:7361* **476**(7361), 454–457.
URL: *https://www.nature.com/articles/nature10356*

Hogan, A. B., Winskill, P. & Ghani, A. C. (2020), 'Estimated impact of RTS,S/AS01 malaria vaccine allocation strategies in sub-Saharan Africa: A modelling study', *PLoS medicine* **17**(11).
URL: *https://pubmed.ncbi.nlm.nih.gov/33253211/*

Holt, R. A., Mani Subramanian, G., Halpern, A., Sutton, G. G., Charlab, R., Nusskern, D. R., Wincker, P., Clark, A. G., Ribeiro, J. M., Wides, R., Salzberg, S. L., Loftus, B., Yandell, M., Majoros, W. H., Rusch, D. B., Lai, Z., Kraft, C. L., Abril, J. F., Anthouard, V., Arensburger, P., Atkinson, P. W., Baden, H., de Berardinis, V., Baldwin, D., Benes, V., Biedler, J., Blass, C., Bolanos, R., Boscus, D., Barnstead, M., Cai, S., Center, A., Chatuverdi, K., Christophides, G. K., Chrystal, M. A., Clamp, M., Cravchik, A., Curwen, V., Dana, A., Delcher, A., Dew, I., Evans, C. A., Flanigan, M., Grundschober-Freimoser, A., Friedli, L., Gu, Z., Guan, P., Guigo, R., Hillenmeyer, M. E., Hladun, S. L., Hogan, J. R., Hong, Y. S., Hoover, J., Jaillon, O., Ke, Z., Kodira, C., Kokoza, E., Koutsos, A., Letunic, I., Levitsky, A., Liang, Y., Lin, J. J., Lobo, N. F., Lopez, J. R., Malek, J. A., McIntosh, T. C., Meister, S., Miller, J., Mobarry, C., Mongin, E., Murphy, S. D., O'Brochta, D. A., Pfannkoch, C., Qi, R., Regier, M. A., Remington, K., Shao, H., Sharakhova, M. V., Sitter, C. D., Shetty, J., Smith, T. J., Strong, R., Sun, J., Thomasova, D., Ton, L. Q., Topalis, P., Tu, Z., Unger, M. F., Walenz, B., Wang, A., Wang, J., Wang, M., Wang, X., Woodford, K. J., Wortman, J. R., Wu, M., Yao, A., Zdobnov, E. M., Zhang, H., Zhao, Q., Zhao, S., Zhu, S. C., Zhimulev, I., Coluzzi, M., della Torre, A., Roth, C. W., Louis, C., Kalush, F., Mural, R. J., Myers, E. W., Adams, M. D., Smith, H. O., Broder, S., Gardner, M. J., Fraser, C. M., Birney, E., Bork, P., Brey, P. T., Craig Venter, J., Weissenbach, J., Kafatos, F. C., Collins, F. H. & Hoffman, S. L. (2002), 'The genome sequence of the malaria mosquito Anopheles gambiae', *Science* **298**(5591), 129–149.
URL: *https://www.science.org/doi/10.1126/science.1076181*

Hozumi, A., Yoshida, R., Horie, T., Sakuma, T., Yamamoto, T. & Sasakura, Y. (2013), 'Enhancer activity sensitive to the orientation of the gene it regulates in the chordategenome', *Developmental Biology* **375**(1), 79–91.

Hsu, P. D., Scott, D. A., Weinstein, J. A., Ran, F. A., Konermann, S., Agarwala, V., Li, Y., Fine, E. J., Wu, X., Shalem, O., Cradick, T. J., Marraffini, L. A., Bao, G. & Zhang, F. (2013), 'DNA targeting specificity of RNA-guided Cas9 nucleases', *Nature Biotechnology 2013 31:9* **31**(9), 827–832.
URL: *https://www.nature.com/articles/nbt.2647*

Hurst, G. D. & Werren, J. H. (2001), 'The role of selfish genetic elements in eukaryotic evolution', *Nature Reviews Genetics 2001 2:8* **2**(8), 597–606.
URL: *https://www.nature.com/articles/35084545*

Ibrahim, M. M., Karabacak, A., Glahs, A., Kolundzic, E., Hirsekorn, A., Carda, A., Tursun, B., Zinzen, R. P., Lacadie, S. A. & Ohler, U. (2018), 'Determinants of promoter and enhancer transcription directionality in metazoans', *Nature Communications 2018 9:1* **9**(1), 1–15.
URL: *https://www.nature.com/articles/s41467-018-06962-z*

James, A. A. (2005), 'Gene drive systems in mosquitoes: rules of the road', *Trends in parasitology* **21**(2), 64–67.
URL: *https://pubmed.ncbi.nlm.nih.gov/15664528/*

Jeffries, C. L., Lawrence, G. G., Golovko, G., Kristan, M., Orsborne, J., Spence, K., Hurn, E., Bandibabone, J., Tantely, L. M., Raharimalala, F. N., Keita, K., Camara, D., Barry, Y., Wat'senga, F., Manzambi, E. Z., Afrane, Y. A., Mohammed, A. R., Abeku, T. A., Hedge, S., Khanipov, K., Pimenova, M., Fofanov, Y., Boyer, S., Irish, S. R., Hughes, G. L. & Walker, T. (2018), 'Novel Wolbachia strains in Anopheles malaria vectors from Sub-Saharan Africa', *Wellcome Open Research* **3**, 113.
URL: *https://pubmed.ncbi.nlm.nih.gov/30483601/*

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A. & Charpentier, E. (2012), 'A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity', *Science* **337**(6069), 816–821.
URL: *https://doi.org/10.1126/science.1225829*

Joung, J. K. & Sander, J. D. (2012), 'TALENs: a widely applicable technology for targeted genome editing', *Nature Reviews Molecular Cell Biology 2012 14:1* **14**(1), 49–55.
URL: *https://www.nature.com/articles/nrm3486*

Kaduskar, B., Kushwah, R. B. S., Auradkar, A., Guichard, A., Li, M., Bennett, J. B., Julio, A. H. F., Marshall, J. M., Montell, C. & Bier, E. (2022), 'Reversing insecticide resistance with allelic-drive in Drosophila melanogaster', *Nature Communications 2022 13:1* **13**(1), 1–8.
URL: *https://www.nature.com/articles/s41467-021-27654-1*

Kaiser, M. L., Wood, O. R., Damiens, D., Brooke, B. D., Koekemoer, L. L. & Munhenga, G. (2021), 'Estimates of the population size and dispersal range of Anopheles arabiensis in Northern KwaZulu-Natal, South Africa: implications for a planned pilot programme to release sterile male mosquitoes', *Parasites and Vectors* **14**(1), 1–18.
URL: *https://parasitesandvectors.biomedcentral.com/articles/10.1186/s13071-021-04674-w*

