Imperial College London

Targeting sex determination for genetic control of the malaria mosquito

---

Kyros Kyrou

Department of Life Sciences

Imperial College London

A thesis submitted for the degree of

Doctor of Philosophy of Imperial College London

and the Diploma of Imperial College

February 2021

Acknowledgements

This thesis is the collective result of four years of my life and during that time many people have helped me in achieving this accomplishment.

I am very grateful to my supervisors, Professor Andrea Crisanti and Dr. Tony Nolan who were always there for me and have always pushed me into becoming a better scientist.

I would like to thank my progress review panel, Professor Austin Burt and Dr. Giorgio Gilestro for their guidance during the assessment process of my degree.

To all my friends, even though I cannot name you one by one, I would like to thank you for putting up with me during this time of my life...
...and to the one person who without his guidance, advice, but most importantly friendship, I would not be where I am today...
Drew, thank you for all that you have done for me during the last six years.

Finally, to my family, who were always there for me no matter what, I would like to say that I love you and I miss you greatly.

Contents

List of Figures 9

List of Tables 13

List of Acronyms 15

Declaration of Originality 17

Copyright Declaration 19

Publications 21

Abstract 23

Glossary of Terms 25

1 Introduction 29

1.1 Mosquitoes and malaria . . . 29

1.2 The malaria burden . . . 29

1.3 Malaria control . . . 31

1.3.1 Anti-malarial drugs . . . 31

1.3.2 Vaccination . . . 33

1.4 Vector control . . . 34

1.4.1 Mosquito life cycle and anatomy . . . 34

1.4.2 Insecticide-treated Nets (ITNs) and Indoor Residual Spraying (IRS) . . . 37

1.4.3 Chemical treatment and habitat removal . . . 38

1.5 Genetic engineering for vector control . . . 40

1.5.1 Sterile insect technique - SIT . . . 41

1.5.2 Release of insects carrying a dominant lethal gene - RIDL . . . 42

1.5.3 Maternal-effect dominant embryonic arrest - Medea . . . 43

1.5.4 Genetic underdominance - GU 44
1.5.5 Heritable microorganisms - *Wolbachia* 45
1.5.6 Heritable microorganisms - *Microsporidia* 47
1.5.7 Sex distorters 48
1.5.8 Homing-based gene drives 49
1.5.8.1 The CRISPR/Cas9 system 52
1.5.8.2 Current status and challenges 54

2 Sex Determination in Insects **57**
2.1 Sex determination in the animal kingdom 57
2.2 Sex determination in insects 64
2.2.1 The hourglass model of evolution 66
2.2.2 Genotypic sex determination 68
2.2.3 Sex determination in *Drosophila melanogaster* 69
2.2.4 Haplo-diploidy and Paternal genome elimination 71
2.2.5 Environmental sex determination, monogeny and parthenogenesis 72
2.3 Conserved genes in insect sex determination 74
2.3.1 The *transformer* gene 74
2.3.2 The *fruitless* gene 75
2.3.3 The *doublesex* gene 76
2.3.3.1 Discovery of the *Dmrt* family 76
2.3.3.2 Gene structure 77
2.3.3.3 *doublesex* conservation 77
2.3.3.4 Role in sex determination and sex differentiation 78
2.4 Sex determination in mosquitoes 80
2.5 Concluding remarks 82

3 Materials & Methods **85**
3.1 Plasmid preparation 85
3.1.1 Bacterial transformation – Miniprep and Maxiprep isolation of plasmid DNA 85
3.1.2 Colony-PCR 85
3.2 Genomic DNA extraction 86
3.3 PCR amplification and gel analysis 86
3.4 Genomic databases and graph representation 87

3.5 Generation of CRISPR helper and donor plasmids 88
3.5.1 p174 gene drive vector 88
3.5.2 p16510 and p17410 CRISPR helper plasmids 89
3.5.3 K101 donor plasmid 89
3.6 Microinjection of mosquito embryos and selection of transformants 90
3.7 Molecular confirmation of gene targeting 90
3.7.1 Confirmation of gene targeting and recombinase-mediated cassette exchange events 90
3.7.2 Containment of mosquitoes 92
3.8 Phenotypic characterisation and microdissections 92
3.9 Phenotype assays 92
3.9.1 Fertility assays 92
3.9.2 Longevity assays 93
3.10 Measuring spread of gene drive in a caged population 94
3.11 Analysis of resistance at the gene drive target site 94
3.11.1 Bioinformatic analysis of the alleles generated during the population experiments . . 95
3.12 *In vitro* cleavage assay of SNP found in wild populations in Africa 95
3.12.1 Plasmid preparation 95
3.12.2 RNP digestion and gel analysis 96

4 Understanding the role of *doublesex* in mosquitoes **97**
4.1 The *doublesex* gene in *Anopheles gambiae* 97
4.2 Using the CRISPR/Cas9 system to disrupt the *doublesex* gene 99
4.2.1 Selection of suitable gRNA 103
4.2.2 Creation of vectors for targeted mutagenesis on exon 5 105
4.3 Disruption of the female-specific isoform of *doublesex* 107
4.3.1 The *doublesex* exon 5 disruption is not lethal 108
4.3.2 Females *dsxF* $^{-/-}$ present sexual developmental abnormalities 108
4.3.3 Female *dsxF* $^{-/-}$ are sterile and have short longevity 110
4.4 Discussion 112

5 A gene drive targeting *doublesex* for genetic control **115**
5.1 Development of a gene drive system 115
5.2 Design and creation of a CRISPR-based gene drive vector targeting exon 5 of the *doublesex* gene 116

5.3 Generation of a gene drive line on exon 5 of the *doublesex* gene 119
5.3.1 The gene drive poses a fertility cost on the female mosquitoes - the parental effect . . 120
5.3.2 Heterozygous mosquitoes transmit the gene drive to more than half of the progeny . 124
5.4 Discussion 126

6 Testing the *doublesex* gene drive system in caged mosquitoes 129
6.1 Gene drive systems for population control 129
6.2 Assessment of the $dsxF^{CRISPRh}$ gene drive system in caged *An. gambiae* 130
6.2.1 The gene drive imposes a reproductive load on the populations until they collapse . . 130
6.2.2 Target site resistance on the *doublesex* locus was not detected within the populations 132
6.2.3 SNP found in wild-type populations can be cleaved by the Cas9 *in vitro* 137
6.3 Controlled releases in large scale set ups successfully suppress mosquito populations 141
6.4 Discussion 144

Discussion and Future Directions 147

Bibliography 153

A Population Experiment Data for *dsx* Gene Drive 181

B Licences and Permissions 185

C Publications 213

List of Figures

1.1 Map of malaria case incidence rate per country in 2018 30
1.2 Malaria global cases and deaths in relation to global funding spent from 2010 to 2018 31
1.3 Malaria projected distribution between the years 2015 and 2030 32
1.4 The life cycle of the mosquito 36
1.5 The male and female pupae of *Anopheles gambiae* 37
1.6 The adult male and female *Anopheles gambiae* mosquitoes 38
1.7 The effect of interventions like ITN, ACT and IRS on infection prevalence and cases across endemic Africa 39
1.8 The Medea and GU systems 45
1.9 The *Wolbachia* system and cytoplasmic incompatibility 46
1.10 Homing endonuclease genes for malaria control and the homing reaction 50
1.11 A gene drive targeting female fertility or survival 51
1.12 The CRISPR/Cas9 system 53

2.1 The diversity of sex determination systems in the animal and plant kingdoms 58
2.2 The Y chromosome degeneration hypothesis 59
2.3 Tree-based representation of sex determining systems found in vertebrates 61
2.4 Dated phylogenetic tree of the Hexapoda subphylum 65
2.5 Tree-based representation of sex determining systems found in invertebrates 67
2.6 The hourglass model of evolution outlining the sex development in insects 68
2.7 Common sex determination systems in insects 73
2.8 The DSX protein isoforms 78

3.1 Expression pattern of the *3xP3* promoter 91
3.2 The abdomen and spermatheca of the female mosquito 93

4.1 Transcript comparison of the *doublesex* gene in *Anopheles gambiae* 98

4.2 Transcript comparison between the *doublesex* gene in *An. gambiae* and *D. melanogaster* . . 99
4.3 Representation of the DSX^{M} and DSX^{F} protein domains and their similarity with the *Dmdsx* 100
4.4 Comparison of genetic sequences of OD1 and OD2 domains within the Anophelinae subfamily101
4.5 Comparison of genetic sequences of OD^{M} and OD^{F} domains within the Anophelinae subfamily 102
4.6 Putative gRNA target sites in *An. gambiae* and their alignment in relation to the CDS of exon 5 within the Anophelinae subfamily 104
4.7 The target site of gRNA-T1 in relation to the two sex-specific isoforms of *Agdsx* 106
4.8 The knock-in strategy using K101 donor-plasmid and verification of the integration 107
4.9 Phenotypic comparison of the male and female mosquitoes of the *dsxF*$^{-}$ line 109
4.10 Closer inspection of the antenna system and genitalia of *dsxF*$^{-}$ mosquitoes 110
4.11 Comparison of the internal reproductive tissues of *dsxF*$^{-}$ mosquitoes 111
4.12 Clutch size and larval output of the *dsxF*$^{-}$ line 112
4.13 Assay comparing the longevity of homozygous *dsxF*$^{-}$ and wild-type mosquitoes 113

5.1 The CRISPR^{h} cassette used to build the gene drive against exon 5 of the *doublesex* gene . . . 116
5.2 Map of the p174 CRISPR gene drive vector 118
5.3 Creation of the *dsxF*^{CRISPRh} line using RMCE and verification of the integration 120
5.4 The homing process of the CRISPR^{h} cassette 121
5.5 Clutch size and fecundity of the *dsxF*^{CRISPRh} gene drive line 122
5.6 Comparison of fecundity of the *AGAP007280*^{CRISPRh} and *dsxF*$_{CRISPRh}$ gene drive lines 123
5.7 Partial conversion of heterozygous gene drive females to the intersex phenotype 124
5.8 Transgenic rates of heterozygous *dsxF*^{CRISPRh} individuals 126
5.9 Microhomology-mediated end-joining and partial homing at the target site 127

6.1 Probability of the stochastic loss of the gene drive from the caged populations in a male-only release scenario 131
6.2 Dynamics of the *dsxF*^{CRISPRh} gene drive in five caged mosquito populations 132
6.3 Dynamics of the most prevalent indels detected in the cage trial populations 133
6.4 In-frame deletions and substitutions detected within the caged populations and their impact on the coding sequence of exon 5 135
6.5 Conservation comparison between the *dsx* and *nudel* gene drive target sites 136
6.6 Single nucleotide variance analysis across the cut site of the *dsxF*^{CRISPRh} gene drive 138

6.7 Distribution of the G>A SNP found in the gRNA-T1 sequence in mosquitoes from sub-Saharan Africa . . . 139
6.8 *In vitro* cleavage assay testing the efficiency of the $dsxF^{CRISPRh}$ gene drive against the SNP found in mosquito populations from Africa . . . 141
6.9 Gene drive dynamics and egg output in large cage experiments . . . 143
6.10 In-frame indel detected in the large cages and its impact on the coding sequence of exon 5 . 144

List of Tables

3.1 Mosquito genome databases and assemblies used in this study 87

3.2 Primers used to amplify the fragments necessary to build p174 88

3.3 The gRNA-T1 sequence and target site 89

3.4 Primers used to amplify the fragments necessary to build K101 90

3.5 Primers used for verification of gene targeting and RMCE events 91

3.6 Primers used for deep amplicon sequencing of the gene drive target site in *dsx* 95

3.7 Oligonucleotides used for spacer assembly for *in vitro* cleavage assay 96

4.1 List of target sites and the gRNAs with their properties found on exon 5 of *dsx* 105

5.1 Transient and transformant recovery from embryo microinjected to promote RMCE and create the gene drive line 119

5.2 Screening results from the gene drive line targeting the *doublesex* gene 125

6.1 Distribution of the G>A SNP found in the gRNA-T1 sequence in mosquitoes from sub-Saharan Africa 140

A.1 Population experiment - Cage trial 1 181

A.2 Population experiment - Cage trial 2 182

A.3 Population experiment - Cage trial 3 182

A.4 Population experiment - Cage trial 4m 183

A.5 Population experiment - Cage trial 4r 183

B.1 Licences and permissions for materials and figures used in this report 185

B.2 Licences and permissions for materials and figures from authored publications 186

B.3 Written permissions for materials and figures used in this report 186

List of Acronyms

aa amino acid

ACTs Artemisinin-based Combination Therapies

Agam *Anopheles gambiae*

Agdsx *Anopheles gambiae doublesex*

bp base pairs

Cas CRISPR-associated protein

CDS Coding Sequence

CI Cytoplasmic Incompatibility

CRISPR Clustered Regularly Interspaced Short Palindromic Repeats

Cys Cysteine

Dmel *Drosophila melanogaster*

Dmrt *dsx/mab-3* related family of trascription factors

DSB Double-Strand Break

dsx *doublesex*

eGFP enhanced Green Fluorescent Protein

ESD Environmental sex determination

EST Expressed Sequence Tag

gRNA guide RNA

GSD Genotypic sex determination

GU Genetic Underdominance

HDR Homology-Directed Repair

HEG Homing Endonuclease Gene

His Histidine

IRS Indoor Residual Spraying

ITNs Insecticide-Treated Nets

LH Left Homology

MAGs Male Accessory Glands

MMEJ Microhomology-Mediated End-Joining

NHEJ Non-Homologous End-Joining

OD Oligomerisation Domain

pHDR Partially Homology-Directed Repair

RDT Rapid Diagnostic Tests

RH Right Homology

Statement of Originality

I hereby declare that the project presented in this thesis is entirely my own work and it is the result of grouped discussions with my supervisors and my progress review panel. Additional sources of information that have been used are duly cited.

Copyright Statement

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

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

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.

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.

Publications

This thesis is following and building upon the results of Hammond et al. (2020a). The main results of this thesis were published in Kyrou et al. (2018) and additional experiments carried out alongside the work of this thesis have been published in Hammond et al. (2017). These two publications are attached at the end of the thesis. Additional work that is following upon the main results of this thesis is in preparation for publication (Hammond et al. 2020b).

Hammond, A.M., Karlsson, X., Morianou, I., Kyrou, K., Beaghton, A., Gribble, M., Kranjc, N., Roberto, G., Burt, Austin., Crisanti, A., and Nolan, T. Regulation of gene drive expression increases invasive potential and mitigates resistance. *bioaRxiv*, 2020a. doi:10.1101/360339

Kyrou, K., Hammond, A.M., Galizi, R., Kranjc, N., Burt, A., Beaghton, A.K., Nolan, T. and Crisanti, A. A CRISPR-Cas9 gene drive targeting *doublesex* causes complete population suppression in caged *Anopheles gambiae* mosquitoes. *Nat. Biotechnol.*, **36**(11):1062–1066, 2018. ISSN 15461696. doi: 10.1038/nbt.4245

Hammond, A.M., Kyrou, K., Bruttini, M., North, A., Galizi, R., Karlsson, X., Carpi, F.M., D'Aurizio, R. Crisanti, A. and Nolan, T. The creation and selection of mutations resistant to a gene drive over multiple generations in the malaria mosquito. *PLoS Genet.*, **13**(10):e1007039, 2017. ISSN 15537404. doi: 10.1371/journal.pgen.1007039

Hammond, A.M., Pollegioni, P., Persampieri, T., North, A., Minuz, R., Trusso, A., Bucci, A., Kyrou, K., Morianou, I., Simoni, A., Nolan, T., Müller, R., and Crisanti, A. *Successful suppression of age-structured large cage populations of the malaria vector using the Ag(QFS)1 suppression gene drive.* Manuscript in preparation. 2020.

Abstract

Malaria is a devastating disease that causes more than 400,000 deaths each year, primarily in underprivileged regions of sub-Saharan Africa. During the last two decades, mortality caused by the disease has been reduced by half, largely driven by coordinated vector control based on the use of insecticides and bed nets. Nevertheless, the declining trend in malaria cases appears to have stalled recently and there is a growing concern that new interventions will be needed to reach widespread malaria elimination. Gene drive systems are potentially transformative in this endeavour because they allow rapid, self-sustaining and species-specific control of the mosquito vector through the limited release of genetically modified mosquitoes. While proof-of-principle studies have demonstrated the feasibility of the approach, none of them have managed to fulfil the requirements needed to progress in field or semi-field testing, largely due to strong fitness costs and genetic resistance to gene drive. This thesis describes the first gene drive system demonstrated to spread in caged populations of *Anopheles gambiae* mosquitoes, unimpeded by resistance or fitness constraints. By targeting an ultra-conserved locus in the gene *doublesex*, this strategy is able to thwart target site resistance in caged experiments whilst driving complete population suppression through the conversion of genetic females to sterile intersex. In this thesis, I demonstrate complete population elimination of caged populations from single releases of gene drive mosquitoes using a 12.5% initial allele release frequency, and crucially, I demonstrate its effectiveness in large cage semi-field conditions designed to reveal complex behaviours otherwise absent in small scale testing. The complete suppression of vector populations using a gene system is a landmark achievement and brings the gene drive technology closer to be implemented in the wild to complement current interventions against malaria.

Glossary of Terms

Arrhenogeny A branch of monogeny in which females produce only male offspring

Androdioecy A sex reproducing system that includes a mixture of male and hermaphrodite individuals

Autosomes Chromosomes that do not determine the sex of an individual

Blastoderm A single layer of cells formed from the oocyte membrane after fertilisation that surrounds the blastocoel

Cellularisation The process in which the zygote stops being a multi-nucleate cell and is separated into distinct cells

Cytoplasmic sex determination The situation in which sex is regulated by cytoplasmic elements such as mitochondria or external parasites

Diaspidoid system A branch of paternal genome elimination in which the paternal genome is eliminated in embryos that are destined to be males resulting in haploid males and diploid females

Dichogamy A situation evolved to avoid self-fertilisation in hermaphrodites in which the male and female gonads develop in different periods during the organisms lifetime

Dioecy A situation found in plants in which the male and female reproductive organs are separated in different plants or trees resulting in male and female individuals

Environmental sex determination A sex determination mode in which sex development is controlled by external environmental factors

Female heterogamety A genotypic sex determination mode in which the females (termed ZW) are heterozygous for the sex-determining locus or chromosomes

Gametes The reproductive cells of an organisms that are derived through meiosis and are fused together in sexual reproduction to form a new organism

Genital imaginal disc Group of cells that originate from tail segments A8, A9 and A10, and produce the sexual dimorphic genitalia and the end of the abdomen

Genital ridge The precursor cells that will form the gonads in the developing embryo

Genotypic sex determination A sex determination mode in which the sex is determined solely by its genetic composition

Gonads The male (testicle) and female (ovaries) gamete-producing organs

Gynodioecy A sex reproducing system that includes a mixture of female and hermaphrodite individuals

Haplodiploidy A genotypic sex determination mode in which males are haploid and derive from unfertilised eggs whereas females are diploid and derive from fertilised eggs

Hermaphroditism The situation in which an individual carries both male and female reproductive organs and gametes

Heteromorphic sex chromosomes Chromosomes that determine the sex of an individual and have a different and distinct karyotypic morphology

Homomorphic sex chromosomes Chromosomes that determine the sex of an individual and have a very different and indistinguishable karyotypic morphology

Lecanoid system A branch of paternal genome elimination in which the paternal genome becomes heterochromatic in embryos that are destined to be males resulting in both diploid males and females

Male heterogamety A genotypic sex determination mode in which the males (termed XY) are heterozygous for the sex-determining locus or chromosomes

Meiosis The cell division process in which the daughter cells result in half the number of sets of chromosomes of the parental cell

Monoecy A branch of hermaphroditism found in plants in which the male and female reproductive organs are separated in different flowers on the same plant or tree

Monogeny The situation in which the offspring of a female are either all males or females

Neo-chromosomes Chromosomes that are formed by fusion of autosomes with the ancestral sex chromosomes

Parthenogenesis A mode of reproduction in which a female produces offspring through unfertilised eggs

Paternal genome elimination A genotypic sex determination mode in which the paternal chromosomes are inactivated or lost following fertilisation resulting in haploid males

Polygenic sex determination A genotypic sex determination mode in which the sex of an individual is determined by multiple loci or genes

Sequential hermaphrodite An individual that changes sex during their life

Sexual reproduction A mode of reproduction in which gametes that have derived through meiosis from two different organisms are fused to produced a new organism

Simultaneous hermaphrodite A branch of hermaphroditism in which an individual carries simultaneously both male and female mature reproductive organs and gametes

Thelygeny A branch of monogeny in which females produce only female offspring

Viviparous The situation in which the offspring are developed in inside the body of the parent

Wolffian duct The embryonic structures that form the internal reproductive organs in males

Zygote A cell that derives after the fusion of the male and female gametes

Zygotic transcription The time point in which the zygote starts producing its own mRNA and proteins through the transcription and translation of its DNA

Chapter 1

Introduction

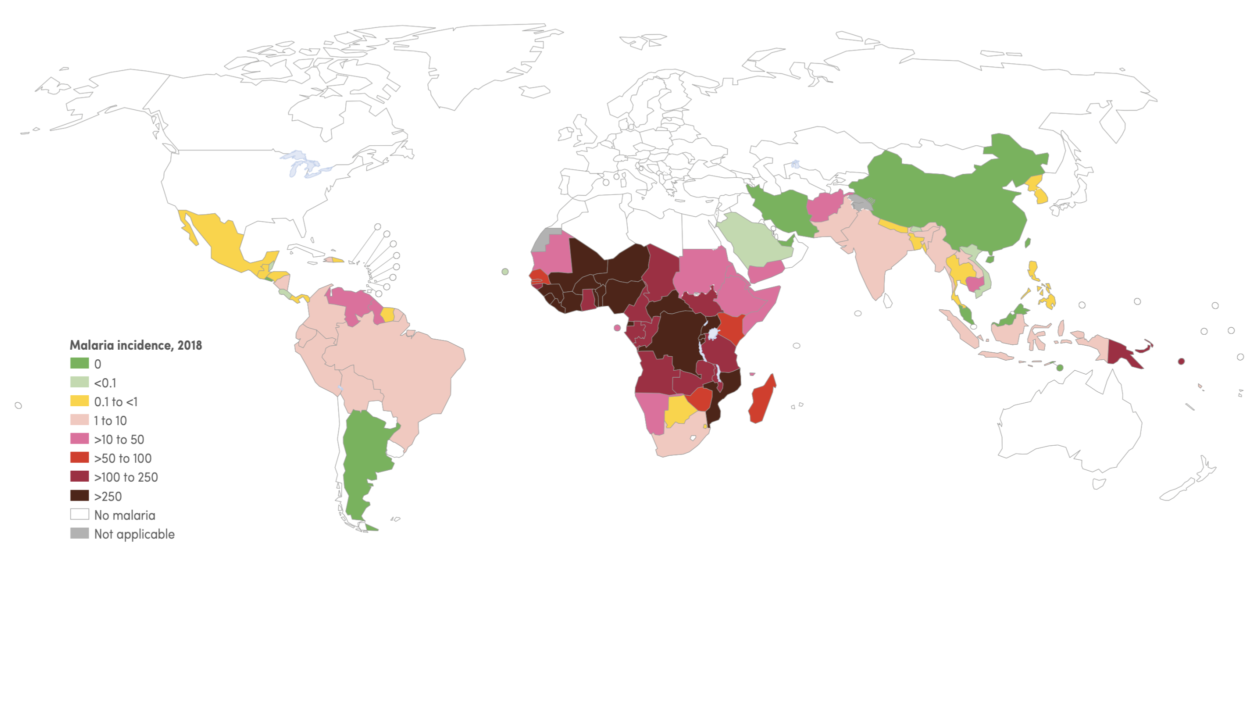
1.1 Mosquitoes and malaria

Mosquitoes are the deadliest animals on this planet. They carry and transmit a plethora of pathogens which cause more than 700,000 of human deaths every year (WHO 2020). They are responsible for the transmission of major vector-borne diseases like Zika, dengue, yellow fever, chikungunya and of course malaria, which alone claims the lives of more than 400,000 people annually. Vector-borne diseases are caused by parasites, bacteria or viruses, and they account for more than 17% of all infectious diseases (WHO 2020). Almost 80% of the world's population is at risk of contracting at least one vector-borne disease, whereas in 2018 half the population of the planet was at risk of contracting malaria (WHO 2019).

1.2 The malaria burden

Malaria is caused by parasites of the genus *Plasmodium*, which are transmitted to humans through the bite of infected female mosquitoes of the Anophelinae subfamily. The disease caused an estimated 228 million cases in 2018, and despite being preventable and curable, it claimed the lives of an estimated 405,000 people (WHO 2019). The African continent is at the centre of the epidemic, carrying 93% of all malaria cases and 94% of all malaria deaths globally (Figure 1.1). Moreover, within the continent the burden of malaria is not evenly distributed, with only six countries to account for more than half the malaria cases worldwide. Nigeria, the heaviest devastated country by the disease, carries a quarter of malaria cases worldwide, followed by the Democratic Republic of the Congo with 12%, Uganda with 5% and Côte d'Ivoire, Mozambique and Niger with 4% each. Infants and children under the age of 5 belong among the high-risk populations of contracting the disease and largely they accounted for 67% of all the deaths caused by malaria in 2018. Other high-risk groups are pregnant women, patients with HIV and

non-immune migrants and travellers (WHO 2019).



WHO: World Health Organization.

Figure 1.1 – Map of malaria case incidence rate per country in 2018. The cases are represented per 1000 population at risk. Estimates of the World Health Organization (WHO). World malaria report 2019 (WHO 2019). Geneva: World Health Organization; 2019. Licence: CC BY-NC-SA 3.0 IGO.

Malaria has been associated with poverty and a decreased quality of life due to the economic costs it brings upon the affected countries. An estimated 2.7 billion US dollars have been allocated in 2018 for the containment and elimination of the disease, a reduction from the 3.2 billion invested in 2017, with endemic countries contributing 30% of the total funding (WHO 2019). Allocation of the funds to research, drugs, diagnostic tests and vector control measures have managed to reduce the annual cases every year. Since 2010, malaria cases have been reduced from 71 cases per 1000 population at risk to 57 in 2014. From 2014 though, the decrease in cases flattened dramatically and even increased the upcoming years (Figure 1.2). This stall in case reduction can be attributed to many factors such as the reduction in funding and delivery of malaria commodities, the emergence of resistance against insecticides and anti-malarial drugs and the evolution of the parasite to avoid detection using rapid diagnostic tests (RDTs) (WHO 2019).

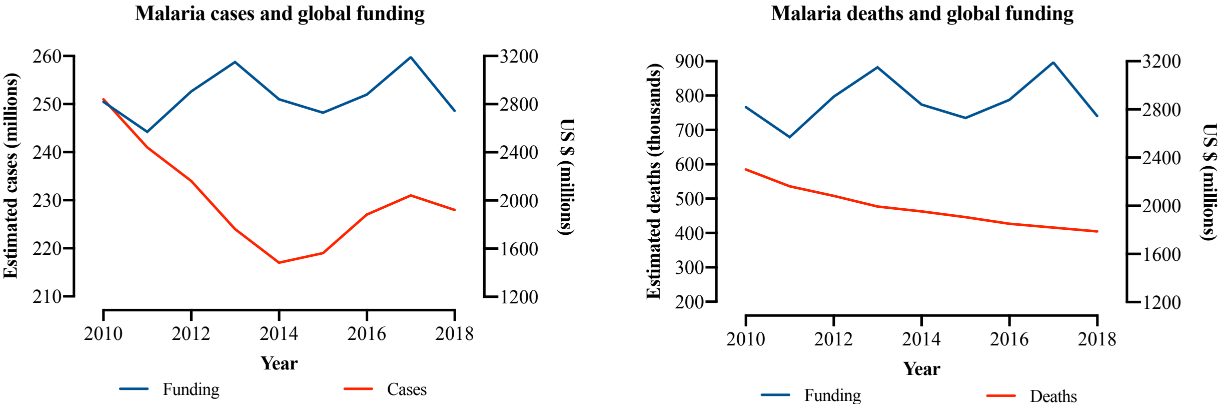


Figure 1.2 – Malaria global cases and deaths in relation to global funding spent from 2010 to 2018. Until 2014 the malaria cases (left panel) were in decline but factors such as reduction in funding and emergence of resistance to insecticides and diagnostic tests have stalled and even reverted the trend. The deaths caused by the disease (right panel) are still in decline but the reduction observed yearly is decreasing. Data from WHO (2019).

1.3 Malaria control

Since the set of the Millennium Development Goals, which aimed among others to reduce poverty and combat malaria worldwide, the World Health Organisation (WHO) has put a new plan in place to achieve a 90% reduction of the global malaria burden by the year 2030 (WHO 2015*a*). Use of anti-malarial drugs, such as artemisinin-based combination therapies (ACTs), insecticide-treated bed nets (ITNs) and indoor residual spraying (IRS) have been used since then as the primary methods to combat the transmission of the disease. In 2018 alone, 214 million ACT treatment courses and 412 million RDTs have been delivered in malaria endemic countries, of which 98% and 64% respectively were delivered in the African region. The use of ITNs has also been increased and currently half the population at risk of malaria in the African region are sleeping under ITNs, a 21% increase from 2010 (WHO 2019). Modelling predicts the elimination of the disease in large areas of South America and Southeast Asia, along with a reduction in transmission in sub-Saharan Africa by 2030, if the coverage of current interventions and access to first-line treatment is increased to 90% (Griffin et al. 2016) (Figure 1.3).

1.3.1 Anti-malarial drugs

Artemisinin and related compounds used in ACTs are the most widely used therapies for people infected with the malaria parasite and are currently used since 2005 as the first line of defence for susceptible groups such as infants and pregnant women. Heavy use of this particular group of drugs has led to the

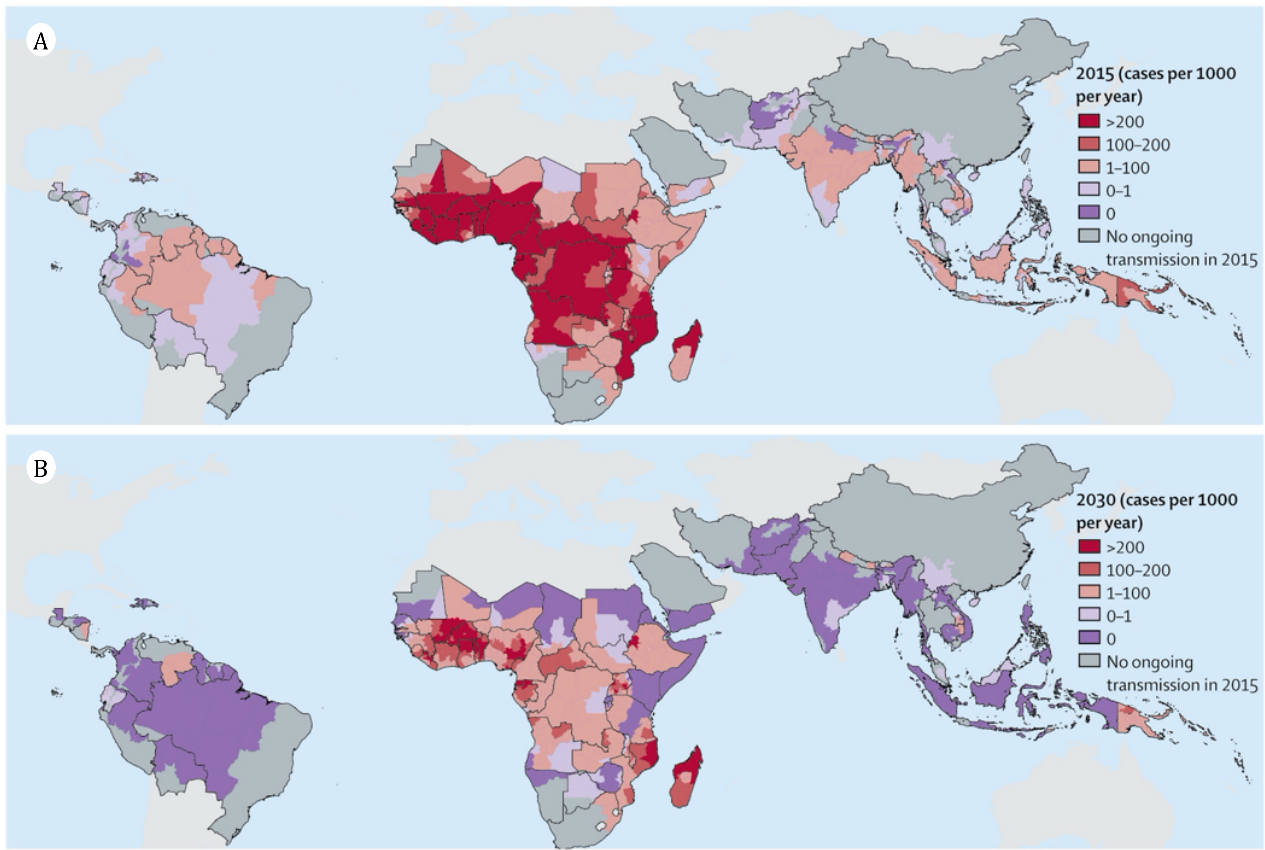


Figure 1.3 – Malaria projected distribution between the years 2015 (A) and 2030 (B). If the coverage of current interventions and access to first-line treatment is increased to 90%, large areas of South America and Southeast Asia will be malaria free by 2030 (purple). The transmission rate is predicted to substantially decrease in sub-Saharan Africa and global mortality rates are predicted to fall by 74%. Over the 15-year period an estimated 2.9 billion cases and 10.4 million deaths will be averted. Image reproduced with permission of the rights holder, Copyright © 2016, Elsevier. Griffin et al. (2016).

emergence of resistance, particularly in south east Asia. Since first detected in Cambodia in 2007 (Noedl et al. 2008, Dondorp et al. 2009), fears for the spread of resistance outside the region has led WHO to launch the Emergency Response to Artemisinin Resistance (ERAR), which aimed to contain and limit the spread of the drug-resistant parasite outside the region. In 2014, a major study under the Tracking Resistance Artemisinin Collaboration (TRAC) attempted to map 15 sites in both Asia and Africa, confirming establishment of the resistant parasite in south-east Asia (Ashley et al. 2014). Mutations and Single Nucleotide Polymorphisms (SNPs) on a single gene, *pfkelch13*, have been identified and associated with parasite resistance (Ariey et al. 2014, Ghorbal et al. 2014, Straimer et al. 2015) and since then they serve as markers for resistance detection. In Cambodia, a small number of resistant cases have been reported without any mutations or SNPs on that particular gene (Mukherjee et al. 2017), indicating that our knowledge and understanding of the molecular basis of parasite resistance is limited.

In Africa, mutations of the *pfkelch13* gene have been found in Rwanda, but up until recently they were not shown to be correlated with resistance against ACTs (Tacoli et al. 2016), suggesting the presence of other genetic factors in local parasite populations that could better be associated with the emergence of resistance. However, a very recent study has shown early warning signs of resistance against artemisinin that seems to be spreading between localities in the country (Uwimana et al. 2020). These de novo indigenous cases confirm that local emergence of resistance is possible in Africa and contradict previous long-standing theories suggesting that resistance in the continent is the result of westward spread of resistant parasites from Southeast Asia (Blasco et al. 2017). Even though assessment of the resistant parasite population was only validated *in vitro*, the appearance of artemisinin-resistant parasites could potentially compromise the continued success of ACTs interventions and pose a major public health threat to the continent. That is why understanding the nature of resistance against ACTs is urgent and of major importance since we currently do not possess alternative compounds with such high efficacy against the parasite. New attempts to replace artemisinin with new alternatives are already undergoing (Flannery et al. 2013) and some new chemicals appear promising in blocking the transmission of the disease (Baragaña et al. 2015).