Kandul, N. P., Liu, J., Buchman, A., Gantz, V. M., Bier, E. & Akbari, O. S. (2020), 'Assessment of a Split Homing Based Gene Drive for Efficient Knockout of Multiple Genes', *G3 Genes—Genomes—Genetics* **10**(2), 827–837.
URL: *https://academic.oup.com/g3journal/article/10/2/827/6026312*

Kandul, N. P., Liu, J., Sanchez C, H. M., Wu, S. L., Marshall, J. M. & Akbari, O. S. (2019), 'Transforming insect population control with precision guided sterile males with demonstration in flies', *Nature*

Communications 2019 10:1 **10**(1), 1–12.
URL: *https://www.nature.com/articles/s41467-018-07964-7*

Khatri, B. S. & Burt, A. (2022), 'Weakly deleterious natural genetic variation greatly amplifies probability of resistance in multiplexed gene drive systems', *bioRxiv* p. 2021.12.23.473701.
URL: *https://www.biorxiv.org/content/10.1101/2021.12.23.473701v2*

Khatri, B. S., Burt, A. & Kim, Y. (2019), 'Robust Estimation of Recent Effective Population Size from Number of Independent Origins in Soft Sweeps', *Molecular Biology and Evolution* **36**(9), 2040.
URL: */pmc/articles/PMC6736332/*

Kim, J. S. (2016), 'Genome editing comes of age', *Nature Protocols 2016 11:9* **11**(9), 1573–1578.
URL: *https://www.nature.com/articles/nprot.2016.104*

Kistler, K. E., Vosshall, L. B. & Matthews, B. J. (2015), 'Genome engineering with CRISPR-Cas9 in the mosquito Aedes aegypti', *Cell reports* **11**(1), 51–60.
URL: *https://pubmed.ncbi.nlm.nih.gov/25818303/*

Kleinschmidt, I. & Rowland, M. (2021), Insecticides and malaria, *in* C. J.M. Koenraadt, J. Spitzen & W. Takken, eds, 'Innovative strategies for vector control', Vol. 6, Wageningen Academic, chapter 2, pp. 17–32.

Kranjc, N. (2022), Expanding the flexibility of genome editing approaches for population control of the malaria mosquito, PhD thesis, Imperial College London, London.
URL: *https://doi.org/10.25560/96597*

Kranjc, N., Crisanti, A., Nolan, T. & Bernardini, F. (2021), 'Anopheles gambiae Genome Conservation as a Resource for Rational Gene Drive Target Site Selection', *Insects* **12**(2), 1–12.
URL: */pmc/articles/PMC7911984/*

Krzywinska, E., Dennison, N. J., Lycett, G. J. & Krzywinski, J. (2016), 'A maleness gene in the malaria mosquito Anopheles gambiae', *Science* **353**(6294), 67–69.
URL: *https://doi.org/10.1126/science.aaf5605*

Krzywinska, E., Ferretti, L. & Krzywinski, J. (2022), 'Establishment and a comparative transcriptomic analysis of a male-specific cell line from the African malaria mosquito Anopheles gambiae', *Scientific Reports 2022 12:1* **12**(1), 1–8.
URL: *https://www.nature.com/articles/s41598-022-10686-y*

Krzywinska, E., Ferretti, L., Li, J., Li, J. C., Chen, C. H. & Krzywinski, J. (2021), 'femaleless Controls Sex Determination and Dosage Compensation Pathways in Females of Anopheles Mosquitoes', *Current Biology* **31**(5), 1084–1091.
URL: *https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7955153/*

Kyrou, K. (2021), Targeting sex determination for genetic control of the malaria mosquito, PhD thesis, Imperial College London, London.
URL: *https://doi.org/10.25560/96822*

Kyrou, K., Hammond, A. M., Galizi, R., Kranjc, N., Burt, A., Beaghton, A. K., Nolan, T. & Crisanti, A. (2018), 'A CRISPR–Cas9 gene drive targeting doublesex causes complete population suppression in caged Anopheles gambiae mosquitoes', *Nature Biotechnology* .
URL: *https://doi.org/10.1038/nbt.4245*

LaManna, C. M. & Barrangou, R. (2018), 'Enabling the Rise of a CRISPR World', *https://home.liebertpub.com/crispr* **1**(3), 205–208.
URL: *https://www.liebertpub.com/doi/abs/10.1089/crispr.2018.0022*

Lee, A. H., Symington, L. S. & Fidock, D. A. (2014), 'DNA repair mechanisms and their biological roles in the malaria parasite Plasmodium falciparum', *Microbiology and molecular biology reviews : MMBR* **78**(3), 469–486.
URL: *https://pubmed.ncbi.nlm.nih.gov/25184562/*

Levi, T., Sloutskin, A., Kalifa, R., Juven-Gershon, T. & Gerlitz, O. (2020), 'Efficient in Vivo Introduction of Point Mutations Using ssODN and a Co-CRISPR Approach', *Biological Procedures Online* **22**(1), 1–12.
URL: *https://biologicalproceduresonline.biomedcentral.com/articles/10.1186/s12575-020-00123-7*

Li, M., Yang, T., Kandul, N. P., Bui, M., Gamez, S., Raban, R., Bennett, J., Sánchez C, H. M., Lanzaro, G. C., Schmidt, H., Lee, Y., Marshall, J. M. & Akbari, O. S. (2020), 'Development of a confinable gene drive system in the human disease vector aedes aegypti', *eLife* **9**.
URL: *https://elifesciences.org/articles/51701*

Lin, J. & Vogt, V. M. (1998), 'I-PpoI, the Endonuclease Encoded by the Group I Intron PpLSU3, Is Expressed from an RNA Polymerase I Transcript', *Molecular and Cellular Biology* **18**(10), 5809.
URL: */pmc/articles/PMC109167/*

Liu, W.-L., Id, H.-Y. Y., Chen, Y.-X., Chen, B.-Y., Leaw, S. N., Linid, C.-H., Suid, M.-P., Tsai, L.-S., Chen, Y., Shiaoid, S.-H., Xiid, Z., Jang, A. C.-C., Chenid, C.-H. & Yu, H.-Y. (2022), 'Lab-scale characterization

and semi-field trials of Wolbachia Strain wAlbB in a Taiwan Wolbachia introgressed Ae. aegypti strain', *PLOS Neglected Tropical Diseases* **16**(1), e0010084.
URL: *https://journals.plos.org/plosntds/article?id=10.1371/journal.pntd.0010084*

Lofgren, C. S., Dame, D. A., Breeland, S. G., Weidhaas, D. E., Jeffery, G., Kaiser, R., Ford, H. R., Boston, M. D. & Baldwin, K. F. (1974), 'Release of chemosterilized males for the control of Anopheles albimanus in El Salvador. 3. Field methods and population control', *The American journal of tropical medicine and hygiene* **23**(2), 288–297.
URL: *https://pubmed.ncbi.nlm.nih.gov/4817674/*

López Del Amo, V., Bishop, A. L., Sánchez C, H. M., Bennett, J. B., Feng, X., Marshall, J. M., Bier, E. & Gantz, V. M. (2020), 'A transcomplementing gene drive provides a flexible platform for laboratory investigation and potential field deployment', *Nature Communications 2020 11:1* **11**(1), 1–12.
URL: *https://www.nature.com/articles/s41467-019-13977-7*

Lutrat, C., Burckbuchler, M., Olmo, R. P., Beugnon, R., Fontaine, A., Akbari, O. S., Argilés-Herrero, R., Baldet, T., Bouyer, J. & Marois, E. (2022), 'Combining two Genetic Sexing Strains allows sorting of non-transgenic males for Aedes genetic control Authors', *bioRxiv* .
URL: *https://doi.org/10.1101/2022.03.11.483912*

Lynch, K. W. & Maniatis, T. (1995), 'Synergistic interactions between two distinct elements of a regulated splicing enhancer', *Genes and Development* .
URL: *https://doi.org/10.1101/gad.9.3.284*

Mali, P., Esvelt, K. M. & Church, G. M. (2013), 'Cas9 as a versatile tool for engineering biology', *Nature Methods* **10**(10), 957–963.
URL: *https://www.nature.com/articles/nmeth.2649*

Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., DiCarlo, J. E., Norville, J. E. & Church, G. M. (2013), 'RNA-guided human genome engineering via Cas9', *Science (New York, N.Y.)* **339**(6121), 823–826.
URL: *https://pubmed.ncbi.nlm.nih.gov/23287722/*

Marois, E., Scali, C., Soichot, J., Kappler, C., Levashina, E. A. & Catteruccia, F. (2012), 'High-throughput sorting of mosquito larvae for laboratory studies and for future vector control interventions', *Malaria Journal* **11**(302).
URL: *https://doi.org/10.1186/1475-2875-11-302*

Marr, E. & Potter, C. J. (2021), 'Base Editing of Somatic Cells Using CRISPR-Cas9 in Drosophila', *CRISPR Journal* **4**(6), 836–845.
URL: *https://www.biorxiv.org/content/10.1101/2021.03.24.436868v1*

Marraffini, L. A. & Sontheimer, E. J. (2008), 'CRISPR Interference Limits Horizontal Gene Transfer in Staphylococci by Targeting DNA', *Science (New York, N.Y.)* **322**(5909), 1843.
URL: */pmc/articles/PMC2695655/*

Marshall, J. M., Buchman, A., Sánchez, C. H. M. & Akbari, O. S. (2017), 'Overcoming evolved resistance to population-suppressing homing-based gene drives', *Scientific Reports* **7**(3776).
URL: *https://doi.org/10.1038/s41598-017-02744-7*

Marshall, J. M., Raban, R. R., Kandul, N. P., Edula, J. R., León, T. M. & Akbari, O. S. (2019), 'Winning the tug-of-war between effector gene design and pathogen evolution in vector population replacement strategies', *Frontiers in Genetics* **10**(OCT), 1072.
URL: *https://doi.org/10.3389/fgene.2019.01072*

McCarty, N. S., Graham, A. E., Studená, L. & Ledesma-Amaro, R. (2020), 'Multiplexed CRISPR technologies for gene editing and transcriptional regulation', *Nature Communications 2020 11:1* **11**(1), 1–13.
URL: *https://www.nature.com/articles/s41467-020-15053-x*

McMeniman, C. J., Lane, R. V., Cass, B. N., Fong, A. W., Sidhu, M., Wang, Y. F. & O'Neill, S. L. (2009), 'Stable introduction of a life-shortening Wolbachia infection into the mosquito Aedes aegypti', *Science (New York, N.Y.)* **323**(5910), 141–144.
URL: *https://pubmed.ncbi.nlm.nih.gov/19119237/*

Meccariello, A., Krsticevic, F., Colonna, R., Del Corsano, G., Fasulo, B., Papathanos, P. A. & Windbichler, N. (2021), 'Engineered sex ratio distortion by X-shredding in the global agricultural pest Ceratitis capitata', *BMC Biology* **19**(1), 1–14.
URL: *https://bmcbiol.biomedcentral.com/articles/10.1186/s12915-021-01010-7*

Mendis, K., Rietveld, A., Warsame, M., Bosman, A., Greenwood, B. & Wernsdorfer, W. H. (2009), 'From malaria control to eradication: The WHO perspective', *Tropical Medicine & International Health* **14**(7), 802–809.
URL: *https://onlinelibrary.wiley.com/doi/full/10.1111/j.1365-3156.2009.02287.x*

Metzloff, M., Yang, E., Dhole, S., Clark, A. G., Messer, P. W. & Champer, J. (2021), 'Experimental demonstration of tethered gene drive systems for confined population modification or suppression', *BioRxiv* .
URL: *https://doi.org/10.1101/2021.05.29.446308*

Miles, A., Harding, N. J., Bottà, G., Clarkson, C. S., Antão, T., Kozak, K., Schrider, D. R., Kern, A. D., Redmond, S., Sharakhov, I., Pearson, R. D., Bergey, C., Fontaine, M. C., Donnelly, M. J., Lawniczak, M. K., Ayala, D., Besansky, N. J., Burt, A., Caputo, B., Torre, A. D., Godfray, H. C. J., Hahn, M. W., Midega, J., Neafsey, D. E., O'Loughlin, S., Pinto, J., Riehle, M. M., Vernick, K. D., Weetman, D., Wilding, C. S., White, B. J., Troco, A. D., Diabaté, A., Costantini, C., Rohatgi, K. R., Elissa, N., Coulibaly, B., Dinis, J., Mbogo, C., Bejon, P., Mawejje, H. D., Stalker, J., Rockett, K., Drury, E., Mead, D., Jeffreys, A., Hubbart, C., Rowlands, K., Isaacs, A. T., Jyothi, D., Malangone, C., Vauterin, P., Jeffery, B., Wright, I., Hart, L., Kluczyński, K., Cornelius, V., Macinnis, B., Henrichs, C., Giacomantonio, R. & Kwiatkowski, D. P. (2017), 'Genetic diversity of the African malaria vector anopheles gambiae', *Nature* **552**(7683), 96–100.
URL: *https://www.nature.com/articles/nature24995*