1.3.2 Vaccination

A vaccine against the malaria parasite is a long-sought goal and efforts to create one were boosted with the publication of the first parasite genome in the late 1990s (Hoffman et al. 1998). An efficient vaccine against the parasite has the potential to eliminate the disease if combined with existing interventions (Blagborough et al. 2013). A number of vaccines have been developed against different stages of the parasite but only one has managed to undergo large-scale phase III trials (Conway 2015). The candidate vaccine, called RTS,S AS01, has been identified through an experimental challenge study (Stoute et al. 1997) and after 30 years in development it is the first to offer protection in children in Africa. The efficacy of the vaccine though has not met the expectations, since it showed to prevent only 30% of serious cases of malaria (Mahmoudi & Keshavarz 2017). Furthermore, the vaccine had been associated with an increase in meningitis cases (RTS 2015) and a sex-specific increase in mortality among young girls that received the vaccine (Klein et al. 2016). The low efficacy and safety issues accompanying the vaccine is making it difficult to obtain worldwide approvals and be established as a standard measure for combating the disease. A new alternative vaccine is slowly being brought forward and is less expensive and can be administered in one-fifth the dose than RTS,S (Venkatraman et al. 2019). The new vaccine, called R21, has already begun phase II trials in Burkina Faso and results about its efficiency to block the transmission of the disease are expected in 2020. If this promising vaccine proves efficient against the parasite and safe

to use, it will provide susceptible populations such as children with the immunity needed to withstand malaria infections. If this immunity can be maintained during the majority of the person's life, it will have a tremendous impact on the death toll of the disease and will eventually reduce the burden and severity of malaria.

1.4 Vector control

In 1897, Dr. Ronald Ross, a military doctor serving in the Indian Army, has demonstrated the model for the transmission of malaria for the first time. Dr. Ross had demonstrated that mosquitoes of the genus *Anopheles* were responsible for the transmission of the disease in humans. He theorised that mosquitoes act as an intermediate host for malaria and used a bird-malaria system to prove his theory. "For his work on malaria", Dr. Ross was awarded the Nobel Prize for Medicine in 1902 (CDC 2015). Ross's work has since put the mosquito vector at the centre of the malaria problem and sparked widespread efforts to disrupt transmission of the disease by removing vector populations from the vicinity of humans and by blocking the contact between mosquito and the human host. It is now well known that malaria is exclusively transmitted by female mosquitoes of the Anophelinae subfamily. The dual-host parasite uses both the mosquito and humans to complete its cycle and it enters the human host during a blood meal, which the female mosquitoes perform in order to gather nutrients and produce their eggs. Approximately forty species are considered to be significant transmitters of the disease (Hay et al. 2010) and four of them, *Anopheles gambiae, Anopheles coluzzii, Anopheles arabiensis* and *Anopheles funestus*, transmit the majority of the disease in Africa (Coetzee et al. 2013, Hay et al. 2010, Sinka et al. 2010). Understanding the life cycle of the mosquito vector plays an important role in developing intervention methods that aim to interrupt the development of mosquitoes and help us control their numbers. Predicting the behaviour of mosquitoes such as host preferences, egg sites and even mating rituals, are important in order to propose simple measures that ultimately have a big impact in reducing the transmission of the disease (Bhatt et al. 2015).

1.4.1 Mosquito life cycle and anatomy

Anopheles mosquitoes belong to the Diptera order, which translates to 'two wings' (di-pteron) in Greek. The Diptera order of insects belongs within a much larger superorder called Endopterygota. The main characteristic of this superorder is the distinct larval, pupal and adult stages outlining the development of the insect, but also the final 'complete metamorphosis' the pupa undergoes to create the adult insect. *Anopheles* mosquitoes are both aquatic and terrestrial insects; stages like the egg, the larva and pupa require water in order to develop (Figure 1.4). Upon oviposition on the surface of water, the egg requires

approximately 2 days to hatch into the first instar larval stage (L1), and approximately 6 to 7 days to go through 3 more instar larval stages (L2-L4) before pupating. The pupa, the "metamorphosis" stage, uses the nutrient reserves built up during the larval stages to break down the body and rearrange tissues to form a fully developed adult mosquito. The time-frames between each stage are by far stable and can vary significantly based on both the temperature and nutrient availability.

The adult *Anopheles* mosquitoes spend their first hour or so after emerging resting on the surface of the water for their outer cuticle to harden and their flight muscles to prepare for flight. The male mosquitoes usually emerge first and become sexually mature within two to three days. Mature males form mating swarms preferably at dusk, attracting females to enter the swarm to mate. Female *Anopheles* usually mate only once, storing enough sperm in their spermatheca to use for subsequent gonotrophic cycles (blood meal, egg maturation and oviposition). Sexual dimorphism in mosquitoes is first obvious from the pupal stage onwards. At the end of the abdomen, the male pupa has a larger genital lobe in relation to the size of its paddles compared to the female pupa (Becker et al. 2010) (Figure 1.5). This ability to sex pupae before they emerge into adults allows us to perform controlled mating experiments in the lab between different strains and transgenic lines.

Although both sexes require nectar in order to survive, female mosquitoes are also in search of blood meal to nourish the development and maturation of their eggs. Although some *Anopheles* species are almost exclusively zoophilic, like *An. quadriannulatus* (Pates et al. 2006), others are preferably anthropophilic and feed on human blood and also are sometimes endophagic, meaning they bite indoors, mainly during the night when the subject is asleep and restful (Ng'Habi et al. 2011, Coetzee et al. 2000). Upon blood meal, the female seeks a secluded spot to rest for the eggs to develop and mature. Between 48 and 72 hours post blood meal they become active and search for oviposition sites to lay their eggs. Small ponds, rainwater puddles and man-made enclosures like buckets, clay pots and abandoned tyres, provide excellent bodies of calm water for the larvae to develop. Unlike their mating preference, female *Anopheles* will blood feed and produce eggs more than once. Since the female mosquitoes typically have a longer lifespan than the male mosquitoes, they can feed for up to three or four times and produce up to 150 to 200 eggs in each gonotrophic cycle.

Adult male and female mosquitoes can be easily distinguished by the naked eye based on sex-specific phenotypic characteristics that carry predominantly on their head. Moreover, different mosquito genera can be also distinguished based on the length and shapes of some of these characteristics. *Anopheles* mosquitoes carry a long proboscis and two long maxillary palps as part of their sensory organs. Furthermore, male mosquitoes carry two antennae that are characterised by long flagellomeres that make the antennae look bushy (plumose antennae), whereas the female mosquitoes have two antennae with much shorter

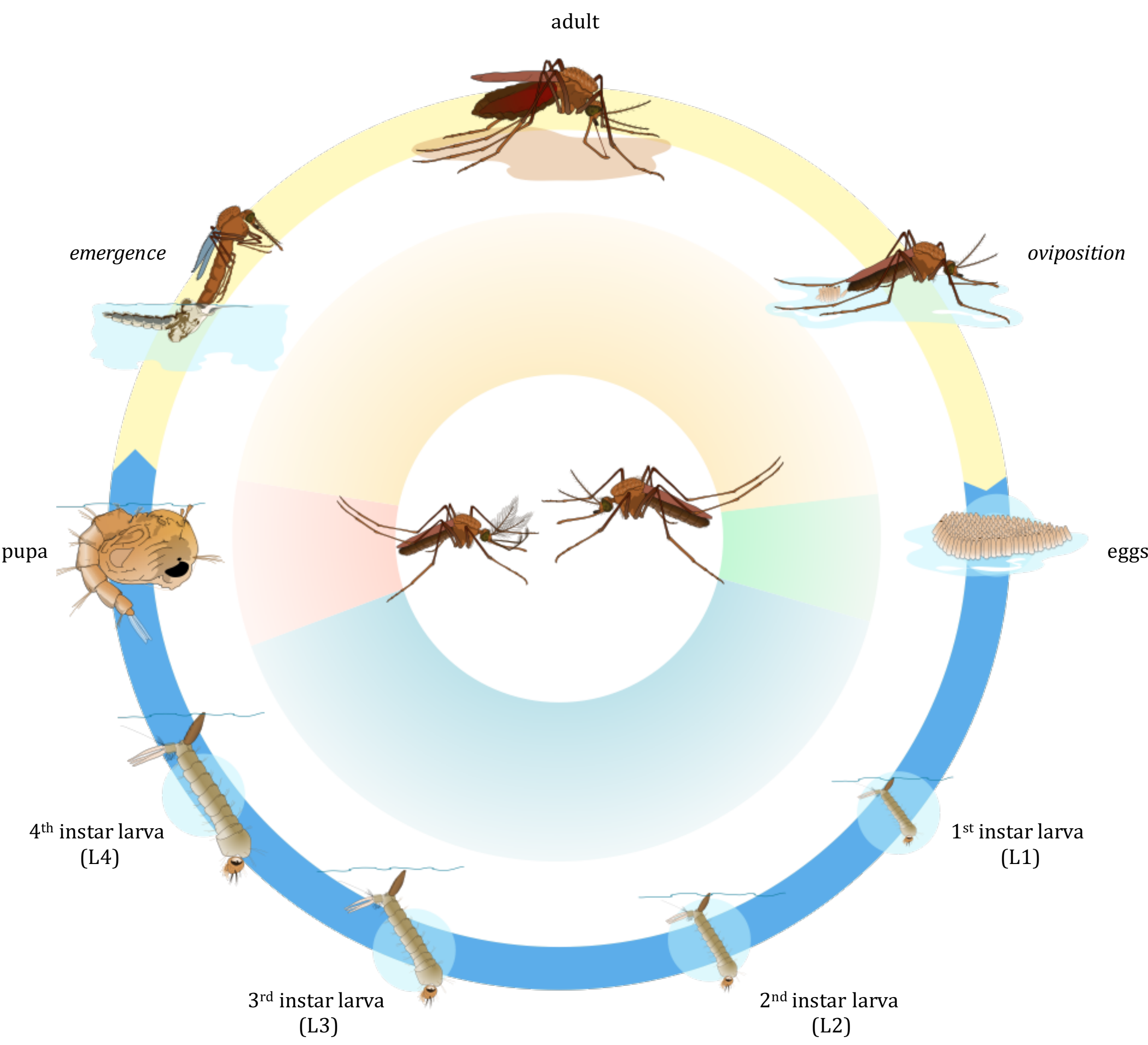


Figure 1.4 – The life cycle of the mosquito. The eggs are laid on the surface of stagnant water bodies and hatch within 24 to 36 hours. The larvae pass through 4 instar stages usually within 6 to 7 days before pupating. The pupa stage is the metamorphosis stage where the body of the insect is broken down and reorganised to form the adult mosquito. The adult emerges from the pupa shell within 36 hours and rests on the surface of the water to prepare for flight. Credit: Mariana Ruiz Villarreal.

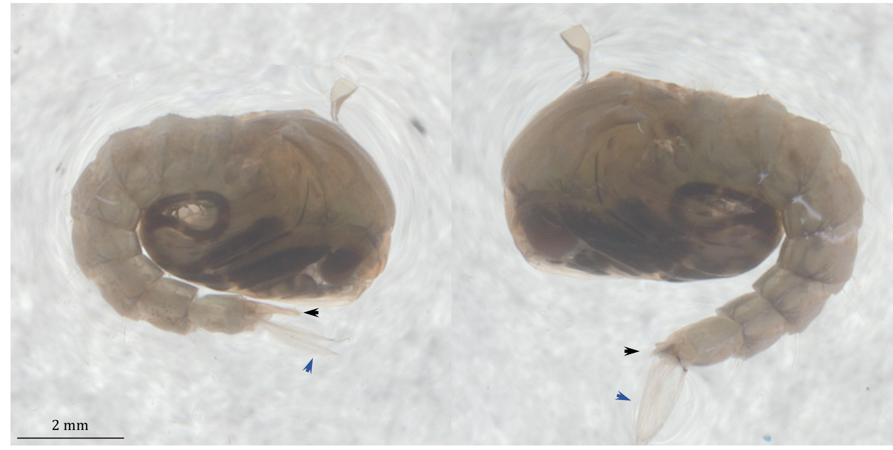


Figure 1.5 – The male and female pupae of *Anopheles gambiae*. The male pupae (left) have a longer genital lobe (black arrow) in relation to the size of the paddles (blue arrow). The female pupae on the other hand (right) have a much shorter genital lobe in comparison.

flagellomeres (pilose antennae). Finally, at the end of the abdomen the male mosquitoes have two claw-like structures called claspers or gonostylus, which are used during the mating process whereas females have two very short cerci (Becker et al. 2010)(Figure 1.6).

1.4.2 Insecticide-treated Nets (ITNs) and Indoor Residual Spraying (IRS)

Historically, interventions that aimed at reducing exposure to the parasite have been at the core of every malaria eradication programme. The use of both ITNs and IRS have achieved significant reductions in malaria cases around the world (Bhatt et al. 2015) (Figure 1.7). Since first introduced 30 years ago (Lines et al. 1987), ITNs proved to be an efficient and affordable intervention against endophagic malaria vectors. By being saturated with insecticides such as organophosphates and pyrethroids, they provide an extra chemical barrier that repels and kills mosquitoes (Lindsay et al. 1991, Miller et al. 1991). When used consistently in wide areas they have the efficacy to reduce the density and survival of mosquito vectors and thus protect the community against the disease (Phillips-Howard et al. 2003). ITNs and their Long-lasting impregnated version (LLINs) can last between 6 months and three years and they have an average annual cost of $2.60 (Yukich et al. 2008). An estimated 578 million ITNs have been delivered globally between 2016 and 2018, with the majority being distributed across the African region (WHO 2019). Although the population that has access to ITNs has been gradually increasing from 33% in 2010 to 57% in 2018, the improvement during the last 3 years has been slow. The same standstill is observed in the percentage of household numbers with at least one ITN for every two people, which increased to 72% in 2016 and remained stable over the past two years. Responsible for this stalling is not only the reduction of funds allocated on malaria programmes in Africa but also the fact that mosquitoes have begun to develop

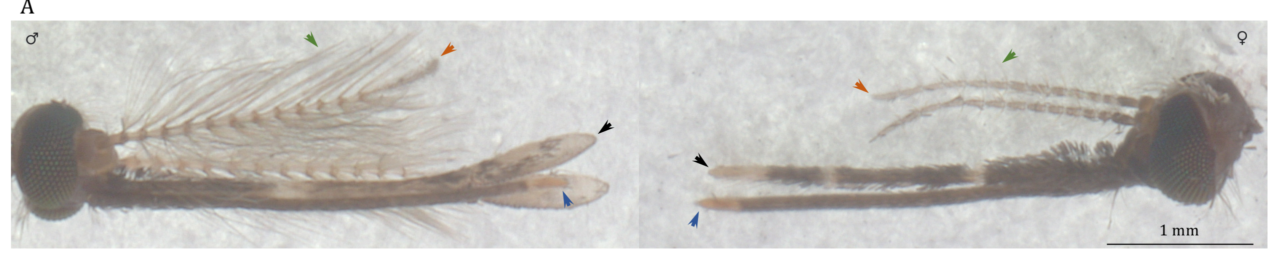


Figure 1.6 – The adult male and female *Anopheles gambiae* mosquitoes. (A) Males carry a plumose antennae system (red arrows) with longer flagellomeres (green arrows), whereas females have a pilose antennae system characterised with shorter flagellomeres. (B) Males carry a pair of claw-like structures at the end of their abdomen (arrow), whereas females carry two short cerci (arrow).

resistance against commonly used cheap insecticides. In just the past 8 years, 81 countries have reported data on insecticide resistance on pyrethroids and organophosphates (WHO 2019), which are commonly used on ITNs.

Similar efficiency against endophagic vectors has the application of spraying interior surfaces of houses every year with long lasting insecticides that repel and kill mosquitoes. Interior surfaces like walls and the roofs provide excellent resting sites for newly blood-fed females that are on the search for secluded areas to rest and develop their eggs. Treating these surfaces with insecticides reduces the lifespan and thus the density of these females and limit their host interaction and the transmission of the disease in the community (WHO 2015*b*). Combined with ITNs, IRS have contributed to a 25% reduction in malaria mortality globally and 33% in the African region during the last decade (WHO 2015*a*). Proper application though requires adequate national programme capacity and infrastructure as well as political will. The difficulty to meet these requirements in large scale areas has hindered the application of IRS and now fewer people at risk are protected by this method (WHO 2015*a*, 2019).

1.4.3 Chemical treatment and habitat removal

Before DTT (dichlorodiphenyltrichloroethane) and other residual insecticides became widely available, efforts to reduce mosquito populations and malaria incidence were essentially carried out by focusing on destroying mosquito habitats and disrupting the mosquito life cycle. Petroleum-derivative chemicals

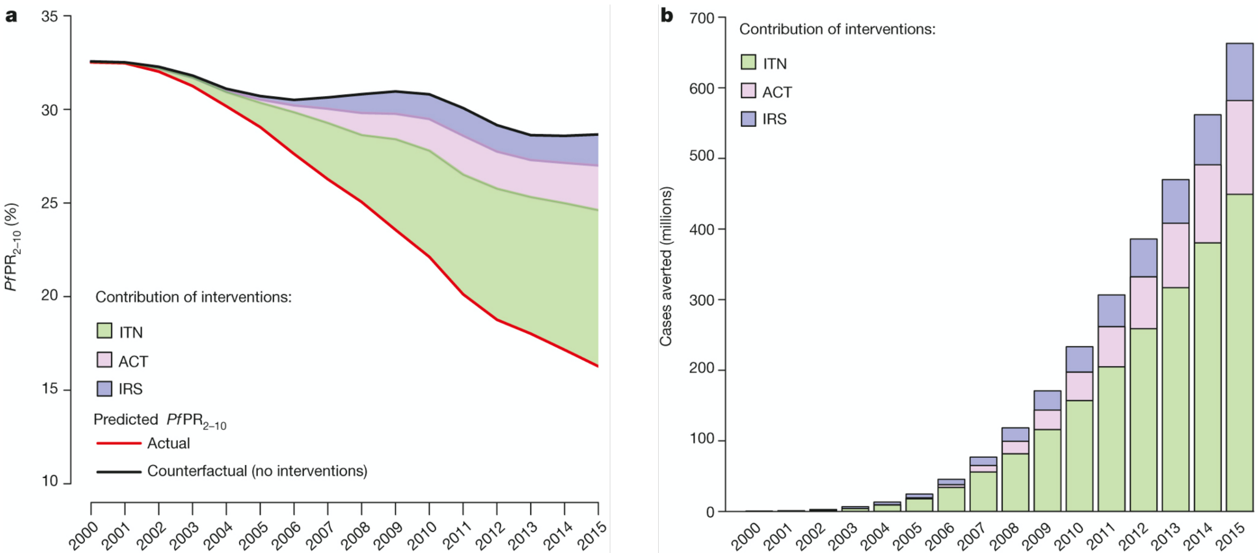


Figure 1.7 – The effect of interventions like ITN, ACT and IRS on infection prevalence and cases across endemic Africa. (A) The malaria infection prevalence in children aged between 2 and 10 years old (*Pf*$PR_{2\text{-}10}$) – indicated as population weighted mean – has halved since 2000. The current situation is indicated with a red line, whereas a hypothetical prediction where no intervention is used is shown with a black line. Contribution of each intervention is indicated with different colours. (B) Predicted number of malaria cases averted by the end of each year since 2000. The contribution of each intervention is indicated with different colours. Reprinted by permission from Springer Nature: Nature, The effect of malaria control on Plasmodium falciparum in Africa between 2000 and 2015, Bhatt et al. (2015), Copyright © 2015, Springer Nature.

like 'Paris Green' and hydrocyanic acid were used widely in the Mediterranean basin to treat the surface of stagnant waters that served as breeding grounds for malaria vectors. Italy, one of the worst affected countries of the basin with more than 2 million cases per year towards the ends of the 19th century (Majori 2012), managed to reduce the malaria transmission early in the 20th century through a combined approach of these larval control measures and free distribution of quinine-based antimalarial drugs (Majori 2012). In 1930, the mass drainage of large geographic areas like the Pontine marshes has caused the malaria incidences in the area to drop almost to zero within a decade, whereas later on, in 1944, the introduction of DDT was used in field tests to treat interior walls of houses and outbuildings against malaria vectors (Majori 2012). Combined these approaches have managed to interfere with the prevalence of the mosquito vector in the country to finally declare Italy malaria free decades later on 1970.

Similar stories are seen in other countries like Israel and the US, where a combination of strong infrastructure of health care services and implementation of interventions that aimed at disrupting the mosquito habitat were successful in eliminating the disease by the middle of the 20th century (Kitron & Spielman 1989). A heavy use of Paris Green and DDT was the core of an eradication campaign that took

place in the small island of Cyprus between 1946 and 1949 and resulted in the complete eradication of the disease from the island by 1967 (Shelley & Aziz 1949). These successful stories are proof that with co-ordination and political will we can achieve malaria eradication using simple interventions to reduce the prevalence of the mosquito vector from large geographic areas. In sub-Saharan Africa, where the problem of malaria is magnitudes higher, applying these methods to combat the disease might not be as efficient and effective as in other countries. Although habitat destruction and fumigation with insecticides do take place in local areas and near human settlements, applying these techniques in a continental scale might not be feasible and realistic. The jungle-like terrain of sub-Saharan African and the abundance of ideal mosquito breeding sites in the area (Munga et al. 2006) are impossible to overcome. Another aspect we need to consider before applying these methods in a such wider scale is that these interventions are not species specific and they tend to affect the entire entomofauna of the local ecosystems in which they are applied, which in the long run is a disadvantage for the local communities (Hulme 2006). With that being said, chemical treatment and habitat removal might prove useful and beneficial in eliminating the disease in the future when they are combined with novel interventions that aim to target the malaria problem at a continental scale.

1.5 Genetic engineering for vector control

Advancements in the field of science, genome engineering and genome sequencing over the last decades have sparked a great interest in the genetic modification of the mosquito vector as an alternative intervention to fight malaria. Both funders and scientific community have shifted their interest to pioneering methods that would block transmission of the disease by reducing the vectorial or reproductive capacity of the mosquito vectors. Many genetic control strategies have been proposed since then that use a variety of tools and genetic heritable elements to interfere with the mosquitoes' ability to either reproduce or transmit the disease. A great advantage of genetic engineering strategies over current interventions, is their sole dependency on vector mating and vertical transmission, which makes these interventions species-specific and environmentally friendly. By also making use of mosquitoes to fight mosquitoes, these interventions can be effective in large geographic areas with inaccessible terrain that otherwise would be difficult to treat with current measures.

Genetic control strategies can be classified into distinct groups using several criteria (reviewed in Alphey 2014). One main simple factor that is used for distinction is their development purpose and desired outcome. Based on this criterion alone, these techniques can be separated in strategies that aim to suppress vector populations and even eliminate them from certain areas (population suppression) or

strategies that aim to replace or enrich vector populations with members that are less effective in transmitting the disease (population replacement /modification). Another criterion used to distinguish between genetic control strategies is their invasiveness and persistence in the populations post implementation. Based on the degree in which they spread and become fixed in the populations after the first release, genetic control strategies can be labelled as 'self-sustaining' or 'self-limiting'. Self-sustaining approaches are designed to spread in populations indefinitely and replace wild-type populations with almost no subsequent releases of the modified vectors. On the contrary, self-limiting approaches are not designed to spread within a population indefinitely but rather to affect the population as long as they are boosted by subsequent periodic releases of modified vectors (Alphey 2014). Although the distinction in the two groups can be definite and unambiguous, the level of persistence and invasiveness can vary between strategies of the same group (Alphey 2014). The sterile insect technique (SIT) and modifications thereof has by far been the most effective group of genetic strategies that have been used across the world to control insect vector and pest populations (reviewed in Alphey 2002). Nonetheless, novel approaches that were made possible after recent advances in targeted genetic engineering have gained a particular interest, and are now under worldwide development from research groups working on vector-borne diseases.

1.5.1 Sterile insect technique - SIT

The sterile insect technique (SIT) is a species-specific, self-limiting strategy for vector control. The technique relies on the continuous release of sterilised males into a wild-type population (Knipling 2015, Krafsur 1998). The idea behind this is that mating between sterile males and wild-type females will result in no progeny, gradually reducing the reproductive capacity of a population. Originally, SIT techniques were using ionizing radiation and chemosterilisation to produce sterile males (Dame et al. 2009). Although these methods affect the mating competitiveness of males, a zero reduction in fitness costs is not essential for a successful SIT programme. The biggest limitation for the intervention is the continuous production and sexing of male insects. For a widespread disease like malaria, the failure to separate males effectively and rapidly has hindered the application of an SIT programme. Advances in sexing strategies that required the use of sex-chromosome-linked selectable markers that granted resistance to insecticides have been used widely to bypass this sexing bottleneck (Kim et al. 1987, Baker et al. 1978). Furthermore, with insect transgenesis it became easier to develop sex-specifically expressed fluorescent markers in early larval stages in the lab in order to theoretically achieve high throughput sorting of mosquitoes for the field (Catteruccia et al. 2005, Smith et al. 2007, Marois et al. 2012). Some went as far and succeeded in tagging sex-specific chromosomes, which historically is very challenging to do. Although the Y chromosome was tagged in

Anopheles gambiae (Bernardini et al. 2014), the fluorescent marker's brightness is very inconsistent and to date has not been used in a sex separation scenario. If we could engineer the Y chromosome of the species to carry a gene that confer resistance to a toxin, we could theoretically create an efficient system to rapidly select for male mosquitoes of the species. Historically though, the Y chromosome has proven very hard to engineer in *Anopheles gambiae* .

SIT and modifications thereof were used in notable cases to combat vector borne diseases and successfully managed to reduce the transmission levels of various diseases to zero or near-zero levels in many cases (reviewed in Alphey 2002, Alphey et al. 2010 and Benedict & Robinson 2003). Although many success stories were in isolated small areas and islands, one notable example of the effects SIT methods can have over large areas is the eradication of the screwworm fly *Cochliomyia hominivorax* from North and Central America (Wyss 2006). In other instances, SIT methods were used to prevent re-invasions of insect species. The vector of the sleeping sickness *Glossina austeni* (tsetse fly), was successfully removed from the small island of Zanzibar over a three-year intensive programme of SIT trials, and after the end of the programme the technique was used to avert re-invasion of the vector from mainland Africa (Vreysen et al. 2000). SIT methods were also used to combat serious outbreaks like the 1989 screwworm outbreak in Libya, which was averted before it spread in Africa and southern Europe (Lindquist et al. 1992). SIT interventions may have helped in containing and eliminating diseases worldwide but they still cannot be implemented in situations where the problem is widely spread over a continent like Africa. For comparison, for the successful eradication of the tsetse fly in the small island of Zanzibar, a total of 8.5 million sterile males were released over the three-year period(Vreysen et al. 2000). When irradiated *Anopheles coluzzii* males were tested in the lab for their effectiveness, it was shown that even when as half as competitive as untreated males, it required 5 to 1 sterile to untreated male ratio to impact the fertility of a female population (Maïga et al. 2014). For malaria the sheer abundance of mosquito vectors in the African region and the limitations we still have in mass producing and sexing genetic mosquitoes are two of the main problems that SIT methods cannot overcome.

1.5.2 Release of insects carrying a dominant lethal gene - RIDL

An advancement on the classic SIT method is the release of insects carrying a dominant lethal gene (RIDL). RIDL is a self-limiting approach that uses toxic effectors that are expressed in the embryo of the insect or later developmental stages (Heinrich & Scott 2000, Thomas et al. 2000). The toxic effector can be designed to be sex-specifically expressed in females only (female-specific RIDL) or it can be designed to affect both sexes (bi-sex RIDL). The idea is that offspring carrying a RIDL genetic element will not develop and this will reduce the reproductive capacity of a population over time and when the RIDL element

reaches high frequencies in the population. To produce these insects in the lab, the RIDL system is under the control of an antidote-responsive element that suppresses the toxic effectors in the presence of the antidote (Heinrich & Scott 2000, Thomas et al. 2000, Phuc et al. 2007). During rearing, both sexes are grown in the presence of the antidote and then, depending on the sex-specificity of the technique, both males and/or females are killed selectively by removing it. In female-specific RIDL systems, males carrying the RIDL element eventually are released in the wild and effectively kill their female progeny when they mate with wild type. Being a self-limiting approach, RIDL transgenic males must be reared and released frequently to successfully maintain the RIDL element in the population in levels that affect the population capacity.

Both female-specific RIDL and bi-sex RIDL systems have been developed for an array of vector and pest species. Bi-sex RIDL systems have been developed for crop pests like the Med fly *Ceratitis capitata* (Gong et al. 2005) and the bollworm *Pectinophora gossypiella* (Morrison et al. 2012). For mosquitoes, bi-sex RIDL systems have been developed and also been tested in field trials around the world with encouraging results. Specifically, a bi-sex RIDL system against the yellow fever mosquito *Aedes aegypti* (Phuc et al. 2007) has been tested in small areas in Thailand (Lacroix et al. 2012), in Brazil (Carvalho et al. 2015) and in the Grand Cayman islands (Harris et al. 2012). In both Brazil and the Grand Cayman, the RIDL programme was able to successfully suppress the local populations of *Aedes aegypti*, encouraging the use of the system to fight local pest and vector populations. Female-specific RIDL systems have also been developed against the vector and in some instances they have resulted in complete population suppression when tested in caged *Aedes aegypti* mosquitoes (Wise de Valdez et al. 2011). Despite the progress and success of the RIDL system in various pests and the mosquito vector *Aedes aegypti*, the technique has yet to be optimised for *Anopheles gambiae* , the major vector of malaria in Africa. The main limitation that hindered the transfer of the technique to the malaria vector is the scale of the problem in the continent. A sustainable suppression of *An. gambiae* in malaria endemic countries using RIDL will require a scale of production and rearing of mosquitoes that would likely exceed current capabilities in the continent.

1.5.3 Maternal-effect dominant embryonic arrest - Medea

Maternal-effect dominant embryonic arrest (Medea) are genetic self-selection systems that rely on the maternal deposition of a toxin in the embryo, which is neutralised by an antidote expressed in the embryo inheriting the Medea element (Figure 1.8). Females heterozygous for the Medea produce progeny that is only viable when they inherit the Medea element from either of their parents. Homozygous females produce 100% viable offspring since they pass the element with the antidote in all the progeny, whereas the viability of male offspring is not affected because the toxin is not expressed in the male germline. In

this two component system, deposition of the toxin will eventually kill offspring that are not inheriting the Medea antidote, increasing its frequency in the wild-type population (Chen et al. 2007, Marshall & Hay 2012).

First discovered in bacteria (Ogura & Hiraga 1983) and flour beetle *Tribolium castaneum* (Beeman et al. 1992), Medea systems and modifications thereof have been tested in *Drosophila* (Chen et al. 2007, Akbari et al. 2014), with a recent study using toxic miRNAs to impair embryonic genes in *Drosophila suzukii* (Buchman et al. 2018). In mosquitoes, the systems have only been proposed theoretically as part of alternative strategies to control the vector populations (Champer et al. 2016). Because Medea bearing chromosomes have an advantage over chromosomes that lack the element, the system can increase in frequency within a population when seeded above a critical threshold (Champer et al. 2016). It has been suggested that by linking an anti-parasite effector on the system, we can use Medea-based replacement strategies to drive the effector in wild-type vector populations (Champer et al. 2016). Medea systems have been difficult to develop and engineer for mosquitoes. Attempts to transfer the systems in vectors such as *Aedes aegypti* were not successful, mostly because of the lack of suitable toxins and antidotes but also because of our limited knowledge in the mosquito embryo biology (Champer et al. 2016).

1.5.4 Genetic underdominance - GU

Genetic underdominance (GU) also known as heterozygote inferiority, was first discovered in 1940 by Alexander Serebrovsky via chromosomal genetic traits that caused infertility or reduced fitness when in heterozygosity but not when in homozygosity (Serebrovsky 1940) (Figure 1.8). Serebrovsky and later on Curtis in 1968, have proposed balanced reciprocal chromosomal translocations as tools for population control. The system has gained interest during the 1970s and it was tested in mosquitoes for its efficiency in both laboratory and field-cage experiments (Baker 1984, Laven et al. 1972). The idea though of GU systems for population control was quickly abandoned because of fitness issues associated with mutagens that were used to induce the translocations. Since then, synthetic GU systems have been proposed for population replacement and suppression using toxins and antidote effectors. In 2001, Davis et al. proposed a synthetic two-element system based on transcriptionally active antidotes that bind the promoter of toxins located on opposite elements, whereas Akbari et al. in 2013 proposed a modification to the system by using maternally expressed toxins and zygotically expressed antidotes. Even though proof-of-principle experiments have been performed in model species like *Drosophila melanogaster* (Guy Reeves et al. 2014), to date neither natural nor synthetic GU systems have been used in the field to suppress pest or vector populations. Novel approaches that suggest the use of RNA-guided endonucleases may provide a more precise and straightforward method to induce sequence-specific translocations (Champer et al.

2016). Given that such GU-based drive systems may require a smaller release threshold compared to SIT and RIDL systems in order to spread in a population, and the fact that the systems can be easily withdrawn from local populations with the release of substantial wild type, the use of GU systems for vector control should be revised and considered (Champer et al. 2016).

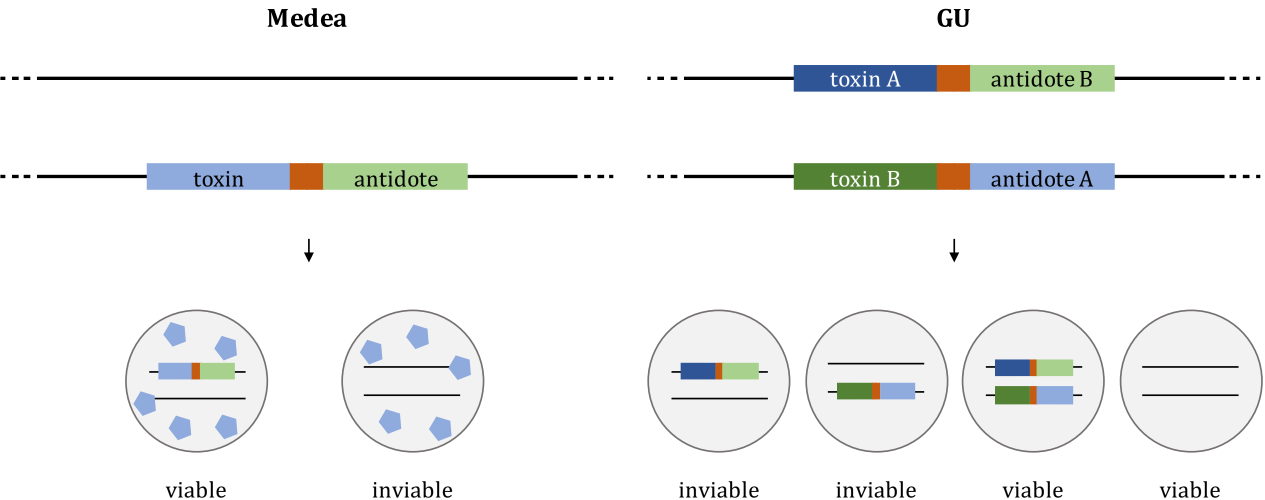


Figure 1.8 – The Medea and GU systems. Both systems use toxins and antidotes to selectively kill offspring that do not contribute in spreading the system. In Medea systems, a female-expressed toxin (blue) is passed maternally to the embryo. Embryos that carry the Medea element can neutralise the toxin and survive, whereas wild-type embryos cannot. In GU systems, embryos that carry both antidotes can neutralise both toxins. If only one element is inherited, the toxin cannot be neutralised and the embryos cannot survive. In both cases, the systems can be linked to anti-parasite effectors (orange), which are carried with the system, increasing in frequency within a population, when the systems are seeded above a threshold frequency.

1.5.5 Heritable microorganisms - *Wolbachia*

Some classes of naturally obligate microorganisms can bias their own inheritance and spread through arthropod populations using an array of methods. The genus *Wolbachia* is an intracellular bacterium parasite of this class that spreads in the progeny using vertical transmission (Werren et al. 2008). First discovered in 1924 (Hertig & Wolbach 1924), the bacteria use a method called cytoplasmic incompatibility (CI) to selectively interfere with the host reproductive capabilities and bias its inheritance in the progeny. *Wolbachia* exploit CI to selectively sterilise uninfected female mosquitoes that otherwise could not transmit the bacteria. Specifically, incompatibility between the sperm and the egg allows infected females to produce progeny when fertilised either by infected or uninfected males but inhibits the production of

progeny when uninfected females mate with infected males (Figure 1.9) (Yen & Barr 1971, Laven 1956). Because the bacteria are transmitted through the egg cytoplasm in almost all of the progeny but are not transmitted through the sperm, CI ensures the production of offspring that only carry the parasite, contributing this way to the spread of the organism.