Mojica, F. J., Díez-Villaseñor, C., García-Martínez, J. & Soria, E. (2005), 'Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements', *Journal of molecular evolution* **60**(2), 174–182.
URL: *https://pubmed.ncbi.nlm.nih.gov/15791728/*

Moreira, L. A., Iturbe-Ormaetxe, I., Jeffery, J. A., Lu, G., Pyke, A. T., Hedges, L. M., Rocha, B. C., Hall-Mendelin, S., Day, A., Riegler, M., Hugo, L. E., Johnson, K. N., Kay, B. H., McGraw, E. A., van den Hurk, A. F., Ryan, P. A. & O'Neill, S. L. (2009), 'A Wolbachia Symbiont in Aedes aegypti Limits Infection with Dengue, Chikungunya, and Plasmodium', *Cell* **139**(7), 1268–1278.
URL: *https://doi.org/10.1016/j.cell.2009.11.042*

Morianou, I., Crisanti, A., Nolan, T. & Hammond, A. (2022), 'CriMCE: A method to introduce and isolate precise marker-less edits via CRISPR-mediated cassette exchange', *The CRISPR Journal* **5**(6).
URL: *https://doi.org/10.1089/crispr.2022.0026*

Moyes, C. L., Athinya, D. K., Seethaler, T., Battle, K. E., Sinka, M., Hadi, M. P., Hemingway, J., Coleman, M. & Hancock, P. A. (2020), 'Evaluating insecticide resistance across african districts to aid malaria control decisions', *Proceedings of the National Academy of Sciences of the United States of America* **117**(36), 22042–22050.
URL: *https://www.pnas.org/content/117/36/22042 https://www.pnas.org/content/117/36/22042.abstract*

Mwakingwe-Omari, A., Healy, S. A., Lane, J., Cook, D. M., Kalhori, S., Wyatt, C., Kolluri, A., Marte-Salcedo, O., Imeru, A., Nason, M., Ding, L. K., Decederfelt, H., Duan, J., Neal, J., Raiten, J., Lee, G., Hume, J. C., Jeon, J. E., Ikpeama, I., Kc, N., Chakravarty, S., Murshedkar, T., Church, L. W., Manoj, A., Gunasekera, A., Anderson, C., Murphy, S. C., March, S., Bhatia, S. N., James, E. R., Billingsley, P. F.,

Sim, B. K. L., Richie, T. L., Zaidi, I., Hoffman, S. L. & Duffy, P. E. (2021), 'Two chemoattenuated PfSPZ malaria vaccines induce sterile hepatic immunity', *Nature 2021 595:7866* **595**(7866), 289–294.
URL: *https://www.nature.com/articles/s41586-021-03684-z*

Nakamura, M., Gao, Y., Dominguez, A. A. & Qi, L. S. (2021), 'CRISPR technologies for precise epigenome editing', *Nature Cell Biology 2021 23:1* **23**(1), 11–22.
URL: *https://www.nature.com/articles/s41556-020-00620-7*

Nash, A., Urdaneta, G. M., Beaghton, A. K., Hoermann, A., Papathanos, P. A., Christophides, G. K. & Windbichler, N. (2019), 'Integral gene drives for population replacement', *Biology open* **8**(1).
URL: *https://pubmed.ncbi.nlm.nih.gov/30498016/*

Ndo, C., Poumachu, Y., Metitsi, D., Awono-Ambene, H. P., Tchuinkam, T., Gilles, J. L. R. & Bourtzis, K. (2018), 'Isolation and characterization of a temperature-sensitive lethal strain of Anopheles arabiensis for SIT-based application', *Parasites and Vectors* **11**(2), 97–105.
URL: *https://parasitesandvectors.biomedcentral.com/articles/10.1186/s13071-018-3216-7*

Neafsey, D. E., Waterhouse, R. M., Abai, M. R., Aganezov, S. S., Alekseyev, M. A., Allen, J. E., Amon, J., Arcà, B., Arensburger, P., Artemov, G., Assour, L. A., Basseri, H., Berlin, A., Birren, B. W., Blandin, S. A., Brockman, A. I., Burkot, T. R., Burt, A., Chan, C. S., Chauve, C., Chiu, J. C., Christensen, M., Costantini, C., Davidson, V. L., Deligianni, E., Dottorini, T., Dritsou, V., Gabriel, S. B., Guelbeogo, W. M., Hall, A. B., Han, M. V., Hlaing, T., Hughes, D. S., Jenkins, A. M., Jiang, X., Jungreis, I., Kakani, E. G., Kamali, M., Kemppainen, P., Kennedy, R. C., Kirmitzoglou, I. K., Koekemoer, L. L., Laban, N., Langridge, N., Lawniczak, M. K., Lirakis, M., Lobo, N. F., Lowy, E., MacCallum, R. M., Mao, C., Maslen, G., Mbogo, C., McCarthy, J., Michel, K., Mitchell, S. N., Moore, W., Murphy, K. A., Naumenko, A. N., Nolan, T., Novoa, E. M., O'Loughlin, S., Oringanje, C., Oshaghi, M. A., Pakpour, N., Papathanos, P. A., Peery, A. N., Povelones, M., Prakash, A., Price, D. P., Rajaraman, A., Reimer, L. J., Rinker, D. C., Rokas, A., Russell, T. L., Sagnon, N., Sharakhova, M. V., Shea, T., Simão, F. A., Simard, F., Slotman, M. A., Somboon, P., Stegniy, V., Struchiner, C. J., Thomas, G. W., Tojo, M., Topalis, P., Tubio, J. M., Unger, M. F., Vontas, J., Walton, C., Wilding, C. S., Willis, J. H., Wu, Y. C., Yan, G., Zdobnov, E. M., Zhou, X., Catteruccia, F., Christophides, G. K., Collins, F. H., Cornman, R. S., Crisanti, A., Donnelly, M. J., Emrich, S. J., Fontaine, M. C., Gelbart, W., Hahn, M. W., Hansen, I. A., Howell, P. I., Kafatos, F. C., Kellis, M., Lawson, D., Louis, C., Luckhart, S., Muskavitch, M. A., Ribeiro, J. M., Riehle, M. A., Sharakhov, I. V., Tu, Z., Zwiebel, L. J. & Besansky, N. J. (2015), 'Highly evolvable malaria vectors: The genomes of 16 Anopheles mosquitoes', *Science* **347**(6217).
URL: *https://doi.org/10.1126/science.1258522*

Nguyen, T. H., Nguyen, H. L., Nguyen, T. Y., Vu, S. N., Tran, N. D., Le, T. N., Vien, Q. M., Bui, T. C., Le, H. T., Kutcher, S., Hurst, T. P., Duong, T. T., Jeffery, J. A., Darbro, J. M., Kay, B. H., Iturbe-Ormaetxe, I., Popovici, J., Montgomery, B. L., Turley, A. P., Zigterman, F., Cook, H., Cook, P. E., Johnson, P. H., Ryan, P. A., Paton, C. J., Ritchie, S. A., Simmons, C. P., O'Neill, S. L. & Hoffmann, A. A. (2015), 'Field evaluation of the establishment potential of wmelpop Wolbachia in Australia and Vietnam for dengue control', *Parasites and Vectors* **8**(1), 1–14.
URL: *https://link.springer.com/articles/10.1186/s13071-015-1174-x https://link.springer.com/article/10.1186/s13071-015-1174-x*

Niki, Y. (1984), 'Developmental analysis of the grandchildless (gs(1)N26) mutation in Drosophila melanogaster: Abnormal cleavage patterns and defects in pole cell formation', *Developmental Biology* **103**(1), 182–189.
URL: *https://doi.org/10.1016/0012-1606(84)90019-8*

Nolan, T. (2021), 'Control of malaria-transmitting mosquitoes using gene drives', *Philosophical Transactions of the Royal Society B* **376**(1818), 20190803.
URL: *https://royalsocietypublishing.org/doi/abs/10.1098/rstb.2019.0803*

Nolan, T., Bower, T. M., Brown, A. E., Crisanti, A. & Catteruccia, F. (2002), 'piggyBac-mediated germline transformation of the malaria mosquito Anopheles stephensi using the red fluorescent protein dsRED as a selectable marker', *The Journal of biological chemistry* **277**(11), 8759–8762.
URL: *https://pubmed.ncbi.nlm.nih.gov/11805082/*

North, A. R., Burt, A. & 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), 1–14.
URL: *https://bmcbiol.biomedcentral.com/articles/10.1186/s12915-020-00834-z*

Nsanzabana, C. (2019), 'Resistance to Artemisinin Combination Therapies (ACTs): Do Not Forget the Partner Drug!', *Tropical Medicine and Infectious Disease* **4**(1), 26.
URL: */pmc/articles/PMC6473515/*

Oberhofer, G., Ivy, T. & Hay, B. A. (2018), 'Behavior of homing endonuclease gene drives targeting genes required for viability or female fertility with multiplexed guide RNAs', *Proceedings of the National Academy of Sciences of the United States of America* **115**(40), E9343–E9352.
URL: *www.pnas.org/cgi/doi/10.1073/pnas.1805278115*

Oberhofer, G., Ivy, T. & Hay, B. A. (2019), 'Cleave and Rescue, a novel selfish genetic element and general strategy for gene drive', *Proceedings of the National Academy of Sciences of the United States of America*

116(13), 6250–6259.
URL: *www.pnas.org/cgi/doi/10.1073/pnas.1816928116*

O'Brochta, D. (2003), 'Gene vector and transposable element behavior in mosquitoes', *Journal of Experimental Biology* **206**, 3823–3834.
URL: *https://doi.org/10.1242/jeb.00638*

Okumu, F. O., Kiware, S. S., Moore, S. J. & Killeen, G. F. (2013), 'Mathematical evaluation of community level impact of combining bed nets and indoor residual spraying upon malaria transmission in areas where the main vectors are Anopheles arabiensis mosquitoes', *Parasites and Vectors* **6**(1), 1–13.
URL: *https://parasitesandvectors.biomedcentral.com/articles/10.1186/1756-3305-6-17*

O'Loughlin, S. M., Magesa, S., Mbogo, C., Mosha, F., Midega, J., Lomas, S. & Burt, A. (2014), 'Genomic Analyses of Three Malaria Vectors Reveals Extensive Shared Polymorphism but Contrasting Population Histories', *Molecular Biology and Evolution* **31**(4), 889.
URL: */pmc/articles/PMC3969563/*

O'leary, S. & Adelman, Z. N. (2020), 'CRISPR/Cas9 knockout of female-biased genes AeAct-4 or myofem in Ae. aegypti results in a flightless phenotype in female, but not male mosquitoes', *PLOS Neglected Tropical Diseases* **14**(12), e0008971.
URL: *https://journals.plos.org/plosntds/article?id=10.1371/journal.pntd.0008971*

Papathanos, P. A., Windbichler, N., Menichelli, M., Burt, A. & Crisanti, A. (2009), 'The vasa regulatory region mediates germline expression and maternal transmission of proteins in the malaria mosquito Anopheles gambiae: A versatile tool for genetic control strategies', *BMC Molecular Biology* **10**(65).
URL: *https://doi.org/10.1186/1471-2199-10-65*

Pham, T. B., Phong, C. H., Bennett, J. B., Hwang, K., Jasinskiene, N., Parker, K., Stillinger, D., Marshall, J. M., Carballar-Lejarazú, R. & James, A. A. (2019), 'Experimental population modification of the malaria vector mosquito, Anopheles stephensi', *PLOS Genetics* **15**(12), e1008440.
URL: *https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1008440*

Phuc, H., Andreasen, M. H., Burton, R. S., Vass, C., Epton, M. J., Pape, G., Fu, G., Condon, K. C., Scaife, S., Donnelly, C. A., Coleman, P. G., White-Cooper, H. & Alphey, L. (2007), 'Late-acting dominant lethal genetic systems and mosquito control', *BMC Biology* **5**(1), 1–11.
URL: *https://bmcbiol.biomedcentral.com/articles/10.1186/1741-7007-5-11*

Price, D. C., Egizi, A. & Fonseca, D. M. (2015), 'The ubiquity and ancestry of insect doublesex', *Scientific Reports 2015 5:1* **5**(1), 1–9.
URL: *https://www.nature.com/articles/srep13068*

Purdy, M., Robinson, M., Wei, K. & Rublin, D. (2013), 'The Economic Case for Combating Malaria', *The American Journal of Tropical Medicine and Hygiene* **89**(5), 819.
URL: */pmc/articles/PMC3820322/*

Randhawa, S. & Sengar, S. (2021), 'The evolution and history of gene editing technologies', *Progress in Molecular Biology and Translational Science* **178**, 1–62.
URL: *https://doi.org/10.1016/bs.pmbts.2021.01.002*

Rashid, I., Campos, M., Collier, T., Crepeau, M., Weakley, A., Gripkey, H., Lee, Y., Schmidt, H. & Lanzaro, G. C. (2022), 'Spontaneous mutation rate estimates for the principal malaria vectors Anopheles coluzzii and Anopheles stephensi', *Scientific Reports 2022 12:1* **12**(1), 1–11.
URL: *https://www.nature.com/articles/s41598-021-03943-z*

Ribeiro, J. M. & Kidwell, M. G. (1994), 'Transposable elements as population drive mechanisms: specification of critical parameter values', *Journal of medical entomology* **31**(1), 10–16.
URL: *https://pubmed.ncbi.nlm.nih.gov/8158612/*