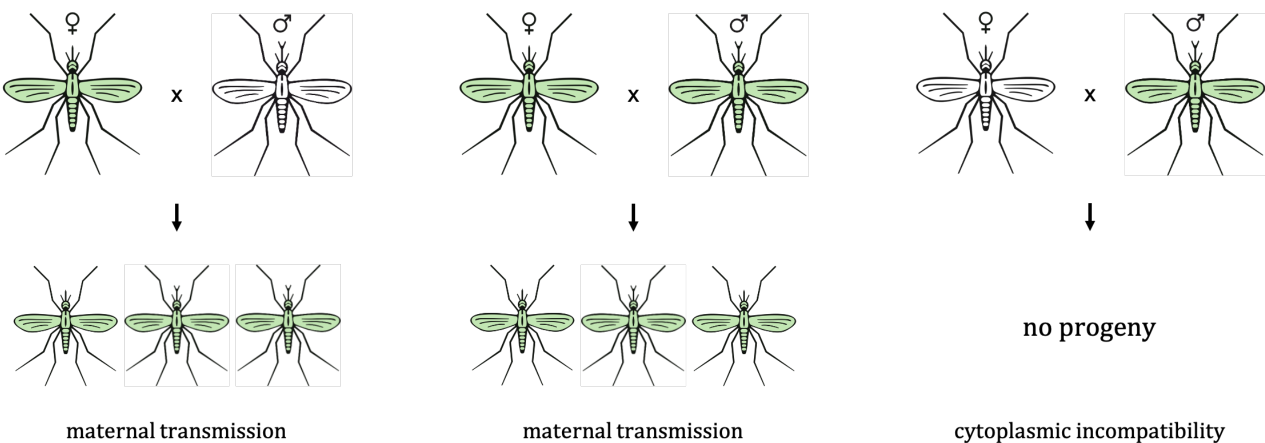


Figure 1.9 – The *Wolbachia* system and cytoplasmic incompatibility. The *Wolbachia* bacteria are transmitted through infected females to the entire progeny, increasing their frequency in the population. The infection status of the male mosquitoes does not affect the fertility of infected females. On the contrary, infected males sterilise uninfected females due to cytoplasmic incompatibility between the egg and the sperm, contributing indirectly to the spread of the bacteria. Infected mosquitoes are annotated in green.

Wolbachia infections have been considered as an intervention method to control transmission of vector-borne diseases since they have been associated in mosquitoes with reduced viability (McMeniman et al. 2009) and vector competence (Moreira et al. 2009, Bian et al. 2013, Sinkins 2013). After vigorous validation processes, one strain has been selected that successfully invades *Aedes* populations and also decreases their infectious capacity to transmit diseases like yellow fever, dengue and chikungunya (Walker et al. 2011, Hoffmann et al. 2011, Van den Hurk et al. 2012, Aliota et al. 2016). The strain has been tested in multiple sites around the world and showed in some cases to sustain an infection rate of almost 90% over the span of multiple years after the end of the release phase (Hoffmann et al. 2014). Since then, the international initiative “World Mosquito Programme”, www.worldmosquitoprogram.org, has made it a goal to use the bacteria and the intervention to fight *Aedes*-borne diseases in multiple countries around the world.

The ability of *Wolbachia* bacteria to successfully infect *Aedes* mosquitoes and also block the transmission of diseases is not universal and not the case for other vector species. When *Wolbachia* strains were

transferred from *Aedes* in *Anopheles* mosquitoes, a reduced efficacy in producing sporozoites has been observed but studies have yet to determine the prevalence of those parasite stages (Killeen 2013, Gomes & Barillas-Mury 2018). In addition, this 'in-transient' infection of *Anopheles* mosquitoes is also shown to affect significantly the fitness and fertility of the mosquitoes (Bian et al. 2013, Hughes et al. 2011). Although there are some reports of natural *Wolbachia* infections in *Anopheles* species (Baldini et al. 2014, Jeffries et al. 2018, Gomes et al. 2017), the strains identified to date do not impose CI and were shown to compete with other members of the mosquito microbiome, hindering in this way their vertical transmission (Hughes et al. 2014). In addition, infected *Anopheles* mosquitoes presented a reduced fitness and fertility (Shaw et al. 2016), which further applies a negative selection pressure on the bacteria and its inheritance. Nevertheless, it was suggested that the presence of *Wolbachia* may be negatively correlated with the presence of *Plasmodium* in some cases (Wang et al. 2017, Gomes et al. 2017), and if this remains the case in heavily infected areas stills needs to be determined. Although promising, the limitations and challenges faced when implementing the system in *Anopheles* mosquitoes may delay the development of a *Wolbachia*-based intervention against malaria.

1.5.6 Heritable microorganisms - *Microsporidia*

Very recently, a new species of *Microsporidia* was identified that seems to completely block the development of the parasite in *Anopheles arabiensis* mosquitoes, which are a major vector of the disease in Africa (Herren et al. 2020). *Microsporidia* are obligately intracellular simple eukaryotes and are closely related to fungi (Han & Weiss 2017). Although *Microsporidia* have been found previously in mosquitoes and have been shown to interfere with the development and infection of the parasite, they also were associated with substantial mosquito larval mortality and a reduced adult survival and fecundity (Fox & Weiser 1959, Undeen & Alger 1975, Haq et al. 1981, Gajanana et al. 1979). The new species of *Microsporidia*, designated *Microsporidia MB* belongs to a separate clade of previously characterised species (Vossbrinck & Debrunner-Vossbrinck 2005) and was shown to have no observable effect on mosquito development, survival and fecundity (Herren et al. 2020). The species was found to occur in natural *Anopheles arabiensis* populations in Kenya with up to 9% prevalence in some cases and found to be maternally transmitted from 45% up to 100% in the progeny. Every mosquito captured in the wild that carried *Microsporidia* was found to be free of the *Plasmodium* parasite, indicating a negative correlation. Even when the *Plasmodium* prevalence was as high as 6% in the sampled populations, co-infection of the two microorganisms was never observed. When mosquitoes were actively infected in the lab, the team was not able to detect the *Plasmodium* parasite in the head and thorax of the mosquito, indicating that *Microsporidia MB* prevents the salivary glands to be colonised by sporozoites and thus reducing the vectorial capacity

of the mosquito (Herren et al. 2020). The mechanism behind this behaviour was attributed to the increase of the expression of an array of genes involved in digestion and immunity, like the *gambicin* gene, which has been shown to be an anti-*Plasmodium* effector in *An. gambiae* (Vizioli et al. 2001). The discovery of the *Microsporidia MB* species and its anti-malaria capabilities is a significant finding in terms of malaria epidemiology and transmission. The use of this natural endosymbiont species as a candidate transmission-blocking tool in population replacement strategies is attractive and could prove to be an excellent intervention to tackle malaria. Of course, because of its novelty and nature, this new strategy needs to be thoroughly examined and assessed prior to any field release scenario.

1.5.7 Sex distorters

Selfish genetic elements that bias the sex ratio of the progeny have been proposed as methods to control vector populations since they were discovered in the vector *Aedes aegypti* in the 1960s (Craig et al. 1960, Hickey & Craig Jr. 1966, Hamilton 1967) and *Culex pipiens* in the 1970s (Sweeny & Barr 1978). A sex chromosome or locus that has the ability to bias its own inheritance during meiosis could in theory interfere with the reproductive capacity of a population and lead to population crash due to the lack of females. Such scenarios have been proposed for the control of the malaria vector using an artificial genetic construct located on the Y chromosome of the species that actively destroys the X chromosome during spermatogenesis (Burt 2003). If the X-bearing gametes die before fertilisation of the egg, more male progeny will be generated at the expense of the females. If such a system does not pose any other fitness or fertility cost to the mosquitoes carrying the nuclease, they can theoretically be used to suppress vector populations (Deredec et al. 2008).

A naturally occurring nuclease discovered in the acellular slime mold *Physarum polycephalum* (Muscarella & Vogt 1989, Muscarella et al. 1990) has shown to cleave with high efficiency human ribosomal DNA (rDNA) repeats, which are conserved in eukaryotic species (Monnat et al. 1999). By serendipity, in *An. gambiae* these repeats are exclusively located on the X chromosome, meaning that the nuclease, called I-PpoI, could be evaluated for its efficiency to create a sex distorder system in the vector (Windbichler et al. 2007). The nuclease was expressed in the male germline of the mosquito, leading to a selective bias towards male embryos by "shredding" the X chromosome. Unfortunately, the nuclease was also transmitted to the progeny via the sperm, cleaving the maternal X chromosome and affecting the survival of the embryos (Windbichler et al. 2008). Since then, attempts to mutate the protein have managed to produce unstable versions that were depleted before reaching the embryo, allowing the survival of the progeny, which was up to 95% male biased (Galizi et al. 2014). Finally, it has been shown that alternative systems like the CRISPR/Cas9 can be also designed to target the same sequences in *Anopheles gambiae* and used

to bias the male sex in the progeny (Galizi et al. 2016).

An efficient and self-sustaining approach using such a system can be successful in invading a population if is placed preferably on the Y chromosome and expressed during meiosis using a germline-specific promoter to achieve a biased inheritance (Burt 2003, Deredec et al. 2008). This way, all male progeny will carry the distorter, increasing the frequency of the system in the population. Historically, the Y chromosome in *An. gambiae* has been very hard to engineer because it consists mainly of heterochromatic regions that are difficult to access and manipulate (Holt et al. 2002). Transgenic expression from the chromosome has only been reported once (Bernardini et al. 2014) and even then, expressing any germline-acting nuclease from any of the two sex chromosomes has been unsuccessful due to meiotic sex chromosome inactivation (MSCI), a situation in which transcription levels are dramatically reduced from both the sex chromosomes during meiosis (Turner 2007). Efforts to engineer nucleases that can escape MSCI and be expressed in the germline by using gypsy insulators (Chen & Corces 2001) are currently underway (unpublished with permission of Dario Meacci).

1.5.8 Homing-based gene drives

Homing endonuclease genes or HEGs, are natural genetic elements that can be found in a large range of organisms (reviewed in Hafez & Hausner 2012). HEGs have the ability to bias their own inheritance in a population by copying themselves between chromosomes and converting heterozygotes into homozygotes. HEGs code for nucleases that recognise and cleave specific DNA sequences of 15 to 30 bp that usually occur once in the genome. Because the HEG itself is located between its own recognition sequence it cannot recognise and cleave the chromosome bearing the HEG (HEG^+). On the contrary, the intact sequence on the homologous chromosome not bearing the HEG (HEG^-) is recognised and cleaved by the nuclease. When that happens, the cell tries to repair the cut using homology-directed repair (HDR) and the homologous chromosome as a template. In this case, the homologous chromosome carries the HEG and thus the HEG itself is copied across to the HEG^- chromosome as a by-product of the repair. This copy and paste reaction is called 'homing' and converts the cell into a homozygote for the HEG (Figure 1.10). By enhancing the homing process in the germline of an organism, the HEG inheritance will be strongly biased among the progeny. Because the homing process is exclusively a by-product of the cell repair mechanisms and is unrelated to the HEG element, any site-specific nuclease could be re-purposed in theory as a homing-based gene drive (Burt 2003).

Homing-based gene drive systems have been suggested as alternative tools that can be used in population suppression or population replacement strategies against malaria vectors (Burt 2003). For population replacement scenarios, the gene drive system can be engineered to carry anti-pathogen effectors or

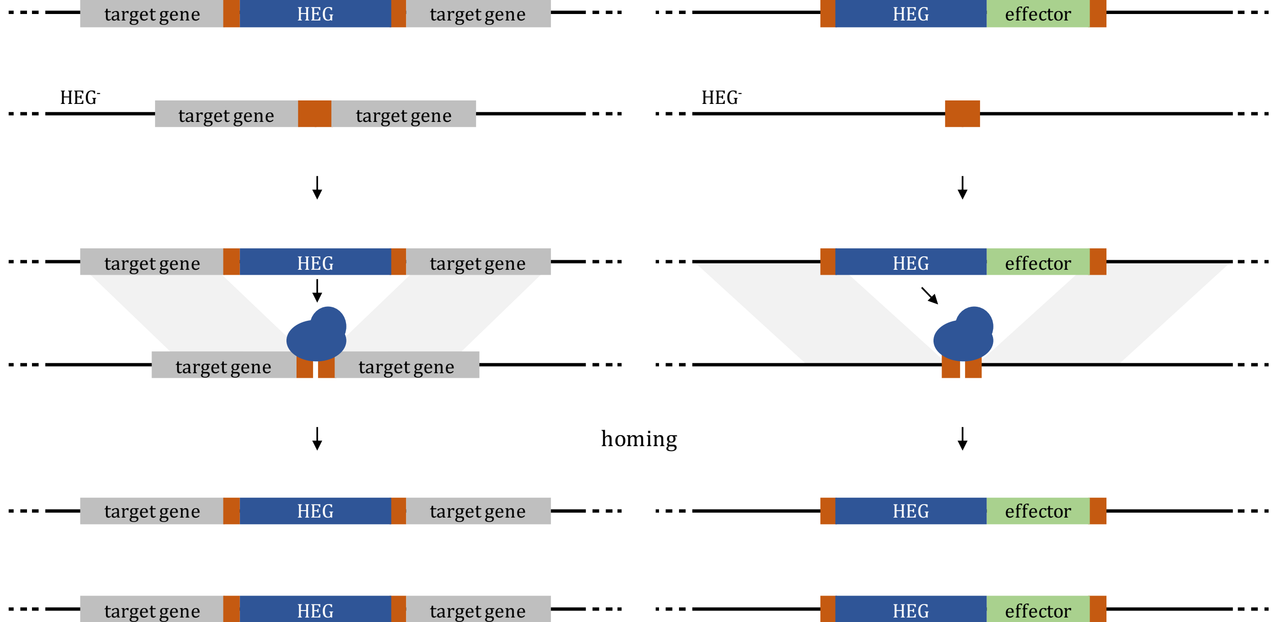


Figure 1.10 – Homing endonuclease genes for malaria control and the homing reaction. The HEG is placed inside its own recognition sequence (orange) and when expressed it cleaves the wild-type sequence on the homologous chromosome not bearing the HEG (HEG-). If the cell's repair machinery uses HDR to fix the cut, the HEG will be copied across to the wild-type chromosome converting the cell to a homozygous for the HEG and biasing its inheritance. If the HEG is designed to cleave a sequence within a target fertility gene (left), the homing process will create homozygous deletions for that gene that will spread in the population and impair the reproductive output of that population over time. On the other hand, if the HEG is linked to an anti-parasite effector (right), its frequency will be biased in the progeny and increase in the population granting resistance to the parasite.

it can be designed to target mosquito genes that are required for parasite development or survival. The idea is that the parasite-resistant mosquitoes will slowly increase in frequency with every generation, reducing the vectorial capacity of the mosquito, i.e. the efficiency with which the vector can transmit the pathogen. Another strategy to reduce the vectorial capacity of the mosquito population without affecting its size, is to design the HEG to target genes that are required for adult longevity. The idea behind this is to disrupt the parasite's life cycle by reducing the longevity of the female mosquitoes and thus the number of gonotrophic cycles, which would lead in a reduced number of progeny per female, something already proposed using *Wolbachia* infections (McMeniman et al. 2009). Because the parasite requires approximately 14 days to mature in the females (Warrell & Gilles 2002) but the females need only 48 to 72 hours for one gonotrophic cycle, this method can actually break the malaria cycle and reduce the pathogen prevalence in the populations (Deredec et al. 2011).

On the other hand, when a gene drive system is designed to suppress vector populations instead, the criteria for the target site are different. For population suppression, the HEG is preferably designed to

target genes that are either required for female fertility or viability, although genes that affect both sexes when mutated in homozygosity could also be used in principle for such a scenario. In any case, the genes need to be haplosufficient so that heterozygous mosquitoes can be fertile and contribute in transmitting the HEG to the next generations. For that to happen, the HEG must be designed to function and home only in the germline of the mosquito. This way, the somatic tissues of the mosquito will not be converted to homozygosity, allowing the mosquito to be fertile (Figure 1.11). If the gene is also required for the function of the germline, the conversion to homozygosity will have negative effects on the fertility of the mosquitoes and the spread of the gene drive. When designed and functioned properly in the germline, gene drive systems can home in the germline of 'carrier' individuals, resulting heterozygous mosquitoes that transmit the HEG to the majority of the progeny, biasing the inheritance of the system in the population.

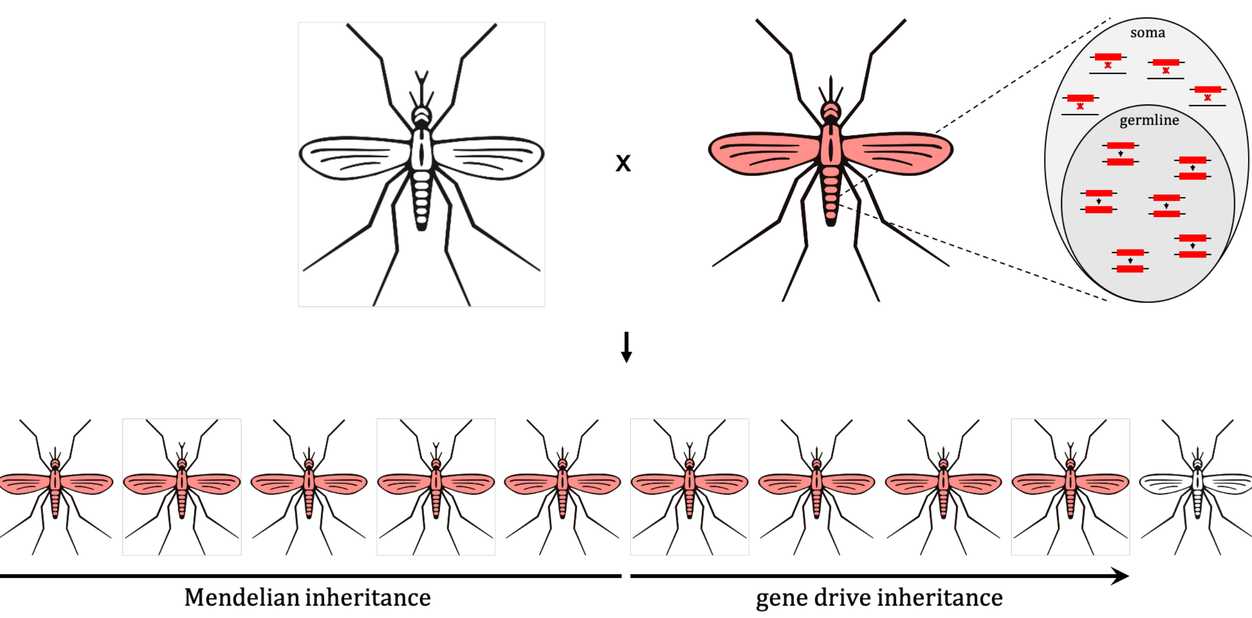


Figure 1.11 – A gene drive targeting female fertility or survival. When the target of a gene drive system is a haplosufficient gene that affects female fertility or survival, the homing reaction has to be restricted in the germline of the mosquito using germline-specific promoters to regulate the expression of the nuclease. This way, the somatic tissues, where the gene is needed, will remain heterozygous and function properly, whereas the germline will convert to homozygous. Female mosquitoes carrying the gene drive in heterozygosity will be able to produce progeny in this way and also spread the gene drive to more than half of the progeny. If the gene drive was not homing in the germline, the gene drive inheritance – or homing rate, would be zero and thus the gene drive would only be inherited by half of the progeny, as stated by the Mendelian rules of heredity.

The effect of the gene drive on the population mostly depends on the rate in which it copies itself in the germline (homing rate) and the fitness cost it imposes on heterozygous carrier individuals (Deredec et al. 2011). A perfect gene drive system will home efficiently in the germline of heterozygous male and female carriers, imposing zero fertility or fitness issues upon those mosquitoes. This way, when mated with wild type, they will always produce progeny that carries the gene drive, doubling its frequency in the population. When the frequency of the gene drive increases substantially within the population, the probability of two heterozygous mating with each other increases, leading to the production of homozygous mosquitoes. If the target gene is essential for mosquito survival or fertility, the reproductive capacity of the population will keep decreasing with the increase of homozygous, resulting gradually in population suppression.

Naturally occurring HEGs have been used and assessed in proof-of-principle experiments in the model organism *Drosophila melanogaster* (Chan et al. 2011) and *Anopheles gambiae* (Windbichler et al. 2011). In mosquitoes, artificial introduction of the native target site of a HEG into the mosquito genome showed that homing process could occur in the organism. However, the re-engineering of natural HEGs to target endogenous genes in the mosquito genome has proven challenging, with only a few notable examples (Chan et al. 2013, Thyme et al. 2014). This is because the 3D structure of the protein and amino acid composition of its active centre are responsible for the HEG specificity. To switch target sites means that the HEG has to be re-engineered at the DNA level and assessed through a process of trial and error to find the one with the right structure that recognise the new site of interest. Over the years, a number of novel gene editing technologies have been used to replace native HEGs with more easily programmable nucleases. Zinc finger nucleases (ZFN) (Kim et al. 1996, Urnov et al. 2010) and transcription activator-like effector nucleases (TALENs) (Miller et al. 2011) are easier to engineer than HEGs and have also been tested for their ability to function in the mosquito (Smidler et al. 2013) but also for their efficiency when used in a gene drive scenario (Simoni et al. 2014). Finally, the more recently discovered CRISPR/Cas9 system has been highly considered as a potential candidate to replace current HEG technologies.

1.5.8.1 The CRISPR/Cas9 system

A revolution in genome engineering brought the discovery of the CRISPR system (clustered regularly interspaced short palindromic repeats) (Figure 1.12, which plays a major role in the immunity of bacteria and archaea against viruses (reviewed in Bhaya et al. 2011 and Wiedenheft et al. 2012). The system uses CRISPR-associated (Cas) nucleases and short RNA sequences, 17 to 20 bp long, to recognise and digest viral DNA that enters the microbial cell. The way the system works in bacteria is by incorporating in their genome pieces of the viral DNA from past infections, creating a molecular 'memory' that is used to fight

new infections. The viral DNA sequence is used to produce small RNA molecules – called guide RNAs (gRNAs), which complex with Cas nucleases to digest viral DNA and stop an infection after a virus infects the cell. The specificity of the CRISPR/Cas system to the target site is defined by the 'spacer' region, which is the region that is identical to the gRNA target site, but also from the protospacer adjacent motif (PAM), which is a 3-bp sequence adjacent to the 3' end of the spacer sequence and is only present on the target DNA. The PAM is specific to the Cas nuclease and is located only on the target site in the viral genome. Since this 3-bp sequence is essential for nuclease activity and is absent from the gRNA loci in the bacterial genome, it helps the system distinguish the viral DNA from the host DNA.

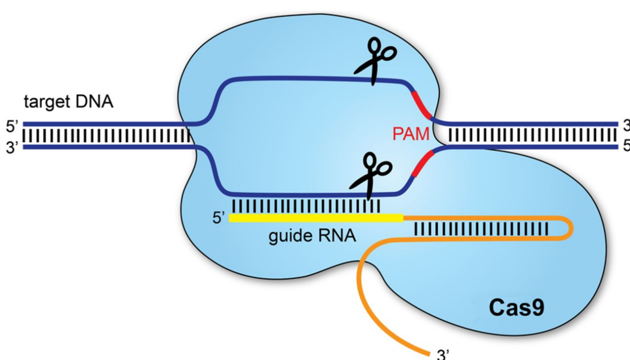


Figure 1.12 – The CRISPR/Cas9 system. The two-component system consists of a Cas9 nuclease and a gRNA molecule, which are combined together to form ribonucleoprotein particles. The specificity of the system is regulated by the gRNA sequence and the PAM site. The PAM site is required for Cas9 activity and is only present in the target locus adjacent to the gRNA. Image reproduced with permission of the rights holder, Copyright © 2016, BMJ Publishing Group Ltd. Redman et al. 2016.

The CRISPR/Cas system has been rapidly adopted by the scientific community and replaced almost all other genetic engineering tools because of its simplicity and easy manipulation. The simple design of a gRNA is all that it takes for the system to change its specificity among different targets, which is much faster than re-engineering HEGs, ZFN or TALEN nucleases. The most commonly used variation of the system, the CRISPR/Cas9 system that derived from *Streptococcus pyogenes* (Jinek et al. 2012), has been immediately used as an alternative tool in homing-based gene drive systems. Proof-of-principle experiments have been performed in yeast (Dicarlo et al. 2013) and fruit fly (Gantz & Bier 2015), before the system was adapted and tested in gene drives for mosquito population replacement or suppression. In 2015, Gantz et al. used the CRISPR/Cas9 system to bias the inheritance of anti-malaria effectors in the progeny of heterozygous *Anopheles stephensi* mosquitoes. Although the system has not been tested in caged populations in that study, the gene drive could home in the germline of the mosquito with up to

99% efficiency.

1.5.8.2 Current status and challenges

In our lab, we use the CRISPR/Cas9 system to target female fertility genes in the malaria vector *Anopheles gambiae* to find suitable targets for gene drive systems aiming at population suppression. We have designed gene drive systems targeting three female fertility genes and have demonstrated the ability of the system to function in the germline of the mosquito and bias its inheritance in the offspring (Hammond et al. 2016). One of those gene drives, targeting the gene *AGAP007280* – the *nudel* orthologue of *D. melanogaster*, has been released in caged mosquito populations to study the dynamics of the system at the locus. The gene drive was released at a 50% transgenic frequency using heterozygous gene drive mosquitoes. As expected from model predictions, the gene drive increased in frequency within 6 generations, reaching 75% to 80% transgenic ratio, but failed to sustain those frequencies and ultimately impose a reproductive load on the population of the cage (Hammond et al. 2017). The reason behind this performance is the selective pressure the gene drive poses on resistant alleles that were created as the result of the nuclease activity on the *AGAP007280* locus. Mutations at the target site that restored the function of the *AGAP007280* gene had a tremendous selective advantage over the gene drive in the population of the cages, gradually increasing in frequency and counteracting the function and spread of the gene drive. The creation and selection of resistant alleles is a major issue for gene drive technology that must be considered when designing such systems, especially when are destined to impose a selective pressure on a population.

Although the gene drive targeting *AGAP007280* was not developed for population suppression but rather to study the dynamics of the system, one of the main reasons it was prone to fail because of resistance was the low conservation that characterised the target sequence of the gene. A high level of variation implies that a locus is not functionally constrained and most probably can handle in-frame mutations, like the ones observed in the cage population experiment (Hammond et al. 2017). A number of measures has been proposed to reduce the impact of the resistant alleles created at the target site in future gene drive systems. The use of multiple gRNAs to target multiple loci in close proximity has been suggested as an alternative way to design gene drives and overcome the problem of resistance. 'Multiplexing' gRNAs like this ensures that even if resistant alleles are formed in some of the targeted loci they will not impair the homing process. This is because in theory only one gRNA, i.e. one double strand break, is needed for the homing process to occur. The idea behind the use of multiple gRNAs ensures that a gene drive system has enough time to invade the population and cause a significant reduction before all the target sites confer resistance.

Finally, the choice of the target site could also play an important role controlling the formation of resistant alleles. Sites like the one chosen for *AGAP007280* allow the creation of resistance due to their low conservation and small functional constraints. By selecting a target site that is highly conserved, not only among closely related species but maybe among distant taxa, may have a dramatic outcome in the formation of resistant alleles. Sites with such characteristics probably code for genes that are functionally constrained or even serve as binding sites for factors that are highly sequence specific. Of course, this additional requirement limits the number of target sites that can be used in gene drive systems targeting female fertility. To find a gene that is sex-specifically expressed and is also highly conserved, we have to look at pathways that regulate fertility and/or sex development. One high candidate is the sex determination pathway, which includes genes that are known to be highly conserved. Within such a pathway the possibility of genes to be also sex-specifically expressed is high, making the sex determination pathway an ideal place to begin. In the next chapter, I outline the sex determination mechanisms used within the insect class and I focus on the *doublesex* gene, which is a highly conserved transcription factor that is also sex-specifically expressed. In chapter 3, I outline the materials and methods used in this project. In chapter 4, I describe and characterise the *doublesex* gene in *An. gambiae*, whereas in chapter 5, I test its potential to be used in a gene drive system. Finally, in chapter 6, I test the gene drive targeting *doublesex* for its efficiency to spread in caged mosquito populations and its ability to handle resistant mutations.

Chapter 2

Sex Determination in Insects

Given that the main research goal of my thesis was to interfere with the sex determination pathway as a novel form for insect control, it is prudent to survey what is known about this process and the pathways that control the development of sex. This is in order to understand *a priori* the impact and effects of disrupting key factors of this pathway and how broadly applicable such an approach might be to implement across the insect class.

2.1 Sex determination in the animal kingdom

The mixing of genomes from different individuals and the combination of alleles that carry beneficial genetic variance has been at the core of the evolution of eukaryotes. Through trial and error, *sexual reproduction*[1] has driven the evolution of species by creating individuals that ultimately combine traits that have survived the enormous selective pressure posed by nature. One great misconception is that the familiar X and Y sex determining systems found in humans and few other model species like *Drosophila melanogaster*, are examples of one conserved sex mechanism, and that the same X and Y primary determining signals are adopted throughout the tree of life. These examples though represent only a tiny fraction of the group of species that make up the tree of life and are not conserved as the primary sex mechanisms that determine the male and female sex (Figure 2.1). There is an enormous diversity of sex determining signals found in eukaryotes and they are rapidly evolving to such a degree that sometimes even closely related species do not share similarities in the way sex development is triggered. Nonetheless, although the primary sex determining mechanisms are diverse within the tree of life, the pathways that lie downstream of the trigger signals that regulate male and female sex development and differentiation, are more conserved and usually follow a common theme.

[1] Sex determining glossary terms emphasised in italics are outlined in Glossary of Terms

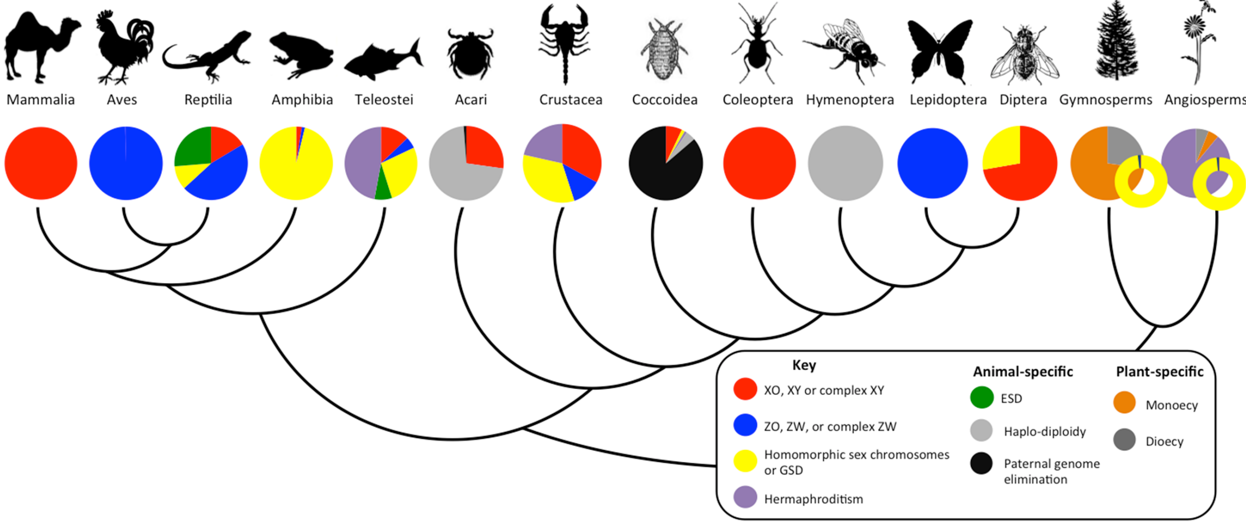


Figure 2.1 – The diversity of sex determination systems in the animal and plant kingdoms. The sex determining signals are far from conserved between different clades. For explanation of sex determining mechanisms see text below. The extra graph in the plant clades shows the proportion of species with sex chromosomes in plants that have separate sexes. Image reproduced with permission of the rights holder, Copyright © 2014, Bachtrog et al. 2014.

Sex in humans is determined by *heteromorphic sex chromosomes* X and Y, which have differentiated early in mammalian evolution from ordinary *autosomes* (Lahn & Page 1999, Ross et al. 2005). The accumulation of an early sex-determining locus on the Y chromosome and subsequent sexually antagonistic mutations have allowed chromosomal inversions to evolve and block recombination from happening between the ancestral X and Y, which eventually led to the complete differentiation of the two sex chromosomes (reviewed in Bachtrog 2013). The inability to recombine and share genetic information with the X chromosome, has allowed the Y chromosome to shrink in size and reduce the number of genes it carries to just fewer than a hundred, while the X chromosome carries almost a thousand (Bachtrog 2013). Even so, the Y chromosome – and more specifically a single gene, the *Sry* gene – is solely responsible for sex determination in humans. The presence or absence of the gene during the sixth week of embryo development determines whether gonadal tissue will differentiate into testes or ovaries respectively (Eggers & Sinclair 2012). When the gonadal tissue is formed in turn, through the release of hormones, it initiates the sex differentiation process of the non-gonadal tissues of the embryo to establish the male or female sex. It is hypothesised by many that the Y chromosome will eventually disappear in the future as part of the continuous evolution of heteromorphic sex chromosomes (Steinemann & Steinemann 2005, Charlesworth & Charlesworth 2000) (Figure 2.2). It is theorised that the system will eventually be replaced by an XO (X null) sex determining system, where the presence/absence or the number of X chromosomes

would determine the male sex, as it is observed in many species (Castillo et al. 2010). However, empirical observations on *neo-chromosomes* of some *Drosophila* species (Lucchesi 1978), combined with model predictions, suggest that sex chromosome degeneration is not a straightforward process and not the final fate of all sex chromosomes (Bachtrog 2013).

The X and Y sex determining system is a great example of *male heterogamety* where the heteromorphic XY individual is developed to be a male whereas the *homomorphic* XX animal developed as a female. This sex determination mechanism is widely found in all mammals, the majority of insect species within the Coleoptera and Dipteran orders and by some fish and a small portion of reptiles (Ashman et al. 2014) (Figure 2.2). *Female heterogamety*, the system where the heteromorphic individual is female and the homomorphic individual a male, is also widespread in many classes and orders such as birds, snakes, crustaceans, arachnids and species of the Lepidoptera order. In this system, named ZW for pure distinction from the XY system, the heteromorphic female individual (ZW) determines whether the embryo would develop to be a female or a male (ZZ). As with the XY system, the ZW system is believed to have evolved from a pair of autosomes that gained a locus or gene that determined the female sex and that subsequent loss of recombination diverged the two into separate chromosomes (reviewed in Ezaz et al. 2006).

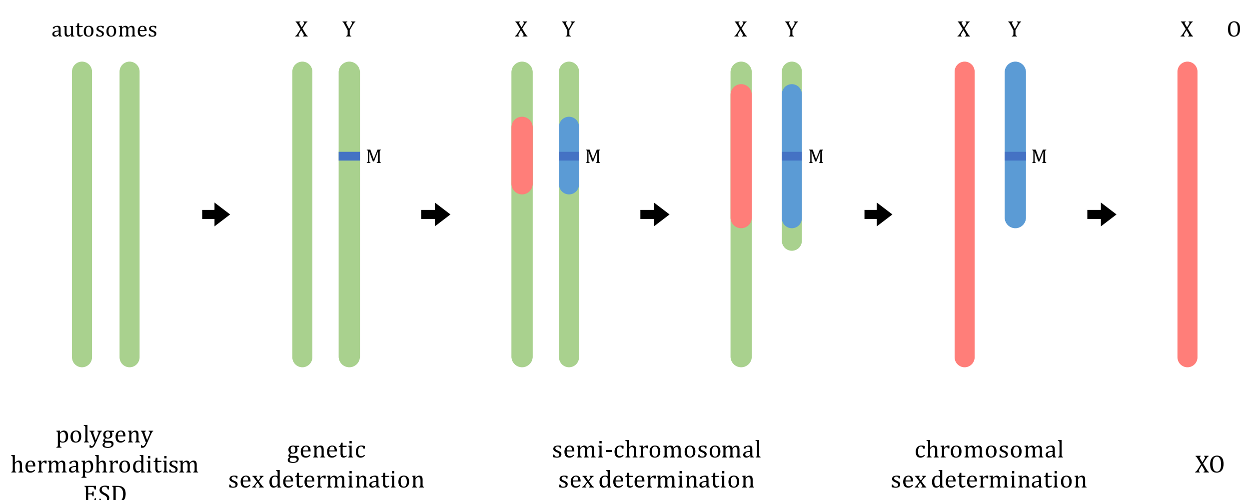


Figure 2.2 – The Y chromosome degeneration hypothesis. The heteromorphic XY sex chromosomes have derived from a pair of autosomes (green) that have acquired a male sex determining locus (M). The M locus restricts recombination at the site, expanding gradually (red and blue) and leading to semi-chromosomal sex determination. Complete inhibition of recombination across the two chromosomes generates two heteromorphic sex chromosomes and chromosomal sex determination. Continuous evolution will eventually lead to a 'genetic erosion' of the Y chromosome and the establishment of an XO sex determining system in some instances. Modified from Steinemann & Steinemann 2005, with permission of the rights holder, Copyright © 2005, Wiley Periodicals, Inc.