Ricci, F. (2012), 'Social Implications of Malaria and Their Relationships with Poverty', *Mediterranean Journal of Hematology and Infectious Diseases* **4**(1).
URL: */pmc/articles/PMC3435125/*

Rosen, L. E., Morrison, H. A., Masri, S., Brown, M. J., Springstubb, B., Sussman, D., Stoddard, B. L. & Seligman, L. M. (2006), 'Homing endonuclease I-CreI derivatives with novel DNA target specificities', *Nucleic Acids Research* **34**(17), 4791.
URL: *https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1635285/*

Rouet, P., Smih, F. & Jasin, M. (1994), 'Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease.', *Molecular and Cellular Biology* **14**(12), 8096.
URL: */pmc/articles/PMC359348/?report=abstract https://www.ncbi.nlm.nih.gov/pmc/articles/PMC359348/*

RTS S Clinical Trials Partnership (2012), 'A Phase 3 Trial of RTS,S/AS01 Malaria Vaccine in African Infants', *New England Journal of Medicine* **367**(24), 2284–2295.
URL: *https://www.nejm.org/doi/full/10.1056/NEJMoa1208394*

RTS S Clinical Trials Partnership (2015), 'Efficacy and safety of RTS,S/AS01 malaria vaccine with or without a booster dose in infants and children in Africa: final results of a phase 3, individually randomised, controlled trial', *Lancet (London, England)* **386**(9988), 31–45.
URL: *https://pubmed.ncbi.nlm.nih.gov/25913272/ https://pubmed.ncbi.nlm.nih.gov/25913272/?dopt=Abstract*

Rudin, N. & Haber, J. E. (1988), 'Efficient repair of HO-induced chromosomal breaks in Saccharomyces cerevisiae by recombination between flanking homologous sequences', *Molecular and cellular biology* **8**(9), 3918–3928.
URL: *https://pubmed.ncbi.nlm.nih.gov/3065627/*

Sachs, J. & Malaney, P. (2002), 'The economic and social burden of malaria', *Nature 2002 415:6872* **415**(6872), 680–685.
URL: *https://www.nature.com/articles/415680a*

Saleh, A., Macia, A. & Muotri, A. R. (2019), 'Transposable elements, inflammation, and neurological disease', *Frontiers in Neurology* **10**(AUG), 894.
URL: *https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6710400/*

Samantsidis, G. R., Panteleri, R., Denecke, S., Kounadi, S., Christou, I., Nauen, R., Douris, V. & Vontas, J. (2020), "What I cannot create, I do not understand': functionally validated synergism of metabolic and target site insecticide resistance', *Proceedings of the Royal Society B: Biological Sciences* **287**(1927).
URL: */pmc/articles/PMC7287358/*

Sander, J. D. & Joung, J. K. (2014), 'CRISPR-Cas systems for genome editing, regulation and targeting', *Nature biotechnology* **32**(4), 347.
URL: */pmc/articles/PMC4022601/*

SanMiguel, P. (1996), 'Nested retrotransposons in the intergenic regions of the maize genome', *Science* **274**, 765–768.
URL: *https://doi.org/10.1126/science.274.5288.765*

Sapranauskas, R., Gasiunas, G., Fremaux, C., Barrangou, R., Horvath, P. & Siksnys, V. (2011), 'The Streptococcus thermophilus CRISPR/Cas system provides immunity in Escherichia coli', *Nucleic acids research* **39**(21), 9275–9282.
URL: *https://pubmed.ncbi.nlm.nih.gov/21813460/*

Scali, C., Catteruccia, F., Li, Q. & Crisanti, A. (2005), 'Identification of sex-specific transcripts of the Anopheles gambiae doublesex gene', *Journal of Experimental Biology* **208**(19), 3701–3709.
URL: *https://doi.org/10.1242/jeb.01819*

Shelly, T. & McInnis, D. (2016), 'Sterile Insect Technique and Control of Tephritid Fruit Flies: Do Species With Complex Courtship Require Higher Overflooding Ratios?', *Annals of the Entomological Society of America* **109**(1), 1–11.
URL: *https://academic.oup.com/aesa/article/109/1/1/2195075*

Sherrard-Smith, E., Hogan, A. B., Hamlet, A., Watson, O. J., Whittaker, C., Winskill, P., Ali, F., Mohammad, A. B., Uhomoibhi, P., Maikore, I., Ogbulafor, N., Nikau, J., Kont, M. D., Challenger, J. D., Verity, R., Lambert, B., Cairns, M., Rao, B., Baguelin, M., Whittles, L. K., Lees, J. A., Bhatia, S., Knock, E. S., Okell, L., Slater, H. C., Ghani, A. C., Walker, P. G., Okoko, O. O. & Churcher, T. S. (2020), 'The potential public health consequences of COVID-19 on malaria in Africa', *Nature Medicine 2020 26:9* **26**(9), 1411–1416.
URL: *https://www.nature.com/articles/s41591-020-1025-y*

Simoni, A., Hammond, A. M., Beaghton, A. K., Galizi, R., Taxiarchi, C., Kyrou, K., Meacci, D., Gribble, M., Morselli, G., Burt, A., Nolan, T. & Crisanti, A. (2020), 'A male-biased sex-distorter gene drive for the human malaria vector Anopheles gambiae', *Nature Biotechnology 2020 38:9* **38**(9), 1054–1060.
URL: *https://www.nature.com/articles/s41587-020-0508-1*

Simoni, A., Siniscalchi, C., Chan, Y. S., Huen, D. S., Russell, S., Windbichler, N. & Crisanti, A. (2014), 'Development of synthetic selfish elements based on modular nucleases in Drosophila melanogaster', *Nucleic Acids Research* **42**(11), 7461–7472.
URL: *https://academic.oup.com/nar/article/42/11/7461/1452254*

Sinka, M. E., Bangs, M. J., Manguin, S., Rubio-Palis, Y., Chareonviriyaphap, T., Coetzee, M., Mbogo, C. M., Hemingway, J., Patil, A. P., Temperley, W. H., Gething, P. W., Kabaria, C. W., Burkot, T. R., Harbach, R. E. & Hay, S. I. (2012), 'A global map of dominant malaria vectors', *Parasites and Vectors* **5**(1), 1–11.
URL: *https://parasitesandvectors.biomedcentral.com/articles/10.1186/1756-3305-5-69*

Sinkins, S. P. & Gould, F. (2006), 'Gene drive systems for insect disease vectors', *Nature Reviews Genetics 2006 7:6* **7**(6), 427–435.
URL: *https://www.nature.com/articles/nrg1870*

Snow, R. W., Craig, M. H., Newton, C. R. J. C., Steketee, R. W. & Snow, B. (2003), 'The Public Health Burden of Plasmodium falciparum Malaria in Africa: Deriving the Numbers', *Disease Control Priorities Project* **11**.
URL: *www.fic.nih.gov/dcpp*

Sorek, R. & Ast, G. (2003), 'Intronic sequences flanking alternatively spliced exons are conserved between human and mouse', *Genome Research* **13**(7), 1631–1637.
URL: *https://doi.org/10.1101/gr.1208803*

Stoddard, B. L. (2005), 'Homing endonuclease structure and function', *Quarterly reviews of biophysics* **38**(1), 49–95.
URL: *https://pubmed.ncbi.nlm.nih.gov/16336743/*

Stouthamer, R., Breeuwer, J. A. & Hurst, G. D. (1999), 'Wolbachia Pipientis: Microbial Manipulator of Arthropod Reproduction', *http://dx.doi.org/10.1146/annurev.micro.53.1.71* **53**, 71–102.
URL: *https://www.annualreviews.org/doi/abs/10.1146/annurev.micro.53.1.71*

Suzuki, M. G., Funaguma, S., Kanda, T., Tamura, T. & Shimada, T. (2005), 'Role of the male BmDSX protein in the sexual differentiation of Bombyx mori', *Evolution & development* **7**(1), 58–68.
URL: *https://pubmed.ncbi.nlm.nih.gov/15642090/*

Tan, A., Fu, G., Jin, L., Guo, Q., Li, Z., Niu, B., Meng, Z., Morrison, N. I., Alphey, L. & Huang, Y. (2013), 'Transgene-based, female-specific lethality system for genetic sexing of the silkworm, Bombyx mori', *Proceedings of the National Academy of Sciences of the United States of America* **110**(17), 6766–6770.
URL: *https://www.pnas.org/content/110/17/6766 https://www.pnas.org/content/110/17/6766.abstract*

Taxiarchi, C., Kranjc, N., Kriezis, A., Kyrou, K., Bernardini, F., Russell, S., Nolan, T., Crisanti, A. & Galizi, R. (2019), 'High-resolution transcriptional profiling of Anopheles gambiae spermatogenesis reveals mechanisms of sex chromosome regulation', *Scientific Reports 2019 9:1* **9**(1), 1–12.
URL: *https://www.nature.com/articles/s41598-019-51181-1*

Tennessen, J. A., Ingham, V. A., Toé, K. H., Guelbéogo, W. M., Sagnon, N., Kuzma, R., Ranson, H. & Neafsey, D. E. (2021), 'A population genomic unveiling of a new cryptic mosquito taxon within the malaria-transmitting Anopheles gambiae complex', *Molecular ecology* **30**(3), 775.
URL: */pmc/articles/PMC7858241/*

Terradas, G., Buchman, A. B., Bennett, J. B., Shriner, I., Marshall, J. M., Akbari, O. S. & Bier, E. (2021), 'Inherently confinable split-drive systems in Drosophila', *Nature Communications 2021 12:1* **12**(1), 1–12.
URL: *https://www.nature.com/articles/s41467-021-21771-7*

The Anopheles gambiae 1000 Genomes Consortium (2021), 'Ag1000G phase 3 SNP data release'.
URL: *https://www.malariagen.net/data/ag1000g-phase3-snp*

The Global Health Observatory (2019), 'Malaria Deaths'.
URL: *https://www.who.int/data/gho/data/themes/topics/topic-details/GHO/deaths*

The Project Wolbachia – Singapore Consortium & Ching, N. L. (2021), 'Wolbachia-mediated sterility suppresses Aedes aegypti populations in the urban tropics', *medRxiv* p. 2021.06.16.21257922.
URL: *https://www.medrxiv.org/content/10.1101/2021.06.16.21257922v1*

Thomas, D. D., Donnelly, C. A., Wood, R. J. & Alphey, L. S. (2000), 'Insect population control using a dominant, repressible, lethal genetic system', *Science (New York, N.Y.)* **287**(5462), 2474–2476.
URL: *https://pubmed.ncbi.nlm.nih.gov/10741964/*

Uguru, N. P., Onwujekwe, O. E., Uzochukwu, B. S., Igiliegbe, G. C. & Eze, S. B. (2009), 'Inequities in incidence, morbidity and expenditures on prevention and treatment of malaria in southeast Nigeria', *BMC International Health and Human Rights* **9**(1), 1–8.
URL: *https://bmcinthealthhumrights.biomedcentral.com/articles/10.1186/1472-698X-9-21*

Urnov, F. D., Rebar, E. J., Holmes, M. C., Zhang, H. S. & Gregory, P. D. (2010), 'Genome editing with engineered zinc finger nucleases', *Nature Reviews Genetics 2010 11:9* **11**(9), 636–646.
URL: *https://www.nature.com/articles/nrg2842*

Uwimana, A., Legrand, E., Stokes, B. H., Ndikumana, J. L. M., Warsame, M., Umulisa, N., Ngamije, D., Munyaneza, T., Mazarati, J. B., Munguti, K., Campagne, P., Criscuolo, A., Ariey, F., Murindahabi, M., Ringwald, P., Fidock, D. A., Mbituyumuremyi, A. & Menard, D. (2020), 'Emergence and clonal expansion of in vitro artemisinin-resistant Plasmodium falciparum kelch13 R561H mutant parasites in Rwanda', *Nature Medicine 2020 26:10* **26**(10), 1602–1608.
URL: *https://www.nature.com/articles/s41591-020-1005-2*

Uwimana, A., Umulisa, N., Venkatesan, M., Svigel, S. S., Zhou, Z., Munyaneza, T., Habimana, R. M., Rucogoza, A., Moriarty, L. F., Sandford, R., Piercefield, E., Goldman, I., Ezema, B., Talundzic, E., Pacheco, M. A., Escalante, A. A., Ngamije, D., Mangala, J. L. N., Kabera, M., Munguti, K., Murindahabi, M., Brieger, W., Musanabaganwa, C., Mutesa, L., Udhayakumar, V., Mbituyumuremyi, A., Halsey, E. S. & Lucchi, N. W. (2021), 'Association of Plasmodium falciparum kelch13 R561H genotypes with delayed parasite clearance in Rwanda: an open-label, single-arm, multicentre, therapeutic efficacy study', *The Lancet. Infectious diseases* **21**(8), 1120–1128.
URL: *https://pubmed.ncbi.nlm.nih.gov/33864801/ https://pubmed.ncbi.nlm.nih.gov/33864801/?dopt=Abstract*

Valid Species — Mosquito Taxonomic Inventory (2021).
URL: *https://mosquito-taxonomic-inventory.myspecies.info/valid-species-list*