Whereas systems like the XY and ZW rely on the presence of individual chromosomes and genes to initiate the development of sex, there are instances where multiple loci *(polygenic sex determination)*, or even the whole genome is responsible for determining sex. In the zebra fish *Danio rerio*, sex is shown to be determined and controlled by multiple regions in the genome (Liew et al. 2012, Bradley et al. 2011) with some of them found to be grouped in one chromosome (Anderson et al. 2012). On the other hand, *haplo-diploidy* and *paternal genome elimination* (PGE) are examples where the *ploidy* is the primary signal for sex development. Haplo-diploidy is found in about 12% of animal species (Bachtrog et al. 2014) and is the only system characterised to date to determine the sex of species of the Hymenoptera class. Bees for example develop as males from unfertilised haploid eggs and females from fertilised diploid eggs (Cook 1993). Finally, PGE is observed in scale insects and is characterised by the selective silencing of the paternal genome in male individuals.

Alongside with the XY and ZW systems, polygenic sex determination, haplo-diploidy and PGE are all examples of *genotypic sex determination* (GSD) in which the genotype of the embryo is solely responsible for the development of sex. In the tree of life, there are instances where external factors are in control of the sexual fate of some species. For example, all crocodiles, most turtles and some fish have developed temperature-dependent sex determination (Ciofi & Swingland 1997, Merchant-Larios & Díaz-Hernández 2013, Quinn et al. 2007) (Figure 2.3). The sex of marine amphipods and some barnacles is shown to be associated with the intensity and period of light during the yearly cycle (photoperiod) (Walker 2005, Guler et al. 2012), whereas social factors like the absence or presence of females affect male development of many coral-dwelling fish and limpets (Warner et al. 1996, Kobayashi et al. 2013). For example, the Bonellidae marine worm family larvae develop as males only if they encounter females, whereas some fish and snails change their sex during their lifetime in response to social cues (Munday et al. 2006). These mechanisms are all examples of *environmental sex determination* (ESD). Exactly how environmental factors like the temperature, light and population size affect the sex of an individual is largely unknown. Nonetheless, a study on the European sea bass *Dicentrarchus labrax*, have shed some light on how temperature affects the methylation of genes that express enzymes that convert androgens into oestrogens (Navarro-Martín et al. 2011).

For some species the separation between GSD and ESD is not definite. Environmental conditions have been shown to affect and alter the sex development of different species that otherwise would have developed into males or females based on their embryonic genetic composition. Usually, these changes occur when the environmental conditions grant some benefits to either of the sexes. For example, some fish like gobies, begin their life as small females and only become males when they are big enough to defend their territory and nesting sites. This way they increase their chances of procreating throughout the

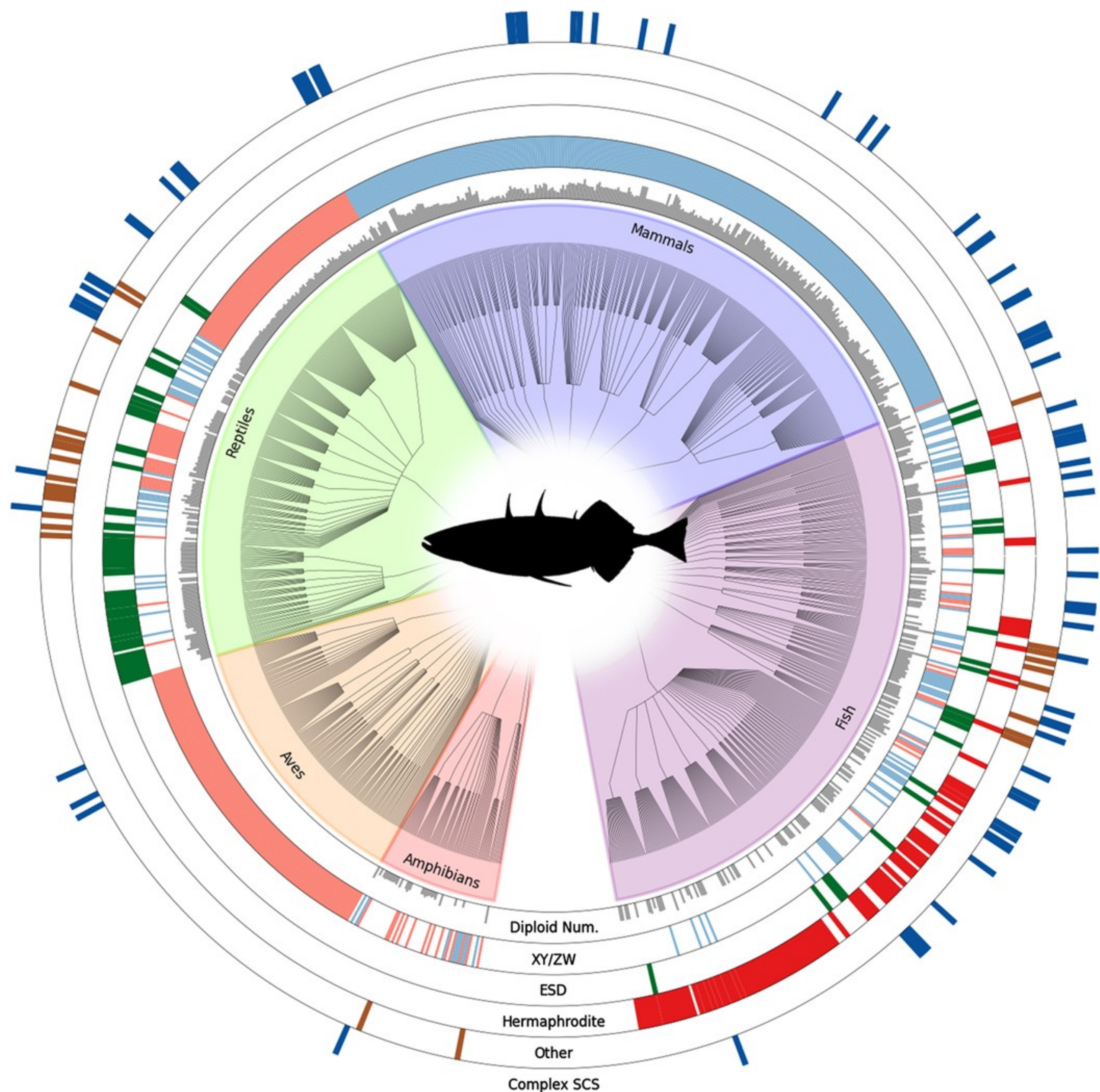


Figure 2.3 – Tree-based representation of sex determining systems found in vertebrates. Each tip represents a genus of species and groups of genera are taxonomically adjusted. The rings show the presence or absence of the respective sex determination system, whereas the height of the bar in the 'Diploid Num.' ring indicates the diploid chromosome number. The blue and red colours in the 'XY/ZW' ring indicate the XY and ZW sex determining system respectively. Minor sex determining systems such as parthenogenesis, gynogenesis and hybridogenesis are grouped in the 'Other' ring and species with complex sex chromosome karyotypes are shown in the 'Complex SCS' ring. The tree was reconstructed using data from 2,145 entries. Image reproduced with permission of the rights holder, Copyright © 2014, Springer Nature, Ashman et al. 2014.

length of their lives and aid their reproductive success (Munday et al. 2006). This *sequential hermaphroditism* is documented in about 2% of all fish, including the famous clown fish and wrasses like the Kobudai. Another example where ESD is altering the sex of individuals is the case of the tonguefish. Tonguefish have adopted a ZW sex determining system, where the females are heteromorphic ZW and males are homomorphic ZZ. It has been shown that changes in the temperature can alter the sex of ZW embryos and force them to develop as males. Like with the European sea bass, these changes have been attributed to modifications of the methylation profile of sex determining genes (Shao et al. 2014).

Some species developed a 'pick and choose' system that allows them to switch between GSD and ESD, based on which is more beneficial for a given situation. Snow skinks for example, have adapted to live within a range of altitudes with climatic extremes. These *viviparous* lizards present different sex determination systems, with GSD occurring at highlands and ESD occurring at lowlands (Pen et al. 2010). This way, in high altitudes where the temperature variance is high, GSD prevents the establishment of extreme sex ratios, which may be detrimental for the populations. On the contrary, at low altitudes where the temperature variance is low, ESD produces optimal sex ratios and also provides a beneficial 'early start' for females (Pen et al. 2010). Although ESD may provide species with adaptations that are beneficial for certain environments, it also poses a great threat to their own existence. Climate change has been altering the average temperatures of our planet, fluctuating rapidly the temperatures of the oceans and creating unpredictable climates weather. Species that fail to adapt to this fast-changing reality may soon face extinction because of skewed sex ratios that push populations towards demographic collapse (Valenzuela et al. 2019). One staggering example is the green sea turtle populations on Raine Island, which is the biggest and most important nesting ground for the species in the Pacific Ocean. In 2018, sand temperatures on the island increased at record levels that the female-to-male sex ratio among the hatchlings was 116 to 1! Combining data from multiple years, scientists predict that the complete feminization of the species in the region is imminent in the near future (Jensen et al. 2018).

Separation of the sexes in two different individuals is not universal in the tree of life. Organisms that are unable to move to find a mate have evolved a different approach of reproducing to ensure the survival of their species. *Hermaphroditism*, the situation where both male and female reproductive organs are located in a single individual is widely adopted within sessile animals and plants. Almost 94% of all flowering plants – or angiosperms – have both the male and female sex structures, stamens and pistils respectively, present in a single individual in different arrangements. Very often, both stamens and pistils are present within the same flower *(simultaneous hermaphrodite)* but sometimes the sex organs are separated in different flowers but on the same plant or tree *(monoecy)*. The situation where both reproductive organs are separate in two different individuals *(dioecy)* is quite rare among angiosperms and only observed in 6% of

all species (Renner & Ricklefs 1995). On the contrary, dioecy is observed in almost 25% of species within gymnosperms. Gymnosperms include seed-producing plants and trees like conifers, that do not produce flowers in order to reproduce. The absence of flowers from this group is an indicator that hermaphrodite structures are quite rare. Indeed, the rest 75% of gymnosperms species have adopted a monoecious way of separating the male and female reproductive organs (Renner & Ricklefs 1995, Bachtrog et al. 2014).

While hermaphroditism is widely adopted in the plant kingdom, the situation observed in the animal kingdom is quite different. Because hermaphroditism is significantly rare from the Arthropod phylum – the largest group of species in the animal kingdom, which also includes insects – hermaphroditism is only encountered in only 5% of all species in the kingdom when taken as a whole. Invertebrates like corals, snails, barnacles, nematodes, trematodes, echinoderms, but also some fish, use this way of reproducing when situations make it difficult to find a mate (Jarne & Auld 2006, Eppley & Jesson 2008). Although hermaphroditism and self-reproduction is a solution for sessile animals like corals, the costs associated with self-fertilisation outweigh the assurance of reproduction in the long term. The continuous production of low fitness progeny through self-fertilisation is a downside for species and an evolutionary unstable method to sustain a population (Charlesworth & Charlesworth 1978, Charlesworth & Willis 2009). The cost of hermaphroditism becomes obvious when considered that plants and animals were driven to separate the sexes independently many times (Bachtrog et al. 2014). Two strategies have been adopted widely by different species in order to minimise self-fertilisation. Spatial separation of the sexes in an individual is one way of reducing self-reproduction and is observed mostly in plants (monoecy) and the majority of hermaphrodite animals. Separation of the sexes in time is also another method to ensure that the maturation of male and female *gonads* do not overlap and avoid self-fertilisation. For example, the stamens and pistils of some hermaphrodite plants develop in different periods during the season *(dichogamy)* (Bertin & Newman 1993), whereas animals use sequential hermaphroditism to avoid self-reproduction (Munday et al. 2006).

Polygenic sex determination and hermaphroditism are tightly linked since species that have recently separated the male and female sex into different individuals are more likely to have developed multiple ways of suppressing or enhancing male or female development (Bachtrog et al. 2014). This situation is well observed in dioecious plants, in which no master sex determining gene has been discovered and characterised to date (Diggle et al. 2011). In rice for example, over 227 genes that promote male sterility have been identified on all 12 chromosomes of the species. In theory, any of these 12 autosomes can evolve a sex determining character and differentiate into a female chromosome in the future (Cui et al. 2012, Bachtrog et al. 2014). Loss of function of genes that determined the development of male or female gonads in hermaphrodites can promote situations in which the breeding system is either dominated by

males and hermaphrodites *(androdioecy)* or females and hermaphrodites *(gynodioecy)*. These two situations are believed to have preceded the complete separation of the sexes in different species (Muyle et al. 2017, Ming et al. 2011). In papaya for example, the XY sex determining system is believed to have developed from a series of events that first created the conditions for a gynodioecy system before evolving to a male heterogametic XY system (Bachtrog et al. 2014, Aryal & Ming 2014, Wang et al. 2012). Examples like these, but also from other species within the hermaphrodite-rich plant kingdom, provide very good insights on the evolution of the sex determining systems and mechanisms in the tree of life (Charlesworth 2015).

2.2 Sex determination in insects

Insects or Insecta is an extensive class of hexapod invertebrates that contains an estimated 2 million species (Mora et al. 2011). They form the largest group within the Arthropoda phylum, which contains familiar forms such as crabs, spiders, mites, lobsters, centipedes and millipedes. It is believed that the Arthropoda phylum alone contains more than 50% of all species within eukaryotes, and that insects alone account for more than 80% of species within the animal kingdom (Mora et al. 2011). The Insect class, which belongs within the Hexapoda subphylum, is divided in 28 orders with the majority of the species to occupy six major orders (Misof et al. 2014) (Figure 2.4). Coleoptera (beetles) is by far the largest, with more that 350,000 described species. Hymenoptera (ants, wasps and bees) consist of more than 160,000 species, whereas Diptera (flies) and Lepidoptera (butterflies and moths) consist of at least 150,000 species each. Hemiptera, which includes true bugs like coccids, cicadas, aphids and shield bugs have over 100,000 described species, whereas Orthoptera (grasshoppers, locusts, crickets and katydids) have approximately 24,000 species. Smaller orders within the class contain some hundreds to a few thousand species, like Trichoptera (caddisflies), Blattodea (cockroaches and termites) and Dermaptera (earwigs), which consist of 15,000, 7,500 and 2,000 species respectively (Gullan & Cranston 2010). For pure comparison of the magnitude and diversity of insect species, there are almost as many species of birds as in the 'small' Trichoptera order (Gullan & Cranston 2010).

Insects have adopted a plethora of sex determining strategies found throughout the tree of life (Bachtrog et al. 2014) (Figure 2.5). The vast majority of insect species reproduce sexually and have adopted mainly a genotypic sex determination system (GSD) that relies on the genotype of the *zygote* (Normark 2003). Common sex determination practices encountered frequently among insects include male or female heterogamety, paternal genome elimination (PGE) and haplo-diploidy (Blackmon et al. 2017). Environmental sex determination (ESD) is not encountered frequently among insects and it has only been described so

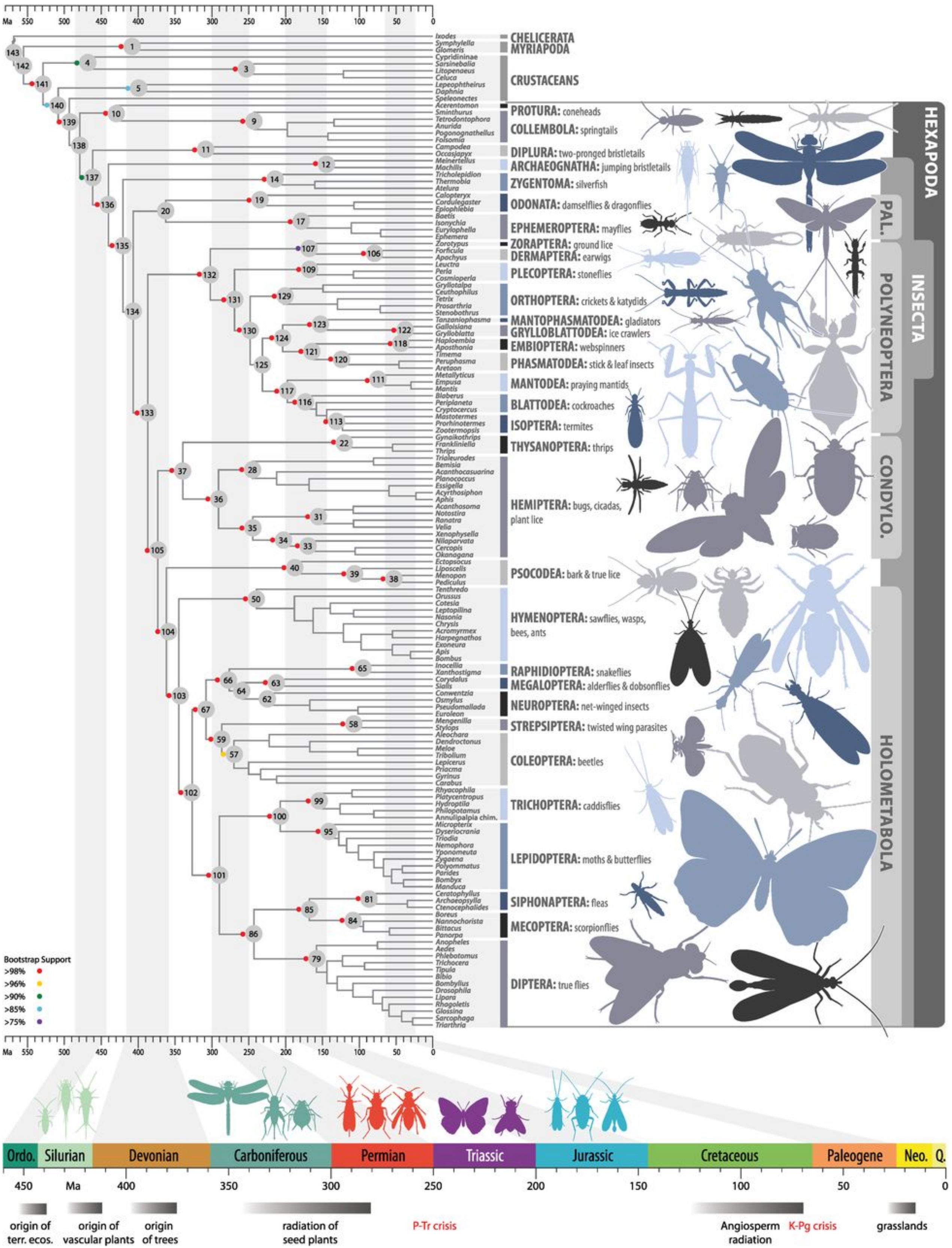


Figure 2.4 – Dated phylogenetic tree of the Hexapoda subphylum. The Insecta class includes 4 major Infraclasses: Palaeoptera, Polyneoptera, Condylognatha and Holometabola. The timeline at the bottom indicates the geological origin of the class and shows major geological and biological events. From Misof et al. 2014. Reprinted with permission from AAAS, Copyright © 2014, American Association for the Advancement of Science.

far in a single *Sciara* fly species (Nigro et al. 2007). Other rare forms of sex determination like *monogeny, parthenogenesis* and *cytoplasmic sex determination* (CSD) are also encountered within the class but usually in a very small number of species. Finally, hermaphroditism is not encountered among insects but is present in a number of non-insect species in the Arthropoda phylum, like Crustaceans (Jarne & Auld 2006). This plethora of sex determining mechanisms adopted by different insect species have been providing important insights into the development of sex and how evolution has adapted these systems over time.

2.2.1 The hourglass model of evolution

While simple cytogenetic techniques have revealed the sex determining mechanisms in many insect species, the specific genes that act as 'switches' to initiate the process for sex development have only been analysed in a handful of species. Characterisation of these genes has revealed a great diversity that underlines the primary sex determining signals within the class even within closely related taxa (reviewed in Gempe & Beye 2011). The activation of these signals initiates a cascade of reactions that involve downstream transcription factors that are alternatively spliced in male or female isoforms that initiate the male or female sex differentiation. In contrast with the diversity of the primary sex determining signals, the factors and genes that lie downstream and at the bottom of the sex determination cascade are surprisingly highly conserved within the insect class. In fact, conservation between species seems to increase dramatically towards the bottom of the sex determination cascade. For example,the gene *transformer (tra)* that lies downstream of the primary sex determining signals is present in the majority of insect species studied to date, whereas genes that lie downstream of tra, like *doublesex (dsx)* and *fruitless (fru)* have shown not only to be present but also highly conserved where they have been characterised. Both *dsx* and *fru* are responsible for the initiation of the sex differentiation in insects. The transcription factors encoded by these two genes bind various targets in the genome and activate cascades of processes that differentiate the male and female sex. The conservation of genes involved in the sex determination and differentiation in insects can be accurately described using the hourglass model of evolution (Figure 2.6). Both *dsx* and *fru* lie at the narrow neck of the hourglass indicating their high conservation among different species. They are the link between the upstream sex determination cascade and the downstream sex differentiation cascade. Genes that lie upstream and downstream of *dsx* and *fru* gradually become more diverse and start to differentiate among taxa.

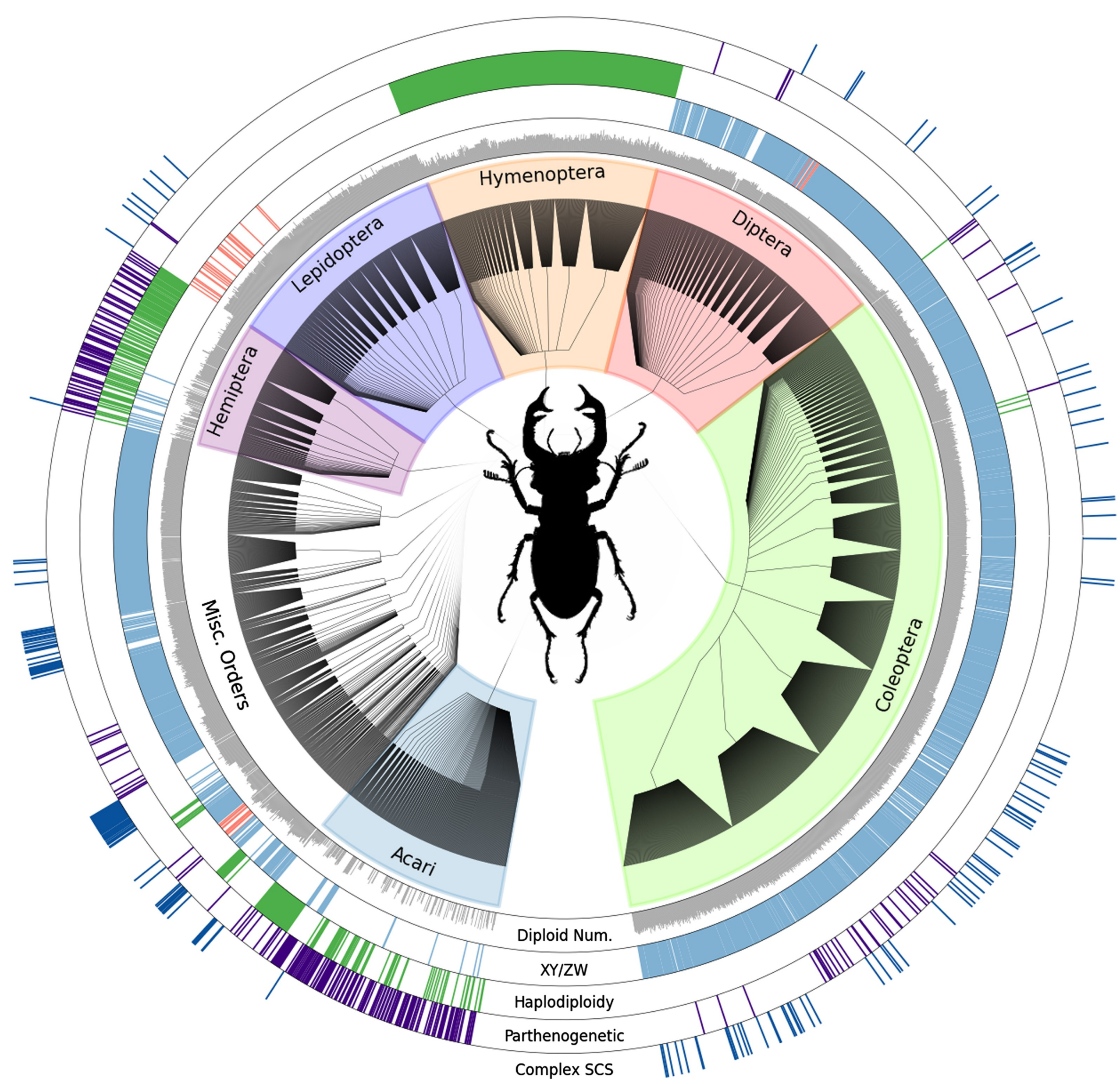


Figure 2.5 – Tree-based representation of sex determining systems found in invertebrates. As with Figure 2.3, each tip represents a genus of species, which are taxonomically aligned. The rings show the presence or absence of the respective sex determination system, whereas the height of the bar in the 'Diploid Num.' ring indicates the diploid chromosome number. The blue and red colours in the 'XY/ZW' ring indicate the XY and ZW sex determining system respectively. Species with complex sex chromosome karyotypes are shown in the 'Complex SCS' ring. The tree was reconstructed using data from 11,556 entries. Image reproduced with permission of the rights holder, Copyright © 2014, Springer Nature, Ashman et al. 2014.

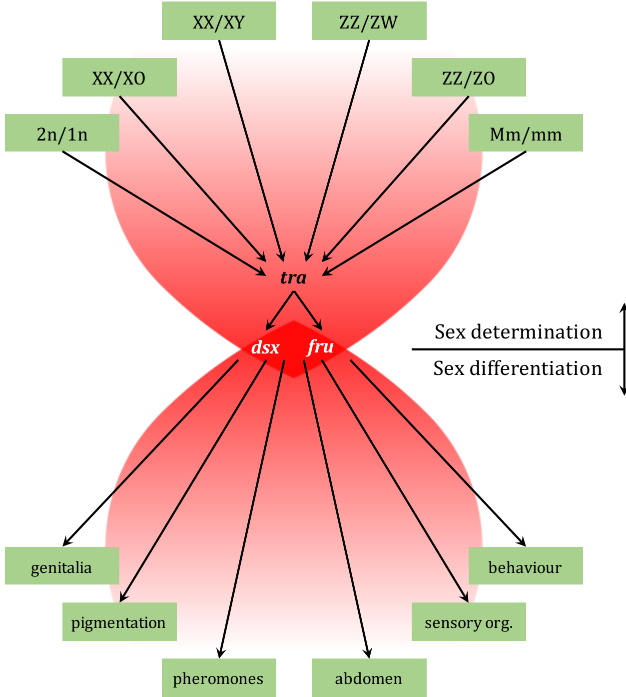


Figure 2.6 – The hourglass model of evolution outlining the sex development in insects. The sex determining signals at the top of the sex determination cascade are diverse among the class. Various gene 'switches' initiate a cascade of reactions that in the majority of species result in the activation of the *transformer (tra)* gene. The gene regulates the alternative splicing of *doublesex (dsx)* and *fruitless (fru)* that lie at the bottom of the sex determination cascade. In contrast with *tra*, *fru* but mostly *dsx*, has been found in all insects that have been characterised to date and is highly conserved among different taxa. The two genes lie at the neck of the hourglass, indicating the lack of diversity at that point in the cascade. Both genes initiate sex differentiation by regulating various downstream targets that affect the development of male and female sex characteristics and processes.

2.2.2 Genotypic sex determination

GSD is one of the most frequent and familiar sex determining systems in the tree of life and is encountered in many clades among eukaryotes. Insect species have widely adopted different types of GSD and the majority are using versions of the familiar male heterogametic XY or female heterogametic ZW systems. Heterogametic systems can predict the sex of an individual but the mechanisms that actively specify the sex differ among species. For example, the presence of a male-dominant sex chromosome like the Y is enough to specify the male sex in species of the Coleoptera order like *Tribolium castaneum* and some genera of the Diptera family Tephritidae like *Ceratitis, Bactrocer* and *Anastrepha* (Sánchez 2008). In these species, the Y chromosome harbours the male determining factors (M-factors) that initiate the development of the male sex, whereas in their absence, the female sex is established. A similar situation is encountered in members

of the Muscidea family. The common house fly, *Musca domestica* uses an XY system to establish the male and female sex but in some populations the male determining M-factors are located on autosomes and not on the Y chromosome. This situation can produce viable males that do not carry a Y chromosome but rather are homozygous for the X chromosome (Dübendorfer et al. 2002). Some populations of the species have even adopted a dominant female-determiner gene that establishes feminisation, even in the presence of the M-factors or the Y chromosome (Hediger et al. 2010). The male determining character of the Y chromosome has also been lost from members of the Drosophilidae family. In *Drosophila melanogaster*, the Y chromosome does not actively specify the male sex. In the family, sex is determined by the number of X chromosomes and more specifically the quantity of X-linked signal elements (XSE) present at certain embryonic developmental nuclear divisions (Erickson & Quintero 2007) (see below).

In some orders, the degeneration of the Y and W sex chromosomes have resulted in a XO or ZO sex determining system. For example, all members of the Blattodea order, which includes cockroaches and termites, and some members of the Orthoptera order, which includes grasshoppers, locusts, crickets and katydids use these systems. Sex in these orders is determined by either the X- or Z-to-autosome ratio or just the numbers of X or Z chromosomes present in the embryo (Castillo et al. 2010, Blackmon et al. 2017). Other orders like Lepidoptera and its sister order Trichoptera, have switched to a female heterogametic system altogether. Sex in butterflies and moths is determined by the presence or absence of the W female-specific sex chromosome. The presence of the W chromosome initiates female development of the embryo whereas in its absence, male sex is established (Sahara et al. 2012). In the silkworm *Bombyx mori*, the sex determining signal has been identified as a Piwi-interacting RNA molecule (piRNA) encoded from the *feminiser* gene located on the female-specific W chromosome (Kiuchi et al. 2014). The piRNA has been shown to bind and neutralise male-specific factors, leading to the feminisation of the embryo. Caddisflies on the other hand have abolished the W sex-specific chromosome and have adopted a ZO sex determining system (Lukhtanov 2000, Marec & Novák 1998). Transitions to different sex determining systems like these are not encountered frequently in insects. By combining data from 359 families, it is clear that male heterogamety is the ancestral mode of sex development within the class. Over 77% of investigated species use a version of male heterogamety, supporting this hypothesis (Blackmon et al. 2017).

2.2.3 Sex determination in *Drosophila melanogaster*

The model species *Drosophila melanogaster* provides an excellent opportunity to study and understand the sex determination pathway in insects. Sex determination in the species has been under vigorous examination and the signals and factors that make up the sex determination cascade have been thoroughly examined. Although both *D. melanogaster* and *An. gambiae* shared a common ancestor 260 million years

ago (Kumar et al. 2017), the high conservation that characterise the sex determination pathway allows us to still use the species as a reference point to study the development of sex in mosquitoes.

As mentioned before, although the species has adopted a male heterogametic XY sex determining system, the Y chromosome does not actively specify the male sex. In the species, male and female sex is established by the concentration of products that are encoded from genes located on the X chromosome (X-linked signal elements - XSEs). Female development is initiated when the concentration of these elements surpasses a threshold when the embryo initiates *zygotic transcription* and the beginning of *cellularization* (Erickson & Quintero 2007). Because females carry two X chromosomes, expression from both chromosomes ensures the quantity of the XSEs is high enough to trigger female development. On the contrary, the presence of only one X chromosome in males is insufficient to produce the quantities needed for female development, and thus the male sex is established.

There are four well characterised XSEs found in *D. melanogaster* that act as sex determining signals: *sisterless A, sisterless B* – also known as *scute, sisterless C* and *runt.* The products of these genes act directly or indirectly on the master regulatory switch that lies downstream in the sex determination cascade, the *sex-lethal* gene *(Sxl). sisterless A, sisterless B* and *runt* encode for transcription factors that regulate directly the expression of the *Sxl* gene (Cline 1988, Duffy & Gergen 1991), whereas *sisterless C* encodes for a *Sxl* transcription factor ligand-activator (Sefton et al. 2000).

Sxl in the species is regulated by two promoters that are sex-specifically active in different developmental stages. A female-specific 'early' or 'establishment' promoter *(Pe)* is expressed in the early female embryo, whereas a 'maintenance' promoter *(Pm)* is expressed in both sexes in all developmental stages from the embryo till adulthood (Bopp et al. 1991, Salz et al. 1989). The transcripts encoded from the *Pm* promoter are alternatively spliced in both males and females to produce a female-specific active SXL_L protein and a truncated inactive male-specific version (SXL_M). The difference between SXL_L and SXL_M is the presence of stop codons in an early exon that is spliced out in females (Bell et al. 1988, Samuels et al. 1991, Bopp et al. 1991). Protein products encoded from the *Pe* promoter (SXL_E) act as splicing regulators that enhance the removal of these exons from the female transcripts. Because both *Pe* and *Pm* promoters are sensitive to upstream XSE transcription factors, the production of SXL_E and SXL_L is only limited to the female sex, whereas males produce only the truncated version (Erickson & Quintero 2007, Cline 1993).

In females, the continuous expression of SXL_L protein is maintained throughout the life of the insect. Because expression from the *Pe* promoter diminishes gradually during development, female-specific alternative splicing of the gene is maintained by the SXL_L protein itself (Bell et al. 1991). Once established, SXL_L acts on the pre-mRNA of the gene, enhancing the removal of the male-specific exons. This way, the gene is locked in an autoregulatory loop that ensures the continuous production of SXL_L and the

maintenance of the female sex determining signal. In male flies, the absence of active SXL_L ensures the splicing of the gene to its male-specific default state and the production of inactive SXL_M. In the sex determination cascade, *Sxl* acts on the downstream *transformer* gene *(tra)* that is also alternatively spliced in a female active and a male inactive form. Active SXL_L protein in females shifts the production towards female-specific transcripts, whereas in males the gene is alternatively spliced to its default inactive form (see below). The *Sxl* gene has been characterised in a number of drosophilid species like *D. virilis* (Bopp et al. 1996) and *D. pseudoobscura* (Penalva et al. 1996). In both species the gene is alternatively spliced to a female active and a male inactive form, mimicking the situation in *D. melanogaster*. The gene has also been characterised in many insect species outside the Drosophilidae family and it was shown that is not regulated in a sex-specific way and involved in the the sex determination pathway (Müller-Holtkamp 1995, Sievert et al. 2000, Meise et al. 1998, Saccone et al. 1998, Lagos et al. 2005). Somewhere during the evolution of the *Drosophila*, the gene has been selected to become the master sex determining regulatory switch of the family. This is a great example of the divergence that characterise the upstream signals of the sex determination pathway in insects.