Vella, M. R., Gunning, C. E., Lloyd, A. L. & Gould, F. (2017), 'Evaluating strategies for reversing CRISPR-Cas9 gene drives', *Scientific Reports 2017 7:1* **7**(1), 1–8.
URL: *https://www.nature.com/articles/s41598-017-10633-2*

Verhulst, E. C. & Van de zande, L. (2015), 'Double nexus—Doublesex is the connecting element in sex determination', *Briefings in Functional Genomics* **14**(6), 396.
URL: */pmc/articles/PMC4652034/*

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. & Marois, E. (2015), 'Tools for anopheles gambiae transgenesis', *G3: Genes, Genomes, Genetics* **5**(6), 1151–1163.
URL: */pmc/articles/PMC4478545/*

Walton, R. T., Christie, K. A., Whittaker, M. N. & Kleinstiver, B. P. (2020), 'Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants', *Science* **368**(6488), 290–296.
URL: *https://www.science.org/doi/abs/10.1126/science.aba8853*

Wang, X., Dong, Q., Chen, G., Zhang, J., Liu, Y. & Cai, Y. (2022), 'Frameshift and wild-type proteins are often highly similar because the genetic code and genomes were optimized for frameshift tolerance', *BMC Genomics* **23**(1), 1–15.
URL: *https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-022-08435-6*

Werther, R., Hallinan, J. P., Lambert, A. R., Havens, K., Pogson, M., Jarjour, J., Galizi, R., Windbichler, N., Crisanti, A., Nolan, T. & Stoddard, B. L. (2017), 'Crystallographic analyses illustrate significant plasticity and efficient recoding of meganuclease target specificity', *Nucleic Acids Research* **45**(14), 8621–8634.
URL: *https://academic.oup.com/nar/article/45/14/8621/3871306*

WHO (2021*a*), 'Malaria FactSheet'.
URL: *https://www.who.int/news-room/fact-sheets/detail/malaria*

WHO (2021*b*), 'World Malaria Report 2021'.
URL: *https://cdn.who.int/media/docs/default-source/malaria/world-malaria-reports/*

Willis, K. & Burt, A. (2021), 'Double drives and private alleles for localised population genetic control', *PLOS Genetics* **17**(3), e1009333.
URL: *https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1009333*

Wiltshire, R. M., Bergey, C. M., Kayondo, J. K., Birungi, J., Mukwaya, L. G., Emrich, S. J., Besansky, N. J. & Collins, F. H. (2018), 'Reduced-representation sequencing identifies small effective population sizes of Anopheles gambiae in the north-western Lake Victoria basin, Uganda', *Malaria Journal* **17**(1), 1–12.
URL: *https://malariajournal.biomedcentral.com/articles/10.1186/s12936-018-2432-0*

Windbichler, N., Menichelli, M., Papathanos, P. A., Thyme, S. B., Li, H., Ulge, U. Y., Hovde, B. T., Baker, D., Monnat, R. J., Burt, A. & Crisanti, A. (2011), 'A synthetic homing endonuclease-based gene drive system in the human malaria mosquito', *Nature 2011 473:7346* **473**(7346), 212–215.
URL: *https://www.nature.com/articles/nature09937*

Windbichler, N., Papathanos, P. A., Catteruccia, F., Ranson, H., Burt, A. & Crisanti, A. (2007), 'Homing endonuclease mediated gene targeting in Anopheles gambiae cells and embryos', *Nucleic Acids Research* **35**(17), 5922.
URL: */pmc/articles/PMC2034484/*

Windbichler, N., Papathanos, P. A. & 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), e1000291.
URL: *https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1000291*

Wolabu, T. W., Cong, L., Park, J. J., Bao, Q., Chen, M., Sun, J., Xu, B., Ge, Y., Chai, M., Liu, Z. & Wang, Z. Y. (2020), 'Development of a Highly Efficient Multiplex Genome Editing System in Outcrossing Tetraploid Alfalfa (Medicago sativa)', *Frontiers in Plant Science* **11**, 1063.
URL: *https://doi.org/10.3389/fpls.2020.01063*

Wood, R. & Newton, M. (1991), 'Sex-ratio distortion caused by meiotic drive in mosquitoes.', *The American Naturalist* **137**(3), 379–391.

Xu, L., Zhao, L., Gao, Y., Xu, J. & Han, R. (2017), 'Empower multiplex cell and tissue-specific CRISPR-mediated gene manipulation with self-cleaving ribozymes and tRNA', *Nucleic Acids Research* **45**(5), e28.
URL: *https://doi.org/10.1093/nar/gkw1048*

Yang, E., Metzloff, M., Langmü Ller, A. M., Xu, X., Clark, A. G., Messer, P. W. & Champer, J. (2022), 'A homing suppression gene drive with multiplexed gRNAs maintains high drive conversion efficiency and avoids functional resistance alleles', *G3 Genes—Genomes—Genetics* **12**(6).
URL: *https://academic.oup.com/g3journal/article/12/6/jkac081/6565321*

Yao, F. A., Millogo, A. A., Epopa, P. S., North, A., Noulin, F., Dao, K., Drabo, M., Guissou, C., Kekele, S., Namountougou, M., Ouedraogo, R. K., Pare, L., Barry, N., Sanou, R., Wandaogo, H., Dabire, R. K.,

McKemey, A., Tripet, F. & Diabaté, A. (2022), 'Mark-release-recapture experiment in Burkina Faso demonstrates reduced fitness and dispersal of genetically-modified sterile malaria mosquitoes', *Nature Communications* **13**(1).
URL: *https://www.nature.com/articles/s41467-022-28419-0*

Zaidi, S. S. e., Mahfouz, M. M. & Mansoor, S. (2017), 'CRISPR-Cpf1: A New Tool for Plant Genome Editing', *Trends in Plant Science* **22**(7), 550–553.
URL: *http://www.cell.com/article/S1360138517300900/fulltext http://www.cell.com/article/S1360138517300900/abstract https://www.cell.com/trends/plant-science/abstract/S1360-1385(17)30090-0*

Zhang, W., Yin, J., Zhang-Ding, Z., Xin, C., Liu, M., Wang, Y., Ai, C. & Hu, J. (2021), 'In-depth assessment of the PAM compatibility and editing activities of Cas9 variants', *Nucleic Acids Research* **49**(15), 8785–8795.
URL: *https://academic.oup.com/nar/article/49/15/8785/6300621*

Zhang, X. H., Tee, L. Y., Wang, X. G., Huang, Q. S. & Yang, S. H. (2015), 'Off-target Effects in CRISPR/Cas9-mediated Genome Engineering', *Molecular Therapy - Nucleic Acids* **4**(11), e264.
URL: *https://doi.org/10.1038/mtna.2015.37*