2.2.4 Haplo-diploidy and Paternal genome elimination

Haplo-diploidy is widely encountered within the Hymenoptera order, which includes bees, wasps, sawflies and ants (Normark 2003). The mechanisms and signals behind sex determination in haplo-diploidy are linked to the complementary sex determiner locus and the *complementary sex determiner* gene *(csd)*, which has been thoroughly characterised in the honeybee *Apis melifera* (Beye et al. 2003). In honeybees, the gene exists in different isoforms due to the presence of a highly variable region within the locus. Females are heterozygous for different *csd* isoforms whereas males are hemizygous since they are haploid. Diploid males that are homozygous for the locus exist but are sterile. The gene is transcribed in both sexes and produces complexes that are only active if they consist of different *csd* isoforms (Beye 2004). Active complexes of the CSD protein, together with the RNA-binding TRA2 protein (Nissen et al. 2012) promote female sex determination by enhancing female-specific splicing of the *feminiser (fem)* gene, which is a *tra* orthologue (Hasselmann et al. 2008, Gempe et al. 2009) (see below).

The haplo-diploidy sex determining rules of the *csd* gene do not apply in all species within the Hymenoptera order. In the parasitic wasp *Nasonia vitripennis*, haplo-diploidy also determines the sex of the embryo, but the *csd* gene is not actively involved in the sex determination cascade. In the species, maternal imprinting and silencing of genes is rather responsible for determining the sex of the embryo (Verhulst et al. 2010). As in other species, female sex is established after activation of the *tra* gene in the zygote (see below). In *Nasonia*, the maternally inherited *tra* allele is silenced by imprinting factors and

cannot be expressed zygotically. On the contrary, the male-inherited tra allele is not imprinted and is actively expressed in the zygote. As a consequence, diploid eggs that carry both a maternal and a paternal *tra* allele develop as females, whereas haploid eggs that carry only the imprinted version of the allele develop as males (Beukeboom et al. 2007, Verhulst et al. 2010).

Another form of haplo-diploidy, the paternal genome elimination (PGE) ensures the development of males in scale insects (coccids) of the Hemiptera order (reviewed in Sánchez 2008) (Figure 2.7). In coccids, two PGE mechanisms have been proposed, the *lecanoid* and *diaspidoid* system. In both systems males and females start their lives from diploid fertilised eggs. The lecanoid system implies that the paternal genome becomes heterochromatic in embryos destined to be males. This way both males and females are structurally diploid, but males are functionally haploid. The diaspidoid mechanism on the other hand implies that the paternal genome is eventually eliminated, resulting in structurally and functionally haploid males. In both the lecanoid and diaspidoid system, the developmental stages where PGE happens varies and is different among species.

2.2.5 Environmental sex determination, monogeny and parthenogenesis

Environmental sex determination (ESD), monogeny and parthenogenesis are also sex determining mechanisms that are encountered within insects but at very low frequencies (Bachtrog et al. 2014, Sánchez 2008). *Sciara* flies have adopted a PGE-ESD sex determination system to establish the sex of the embryo. Both male and female *Sciara* flies start their lives with one set of maternally derived chromosomes and two sets of paternally derived chromosomes. Males eliminate both the paternally derived sets of chromosomes, whereas females eliminate only one. Fluctuations in temperature can shift the production of more females by increasing the numbers of embryos that eliminate only one of the paternally derived sets of chromosomes. This sex distorted ratio is not achieved through killing of the male embryos but rather through their transformation to females by a temperature-sensitive maternal effect that regulates chromosome elimination and silencing (Nigro et al. 2007).

Monogeny is encountered in some flies and some crustaceans and it describes the situation where all offspring of a particular female are either exclusively male or female. In the Hessian fly, *Mayetiola destructor*, a specific gene called *Chromosome maintenance (Cm)* was associated with this behaviour. The arrangement of three *Cm* alleles in the species was shown to be responsible for the production of *thelygenous* females – females that produce only female offspring, or *arrhenogenous* females – females that produced only males. This behaviour has been observed so far in the Cecidomyiidae family but also within the Sciaridae family (Metz 1938, Stuart & Hatchett 1991). Another mode of reproduction, parthenogenesis, has also been reported in some insect species. Parthenogenesis describes the development of a

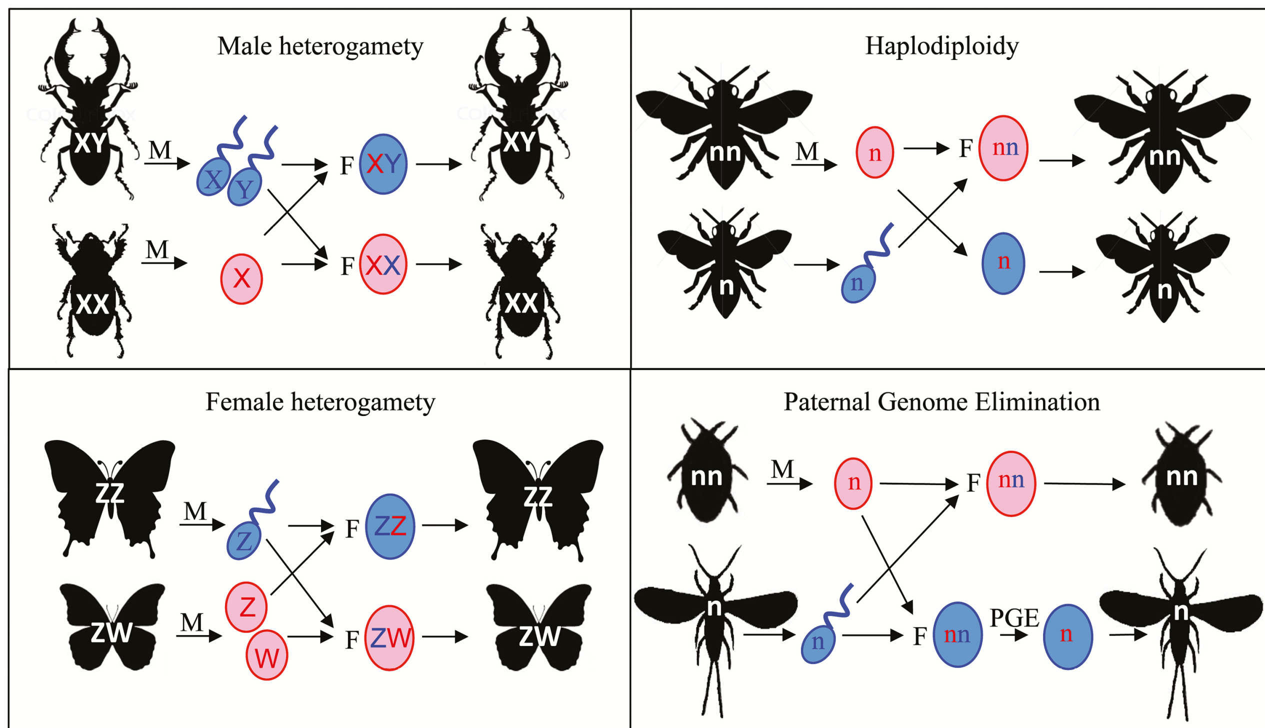


Figure 2.7 – Common sex determination systems in insects. Male heterogamety, female heterogamety, haplo-diploidy and paternal genome elimination (PGE) are four of the main sex determining mechanisms observed in insects. In male and female heterogametic systems, the male and female individuals respectively dictate the sex of the embryo. In haplo-diploidy, diploid eggs develop into females whereas unfertilised eggs develop into males. In PGE, the paternal genome is either silenced or eliminated in embryos destined to be males. Blackmon et al., Sex Determination, Sex Chromosomes, and Karyotype Evolution in Insects, Journal of Heredity, 2017, **108**(1), 78-93, by permission of Oxford University Press. Copyright © 2016, Oxford University Press (Blackmon et al. 2017).

female embryo from unfertilised diploid eggs. This way of reproduction has evolved independently many times within the insect class and is found in almost all orders of insects. Nevertheless, parthenogenesis is always found in less than 1% of insect species, supporting the hypothesis that asexual reproduction is not evolutionary stable and is 'short-lived' (Blackmon et al. 2017).

Insects have developed an array of sex determining mechanisms in order to establish sex. In all cases examined to date, the sex determination pathway follows a common theme that is in line with the hourglass model of evolution. Upstream factors are highly diverse and almost never the same between different taxa, whereas genes that lie downstream on the pathway share more similarities between different species. Three genes that lie at the bottom of the sex determination cascade: *transformer*, *fruitless* and *doublesex* are example of conserved factors that have a significant role in sex development of insects and therefore might be broadly applicable as targets for gene drive systems targeting sex-specific fertility or viability for population control.

2.3 Conserved genes in insect sex determination

2.3.1 The *transformer* gene

The *transformer* gene *(tra)* lies downstream of the primary sex determining signals and it has been characterised in many species within Coleoptera, Hymenoptera, and members of the Drosophilidae and Tephritidae families (reviewed in Sánchez 2008). In a generalised model of insect sex determination, the gene directs female development by regulating the alternative splicing of downstream genes such as *dsx* and *fru*. Although the gene appears to be present in many species, indicating an ancestral mode of evolution, *tra* is much less conserved than both *dsx* and *fru* (Geuverink & Beukeboom 2014). In some species, the sequence of the gene has diverged beyond recognition but it has retained its ability to regulate downstream genes on the sex determination pathway (Pane et al. 2002). In species like *Ceratitis capitata* and *Apis mellifera* for example, homologues of the gene were found based on their structure and ability to bind downstream effectors. *tra* homologues have never been found in mosquito species (Geuverink & Beukeboom 2014), but the presence of TRA-like binding sites in both *dsx* and *fru* indicate the presence of *tra*-like factors that might have diverged beyond recognition and therefore are undetectable by simple homology-based searches (Scali et al. 2005, Gailey et al. 2006, Price et al. 2015a, Salvemini et al. 2011, 2013). In other orders such as Lepidoptera and Strepsiptera, the gene appears to be completely absent. In these orders, factors that regulate the alternative splicing of *dsx* appear to be unrelated to *tra* (Mita et al. 2004, Suzuki et al. 2010). Outside of the Insect class, *tra* has been characterised in *Daphnia magna*, a small crustacean of the Arthropoda phylum, where the gene appears to not be involved in the sex determination pathway (Kato et al. 2010).

In species where the gene has been characterised, *tra* is ubiquitously expressed in both males and females and produces two sex-specific isoforms through alternative splicing. *tra* is predominantly active only in females because the male-specific isoform of the gene contains early exons that are equipped with in-frame stop codons that result in truncated inactive products. In females of some species, consecutive expression of active *tra* is achieved through an autoregulatory loop that ensures the female mode of development is always "on". *tra* products have been shown to bind the *tra* pre-mRNA and promote the production of female-specific isoforms (reviewed in Bopp et al. 2014). In males, the absence of active *tra* ensures the splicing of the pre-mRNA to its 'default' inactive form. Such a positive feedback loop amplifies the upstream sex determining signal and locks the system in a female- or male-specific development.

In *Drosophila melanogaster*, where the system has been thoroughly characterised, the gene is alternatively spliced by the upstream sex regulator *Sex-lethal (Sxl)* (Valcarcel et al. 1993). Binding of the female-specific SXLL on the *tra* pre-mRNA ensures the exclusion of the male-specific exon and the production

of active TRA protein. In the absence of SXL_L, the pre-mRNA of *tra* is alternatively spliced to its default inactive state that includes the male-specific exon (Boggs et al. 1987, Belote et al. 1989). The active TRA protein in females complexes with TRA2, which unlike TRA, is ubiquitously expressed in both sexes. The TRA/TRA2 complex interacts with the splicing machinery and help include or exclude sex-specific exons in downstream alternatively spliced genes like *dsx* and *fru* (Sánchez 2008, Bopp et al. 2014). The complex recognises splice acceptor sites flanking the targeted exons and assembles different factors at the locus that have enhancing or inhibiting splicing properties (Ruiz et al. 2007). A well characterised factor that pairs with the TRA/TRA2 complex is RbpI, which is essential for the recognition of these splice sites by the splicing machinery and the inclusion of the respective exons in the final mRNA (Lynch & Maniatis 1996).

In other species where *Sxl* is absent or not associated with sex development, maternally deposited TRA in the early embryos have been shown to initiate the autoregulatory loop that is essential to maintain *tra* expression. In species like *Ceratitis capitata* (Pane et al. 2002), *Musca domestica* (Hediger et al. 2010), *Bactrocera oleae* (Lagos et al. 2007), *Anastrepha obliqua* (Ruiz et al. 2007) and *Tribolium castaneum* (Shukla & Palli 2012*b*), maternally deposited TRA protein is present in both male and female early embryos. In female embryos, the maternal protein initiates the autoregulatory loop that locks the embryo to its female state. This situation is averted in male embryos by male determining signals. These signals disrupt directly or indirectly the formation of the TRA/TRA2 complex and avoid the activation of the autoregulatory loop. As a result, *tra* is spliced on the default state and the male sex is established (Bopp et al. 2014). It is believed that the initiation of the *tra* autoregulatory loop by maternally deposited elements may reflect the ancestral state of sex establishment in insects, given its presence in different orders (Bopp et al. 2014). Indeed, evolution of female determining signals upstream of *tra*, like the *Sxl* in *D. melanogaster*, have been characterised only within Drosophilidae and the honeybee *Apis melifera*.

2.3.2 The *fruitless* gene

The *fruitless* gene *(fru)* has been characterised in only a handful of species but has been thoroughly examined in the model species *Drosophila melanogaster* (reviewed in Billeter et al. 2006). The gene lies downstream of *tra* in the sex determination pathway and alongside with *doublesex* promotes the sex differentiation processes mainly of neuronal tissues. The gene has 4 promoters (P1-P4) that produce 18 different groups of transcripts through alternative splicing, making it one of the most complex genes in the genome. All transcripts of the gene encode proteins that contain a BTB *(bric-a-brac, tramtrack, broad-complex)* dimerization domain on the N-terminus and a zinc-finger DNA-binding domain on the C-terminus (Billeter et al. 2006a, Ryner et al. 1996, Dalton et al. 2013). Transcripts of the gene are alternatively spliced at their 3'-end between 5 exons that encode for different zinc-finger DNA-binding domains

named A to E (Usui-Aoki et al. 2000). Because of the DNA-binding activity of the gene, it has been suggested that *fru* is a transcription factor and associated with chromatin modifying proteins (Ito et al. 2012), something evaluated *in vitro* (Dalton et al. 2013) and *in vivo* (Neville et al. 2014).

Fruitless is mainly associated with the sex-specific differentiation of the male central nervous system (CNS). Transcripts of the gene that are transcribed from the P1 promoter include an exon with early stop codons that result in immature protein products in females (Usui-Aoki et al. 2000). The exon is present in both males and females and equipped with TRA/TRA2 binding sites that result in the inclusion of the exon in females where the TRA/TRA2 complex is present. Because the exon lies upstream of the BTB and zinc-finger loci of the gene, it ensures the production of incomplete *fru* protein products from this promoter in females. The absence of the TRA/TRA2 complex from males, ensures the production of mature male-specific isoforms that are expressed in 2-5% of all CNS neurons and have been shown to be involved in male courtship behaviours (Lee et al. 2000, Demir & Dickson 2005, Manoli et al. 2005, Kimura et al. 2005). Transcripts form P2-P4 promoters are not sex-specifically expressed and are present in both males and females. These transcripts are expressed in neuronal and non-neuronal tissues from early developmental stages and code for proteins that have vital functions in both sexes (Goodwin et al. 2000, Billeter et al. 2006).

2.3.3 The *doublesex* gene

2.3.3.1 Discovery of the *Dmrt* family

The *dsx* gene was first identified in 1965 in *D. melanogaster* through a recessive mutation that caused male and female individuals to develop into intersex flies that shared an array of external and internal phenotype anomalies (Hildreth 1965). Molecular characterisation of the *dsx* locus revealed later on that the gene is involved in the sex development of the species, and that it promotes sex differentiation through the production of sex-specific proteins that result from alternative splicing of the same gene (Burtis & Baker 1989, Baker & Wolfner 1988). The discovery and characterisation of *dsx* and also *male abnormal-3 (mab-3)* gene in *Caenorhabditis elegans* decades later(Shen & Hodgkin 1988), lead to the identification of a family of genes that appear to have a common evolutionary origin (Erdman & Burtis 1993, Raymond et al. 1998, Zhu et al. 2000). These *dsx/mab-3*-related genes (*Dmrt* genes) carry a specific DNA-binding domain with an atypical zinc-finger $C_2H_2C_4$ Cys-His configuration that binds the DNA minor groove (Erdman & Burtis 1993, Zhu et al. 2000). This domain, called DM domain or *Dmrt* domain, has been identified in all vertebrate and invertebrate genomes studied to date, including humans *(DMRT1)* (Raymond et al. 1999) and both mice and chickens *(Dmrt1)* (Raymond et al. 1999a). The *Dmrt* domain is shared across many

species in the animal kingdom and it is always associated with genes that are involved directly or indirectly in many developmental processes, but most importantly in the sex-specific differentiation of the species (Hong et al. 2007, Kopp 2012).

2.3.3.2 Gene structure

Like with *fru*, the *dsx* gene lies at the bottom of the sex determination cascade and it is alternatively spliced to produce a male-specific (DSX^M) and a female-specific (DSX^F) isoform (Figure 2.8). The gene and all its orthologs studied to date code for transcription factors that have a similar structure that includes two oligomerisation domains(Verhulst & van de Zande 2015). The first oligomerisation domain (OD1) lies at the N-terminus of the gene and it contains the *Dmrt* DNA-binding domain, which is adjacent to an α-helix dimerization domain. The second oligomerisation domain (OD2) is located at the C-terminus of the protein and it is followed by one sex-specific domain that is either male-specific (OD^M) or female-specific (OD^F). Although both OD1 and OD2 domains are sex-independent, the OD2 domain is not shared across the Dmrt family but rather is specific to *dsx*(Erdman et al. 1996, An et al. 1996). The sex specificity of the *dsx* gene comes from both the OD^M and OD^F domains that are present only on the male- and female-specific isoform respectively. Experiments performed on the *dsx* gene of *D. melanogaster* have shown that both the DNA-binding domain and the dimerization domain are required for DNA-binding activity(An et al. 1996), whereas OD2, OD^M and OD^F have no role in DNA-binding activity(Burtis et al. 1991, Cho & Wensink 1997). The fact that the sex specific domains of the protein do not regulate binding to the DNA suggests that both DSX^M and DSX^F bind the same target sequences within the genome, and that sex specific differentiation is achieved via interaction of both OD^M and OD^F with the different components of the transcriptional machinery of the cell (An et al. 1996, Verhulst & van de Zande 2015).

2.3.3.3 *doublesex* conservation

The *Dmrt* gene family belongs in a group of highly conserved factors that are found almost identical among different species. The conservation of the *Dmrt* gene family throughout the animal kingdom was validated when the DSX^M version of *D. melanogaster* was able to rescue *mab-3* mutant phenotypes in *C. elegans* and restore the male function of the gene when ectopically expressed (Raymond et al. 1998), despite the 740 million years that diverge the two species (Kumar et al. 2017). Within insects, the *dsx* gene family has been thoroughly examined for its conservation among different taxa spread among different orders. A major study that analysed expression sequence tags (ESTs) from the 32 orders of Hexapoda revealed a *dsx* gene that appears to be ubiquitous in the subphylum (Price et al. 2015). Orthologs of the gene were recovered from 30 orders with the majority of them showing evidence for alternatively spliced tran-

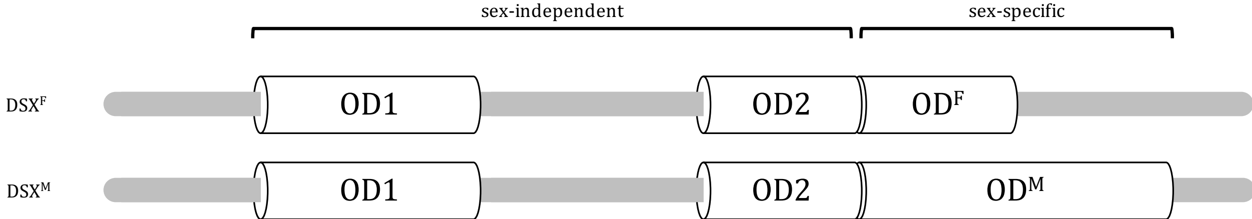


Figure 2.8 – The DSX protein isoforms. The *dsx* gene produces two sex-specific isoforms in all insect species studied to date. The alternative splicing pattern and domain structure of the male-specific (DSX^M) and female-specific (DSX^F) isoforms are highly conserved among taxa. Both DSX^M and DSX^F carry a common sex-independent section that consists of the first and second oligomerisation domains (OD1 and OD2). OD1 is located at the N-terminus of the gene and includes the ultra-conserved *Dmrt* DNA-binding domain of the *Dmrt* gene family and a second α-helix dimerization domain. The OD2 domain is located on the C-terminus of the gene and it is immediately followed by the sex-specific OD^M and OD^F domains. Both these domains are the result of alternative splicing of the *dsx* gene that is controlled by upstream factors like the female-specific *tra*.

scripts that hint to a common ancestor with such characteristics. This hypothesis is strongly supported by the fact that alternative splicing of the gene seems to be absent from the water flea *Daphnia magna* and the western predatory mite *Metaseiulus occidentalis*, members of the Crustacea and Chelicerata subphylum respectively. In these two Hexapoda outgroups, the *dsx* pre-mRNA shows no signs of alternative splicing but rather a sex-biased expression in males (Kato et al. 2011, Pomerantz et al. 2015, Price et al. 2015). With the exception of three orders (Embioptera, Isoptera and Mantophasmatodea), both OD1 and OD2 domains were found together in all Hexapoda orders. The OD1 domain could not be identified in Mantophasmatodea, whereas in Embioptera and Isoptera evidence of the OD2 could not be recovered from ESTs (Price et al. 2015). Nevertheless, the ancestry, ubiquity and conservation of the *dsx* gene in insects is unchallenged and further supports the hourglass model of evolution of the sex determination and development in the class.

2.3.3.4 Role in sex determination and sex differentiation

The *dsx* gene promotes sexual differentiation by regulating the expression of various downstream genes that are involved in many aspects of sex-specificity and development. The *yolk protein* genes in *D. melanogaster* were the first *dsx* targets to be identified in insects (Burtis et al. 1991, Coschigano & Wensink 1993). It was demonstrated that both DSX^M and DSX^F bind the same fat body enhancer of two *yolk protein* genes in the species. The male-specific isoform suppresses the enhancer and thus the expression of the *yolk protein*, whereas the female-specific version enhances its function and the production of *yolk protein*. The *dsx/yolk protein* system has also helped to reveal and characterise the DSX-binding site on the fat

body enhancer of the *yolk protein* genes (Erdman et al. 1996). The 13-bp palindromic motif (GCAACAAT-GTTGC) was used as a probe to find putative DSX targets in other Drosophila and other Dipteran species (Luo et al. 2011). The motif is indeed shared across many insect lineages with some variation flanking the core sequence ACA[A/T]TGT (Luo et al. 2011, Verhulst & van de Zande 2015). This has helped identify and validate putative target genes of DSX in *D. melanogaster* and many other species and new gene discoveries were quick to follow.

In *D. melanogaster* the role of *dsx* in sex differentiation has been thoroughly examined. Apart from the regulation of the *yolk protein* genes, the specific abdominal pigmentation of males was also shown to be regulated by *dsx* (Kopp et al. 2000, Williams et al. 2008). The *bric-a-brac (bab)* gene suppresses abdominal pigmentation and it is active in females. Both DSX^M and DSX^F bind the same regulatory element upstream of *bab*, restricting or enhancing its function respectively. Female-specific pheromones in the species are also regulated by *dsx*. DSX^F was shown to enhance the production of female-specific pheromones by acting on the desaturase *DesatF* gene (Shirangi et al. 2009). The inability of DSX^M to repress expression of the *DesatF* gene in the same experiment revealed that DSX^F and DSX^M do not always act antagonistically. Other targets of the *dsx* gene involve genes that regulate the development of gustatory sense organs and the differentiation of the *genital imaginal disk* and abdomen. The number of gustatory sense organs in the foreleg of the species is regulated by both DSX^M and DSX^F isoforms on a segment-by-segment basis, revealing a distinct spatial function of the gene in nearby tissues (Mellert et al. 2012, Verhulst & van de Zande 2015). On the abdomen, DSX^M promotes the expression of *extramacrochetae (emc)*, which is involved in the reduction of the A7 abdominal segment in males, whereas DSX^F on the other hand suppresses the expression of the emc gene (Foronda et al. 2012). Finally, suppression of *branchless* by DSX^F results in the absence of male-specific structures in females and the development of female genitalia (Müller-Holtkamp 1995).

The gene's involvement in sex differentiation has also been reported in species outside of *Drosophila*. In the silkworm *Bombyx mori*, the DSX^F isoform has been shown to activate the expression of the *vitello-genin* gene by binding to the core of the palindromic ACA[A/T]TGT sequence, whereas DSX^M was shown to repress the expression of the gene (Suzuki et al. 2003). In the species, the development of the A8 ab-dominal segment in males is also regulated by *dsx*. When DSX^M was ectopically expressed in females, an abnormal A8 segment was formed, indicating that DSX^M is essential for the development of the A8 seg-ment in males(Duan et al. 2014). In the beetle *Tribolium castaneum*, *dsx* also regulates the expression of *vitellogenin* as observed in *B. mori*. In the species, the 13-bp DSX-binding sequence was characterised on an additional 8 different genes, indicating a direct relation with *dsx* (Shukla & Palli 2012*a*). In other beetles, like the horned beetles *Onthophagus* and *Trypoxylus*, knockdown of *dsx* resulted in the reduction

of the size of horns in males and the development of small horns in females, which normally do not develop horns (Kijimoto et al. 2012, Ito et al. 2013). In the stag beetle *Cyclommatus metallifer*, knockdown of *dsx* resulted in a dramatic reduction of the mandibles of males and in a slight overdevelopment in females (Gotoh et al. 2014). Finally, in the Hymenoptera order, the *dsx* gene was associated with wing size and development in the parasitic wasp *Nasonia vitripennis* (Loehlin et al. 2010), whereas in Lepidoptera, the gene has been associated with wing patterns variations in the butterfly *Papilio polytes* that play an important role in mimicry within the species (Kunte et al. 2014).

All these examples are proof that *dsx* is the connecting link between sex determination and sex differentiation. In contrast with other genes, like *fru* and *tra*, *dsx* has been found in all insect species studied to date, acting always as a double switch that underpins male and female sex differentiation. The gene has numerous targets in the genome that promote the development of sex characteristics and processes that eventually define the male and female individual. The *dsx* gene is probably one of the most highly conserved genes in the insect class and domains like the *Dmrt* domain are present in many species across the animal kingdom. The fact that domains like the *Dmrt* are shown to play a role in sex determination in all species that have been characterised, is evidence that this family of genes was present in our common ancestors and has barely diverged in millions of years. This high conservation and sex-specificity of *dsx*, makes it an ideal candidate to use in interventions that aim to control populations of insect pests and disease vectors. A highly conserved female-specific isoform would provide an excellent target for gene drive systems aiming at controlling mosquito populations. The *dsx* gene could provide the solution to the problem of resistance recorded in previous gene drive systems and help build a more robust system that can spread in a mosquito population.

2.4 Sex determination in mosquitoes

Mosquitoes belong within the Culicidae family of the Diptera order. The family contains two subfamilies, Culicinae and Anophelinae, which include over 3,000 species in total. The sex determination of mosquitoes has been under thorough investigation due to the high importance mosquitoes have in global health and economy. Over a century ago, Dr. Nettie Stevens showed that mosquitoes share both homomorphic and heteromorphic sex chromosomes (Stevens 1911). She discovered that members of the Anophelinae subfamily, like *Anopheles punctipennis*, have heteromorphic sex chromosomes whereas members of the Culicinae subfamily, like *Culex quinquefasciatus* and *Culex tarsalis*, have homomorphic sex chromosomes. Up to date, various genetic evidence support that observation and suggest that in Culicinae, sex is determined by a dominant M-factor located on homomorphic loci (M-locus), whereas in

Anophelinae, sex is determined by M-factors located on the heteromorphic Y chromosome (reviewed in Biedler & Tu 2016). The divergence of the sex determining systems between the two families trace back 200 million years ago, which creates an opportunity to study the co-evolution of sex determining mechanism in the family (Reidenbach et al. 2009).

Despite efforts to map the sex determination pathway in mosquitoes, the primary sex determining signals that initiate sex development have been discovered only in a handful of mosquito species. In the yellow fever mosquito *Aedes aegypti*, a candidate gene was characterised from the M-locus located on the homomorphic sex chromosomes of the species (Hall et al. 2015). The gene called *Nix* was shown to be expressed before the sex is determined on the onset of the maternal-to-zygotic transcription. The gene is a distant homologue of *tra-2* and the encoded protein has domains necessary for RNA-binding activity. Because of its ability to bind RNA, it was proposed that *Nix* regulates the splicing of *dsx* either directly, or indirectly through complex assemblies with unknown factors. Knockout mutations of *Nix* using the CR-ISPR/Cas9 system resulted in male individuals with female characteristics such as female antennae and genitals, whereas ectopic expression of the gene in female embryos resulted in the development of male characteristics such as testis and external genitalia. Recently, a study has shown that transgenic expression of *Nix* was sufficient to convert female *Aedes aegypti* mosquitoes into males carrying all the male-specific sexual dimorphic traits and male-specific gene expression profile (Aryan et al. 2020). A homologue of the gene with similar maleness properties has been found in other *Aedes* mosquito species like *Aedes albopictus*, indicating a conservation of the sex determining signal within the genus (Hall et al. 2015). Nevertheless, homologues of the gene could not be identified in distant taxa such as *Culex quinquefasciatus* and *Anopheles gambiae*, revealing once again the diversity of the primary sex-determining signals at the top of the sex determination pathway.

Within the *Anopheles* genus, putative M-factors have been identified on the heteromorphic Y chromosome. In *Anopheles stephensi*, a candidate gene *(Guy1)* was selected from four genes discovered on the Y chromosome. The gene is expressed during the onset of the maternal-to-zygotic transcription, earlier and independent of any other Y chromosome genes (Criscione et al. 2013). In *Anopheles gambiae*, a gene called *gYG2* was found to be expressed from the Y chromosome during the onset of maternal-to-zygotic transcription (Hall et al. 2013). The gene, which later on was named *Yob*, has a similar secondary structure with *Guy1* and is transcribed in the embryo before the blastoderm formation and before any other Y-linked genes. Ectopic expression of the gene in embryos proved lethal for females, whereas ectopic delivery of Yob protein in cell lines enhanced the production of the male-specific *dsx* transcripts and the depletion of the female-specific isoform of the gene (Krzywinska et al. 2016). *Yob* is the only Y linked gene present across the *Anopheles gambiae* species complex, a group of at least 8 morphologically indistin-

guishable species (Hall et al. 2016). These characteristics make *Yob* a suitable candidate for representing the M-factor in the species.

An interesting aspect of mosquito sex determination is the failure to detect any homologues of the *transformer* gene. To date, homologues of the gene have never been found in mosquito species (Geuverink & Beukeboom 2014). What is more interesting is that TRA/TRA-2-like binding sites have been found on the *dsx* and *fru* pre-mRNA of three mosquito species *(A. aegypti, An. gambiae* and *C. quinquefasciatus)* (Scali et al. 2005, Gailey et al. 2006, Price et al. 2015a, Salvemini et al. 2011, 2013). The hypothesis is that *tra* has either evolved to a degree that it is beyond detection or that it has been replaced by another factor or an early sex-determining signal e.g. *Nix*. On the contrary, the more conserved *dsx* gene has been characterised in all mosquito species studied to date, supporting the hourglass model of evolution principle of the sex determination pathway in insects. Apart from the sex determining signals that lie at the top of the sex determination pathway and *dsx* and *fru* that lie at the bottom, the sex determination pathway in mosquitoes is largely unknown and uncharacterised. In some species, like *Aedes aegypti*, the ability of the primary sex determining signal to regulate the alternative splicing of *dsx* and *fru* indicates the absence of intervening factors from the pathway. Of course, if this is the case needs to be thoroughly examined.

2.5 Concluding remarks

Sex within eukaryotes has evolved as a way to mix genomes and select the best and fittest individuals. Across the tree of life, sex is established using an array of different sex determining mechanisms and insects capture an incredible variety of these mechanisms. Sex determination in insects follows a common theme. The mechanisms and signals that determine the sex and lie at the top of the sex determination hierarchy are diverse and almost never similar between different taxa. On the contrary, genes and factors that lie downstream in the pathway, gradually become more conserved in such a degree that genes like *dsx*, that lie at the bottom of the sex determination cascade, are highly conserved and found consistently between all insect species studied to date. The *dsx* gene acts as a binary switch between male and female sex development by regulating a plethora of downstream factors and processes that are all involved in sex differentiation. The gene provides an excellent target for gene drive systems aiming to control insect populations by targeting sex development. By interrupting the sex-specific function of *dsx* we could selectively skew the sex ratios of wild insect pest and vector populations and decrease their numbers. An additional great advantage of the gene that makes it a suitable target for a gene drive system is its conserved nature. First generation gene drives have failed to spread efficiently in insect populations because of resistance at the target sites that counteracted the function of the gene drive. A gene like *dsx* could offer

a solution to this problem. The highly conserved gene could provide a more robust target site that does not allow variations in its DNA sequence, which are essential for the establishment of resistance. A gene drive system targeting such a sequence in the malaria vector *Anopheles gambiae* is thoroughly examined in the next chapters for its efficiency and its ability to spread in mosquito populations without being affected by resistance. By targeting the female-specific isoform of the gene in the species we hope to build a system that could be used as a self-sustaining form of population control that could complement current interventions.

Chapter 3

Materials & Methods

3.1 Plasmid preparation

3.1.1 Bacterial transformation – Miniprep and Maxiprep isolation of plasmid DNA

Plasmids were transformed into OneShot™ Top10 competent *Escherichia coli* cells (ThermoFisher Scientific - #C404006) according to manufacturer's protocol. Briefly, 2 µL of each ligation or plasmid assembly reaction were gently inserted in 25 µL freshly-thawed competent cells and left to incubate on ice for 30 minutes. To heat shock the cells, the solutions were incubated for exactly 30 seconds in a water bath set to 42°C and then immediately placed on ice. Practising sterile technique to avoid contamination, 125 µL of pre-warmed S.O.C medium was added in each tube and incubated on a shaking incubator at 37°C for 45 minutes. Practising sterile technique, 100 µL from each tube were spread on agar petri dishes containing the appropriate antibiotic and left overnight to grow in a 37°C incubator. Colonies were carefully selected from the agar using a 10 µL pipette tip to perform colony-PCRs (see below), after they were re-streaked on a new agar petri dish to grow more cell material for subsequent analysis. Up to 30 colonies can be regrown onto the new petri dish by using 1 cm area to re-streak each colony. Colonies with successful colony-PCRs were selected from the new agar using a 10 µL pipette tip and were used to inoculate sterile Luria-Bertani (LB) broth with the appropriate antibiotic for subsequent miniprep or maxiprep preparation. Plasmids were purified using the QIAprep Spin Miniprep (QIAGEN - #27104) and Maxiprep (#12163) kits according to manufacturer's instructions.

3.1.2 Colony-PCR

Colony-PCRs were performed after each bacterial transformation to verify successfully transformed cells. Colonies from the agar plates were selected using a 10 µL pipette tip and diluted in 5 µL sterile water. 2 µL of the diluted colony was subsequently used as template for PCR reactions. The reactions were performed

using the QIAGEN Fast Cycling PCR kit (#203743) and modified to a 10 µL total reaction volume. PCR reactions were loaded in 200 µL PCR tubes and left to incubate in a Veriti™ Thermal Cycler (ThermoFisher Scientific - #4375305). For the incubation, the initial denaturation step was adjusted to five minutes at 90°C to help rupture the bacteria cells and release the DNA into the solution. The initial denaturation was followed by a 5-second denaturation step at 96°C, a 5-second annealing step and an elongation step at 68°C, which were repeated for 40 cycles. A final one-minute extension step at 72°C was used at the end to finalise the reaction. Because the master mix contains CoralLoad® Dye, half the reaction volume was immediately used for gel electrophoresis on 1% agarose gel. The electrophoresis was allowed to run at 110 V for 45 minutes using a Wide Mini-Sub Cell GT system (Bio-Rad - #1704468EDU) and PCR fragments were verified using a Gel Visualisation System (Azure Biosystems - #c280).

3.2 Genomic DNA extraction

Mosquito DNA was extracted using the Wizard® Genomic DNA Purification kit (Promega - #A1120). Mosquitoes were homogenised using micropestles in 200 µL nuclei lysis solution in microcentrifuge tubes and incubated for 30 minutes at 65°C before following the manufacturer's protocol. Briefly, 200 µL of the lysate were mixed with 70 µL of protein precipitation solution and incubated for five minutes on ice before being vortexed and centrifuged for one minute at 13,000 g on a Benchtop Centrifuge (Eppendorf™ - #5424). The supernatant containing the genomic material was transferred to a new microcentrifuge tube and mixed with 200 µL isopropanol (Sigma-Aldrich, #I9516) before centrifuged at 4°C for 10 minutes at 13,000 g on a cold table centrifuge (Hettick® - #Z652113). The supernatant was removed and the pellet containing the genomic material was washed with 200 µL of 70% Molecular Biology Grade ethanol, Absolute (FisherScientific - #100644795). After a final centrifugation at 13,000 g in the cold centrifuge, the ethanol was discarded carefully to avoid loss of genomic material and the pellet was air dried on a flat surface for 10 minutes. The genomic material was suspended in 30 µL UltraPure DNase/RNase-free Distilled Water (Invitrogen - #11538646) and quantified using a NanoDrop spectrophotometer (ThermoFisher Scientific - #ND2000).

3.3 PCR amplification and gel analysis

Standard PCRs were performed using the Phusion® High-Fidelity PCR kits (NEB UK - #E0553S). PCR reactions were loaded in 200 µL PCR tubes and left to incubate in a Veriti™ Thermal Cycler following the manufacturer's recommendations. An initial 30-second denaturation step at 98°C was followed by a 10-second denaturation step at 98°C, a 30-second annealing step and an elongation step at 72°C, all three

Species	Strain	Assembly	Assembly level
Anopheles gambiae	PEST	AgamP4	chromosome
Anopheles coluzzii	Mali-NIH	AcolM1	scaffold
Anopheles merus	MAF	AmerM2	scaffold
Anopheles arabiensis	Dongola	AaraD1	scaffold
Anopheles quadriannulatus	SANGWE	AquaS1	scaffold
Anopheles melas	CM1001059_A	AmelC2	scaffold
Anopheles christyi	ACHKN1017	AchrA1	scaffold
Anopheles epiroticus	Epiroticus2	AepiE1	scaffold
Anopheles minimus	MINIMUS1	AminM1	scaffold
Anopheles culifacies	A-37	AculA1	scaffold
Anopheles funestus	FUMOZ	AfunF3	chromosome
Anopheles stephensi	Indian	Astel2	scaffold
Anopheles maculatus	Maculatus3	AgamM1	scaffold
Anopheles farauti	FAR1	AfarF2	scaffold
Anopheles dirus	WRAIR2	AdirW1	scaffold
Anopheles sinensis	SINENSIS	AsinS2	scaffold
Anopheles atroparvus	EBRO	AatrE3	scaffold
Anopheles darlingi	Coari	AdarC3	contig
Anopheles albimanus	STECLA	AalbS2	chromosome

Table 3.1 – Mosquito genome databases and assemblies used in this study. Sequence alignments and primer design were performed using sequences from current genome databases and assemblies deposited at VectorBase, www.vectorbase.org (Giraldo-Calderón et al. 2015), during the time of the experiments.

repeated for 35 cycles. The reactions was finalised using a a final 10-minute extension at 72°C. DNA templates included plasmid DNA (≈10 pg), PCR products (≈10 ng) and genomic DNA (≈30ng). The PCR products were subsequently visualised on 1% agarose gel using 5 µL loading sample, consisted of 1 µL of the PCR reaction and 4 µL DNA Loading Buffer Blue (Bioline - #BIO-37045). The loaded gel was run using 110 V for 45 minutes and PCR fragments were eventually verified using a Gel Visualisation System.

3.4 Genomic databases and graph representation

All genomic information and sequences for primer design was retrieved from VectorBase, www.vectorbase.org (Giraldo-Calderón et al. 2015), using the *Anopheles gambiae PEST* genome assembly (AgamP4.12). The assemblies used for other species are outlined on Table 3.1. All genomic, oligonucleotide and primer sequences are displayed in a 5' → 3' orientation unless stated otherwise. GraphPad Prism software v9.0, www.graphpad.com, was used to plot data and perform statistical analyses.

3.5 Generation of CRISPR helper and donor plasmids

3.5.1 p174 gene drive vector

The CRISPR gene drive vector p174 was created using fragments and PCR products amplified with overhangs for subsequent Gibson assembly reaction. The *zpg* promoter and terminator fragments were amplified from genomic DNA using the primers listed in Table 3.2 to produce a 1,126 bp and a 1,094 bp fragments respectively, whereas the *hCas9* fragment (4,273 bp) was amplified from plasmid p165 (Hammond et al. 2016). All fragments were purified from PCR reactions using the QIAquick PCR Purification kit (QIAGEN - #28106). The gRNA-spacer cloning site, the *3xP3::RFP* fluorescent marker coding sequence and the backbone of the plasmid with the *attB* recombination sequences were all excised from p165 as one fragment using AscI (ThermoFisher Scientific - #FD1894) and BshTI (#FD1464) restriction digestion enzymes. The digestion was performed according to manufacturer's instructions and all the reaction was subsequently loaded on a 0.7% agarose gel to validate the digestion. The 5,698 bp fragment was excised from the gel and purified using the QIAquick Gel Extraction kit (QIAGEN - #28115). All 4 fragments (the backbone, the *zpg* promoter and terminator and the *hCas9* fragment) were cloned together into one Gibson assembly reaction (Gibson et al. 2009) using the Gibson Assembly® master mix (NEB UK - #E2611L). The reaction was setup and performed according to manufacturer's recommendations and 2 µL were immediately used to transform OneShot™ Top10 competent cells.

Element	**Primer name**	**Primer sequence**
zpg promoter	zpgprCRISPR-F	GCTCGAATTAACCATTGTGGACCGGTCAGCGCTGGCGGTGGGGA
	zpgprCRISPR-R	TCGTGGTCCTTATAGTCCATCTCGAGCTCGATGCTGTATTTGTTGT
zpg terminator	zpgteCRISPR-F	AGGCAAAAAAGAAAAAGTAATTAATTAAGAGGACGGCGAGAAGTAATCAT
	zpgteCRISPR-R	TTCAAGCGCACGCATACAAAGGCGCGCCTCGCATAATGAACGAACCAAAGG
hCas9	hCas9CRISPR-F	ACAACAAATACAGCATCGAGCTCGAGATGGACTATAAGGACCACGA
	hCas9CRISPR-R	ACTTCTCGCCGTCCTCTTAATTAATTACTTTTTCTTTTTTGCCTGGCC

Table 3.2 – Primers used to amplify the fragments necessary to build p174. Gibson primers with overhangs (underlined) were designed to overlap to place the *hCas9* fragment between the *zpg* promoter and terminator fragments.

3.5.2 p16510 and p17410 CRISPR helper plasmids

The CRISPR helper plasmid p16510 and CRISPR gene drive vector p17410 were created using the p165 plasmid (Hammond et al. 2016) and p174 vector respectively as destination vectors for the gRNA-T1 spacer. The spacer was created using 20 bp oligonucleotides (Table 3.3) carrying overhangs for directional cloning into Eco31I-digested plasmids using Golden Gate assembly reaction (Engler et al. 2008). The oligonucleotides were annealed in a 50 µL reaction containing 100 µM of each primer and 2.5 µL NaCl 1M using a thermocycler and a ramp-down setup programme starting from 95°C down to 20°C. The annealed spacers were subsequently cloned into p165 and p174 using a one-step restriction/ligation reaction as outlined in Engler et al. 2008.

Target site	gRNA sequence	Oligonucleotide sequence
GTTTAACACAGGTCAAGCGGTGG	GTTTAACACAGGTCAAGCGG	TGCTGTTTAACACAGGTCAAGCGG AAACCCGCTTGACCTGTGTTAAAC

Table 3.3 – The gRNA-T1 sequence, target site and the oligonucleotides used to anneal the gRNA-T1 spacer. The gRNA-T1 consists of 20 nucleotides that are homologous to the target site in exon 5 of the *dsx* gene. The target site includes the PAM sequence, an additional 3 nucleotides downstream (underlined). The oligonucleotides used to anneal the gRNA-T1 spacer carry overhangs (underlined) for subsequent Golden Gate cloning.

3.5.3 K101 donor plasmid

The donor plasmid K101 was created using Gibson assembly reaction with fragments amplified from genomic DNA and fragments excised from p163 plasmid used previously in Hammond et al. 2016. Two sequences were amplified from the immediate upstream and downstream regions of the cut site of gRNA-T1 on exon 5, using the sets of primers listed in Table 3.4 to produce a 1,938 bp and 2,015 bp fragments, resembling the left (LH) and right homology (RH) arms of the target site respectively. The eGFP fragment and backbone of the plasmid were excised from p163 using BfuI (ThermoFisher Scientific - #ER1501) and MluI (#FD0564) with BshTI (#FD1464) restriction enzymes respectively. As before, both PCR fragments and digestion fragments were PCR and gel purified using the QIAquick PCR Purification and Gel Extraction kits. All 4 components were mixed accordingly and assembled in a Gibson assembly reaction (Gibson et al. 2009) using the Gibson Assembly® master mix and 2 µL were immediately used to transform OneShot™ Top10 competent cells.

Homology	Primer name	Primer sequence
Left	dsxϕ31L-F	GCTCGAATTAACCATTGTGGACCGGTCTTGTGTTTAGCAGGCAGGGGA
	dsxϕ31L-R	CACCAAGACAGTTAACGTATCCGTTACCTTGACCTGTGTTAAACATAAAT
Right	dsxϕ31R-F	GGTGGTAGTGCCACACAGAGAGCTTCGCGGTGGTCAACGAATACTCACG
	dsxϕ31R-R	TCCACCTCACCCATGGGACCCACGCGTGGTGCGGGTCACCGAGATGTTC

Table 3.4 – Primers used to amplify the fragments necessary to build K101. Gibson primers with overhangs (underlined) were designed to amplify the left and right homology arms surrounding the cut site. The Gibson overhangs are homologous to either the eGFP fragment or backbone to correctly align the homology arms on the K101 donor plasmid.

3.6 Microinjection of mosquito embryos and selection of transformants

Microinjection of mosquito embryos was performed according to standard protocols as described before (Fuchs et al. 2013). To create the *dsxF* $^-$ line, *Anopheles gambiae* G3 wild-type mosquito embryos were injected with a solution containing both p16510 and K101 plasmids, each at 300 ng/µL concentration in phosphate-buffered saline solution (PBS) (ThermoFisher - #10010031). After 48 hours, hatched L1 larvae were screened for transient expression of the eGFP marker in the lower abdomen using a fluorescent microscope (Nikon, Eclipse TE200). G_0 larvae were isolated and were grown to adulthood and mated with wild type. Successful integration events were identified as eGFP$^+$ transformants among the G_1 progeny, expressing the marker in the optic lobe of the head and dorsal ganglia that run the length of the larval body (Figure 3.1). To create the *dsxF^{CRISPRh}* gene drive line, embryos from *dsxF* $^-$ were injected with solution containing both p17410 and the *vas2::integrase* helper plasmid (Volohonsky et al. 2015) at 300 ng/µL concentration. Hatched L1 eGFP$^+$ larvae were selected and screened for the presence of the RFP marker in their lower abdomen. G_0 larvae expressing the RFP marker were isolated and left to grow to adulthood and mated with wild type. Successful RMCE events were identified among the G_1 progeny as *3xp3::GFP$^-$/3xP3::RFP$^+$* larvae since the replacement of the two cassettes will result in the removal of the eGFP marker present on the *dsxF* $^-$ line and the expression of the RFP marker present on the CRISPR^h cassette.

3.7 Molecular confirmation of gene targeting

3.7.1 Confirmation of gene targeting and recombinase-mediated cassette exchange events

Both *dsxF* $^-$ and *dsxF^{CRISPRh}* lines were verified genetically for the correct integration of the transgene in the genome. For *dsxF* $^-$, primers were designed to bind outside the regions used to amplify the homology

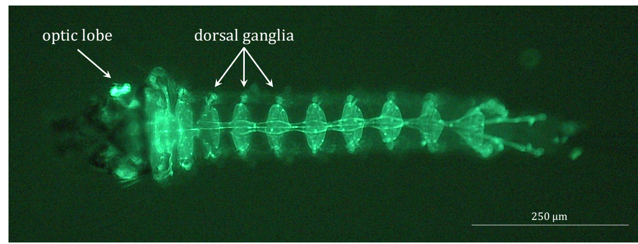


Figure 3.1 – Expression pattern of the *3xP3* promoter. In transgenic larvae, a fluorescent marker that is under the regulation of the *3xP3* promoter is always visible in the optic lobe of the head of the larvae but also the dorsal ganglia that run the length of the larval body. This expression pattern is constant and visible in all developmental stages of the mosquito.

arms for the creation of the K101 plasmid so that only correct integrations within the genome could be identified (Figure 4.8B). Primers 'dsxin3-F' and 'GFP-F' were used to amplify the upstream region of the integration and primers '3xP3' and 'dsxex6-R' were used for the downstream region (Table 3.5). Successful events would generate the expected bands when visualised on a gel at 2,934 bp and 2,310 bp respectively. Primer sets 'dsxex4-F' + 'RFP-R' and 'dsxex5-R1' + 'hCas9-F' were used to verify the left and right side of the orientation of the CRISPR^{h} cassette in the target locus. The primers amplify a 1,380 bp and 1,979 bp respectively. All PCR fragments were purified using the QIAquick PCR Purification kit and were sent for genomic sequencing for verification.

Primer name	Primer Sequence
dsxin3-F	GGCCCTTCAACCCGAAGAAT
GFP-F	GCCCTGAGCAAAGACCCCAA
3xP3	TATACTCCGGCGGTCGAGGGTT
dsxex6-R	GAATTTGGTGTCAAGGTTCAGG
dsxex4-F	GCACACCAGCGGATCGACGAAG
RFP-R	CAAGTGGGAGCGCGTGATGAAC
hCas9-F	CCAAGAGAGTGATCCTGGCCGA
dsxex5-R1	CTTATCGGCATCAGTTGCGCAC

Table 3.5 – Primers used for verification of gene targeting and RMCE events. List of primers used to verify the correct integration of the eGFP cassette in the *dsxF*$^{-}$ line and the RMCE event in the *dsxF*^{CRISPRh} line.

3.7.2 Containment of mosquitoes

Wild-type and transgenic mosquitoes were kept at an underground insectary in dedicated cubicles with controlled temperature and humidity, compliant with Arthropod Containment Guidelines Level 2 (Scott 2005). The insectary is protected with two levels of security card access and at least 7 doors from the external environment. All work is performed in accordance with Imperial College biosafety protocols for genetically modified (GM) organisms and specific risk assessments for the use and containment of GM mosquitoes, which are compliant with a recent report calling for safeguarding the creation, study and use of gene drive technology (Akbari et al. 2015).

3.8 Phenotypic characterisation and microdissections

To confidently separate mosquitoes based on their genotype, a Complex Object Parametric Analyzer and Sorter instrument (COPAS) (Union Biometrica, USA) was used to confidently distinguish between heterozygous and homozygous L1 larvae, based on the intensity of the fluorescent marker of the transgene (Marois et al. 2012). The larvae were left to grow, and adult mosquitoes were collected in 15 mL centrifuge tubes and anaesthetised on ice for 5 minutes prior to phenotypic analysis and dissections. The mosquitoes were observed under a stereomicroscope (Olympus – SZX7) and pictures were taken using a HiChrome-SMII digital mounted camera (GT Vision, UK). Dissections were performed on microscopic slides using PBS buffer and two pairs of forceps. For phenotypic characterisation of the genotypes, the legs of the mosquitoes were removed to achieve the profile orientation for better comparison. The gonads, and spermathecae were carefully isolated from the rest of the carcass and covered with borosilicate glass square coverslip to be levelled and visualised under an EVOS XL Core Cell imaging system (ThermoFisher Scientific - #AMEX1200).

3.9 Phenotype assays

3.9.1 Fertility assays

Progeny from intercrossed *dsxF*$^{-}$ mosquitoes carrying an RFP marker on the Y chromosome (Bernardini et al. 2014) were screened for the intensity of the eGFP marker and were separated into homozygous (*dsxF*$^{-/-}$) and heterozygous (*dsxF*$^{+/-}$) groups. When pupae, the groups were separated into males and females and were mated separately with wild type in groups of 50 in BugDorm-4S1515 rearing cages (BugDorm, Taiwan). The mosquitoes were left to mate for five days before they were blood fed on anesthetised mice for 10 minutes. At least 45 blood-fed females from each cross were allowed to lay individually into

100 mL cups filled with water, lined with filter paper and covered with mesh. The entire egg and larval progeny were counted for each female and to verify the genotype of the parent a minimum of 20 larvae were screened for the presence of the eGFP marker. Females that failed to lay eggs or produce larvae were dissected to check for evidence of mating. For that, females were anaesthetised on ice for five minutes and were dissected to retrieve their spermatheca, a black sphere at the end of their abdomen (Figure 3.2). The spermatheca was ruptured using a cover slip and was checked under an EVOS imaging system for the presence of sperm. Females with no evidence of sperm were excluded from the analysis. For the $dsxF^{CRISPRh}$ line the analysis was performed essentially the same way with some minor differences. Heterozygous mosquitoes ($dsxF^{+/-}$) were screened and separated from both male and female parents to assess the parental effect of the construct on the phenotype. As before, the entire egg and larval output were counted but this time the entire larval progeny was screened for the presence of the CRISPR^{h} RFP marker to assess the transgenic rate of the line. In contrast with before, unmated females were included in the analysis. All group statistical differences were calculated and reported using the Kruskal-Wallis and Mann-Whitney nonparametric tests.

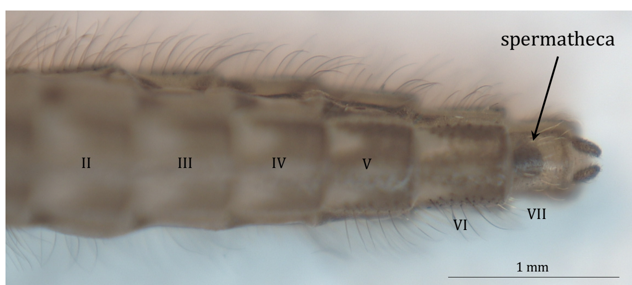


Figure 3.2 – The abdomen and spermatheca of the female mosquito. The spermatheca, a female-specific structure that is used as sperm storage is located at the last segment of the abdomen (VII) and is clearly visible as a black sphere, even through the mosquito cuticle.

3.9.2 Longevity assays

Homozygous ($dsxF^{-/-}$) and wild-type ($dsxF^{+/+}$) larvae were reared separately to pupation and a random 100 pupae from each group were mixed together in BugDorm-4S2222 rearing cages and left to emerge. Dead mosquitoes were collected every day at the same time and were sexed using a stereomicroscope. Female individuals from each group were identified by the presence or absence of inverted claspers, whereas male $dsxF^{-/-}$ mosquitoes were identified using a fluorescent microscope by examining the eyes of the mosquitoes for the presence of the eGFP marker. The experiment was performed in triplicates and the median

($\tilde{\chi}$) survival was analysed for statistical significance using the Log-rank (Mantel-Cox) test.

3.10 Measuring spread of gene drive in a caged population

All cage trials were initiated using 300 wild-type females, 150 wild-type males and 150 *dsxF*^{CRISPRh}/+ male individuals originated from transgenic male parents. All mosquitoes were reared in parallel and both wild type and *dsxF*^{CRISPRh} were screened for the presence of the RFP marker. Mosquitoes were placed in BudDorm-4S2222 rearing cages as pupae and left to emerge. The cages were left to mate for 5 days before blood feeding on mice for 15 minutes. Eggs were collected after 2 days using 300 mL glass bowls filled with water and lined with filter paper. To seed the next generations, a total of 650 randomly-chosen eggs were collected from the egg bowls using strings of filter paper and were gently washed into a rearing tray filled with 500 mL of water and lined with filter paper in order to hatch. The rest of the clutch size from every cage was counted each generation using JMicroVisioin v1.27 software. All the larval progeny hatched from the trays were screened for the presence of the RFP marker to assess the transgenic rate of the generation and were split equally into trays and reared to pupation. All pupae from each cage were counted and were placed into new cages to emerge, marking the start of the next generation. The cycle was repeated until the cage populations were unable to blood feed or produce eggs.

3.11 Analysis of resistance at the gene drive target site

Each generation the entire population of each cage was collected into 15 mL centrifuge tubes and placed in -20°C for storage. Genomic material was obtained *en masse* from the whole population of each cage for every generation using the Wizard® Genomic DNA Purification kit. A PCR reaction was performed on 40 ng of DNA material using the KAPA HiFi HotStart Ready mix PCR kit (Kapa Biosystems - #KR0370). The reactions were performed in 50 mL volume in 200 μL PCR tubes using primers carrying the Illumina Nextera Transposase adapters (Table 3.6) for subsequent deep amplicon sequencing. The PCR reactions were incubated in a Veriti™ Thermal Cycler using the following setup: a three-minute initial denaturation step at 95°C, followed by a 20-second denaturation step at 98°C, a 20-second annealing step at 68°C and a nine-second extension step at 72°C, followed by a final five-minute extension step at 72°C. PCR reactions were performed under non-saturating conditions so as not to distort the relative frequency of the alleles at the target site. To achieve that, the PCR reaction was paused after 20 cycles and half the volume of the reaction was stored at -20°C for subsequent deep amplicon sequence analysis. The rest of the reaction was left to run for an additional 20 cycles and used to verify the PCR amplification on a 1% agarose gel. Successful PCR reactions were identified with a band at 358 bp and the corresponding samples stored at

-20°C were sent for deep amplicon sequencing (Genewiz® Brooks Automation, Germany and PoloGGB, Italy).

Primer Name	Primer sequence
4050-Illumina-F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTACTTATCGGCATCAGTTGCG
4050-Illumina-R	GACTGGAGTTCAGACGTGTGCTCTTCCGATCTGTGAATTCCGTCAGCCAGCA

Table 3.6 – Primers used for deep amplicon sequencing of the gene drive target site in *dsx*. The pair of primers used to amplify the target site on exon 5 to identify non-homologous end joining (NHEJ) and microhomology-mediated end joining (MMEJ) events after the digestion of the target site by the gene drive. The Nextera Transposase adapters area underlined.

3.11.1 Bioinformatic analysis of the alleles generated during the population experiments

Raw deep amplicon sequencing data were analysed using the CRISPResso V2.0 software (Clement et al. 2019). By using the target site as a reference, the software groups alleles based on indels and single nucleotide variance (SNV) around the target site including 20 bp flanking regions. Alleles with less that 3 reads and an average quality score below 30 (phred33) were discarded and were not used in the analysis. The 12 most prevalent alleles and the SNV frequencies across the length of the amplicon were used to plot the graphs.

3.12 *In vitro* cleavage assay of SNP found in wild populations in Africa

3.12.1 Plasmid preparation

The spacers for the original gRNA-T1 target site (WT^{ts}) and target site with the SNP found in Africa (SNP^{ts}) were assembled using reverse complement oligonucleotides (Table 3.7) carrying overhangs for directional cloning using Golden Gate assembly. The oligonucleotides were annealed together in a 50 µL reaction containing 100 µM of each oligonucleotide and 2.5 µL 1M NaCl using a thermocycler and a ramp-down setup programme starting from 95°C down to 20°C. Both WT^{ts} and SNP^{ts} annealed spacers were cloned into DR274, a plasmid carrying an Eco31I spacer cloning site (Hwang et al. 2013), using a one-step restriction/ligation reaction as outlined in Engler et al., 2009(Engler et al. 2009). The new plasmids pWT^{ts} and $pSNP^{ts}$ were purified and linearized using SspI restriction digestion enzyme (ThermoFisher Scientific - #FD0774) and were verified by genomic sequencing using M13-forward universal primer. An extra plas-

mid (pWT^{tsnp}) carrying the wild-type target sequence lacking the -TGG protospacer adjacent motif (PAM) that is necessary for Cas9 activity was also created in parallel to serve as a control. DR274 was a gift from Keith Joung (Addgene plasmid #42250; http://n2t.net/addgene:42250; RRID:Addgene_42250).

Spacer	Target site	Oligonucleotide sequence
WT^{ts}	GTTTAACACAGGTCAAGCGG**TGG**	TAGGGTTTAACACAGGTCAAGCGG**TGG** AAAC**CCA**CCGCTTGACCTGTGTTAAAC
SNP^{ts}	GTTTAACACAGGTCAAGCAG**TGG**	TAGGGTTTAACACAGGTCAAGCAG**TGG** AAAC**CCA**CTGCTTGACCTGTGTTAAA
WT^{tsnp}	GTTTAACACAGGTCAAGCGG	TAGGGTTTAACACAGGTCAAGCGG AAACCCGCTTGACCTGTGTTAAAC

Table 3.7 – Oligonucleotides used for spacer assembly for *in vitro* cleavage assay. List of oligonucleotides used to anneal the WT^{ts}, SNP^{ts} and WT^{tsnp} spacers. The spacers were designed to carry the original target site, the target site with the G>A SNP found in Africa (red) and the target site without the PAM sequence respectively. The oligonucleotides carried overhangs for subsequent Golden Gate cloning. The PAM sequence -TGG and its reverse compliment -CCA are annotated in bold.

3.12.2 RNP digestion and gel analysis

Ready-to-use sgRNA provided by Synthego (USA) was mixed with *Streptococcus pyogenes* Cas9 enzyme (NEB UK - #M0386) at 1:1 molar ratio and according to manufacturer's instructions were left to incubate for 10 minutes at room temperature to form ribonucleoprotein particles (RNPs). Following the protocol, the linearised substrate from the previous step was added to the reaction at a 10:10:1 molar ratio (Cas9:sgRNA:substrate) and left to incubate for 30 minutes at 37°C. The digestion was terminated using 1 µL of Proteinase K (NEB UK - #P8107) followed by a 10-minute incubation at room temperature. 20 µL were used to visualise the fragments on a 0.7% agarose gel.

Chapter 4

Understanding the role of *doublesex* in mosquitoes

4.1 The *doublesex* gene in *Anopheles gambiae*

The *doublesex* gene in *Anopheles gambiae (Agdsx)* was first identified in the species in 2005 using the *Drosophila melanogaster doublesex (Dmdsx)* as a probe to scan available expressed sequence tag (EST) databases for signs of homology (Scali et al. 2005). The *Dmrt* OD1 and OD2 domains of the gene family were quickly identified due to their high conservation and were quickly followed by the discovery of the female-specific locus. The gene, annotated as *AGAP004050*, is located on band 17C on chromosome 2R and it consists of 7 exons spread over an 85 kb region with coordinates 2R:48,703,664 – 48,788,460. Extensive analysis on male and female cDNA libraries revealed the genomic organisation of the gene and the alternative splicing mechanisms that produce the male- and female-specific transcripts. The gene produces two isoforms through alternative splicing, which are characterised by the inclusion or removal of exon 5 (Scali et al. 2005). The *dsx* male-specific isoform (AgdsxM) lacks this exon, which is only present in the female-specific isoform of the gene (AgdsxF) (Figure 4.1). Because this is the only difference between the two isoforms exon 5 is considered to be the female-specific exon of the species.

Both AgdsxM and AgdsxF transcripts appear to share the same 2,335 bp 5' noncoding untranslated region (UTR), which spans exon 1 and part of exon 2. Translation for both isoforms starts at a universal start codon located in exon 2, which also encodes for the highly conserved *Dmrt* OD1 domain. The *dsx* oligomerisation OD2 domain is located in exon 4, which is only 135 bp long, and continues into exon 5 or exon 6 depending on the isoform. Exon 5, a 1,712 bp region in the female-specific isoform produces a 30 amino acid (aa) long product due to an early stop codon located just 90 bp downstream. Exon 6 on the other hand is 1,267 bp long and encodes for a much larger product of 398 aa. The last exon is 2,669 bp

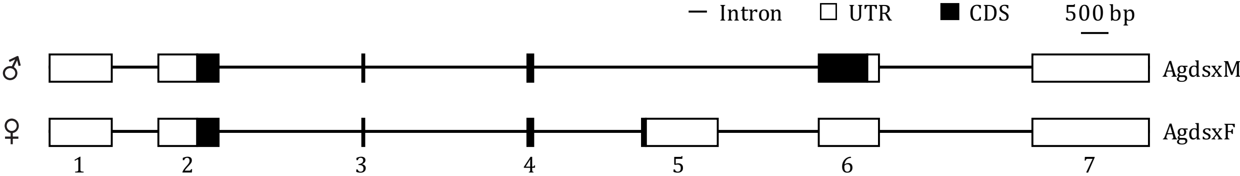


Figure 4.1 – Transcript comparison of the *doublesex* gene in *Anopheles gambiae* . The gene consists of 7 exons and produces two sex-specific isoforms through alternative splicing. Exon 5 is spliced out in the male-specific isoform (AgdsxM) and included in the female-specific isoform (AgdsxF) and thus is considered to be the female-specific exon. Transcription of both isoforms starts at a universal start codon in exon 2. Introns are not drawn to scale. Image reproduced with permission of the rights holder, Copyright © 2018, Springer Nature, Kyrou et al. 2018.

long and serves as part of the 3' UTR of both transcripts. Because in AgdsxF both exon 6 and the last exon serve as part of the UTR, the isoform is characterised by a longer 3' UTR compared to its male counterpart (Scali et al. 2005). This is in contrast with the situation in *Drosophila melanogaster (Dmel)* where the male-specific exon 5 and the last exon are not part of the female-specific transcript (Burtis & Baker 1989) (Figure 4.2).

As in other taxa (Price et al. 2015), the gene retains the conserved domains required for oligomerisation and DNA binding activity. The two isoforms of the gene produce two sex-specific protein products (DSX^M and DSX^F) with distinct protein domains corresponding to the OD1, OD2 and OD sex-specific domains of the *doublesex* family. In both proteins, the *Dmrt* OD1 domain is located on the N-terminus of the protein and carries the characteristics of an atypical DNA-binding zinc-finger $C_2H_2C_4$ Cys-His configuration (Erdman & Burtis 1993, Raymond et al. 1998, Zhu et al. 2000). The domain exhibits 89% identity and 93% similarity at the protein level when compared with *Dmdsx* (Figure 4.3) despite the 250 million years that diverge the two species (Kumar et al. 2017). The OD2 oligomerisation domain, which is closer to the C-terminus of the protein, is the last common domain between DSX^M and DSX^F and exhibits 57% identity and 75% similarity when compared with *Dmdsx*. Finally, the OD^M and OD^F sex-specific domains are located immediately downstream of OD2 and they differ with each other not only in length but in conservation as well. The 30 aa product encoded by the female-specific exon is 80% identical and 87% similar to *Dmdsx*, whereas the longer male-specific product is only 6% identical and 8% similar but can reach 22% identity and 30% similarity when aligned partially.

Within the Anophelinae subfamily, both the *Dmrt* OD1 domain but also the OD2 oligomerisation domain are highly conserved (Figure 4.4). Alignment of sequences of both AgdsxM and AgdsxF obtained from available genetic data from 18 different *Anopheles* species (Neafsey et al. 2015) reveals great conser-

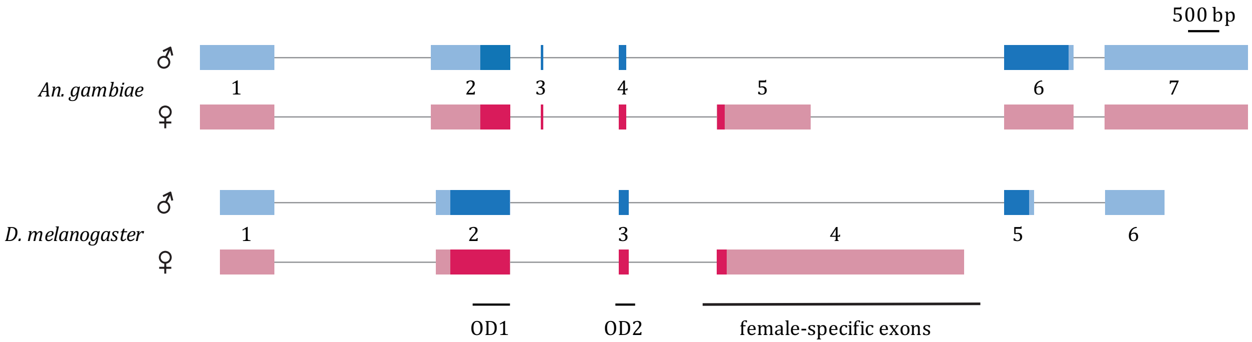


Figure 4.2 – Transcript comparison between the *doublesex* gene in *An. gambiae* and *D. melanogaster*. In both species the exons are alternatively spliced to create the male- and female-specific isoforms. The OD1, OD2 and female specific domains are highly conserved between the two species, whereas the male-specific exons are much more diverged. Light colour indicates untranslated regions. Introns are not drawn to scale

vation throughout the subfamily but in particular within the *Anopheles gambiae* species complex. The complex, also known as *Anopheles gambiae sensu lato (s.l.)*, consists of at least eight morphologically indistinguishable species and members include *An. gambiae, An. arabiensis, An. quadriannulatus, An. merus, An. melas, An. bwambae* (reviewed in White et al. 2011), *An. coluzzii* and *An. amharicus* (Coetzee et al. 2013). The complex is far from complete and new species are added on a regular basis, with *Anopheles tengrela* to be the latest addition to the list (Tennessen et al. 2020). The alignment also reveals a noticeable difference between the degree of conservation between the male- and female-specific coding sequences (CDS) (Figure 4.5). The 90 bp that characterise AgdsxF appears to be much more conserved not only within the *An. gambiae* complex but within the Anophelinae subfamily as well. Specifically, the CDS sequence of the OD^F presents 98.9% (89/90) identity within members of the complex and 84.4% (76/90) within the subfamily. In contrast, the 1,194 bp male-specific CDS of exon 6, although conserved within the complex, is much more diverse in the subfamily. The sequence presents 85.9% (1026/1194) identity within the complex but only 25.8% (303/1194) identity within the subfamily. This high conservation of the AgdsxF isoform could be a solution to the resistance the first-generation gene drives presented when targeting a non-conserved site on the *AGAP007280* gene (Hammond et al. 2017).

4.2 Using the CRISPR/Cas9 system to disrupt the *doublesex* gene

To investigate whether the *doublesex* gene is a suitable target for use in a gene drive system approach to suppress and control populations of *Anopheles gambiae* mosquitoes, I sought to disrupt the gene and study its function and suitability. In the lab, we have been using the CRISPR/Cas9 system since 2015 to

Figure 4.3 – Representation of the DSX^M and DSX^F protein domains and their similarity with the *Dm-dsx*. (**A**) Graphical representation of the OD1, OD2 and OD sex-specific domains of the DSX proteins. DSX^M and DSX^F share the same OD1 and OD2 domains but they differ on the sex-specific OD domains. ODF is smaller than OD^M and much more conserved between species. Domains are not drawn to scale. (**B**) Sequence comparison and similarities between the DSX^F (upper) and DSX^M (lower) proteins of *Anopheles gambiae (Agam)* and *Drosophila melanogaster (Dmel)*. The OD1 (yellow), OD2 (green) and OD^F (red) are much more conserved between the two species, whereas the male specific OD^M (blue) is far more diverged. Colon (:) and period (.) indicate strong and weak conservation of residue properties respectively. Comparison was performed using the Needleman-Wunsch alignment algorithm with EMBOSS Needle tool (EMBL-EBI).

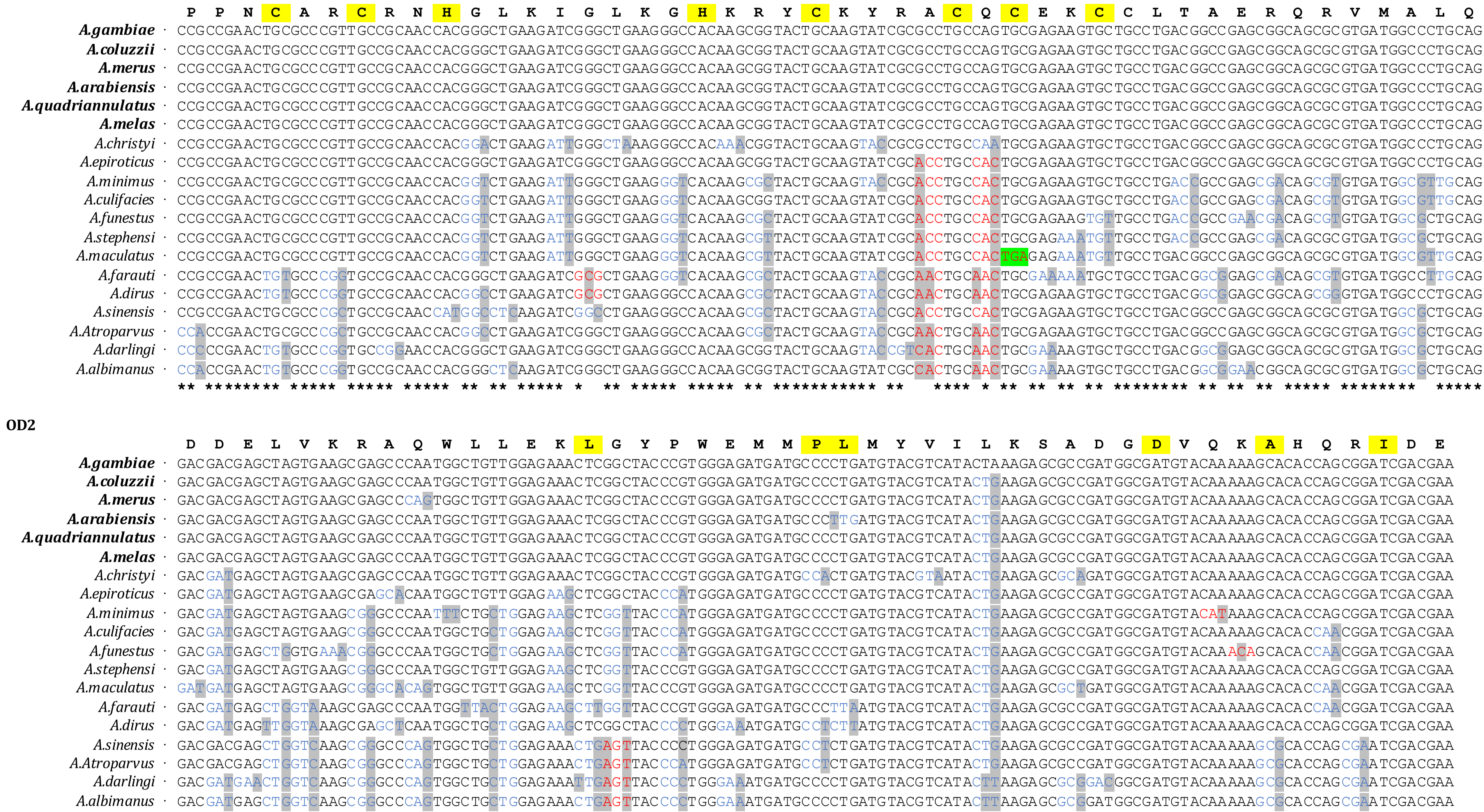


Figure 4.4 – Comparison of genetic sequences of OD1 and OD2 domains within the Anophelinae subfamily. Both OD1 and OD2 domains are conserved within the subfamily but are particularly highly conserved within the *An. gambiae* species complex (bold). The $C_2H_2C_4$ Cys-His motif of the *Dmrt* family and the conserved residues of the OD2 domain are highlighted in yellow. In *An. maculatus*, the OD1 domain is not annotated completely and the particular TGA stop codon (highlighted in green) is most probably the result of poor consensus within the scaffold assembly of the AmacM1 dataset. Variation from the *An. gambiae* sequence is shaded in grey, whereas silent and non-silent changes in the CDS are anotated with blue and red respectively.

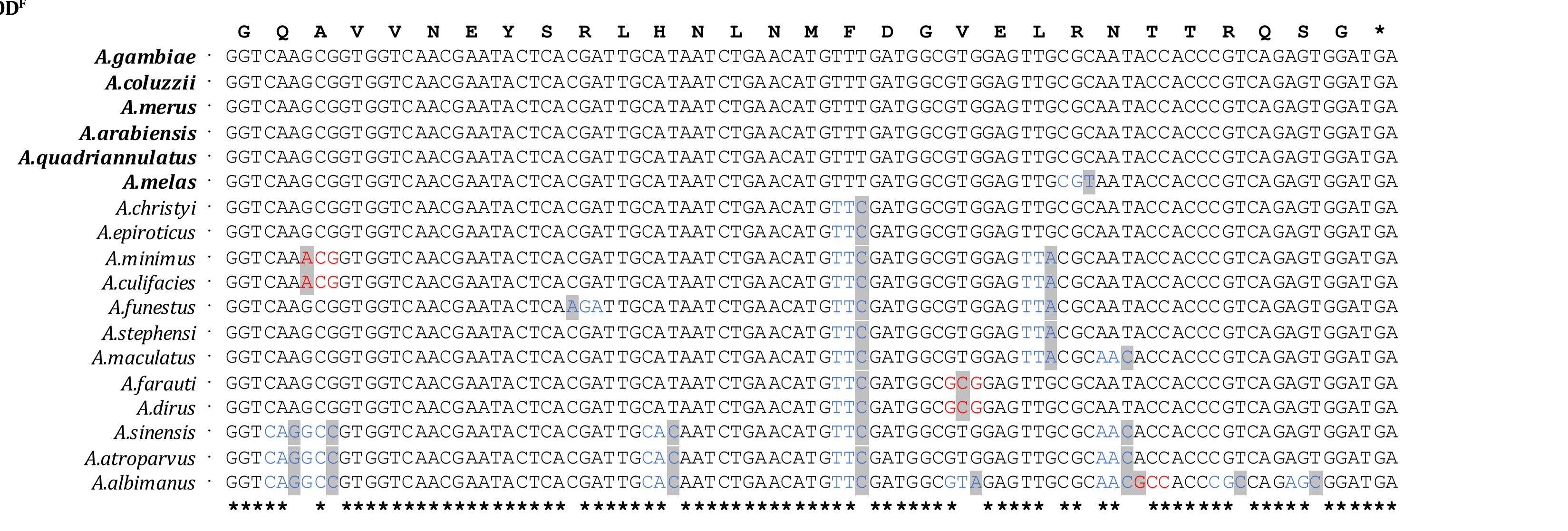


Figure 4.5 – Comparison of genetic sequences of OD^M and OD^F domains within the Anophelinae subfamily. The female-specific domain of the gene is highly conserved not only within the *An. gambiae* species complex (bold) but within the subfamily as well. Even distantly related species like *An. albimanus*, which diverged from the rest of the family 60 million years ago, present 90% identify on the DNA level for this particular domain. On the contrary, the male-specific domain of the gene is not under the same selective pressure as OD^F since it is highly diverged within the subfamily. The domain still retains some of its conservation within the *An. gambiae* complex. For OD^M only the first 123 bp are shown out of the 1,194 bp that encode for the domain. Variation from the *An. gambiae* sequence is shaded in grey, whereas silent and non-silent changes in the CDS are anotated with blue and red respectively.

disrupt and study genes in *An. gambiae* and we have adjusted and perfected this genetic engineering tool to be efficient in the mosquito. We have isolated exon 5 as a primary target because of its high conservation within the *An. gambiae* complex and the Anophelinae subfamily. By only focusing on this exon, present only in the female-specific isoform of the gene, we are also investigating whether the phenotypic effect of disrupting AgdsxF is confined strictly to females, something ideal for a population suppression gene drive system.

4.2.1 Selection of suitable gRNA

The 90 bp of coding sequence of exon 5, including short flanking regions, was scanned to search for gRNAs to disrupt the exon. I have combined data from the web tool ChopChop v3 (Labun et al. 2016) and "Guido" (unpublished with permission of Nace Kranjc), a homemade programme that ranks gRNAs based on their ability to create double-strand breaks (DSB) on loci surrounded by small homologous sequences that can be used to repair the DSB using microhomology-mediated end joining (MMEJ). The advantage of Guido is that it can distinguish the likely aftermath of the repair and predict if MMEJ will result in a sequence that retains the coding frame of the exon or not. This is a crucial criterion to consider when selecting a gRNA for a gene drive system to reduce the chances of creating functional resistant alleles that will counteract the function and the spread of the gene drive (Hammond et al. 2017). By selecting gRNAs that have more chances of creating out of frame mutations through MMEJ the probability of those alleles being selected is substantially reduced.

Ten gRNAs have been identified throughout the coding sequence of exon that can be used with the CRISPR/Cas9 system to knock out the exon (Figure 4.6). Five of these gRNAs score total efficiency above 50% based on Doench et al. 2016, and have no obvious off-target sequences within the *Anopheles gambiae* reference genome (AgamP4 assembly) (Labun et al. 2016). Using Guido, I could also reveal any off-target sequences these gRNAs have in our lab colony *An. gambiae* G3. Our homemade algorithm can also point out any single nucleotide polymorphisms (SNPs) specific to our lab colony that are present within the gRNA sequence and thus help filter out gRNAs that might not be suitable to test a gene drive system in our lab colony. Combining the information from both ChopChop and Guido I have ranked all ten gRNAs based on all these parameters but also based on their distance from the 5' end of the coding sequence of the exon (Table 4.1). The idea is that a gRNA located at the 5' end of the exon will have a greater impact on the transcript of AgdsxF and even maybe on the splicing of the exon. Out of the ten gRNAs I chose the first gRNA, named gRNA-T1 with sequence 5'-GTTTAACACAGGTCAAGCGG-3', to proceed with the disruption of the exon (Kyrou et al. 2018). The cleavage site of this gRNA is just 6 bp downstream of the 5' end of the exon (Figure 4.7).

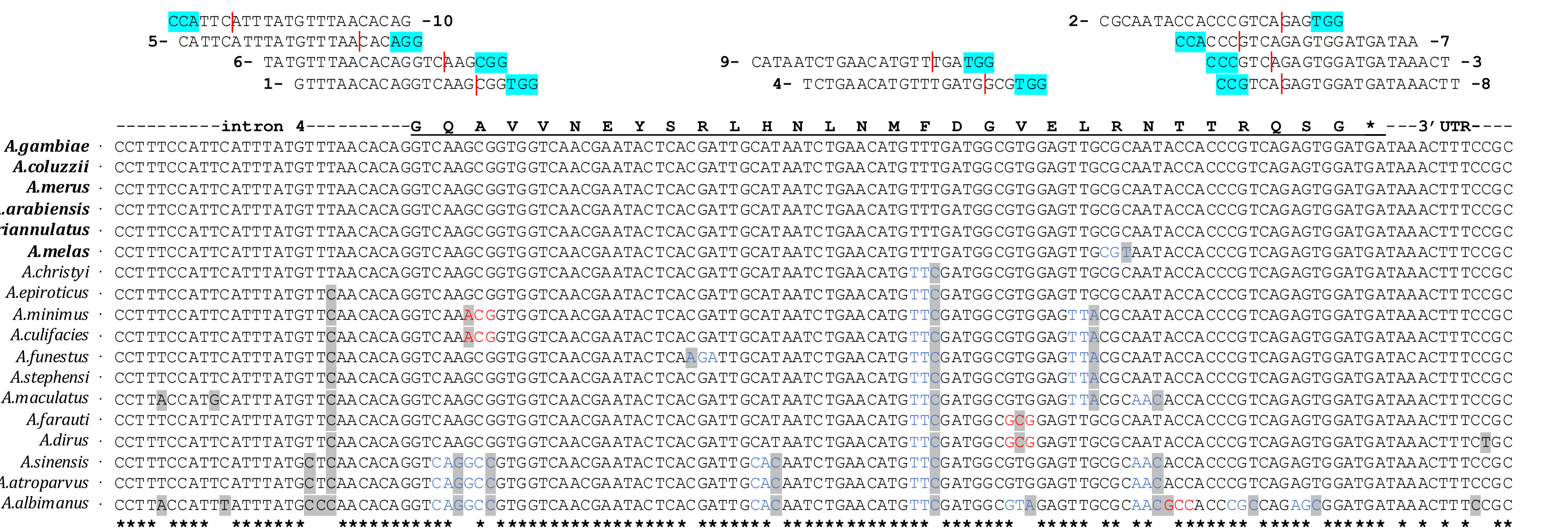


Figure 4.6 – Putative gRNA target sites in *An. gambiae* and their alignment in relation to the CDS of exon 5 within the Anophelinae subfamily. ChopChop and Guido revealed the presence of ten gRNA target sites distributed across different parts of the CDS of the exon, including short flanking regions that extend in intron 4 and the 3' UTR (dashed lines). The PAM sequence of each target site, which is not part of the gRNA itself, is shaded in blue and the cleavage site of each gRNA is annotated with a vertical red dotted line. The gRNAs are numbered in relation to their efficiency and MMEJ score outlined in Table 4.1. Variation from the *An. gambiae* sequence is shaded in grey, whereas silent and non-silent changes in the CDS are anotated with blue and red respectively.

Rank	Target site	gRNA sequence	Location	Strand	Off-targets	Cleavage efficiency	MMEJ score
1	GTTTAACACAGGTCAAGCGGTGG	GTTTAACACAGGTCAAGCGG	2R:48714637	-	0	64.48	88.9
2	CGCAATACCACCCGTCAGAGTGG	CGCAATACCACCCGTCAGAG	2R:48714561	-	0	59.75	40.9
3	AGTTTATCATCCACTCTGACTGG	AGTTTATCATCCACTCTGAC	2R:48714551	+	0	53.32	40.9
4	TCTGAACATGTTTGATGGCGTGG	TCTGAACATGTTTGATGGCG	2R:48714589	-	0	52.57	62.5
5	CATTCATTTATGTTTAACACTGG	CATTCATTTATGTTTAACAC	2R:48714648	-	0	52.09	81.0
6	TATGTTTAACACAGGTCAAGTGG	TATGTTTAACACAGGTCAAG	2R:48714640	-	1(3)	66.29	70.7
7	TTATCATCCACTCTGACGGGTGG	TTATCATCCACTCTGACGGG	2R:48714554	+	1(3)	65	82.9
8	AAGTTTATCATCCACTCTGATGG	AAGTTTATCATCCACTCTGA	2R:48714550	+	1(3)	57.2	40.9
9	CTGTGTTAAACATAAATGAATGG	CTGTGTTAAACATAAATGAA	2R:48714649	+	3(3)	50.08	100
10	CATAATCTGAACATGTTTGATGG	CATAATCTGAACATGTTTGA	2R:48714594	-	3(3)	39.54	83.5

Table 4.1 – List of target sites and the gRNAs with their properties found on exon 5 of *dsx*. The ten gRNAs are ranked based on their properties such as the numbers of off-targets they have within the *An. gambiae* genome, their cleavage efficiency (Doench et al. 2016) and MMEJ score (Kranjc, Nace unpublished). Even though some gRNAs have high cleavage efficiency and a high MMEJ score they are ranked at the end of the list because of the amount of off-targets they have within the genome. The MMEJ score is indicated as a percentage reflecting the length of the overlapping MMEJ sequences and the inability of the resulting product to restore the reading frame of the protein. The number of off-targets are indicated alongside the number of nucleotide mismatches they have (in parenthesis) compared with the respective target site. The PAM sequence of the target sites is underlined.

4.2.2 Creation of vectors for targeted mutagenesis on exon 5

To disrupt the female-specific exon I followed the same principle technique used previously in Hammond et al. 2016. Based on this approach, two plasmids are required to successfully disrupt the sequence of interest and also knock-in a fluorescent marker cassette to track the disruption: a CRISPR/Cas9 vector (referred to as CRISPR-helper plasmid hereafter) that carries the components of the CRISPR/Cas9 system, and a plasmid that provides the template DNA sequence with the fluorescent marker cassette that will be used to repair the double-strand break (referred to as donor-plasmid hereafter). Both vectors have to be co-injected into the posterior of fertilised embryos where the pole cells are located. The idea is to transform these cells that eventually will form the germline of the mosquito so that the transgene can be passed on to the next generation.

The p165 CRISPR-helper plasmid used previously (Hammond et al. 2016) contains all the necessary elements to specifically target and cleave a locus of interest in *An. gambiae* . The plasmid carries a human codon-optimised version of the *Streptococcus pyogenes Cas9* gene under the regulation of the germline-specific *vasa2 (vas2)* promoter and terminator. This will ensure the expression of the nuclease in the germ cells of the mosquito (Papathanos et al. 2009). The plasmid also carries a gRNA spacer-cloning site and backbone (Hwang et al. 2013) under the control of the *An. gambiae U6snRNA* promoter, which is ubiquitously expressed in the mosquito (Konet et al. 2007). This RNA polymerase III-transcribed promoter

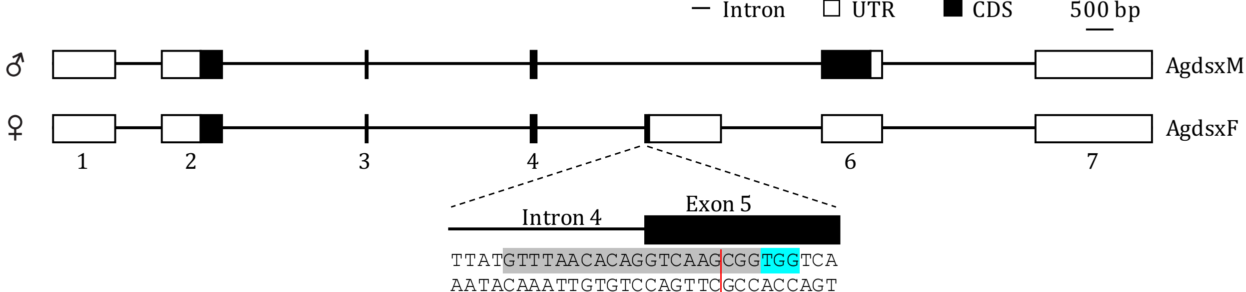


Figure 4.7 – The target site of gRNA-T1 in relation to the two sex-specific isoforms of *Agdsx*. The gRNA targets a sequence (shaded in grey) that spans the intron 4 - exon 5 boundary and induces a DSB 6 bases downstream of the 5' of the exon (vertical red dashed line). The male-specific isoform of the gene should not be affected by this gRNA since the exon is spliced out in AgdsxM. The PAM of the target site is annotated in blue. Image reproduced with permission of the rights holder, Copyright © 2018, Springer Nature, Kyrou et al. 2018.

ensures that the gRNA will be transcribed without a 5' cap and a 3' poly-A tail, an essential requirement for the gRNA to be accessible to complex with the Cas9 protein. Transcription of the gRNA cassette terminates using a small 8-nucleotide long poly(T) tract at the end of the gRNA cassette (Konet et al. 2007). Based on Hwang et al. 2013, I used Golden Gate cloning to modify the spacer-cloning site to contain our gRNA-T1 spacer targeting exon 5. The newly created plasmid was named p165.10.

The donor-plasmid provides the cell with an artificial template sequence to repair the double-strand break (DSB). The plasmid carries sequences homologous to the regions immediately upstream and downstream of the cut site, mimicking the homologous uncut chromosome. Using homology-directed repair (HDR) the repair machinery tries to fix the cut using the donor-plasmid and thus incorporating in the cut site any sequences surrounded by the two homology arms (Figure 4.8A). By flanking a fluorescent marker cassette with the homology arms, we can knock in the cut site an enhanced green fluorescent protein marker (eGFP) to track the mutation. The homology arms of p163, the donor-plasmid used previously (Hammond et al. 2016), were replaced with 2 kb fragments amplified from the regions upstream and downstream of the cut site of gRNA-T1 on exon 5. The homology arms flank an eGFP marker cassette surrounded by two *attP* sequences in reverse orientation, which will be incorporated in the genome after a successful HDR event. The *attP* sequences will be used at a later stage to modify the locus through ϕC31 recombinase-mediated cassette exchange (RMCE) (Bateman et al. 2006) to create a gene drive line (See Chapter 5). The newly created donor-plasmid was named K101 (Kyrou et al. 2018).

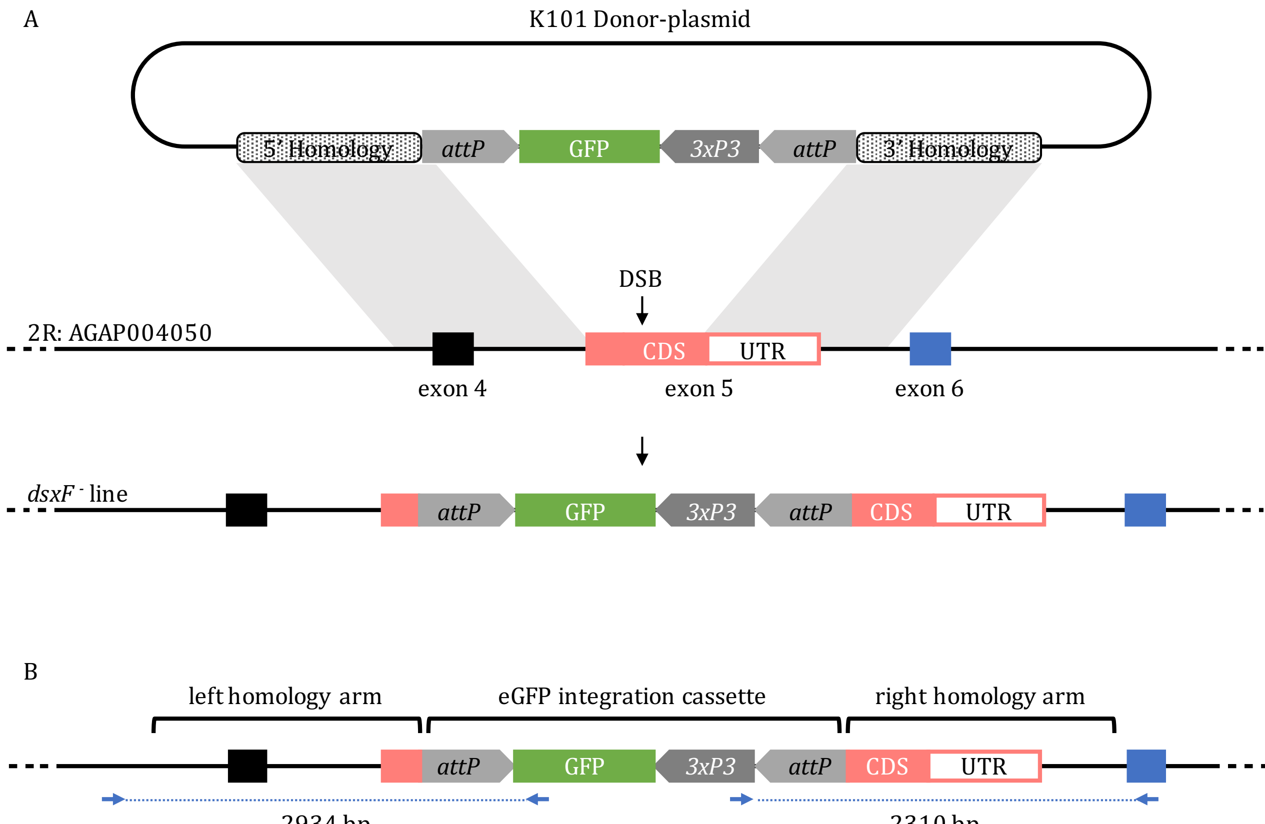


Figure 4.8 – The knock-in strategy using K101 donor-plasmid and verification of the integration. (A) The donor-plasmid contains sequences homologous to the regions upstream and downstream of the gRNA-T1 target site. Upon cleavage, the homology arms will serve as template for HDR and thus incorporate in the target site the cassette containing the *attP* sequences and the eGFP marker. The cassette disrupts the coding sequence of exon 5 (red) generating the *dsxF* $^{-}$ line. (B) Primers (arrows) that bind outside the sequences used to amplify the homology arms and built K101 were used to verify correct integrations of the cassette in the genome. The black and blue boxes represent exon 4 and 6 respectively and the PCRs product length is indicated. Elements are not drawn to scale.

4.3 Disruption of the female-specific isoform of *doublesex*

Both p165.10 and K101 were co-injected into newly laid wild-type embryos and surviving G_0 larvae carrying the plasmids in transient were identified by the expression of the eGFP marker in their lower abdomen and anal papillae (Figure 3.1). Progeny from G_0 mosquitoes, with successfully integrated cassette into their genome, were identified as eGFP$^+$ larvae among the progeny. The *3xP3* promoter regulating the expression of the eGFP marker is active in the optic lobe of the head and the abdomen of the larvae (Figure 3.). G_1 eGFP$^+$ transformants were individually crossed to wild type to separate possible different integration events of the knock-in cassette in the genome. Correct integrations were verified with PCR reactions on G_2 eGFP$^+$ progeny obtained from every G_1 mosquito founder (Figure 4.8B). To detect only the correct

integration within exon 5, the primers used to verify the integration were designed to bind outside of the sequences used to amplify the homology arms for K101 donor-plasmid (Figure 4.8B) (Kyrou et al. 2018).

4.3.1 The *doublesex* exon 5 disruption is not lethal

The knock-in cassette resulted in the complete disruption of the coding sequence of exon 5 of the dsx gene (referred to as *dsxF* $^{-}$ hereafter) (Kyrou et al. 2018). In silico analysis predicts the addition of a random 57-amino acid sequence before the first stop codon is encountered on the integrated cassette. To determine the phenotype of the mutation, I intercrossed heterozygous individuals *(dsxF* $^{+/-}$*)* to obtain all the genotypes to monitor their development and survival. Offspring were separated as L1 larvae into heterozygous *(dsxF* $^{+/-}$*)*, wild type *(dsxF* $^{+/+}$*)*, and homozygous individuals *(dsxF* $^{-/-}$*)* based on the intensity of the fluorescent marker using a Complex Object Parametric Analyzer and Sorter (COPAS) instrument (Marois et al. 2012). A 'strong' intensity of the eGFP reporter indicates the presence of two *dsxF* $^{-}$ alleles, whereas a 'weak' intensity is a sign that the larva carries only one *dsxF* $^{-}$ allele. Lack of the eGFP reporter indicates the absence of *dsxF* $^{-}$ alleles. Initially, a total of 3,197 L1 larvae were separated based on the presence of the eGFP marker. Out of those, 2,387 were eGFP^{+} (74.7%) and 810 were eGFP^{-} (25.3%), indicating that *dsxF* $^{-}$ allele is not lethal up to the L1 developmental stage. A chi-square test using the expected Mendelian ratio 3:1 returned a high P value (P =0.67) supporting that assumption. To support this further, I additionally separated 1,053 L1 larvae based on their genotype resulting in 262 *dsxF* $^{-/-}$ (24.9%), 523 *dsxF* $^{+/-}$ (49.6%) and 268 *dsxF* $^{+/+}$ (25.5%) larvae, which represent the expected 1:2:1 Mendelian ratio (P =0.94).

4.3.2 Females *dsxF* $^{-/-}$ present sexual developmental abnormalities

All genotype groups were allowed to grow separately up to the pupal stage where I could confidently sex them into males and females. I have retrieved male and female ratios close to the expected 1:1 for both the *dsxF* $^{+/-}$ (260:258) and *dsxF* $^{+/+}$ (130:138) groups, whereas I was unable to visibly sex the pupae obtained from the *dsxF* $^{-/-}$ group (n =259) because all the pupae in this group carried the male-specific pupal pattern. When emerged, half of the individuals of this group were normal males and the rest carried both male and female phenotypic characteristics (Kyrou et al. 2018). To confidently separate the males from the females I crossed the *dsxF* $^{-}$ line with another transgenic line that carries a red fluorescent protein (RFP) marker on the Y chromosome (Bernardini et al. 2014). The Y^{RFP} line carries no other genetic alteration in the genome and allowed me to track the Y chromosome and thus separate the male (XY) mosquitoes from the female (XX). I successfully crossed *dsxF* $^{+/-}$ female mosquitoes with Y^{RFP} males and intercrossed the offspring to obtain the *dsxF* $^{-/-}$ genotype. As expected, the mosquitoes that carried both male and female characteristics were all genetically females (Figure 4.9). These 'intersex' mosquitoes, instead of

the pilose female-specific antenna, carry a plumose antenna with longer flagellomeres, which is a male-specific characteristic. All of them also carried claspers, although dorsally rotated, and they do not have the female-specific short cerci. They are also equipped with palps with a tendency to flatten at the edge (Figure 4.10). These phenotypic abnormalities are absent from the females of the *dsxF* $^{+/-}$ group, indicating that the gene may be haplosufficient. Furthermore, all males from both the *dsxF* $^{+/-}$ and *dsxF* $^{-/-}$ classes seemed to be unaffected by this disruption of the AgdsxF isoform (Kyrou et al. 2018), something ideal if we want ot build a gene drive system targeting this exon.

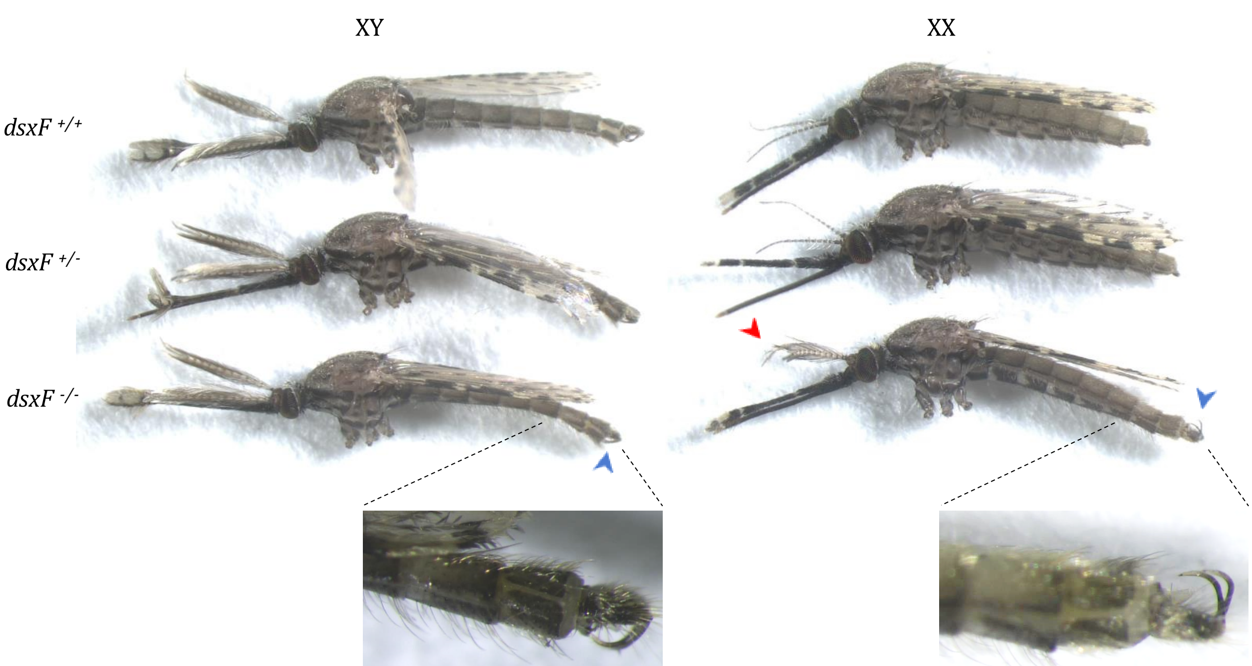


Figure 4.9 – Phenotypic comparison of the male and female mosquitoes of the *dsxF* $^{-}$ line. Homozygous *(dsxF* $^{-/-}$*)*, heterozygous *(dsxF* $^{+/-}$*)* and wild type *(dsxF* $^{+/+}$*)* were separated into males (XY) and females (XX) with the help of a YRFP marker. The XX *dsxF* $^{-/-}$ carried both male- and female-specific characteristics like the plumose antennae (red arrow) and claspers (blue arrows), although dorsally rotated. These characteristics were not present on the XX *dsxF* $^{+/-}$ mosquitoes, hinting that the AgdsxF isoform is haplosufficient. All the male classes seemed not to be affected by the disruption of the female-specific exon. Image reproduced with permission of the rights holder, Copyright © 2018, Springer Nature, Kyrou et al. 2018.

Dissections of intersex mosquitoes revealed the absence of spermatheca and the presence of male accessory glands (MAGs) that are present only in male mosquitoes. The intersex mosquitoes were also checked for the presence of ovaries and I was able only to retrieve a deformed testis-like tissue without any obvious cell structure from 9 out of 50 mosquitoes (≈20%) examined (Figure 4.11). On the contrary,

Figure 4.10 – Closer inspection of the antenna system and genitalia of *dsxF*$^{-}$ mosquitoes. Homozygous *(dsxF*$^{-/-}$*)*, heterozygous *(dsxF*$^{+/-}$*)* and wild type *(dsxF*$^{+/+}$*)* were separated into males (XY) and females (XX) with the help of a Y^{RFP} marker. The XX *dsxF*$^{-/-}$ carried both male- and female-specific characteristics that were mostly visible at the head of the mosquito and the abdomen. Although females, the antenna system resembles that of males and the palps (blue arrows) are not streamlined like in wild-type females. All *dsxF*$^{-/-}$ females carry a pair of dorsally rotated claspers (red arrows) that are a male-specific structure. These characteristics were not present on the XX *dsxF*$^{+/-}$ mosquitoes hinting that the AgdsxF isoform is haplosufficient. All the male classes were seemed not to be affected by the disruption of the female-specific exon.

the female mosquitoes of the *dsxF*$^{+/-}$ group did not present any of these abnormalities but rather were indistinguishable with the *dsxF*$^{+/+}$ group. Finally, all male groups were equipped with MAGs and testicles as expected, indicating again the female specificity of the *dsxF*$^{-}$ mutation.

4.3.3 Female *dsxF*$^{-/-}$ are sterile and have short longevity

An extensive phenotypic analysis was performed on the *dsxF*$^{-}$ line to determine the fecundity and survival rates of the different genotypes, which are important parameters to know to build a successful gene drive system. I crossed separately males and females (n =50) from both the *dsxF*$^{+/-}$ and *dsxF*$^{-/-}$ classes with wild-type male and female mosquitoes (n =50) and set the females of each cross to lay eggs individu-

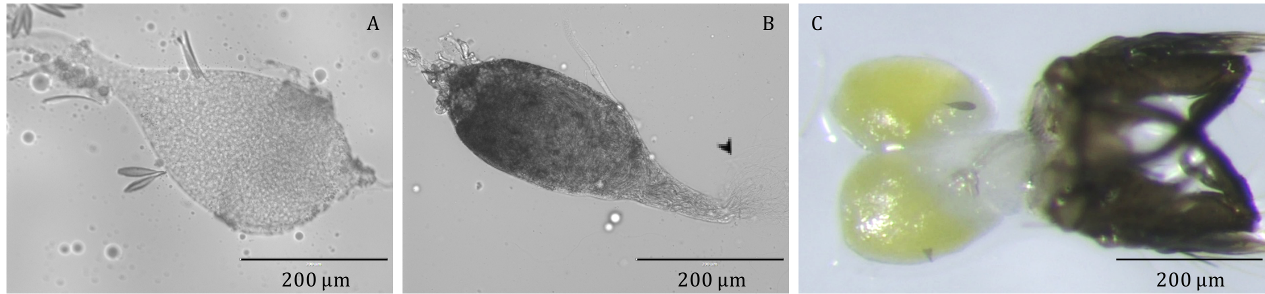


Figure 4.11 – Comparison of the internal reproductive tissues of *dsxF*$^{-}$ mosquitoes. Dissections of intersex individuals revealed the presence of a testis-like structure with no obvious cell organisation (panel A). Although the tissue was the same size and shape with testicles retrieved from *dsxF*$^{+/+}$ males (panel B) they differed in their consistency and structural organisation. Arrow in panel B shows the presence of sperm escaping the sperm duct (arrow). Images were taken using the EVOS imaging system. The *dsxF*$^{-/-}$ females also possess MAGs (panel C), which is a male-specific tissue and absent from female wild-type mosquitoes. Image reproduced with permission of the rights holder, Copyright © 2018, Springer Nature, Kyrou et al. 2018.

ally. I also included a wild-type cross of 50 male and 50 female mosquitoes as a control. An interesting observation was the inability of the intersex females to perform a blood meal. Surprisingly the intersex females were able to sense the blood source and gather to take a blood meal but they were unable to pierce the membrane and feast on the blood. Since the blood meal is an essential pre-requisite for the formation of eggs, the intersex mosquitoes were not able produce any progeny. On the contrary, *dsxF*$^{+/-}$ females and both *dsxF*$^{+/-}$ and *dsxF*$^{-/-}$ males showed wild-type levels of fecundity, indicated by the clutch size and larval output (Figure 4.12). The average larval output per mated female ($n \geq 43$) was 129.5 ±6.3 for the wild-type control, 114.8 ±8.5 for *dsxF*$^{-/-}$ males, 129.2 ±7.1 for *dsxF*$^{+/-}$ males and 129.3 ±5.9 for *dsxF*$^{+/-}$ females – (± indicates standard error of the mean - s.e.m). I performed a Kruskal-Wallis non-parametric test to characterise the difference between the means of the groups that did lay larvae, and as expected it was not significant ($P = 0.89$).

A survival assay was also performed to reveal the effects the disruption of exon 5 has on the longevity of the intersex mosquitoes. For this experiment I tested males and females ($n > 110$) from both the wild type (*dsxF*$^{+/+}$) and homozugous (*dsxF*$^{-/-}$) groups. Both males and females were mixed in three large cages and the number of dead mosquitoes was recorded every day. The sex and genotype of the female and intersex mosquitoes were easily detected based on external phenotypic characteristics using a stereomicroscope, whereas the male mosquitoes were separated using a fluorescent microscope by detecting eGFP expression in the eyes of the insect. From the survival graphs it is clear that intersex mosquitoes do not live as long as wild-type females (Figure 4.13). Their median survival ($\tilde{\chi} = 16$ days) does not represent the

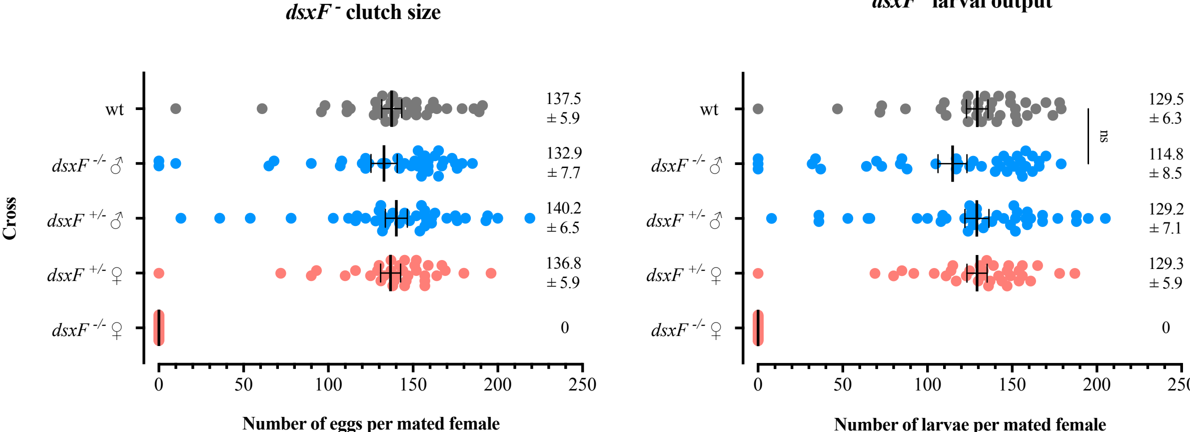


Figure 4.12 – Clutch size and larval output of the *dsxF*$^{-}$ line. Heterozygous *(dsxF*$^{+/-}$*)* females were able to blood feed and produce eggs and larvae equivalent with the wild-type control (wt), whereas homozygous homozygous *(dsxF*$^{-/-}$*)* females were not able to produce eggs, supporting the ideal that the female-specific isoform of dsx is haplosufficient. As expected, male *dsxF*$^{-/-}$ and *dsxF*$^{+/-}$ mutants were unaffected by the disruption of the female-specific exon of *doublesex* . A Kruskal-Wallis test was performed to analyse the variation within the groups and as expected all genotypes apart from the *dsxF*$^{-/-}$ females present no significant differences when compared with the control. Bars and error bars indicate the mean and standard error of the mean (s.e.m) respectively. Image reproduced with permission of the rights holder, Copyright © 2018, Springer Nature, Kyrou et al. 2018.

longevity of the *dsxF*$^{+/+}$ females ($\tilde{\chi}$ =31 days) but is closely mirroring that of the male mosquitoes ($\tilde{\chi}$ =14 days), which can be explained if we consider that the intersex females have more male-specific characteristics. Also, another explanation could be the inability of the intersex mosquitoes to mate, since mated female mosquitoes live longer than unmated ones (Villarreal et al. 2018). Male mosquitoes seem to be unaffected from the disruption of the female-specific exon, indicating once more the sex specificity of the phenotype in the *dsxF*$^{-}$ line.

4.4 Discussion

The *doublesex* gene in *An. gambiae* plays a key role in the sexual development and differentiation of the mosquito. The gene is alternatively spliced to produce a male- and a female-specific isoform through exon shuffling. Here we provide the first ever evidence that the female-specific isoform of the gene is essential for the sexual maturation of the female mosquito and the development of primary and secondary female characteristics. Using a CRISPR/Cas9 system approach, we disrupted the coding sequence of the female-specific exon of the gene producing female mosquitoes with an intersex phenotype that carry both male- and female-specific characteristics. These females lack female-specific tissues such as ovaries and sper-

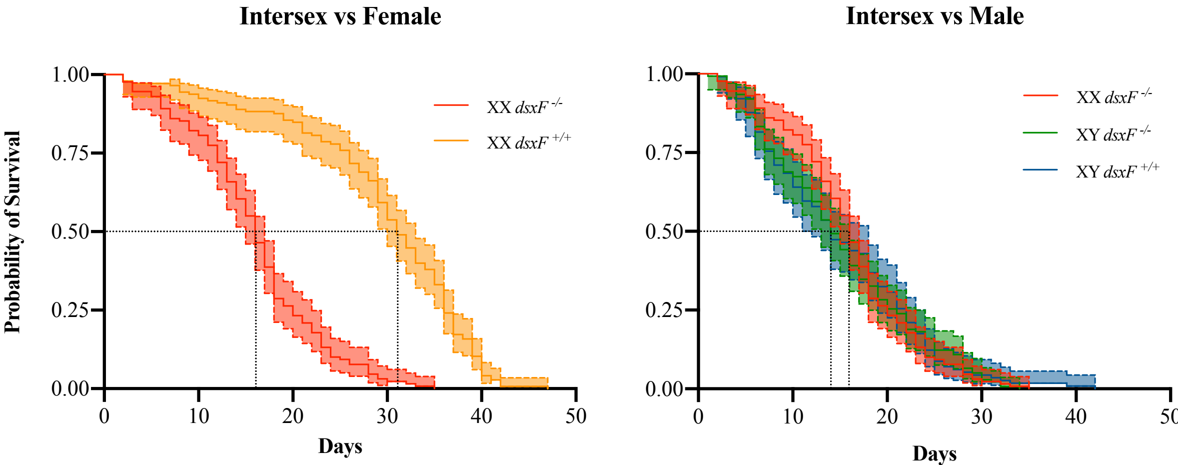


Figure 4.13 – Assay comparing the longevity of homozygous *dsxF* $^{-}$ and wild-type mosquitoes. The longevity of female mosquitoes is substantially reduced when they carry the *dsxF* $^{-}$ allele in homozygosity *(dsxF* $^{-/-}$*)*. Their median survival ($\tilde{\chi}$ =16 days) is significantly different (*P* <0.0001) when compared with wild-type females, which have a median survival of 31 days (left). On the contrary, the median survival of the intersex mosquitoes is not different (*P* =0.96) from the survival of wild-type and *dsxF* $^{-/-}$ male mosquitoes ($\tilde{\chi}$ =14 days) (right), supporting the findings that *dsxF* $^{-/-}$ females have more male-specific characteristics than female-specific. The data were analysed using the Log-rank (Mantel-Cox) test and the graphs represent the fraction of survival with 95% confidence intervals.

matheca and have male-specific antennae system and genitalia, as well as internal male-specific tissues. Furthermore, although these females are attracted to blood source, they cannot pierce the membrane and blood feed to produce progeny. Finally, the intersex females present survival rates that differ substantially from wild type females but are similar to males. The facts that these abnormalities are not shared by females that carry the mutation in heterozygosity and that male mosquitoes seem not to be affected by the mutation, indicate that the isoform is haplosufficient and specific to females. These characteristics are essential properties for a gene in order to be used in a gene drive system aiming at suppressing mosquito populations.

Both isoforms of the *doublesex* gene encode for characteristic domains that are conserved within the *doublesex* family in insects. Specifically, the CDS of the female-specific exon of the gene appears to be highly conserved, not only within the *An. gambiae* species complex but also within the Anophelinae subfamily, which includes species that diverged from each other more than 60 million years ago. This high conservation is an indication that the sequence is under great selective pressure and might not allow indel mutations or small variants that change the sequence but retain the reading frame of the protein. Supporting this hypothesis is the fact that the sequence of this particular exon is up to 80% identical with distant related species such as *D. melanogaster*, which shared a common ancestor with *An. gambiae* over 250

million years ago. Combining this unique characteristic of the female-specific sequence with the female specificity and haplosufficiency, makes the female-specific exon of the doublesex gene an ideal candidate to test a gene drive system for controlling mosquito populations. In the next chapters I build a gene drive system against this exon and test its abilities to overcome resistance and suppress mosquito populations.

Chapter 5

A gene drive targeting *doublesex* for genetic control

5.1 Development of a gene drive system

Gene drive systems are genetic technologies that have the ability to bypass their own inheritance by copying themselves from one chromosome to the homologous one in a process called 'homing'. In theory, gene drives can increase their frequency within fast reproducing populations such as insects, even when starting at very low frequencies. The use of such systems for pest population and vector control has been suggested over 15 years ago (Burt 2003) but recently they have been at the centre of attention with the discovery of easily programmable endonucleases such as the CRISPR/Cas9. The systems can be used to fight vector borne diseases such as malaria, by either reducing the reproductive capacity of vector populations (population suppression) or by hindering their ability to transmit the disease (population replacement /modification). For a complete background on gene drive technology see Chapter 1 – Introduction.

In our lab, we have been designing and testing gene drive systems for population suppression against the main vector for malaria in Africa, *Anopheles gambiae* . We have been testing such systems against haplosufficient female fertility genes that stop the females from producing progeny when they are disrupted in homozygosity. Such target genes do not affect the fertility of males and heterozygous females and thus allow the gene drive system to be passed in the next generations and increase in frequency within a population. By using germline-specific promoters to control the expression of the Cas9 nuclease, we restrict the homing process of the gene drive in the germline of the mosquito and avoid affecting somatic tissues in heterozygous females that might be essential for fertility. The choice of the promoter driving the expression of the nuclease is essential and must be thoroughly considered before designing a gene drive system. In this chapter, I describe how I built a gene drive system against the female isoform of the *dsx*

gene and I assess its efficiency to home in the germline of the mosquito. Finally, I compare the advantages of using a new germline-specific promoter for the system and the implications on the fertility of the mosquitoes, which was a major issue in previous gene drive systems.

5.2 Design and creation of a CRISPR-based gene drive vector targeting exon 5 of the *doublesex* gene

The discovery of easily programmable endonucleases such as the CRISPR/Cas9 has provided the tools for fast and easy engineering of gene drive systems. The Cas9 endonuclease can be easily reprogrammed to target a new sequence in the genome by just changing the sequence of the gRNA. To create a CRISPR-based gene drive I must first design and construct a gene drive cassette housing all the necessary elements for the gene drive system (referred to as CRISPR^h cassette hereafter). Based on previous gene drive systems (Hammond et al. 2016, Gantz et al. 2015, Hammond et al. 2020a), a CRISPR^h cassette should mainly consist of the Cas9 gene and the gRNA sequence targeting the desired locus, but also a fluorescent marker to visualise the spread of the drive (Figure 5.1). I use the same human codon-optimised version of the *Streptococcus pyogenes Cas9* gene *(hCas9)* used in p165 CRISPR-helper plasmid (Hammond et al. 2016) and the gRNA-T1 spacer against exon 5 of the *doublesex* gene used to create the *dsxF*$^-$ knockout line (See Subsection 4.2.1). The complete CRISPR^h cassette needs to be flanked by two *attB* sequences in reverse orientation for it to be incorporated in the genome via ϕC31 recombinase-mediated cassette exchange (RMCE) event (Bateman et al. 2006), using the two *attP* sequences present in the *dsxF*$^-$ line (See Subsection 4.2.2). Finally, the CRISPR^h cassette needs to be cloned into a plasmid backbone to transform and prepare for subsequent microinjections in mosquito embryos.

Figure 5.1 – The CRISPR^h cassette used to build the gene drive against exon 5 of the *doublesex* gene. The CRISPR^h cassette contains a source of nuclease and a gRNA to direct cleavage activity on the locus of interest, but also a fluorescent marker to visualise carrier mosquitoes under a fluorescent microscope. The nuclease is under the expression of a germline-specific promoter, in this case the *zero population growth promoter (zpg)*, in order to restrict the homing process in the germline of the mosquito. The gRNA targeting exon 5 is under the regulation of the *Agam U6snRNA* promoter to ensure the gRNA is transcribed without a 5' cap and a 3' poly-A tail in order to be able to complex with the Cas9 protein. Finally, all the components are flanked by two reverse *attB* sequences for directed RMCE event in the *doublesex* locus. Elements are not drawn to scale.

The p165 CRISPR-helper plasmid used previously to create p165.10 (See Subsection 4.2.2) carries some of the components necessary to create the CRISPR-based gene drive vector targeting *dsx*. The plasmid carries a suitable backbone for bacterial transformation and replication and a CRISPR^{h} cassette used in the gene drive system reported previously in Hammond et al. 2016. The CRISPR^{h} cassette houses a gRNA spacer-cloning site and backbone (Hwang et al. 2013) under the control of the *Agam U6snRNA* promoter, which is a polymerase III-transcribed promoter and ubiquitously expressed in the mosquito (Konet et al. 2007). The cassette also carries a red fluorescent protein (RFP) coding sequence under the expression of the *3xP3* promoter and an *hCas9* nuclease gene under the regulation of the germline-specific *vas2* promoter and terminator. Finally, the CRISPR^{h} cassette is enclosed within two reverse *attB* recombination sequences for subsequent RMCE event.

For the gene drive targeting *dsx*, I could have used p165 and modify it to contain the gRNA-T1 spacer to target exon 5 of the *dsx* gene. Previous studies though have shown that the *vas2* regulatory elements controlling expression of the Cas9 in that plasmid might not be ideal for a gene drive system targeting female fertility. The *vas2* promoter appears to be expressed in a small portion of somatic cells, referred to as 'somatic leakiness', which was related with a significant female fertility cost in mosquitoes (Hammond et al. 2016) and mutant cuticle tissue cells in *D. melanogaster* (Port et al. 2014). This reduction in fertility was associated with the expression of the nuclease in the somatic tissues surrounding the germline, leading to mutations that are detrimental for female fertility. Specifically, when the *vas2* regulatory elements were used previously to control the expression of the gene drive targeting the female fertility gene *AGAP007280*, the carrier females were only 8% fertile when compared with wild-type females (Hammond et al. 2016). To overcome this problem, I have replaced the regulatory elements of the *hCas9* gene with the promoter and terminator of the *zero population growth (zpg)* gene *(AGAP006241)*, the homologous gene to *D. melanogaster zpg* gene.

Zpg is known to play a role in forming gap-junctions in the germline of both the male and female *D. melanogaster* and is shown to be concentrated as mRNA in the pole cells of the embryo (Tazuke et al. 2002). The gene in *An. gambiae* is expressed in the germline of both the male and female mosquitoes (Baker et al. 2011) and plays a key role in the development of the gonads in both sexes, as shown from RNA interference (RNAi) experiments (Magnusson et al. 2011). When the *zpg* regulatory elements were assessed in a gene drive system targeting the *AGAP007280* gene, they were shown to improve the fertility of carrier heterozygous females from 8% observed previously to 54.9% when compared with wild types (Hammond et al. 2020a). This is an indication that the *zpg* regulatory elements, although failing to restore full female fertility perhaps due to a small degree of somatic leakiness, are more germline-restrictive compared to the *vas2* gene and much more suitable to use in a gene drive system targeting female fertility.

I used VectorBase, www.vectorbase.org (Giraldo-Calderón et al. 2015) to look at the genomic location and sequence of the *zpg* gene in *An. gambiae* (AgamP4 assembly) and design suitable primers to amplify both the promoter and terminator of the gene. To ensure to capture all regulatory sequences and UTRs of the gene, a minimum of 1000 bp were amplified upstream and downstream of the coding sequence of the gene (Hammond et al. 2020a). The primers were designed to carry Gibson assembly overhangs to incorporate the promoter and terminator fragments upstream and downstream of the *hCas9* gene in the final gene drive vector. The *hCas9* gene was amplified without the *vas2* regulatory elements from p165, using primers with homologous overhangs to the promoter and terminator of the *zpg* gene. The gRNA-spacer cloning site, the *3xP3::RFP* fluorescent marker coding sequence and the backbone of the plasmid with the *attB* recombination sequences were all excised from p165 as one fragment using restriction digestion. All four fragments were assembled into one vector which I named p174 (Figure 5.2).

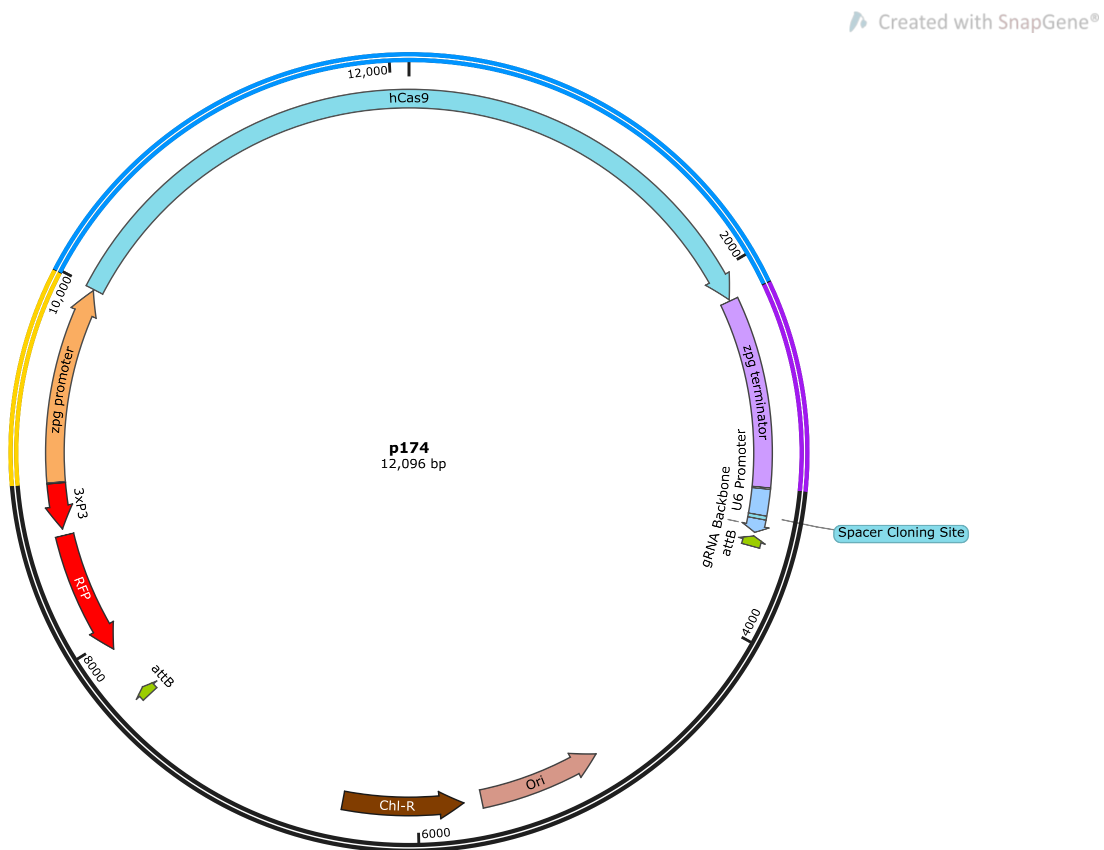


Figure 5.2 – Map of the p174 CRISPR gene drive vector. The plasmid was created using Gibson assembly reaction to combine the backbone (black), the promoter (orange), the terminator (purple) and the *hCas9* (blue) fragments. The two *attB* sequences (green) flank the CRISPR^{h} cassette which will be incorporated in the genome using RMCE.

5.3 Generation of a gene drive line on exon 5 of the *doublesex* gene

To create a gene drive system targeting exon 5 of the *dsx* gene , I have cloned the gRNA-T1 spacer into p174 using Golden Gate cloning as I did before with the *dsx* CRISPR-helper plasmid p165.10 (See Chapter 4.2.2). The newly created plasmid, p174.10, was injected into mosquito embryos derived from the *dsxF* $^-$ line to promote RMCE events using the *attP* and *attB* recombination sites. To initiate the RMCE events, p174.10 was co-injected with a helper plasmid carrying the *Streptomyces phage* ϕC31 *integrase* gene under the expression of the *vas2* promoter (Volohonsky et al. 2015). The idea behind this is that the integrase gene will be expressed in the germline precursor cells of the embryo and initiate a recombination event between the two *attP* sites on the genome and the two *attB* sites on p174.10 plasmid, resulting in the exchange of the two intervening sequences. Successful recombination events were be visible under a fluorescent microscope among the G_1 progeny as eGFP$^-$ and RFP$^+$ larvae because of the exchange of the *3xp3::eGFP* marker of the *dsxF* $^-$ line with the *3xP3::RFP* marker of the CRISPR^h cassette (Figure 5.3A). Specifically, from 216 eggs injected, I retrieved 19 RFP$^+$ G_1 larvae and 4 GFP$^+$/RFP$^+$ individuals that were the result of putative complete and incomplete RMCE (Table 5.1). The 19 RFP$^+$ founders were individually crossed to wild type to investigate and verify successful recombination events. I performed PCR reactions on the RFP$^+$ G_2 progeny of some of the founders using primers that bind on the *hCas9* gene and the RFP marker of the CRISPR^h cassette (Figure 5.3B) to amplify successful integration events within the *dsx* locus (Kyrou et al. 2018).

				G_0 transients		G_1 transformants		
Line injected	**Plasmids injected**	**Eggs injected**	**Total larvae**	***dsxF*$^-$ transgenic**	***dsxF*$^-$ non-transgenic**	**RFP$^+$/GFP$^-$**	**RFP$^+$/GFP$^+$**	**Transient rate**
dsxF$^-$	p174.10 *vasa:integrase*	216	100	7	13	19	4	20%

Table 5.1 – Transient and transformant recovery from embryo microinjected to promote RMCE and create the gene drive line. A total of 216 eggs obtained from the *dsxF* $^-$ line were co-injected with both p174.10 CRISPR-helper plasmid and the *vasa::integrase* plasmid to promote RMCE in the germline. A total of 100 survivors (46.3%) were screened for expression of the RFP plasmid marker that indicates successful injection of the solution in the embryos (transients). Seven of the transients were carrying the *dsxF* $^-$ allele whereas 13 were wild type. Offspring from the seven *dsxF* $^-$ transient mosquitoes were screened for the switch of the eGFP marker of the *dsxF* $^-$ line for the RFP marker of the gene drive cassette. A total of 19 RFP$^+$/GFP$^-$ larvae were recovered and G_2 offspring were verified for successful RMCE.

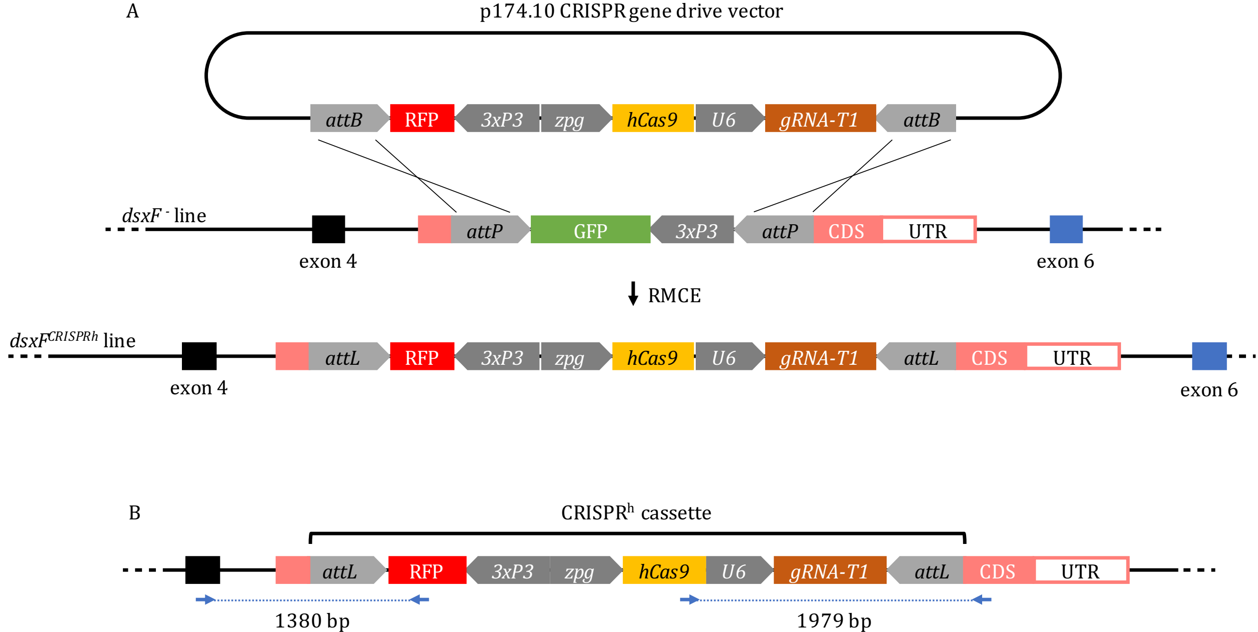


Figure 5.3 – Creation of the *dsxF^{CRISPRh}* line using RMCE and verification of the integration. (A) The *attP* and *attB* recombination sequences, located on the *dsxF$^-$* line and the p174.10 plasmid respectively, were used to replace the eGFP knock-in cassette of the *dsxF$^-$* line with the CRISPR^h cassette of the p174.10 gene drive vector using RMCE. Successful integration events were verified under a fluorescent microscope as RFP$^+$/eGFP$^-$ larvae. (B) Primers (arrows) that bind outside the destined location of the CRISPR^h cassette were used to verify the integration of the CRISPR^h cassette in the exon 5 locus. The black and blue boxes represent exon 4 and 6 respectively and the PCRs product length is indicated. Elements are not drawn to scale.

5.3.1 The gene drive poses a fertility cost on the female mosquitoes - the parental effect

Under perfect circumstances, the *hCas9* should be expressed in the germline of both males and females and complex with the gRNA targeting exon 5, leading to the cut of the wild-type allele and subsequent homing of the gene drive cassette in the wild-type locus (Figure 5.4). Previously, when we measured the fertility and transgenic rates of other gene drive lines, we noticed that the fertility cost and the homing rates associated with the CRISPR^h cassette were not evenly distributed among the progeny. We discovered that the origin of the CRISPR^h cassette played a major role in the fertility and also the homing rates of the offspring. Both male and female heterozygous for the gene drive mosquitoes presented different fertility and homing rates depending on the sex of the parent from which they inherited the CRISPR^h cassette (parental effect) (Hammond et al. 2020a). Different promoters appear to cause the Cas9 to be maternally and/or paternally deposited in the early embryo, either as transcript or protein, stochastically converting cells to homozygous null mutants (mosaicism). It has been proposed that the Cas9 cleaves the sequence of the homologous chromosome inherited from the wild-type parent, causing either early homing in the

embryo or non-homologous end-joining (NHEJ) mutations that affect the homing capabilities and fertility of the adult mosquito (Hammond et al. 2020a).

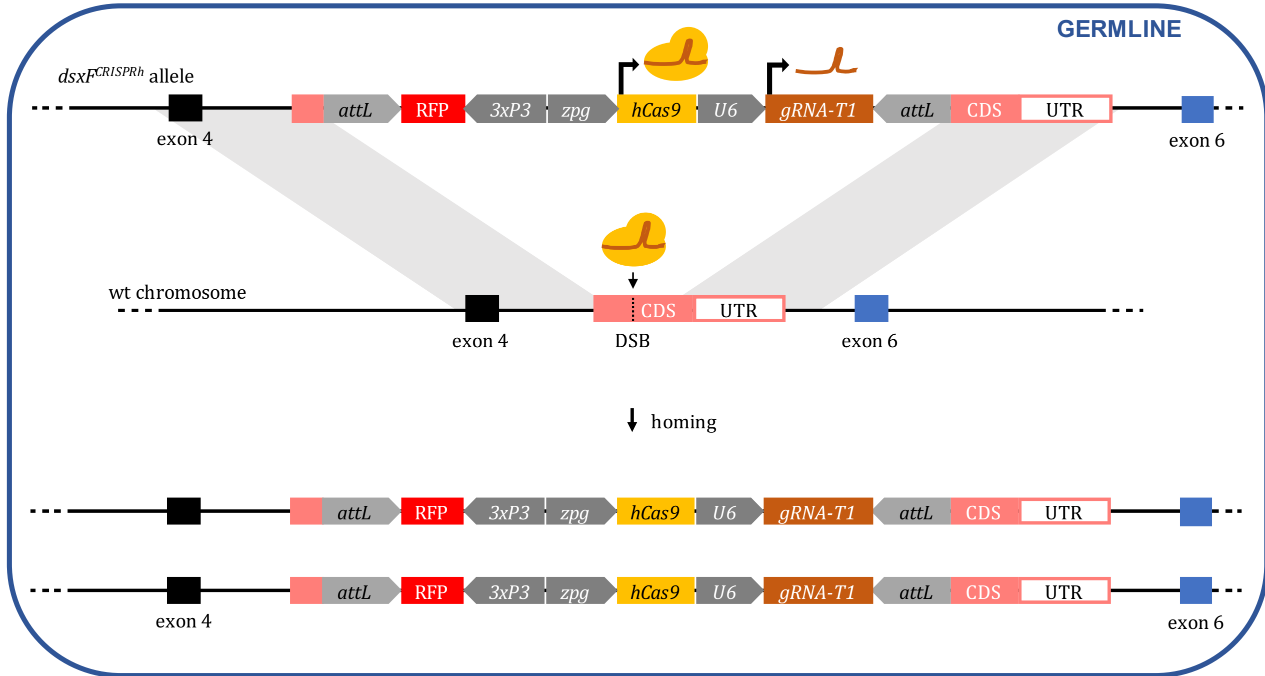


Figure 5.4 – The homing process of the CRISPR^h cassette. The nuclease of the gene drive is expressed in the germline of the mosquito and complexes with the gRNA-T1 to cleave the wild-type (wt) sequence on the wild-type chromosome. Using HDR, the cell attempts to repair the cut using the *dsxF*^{CRISPRh} allele, copying the CRISPR^h cassette in the wild-type chromosome (homing), disrupting the exon in the process. The homing reaction converts the germline cells into homozygous for the *dsxF*^{CRISPRh} allele and for the *Agdsx* knockout.

To assess fertility rates of heterozygous gene drive individuals for the *dsxF*^{CRISPRh} allele (referred to as *dsxF*^{CRISPRh}/+ hereafter), I performed controlled crosses with wild-type mosquitoes. I separated both *dsxF*^{CRISPRh}/+ male and female mosquitoes (n =50) from both *dsxF*^{CRISPRh}/+ male and female parents to assess the parental effect, and crossed them to 50 wild-type females and males respectively. Both the clutch size and larval output of every female were counted to assess the fertility of the line. Larvae from the crosses were also subsequently screened for the presence of the RFP marker to calculate the portion of offspring carrying the gene drive (see below). Clutch size and larvae hatched from at least 42 females from each cross revealed that male *dsxF*^{CRISPRh}/+ mosquitoes were as fertile as the wild-type control, irrespective of the sex of the parent from which they received the CRISPR^h cassette (Figure 5.5). On the contrary, female mosquitoes appeared to be impaired by the somatic leakiness of the *zpg* promoter. The effect was greater on females that received the CRISPR^h cassette paternally than those that received it ma-

ternally. Specifically, when compared with the control, we observed a relative clutch size and larval output of 72.6% and 44.7% respectively from 42 females that received the gene drive allele maternally, and significantly reduced rates of 26.5% and 7.3% from 46 females that received the allele paternally ($P < 0.0001$; Kruskal-Wallis test).

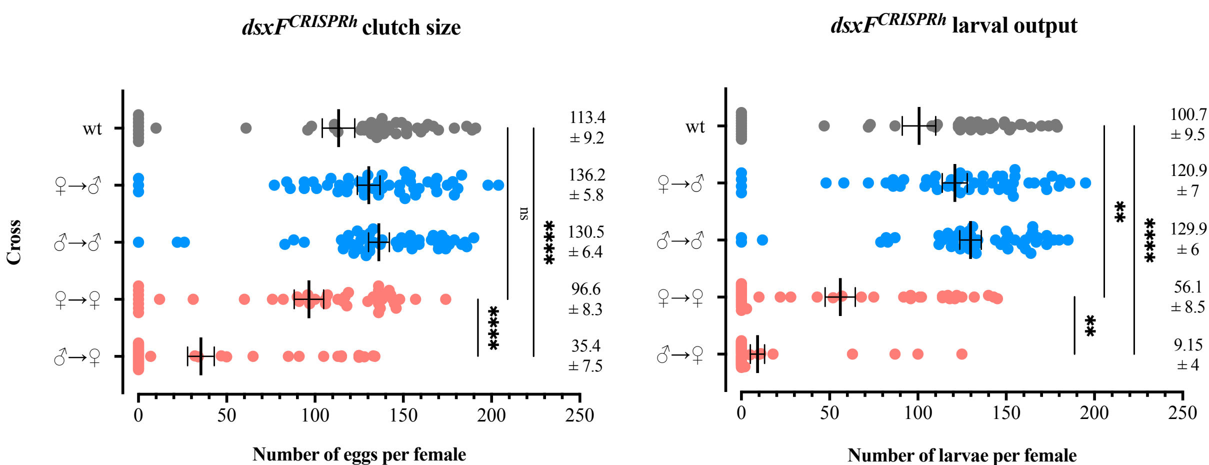


Figure 5.5 – Clutch size and fecundity of the *dsxF*^{CRISPRh} gene drive line. Males and females heterozygous for the gene drive line were assessed for their clutch size (left) and larval output (right) to investigate fertility issues associated with the *zpg* promoter. To reveal any parental effect, the groups were separated based on whether they received the gene drive allele maternally or paternally (e.g. ♀→♂represents a heterozygous male that received the gene drive allele maternally). Although both female groups present a reduction in both the clutch size and larval output that hints to some somatic leakiness caused by the *zpg* promoter, females that received the allele paternally have a significantly reduced fertility compared with females that received the allele maternally. This difference indicates some nuclease activity in the early embryo, probably deposited by the sperm, that causes mosaicism in the female offspring, impairing their ability to produce progeny. Bars and error bars indicate the mean and standard error of the mean respectively (**, $P < 0.01$; ****, $P < 0.0001$; Kruskal-Wallis test).

Although an overall reduction on female fertility was expected by using the *zpg* promoter to drive the expression of the Cas9, the extra fertility cost observed on females that received the CRISPR^{h} cassette paternally was not in line with previous gene drives using the same regulatory elements (Hammond et al. 2020a). When targeting the female fertility gene *AGAP007280*, female mosquitoes were equally fertile irrespective from which parent they received the CRISPR^{h} from (Figure 5.6). Although reduced, when compared with wild types, those females presented a relative larval output of 47.1% and 54.9% when receive the CRISPR^{h} cassette maternally and paternally respectively. Since the only difference between the two gene drives is their location in the genome, this suggests that locus dependency may also affect the

function of the *zpg* promoter. A combination of somatic leakiness and enhanced paternal deposition may be an explanation why females that receive the gene drive targeting *dsx* appear to have a dramatic fertility cost. High deposition rates of the Cas9 in the early female embryo may contribute to more 'embryonic' homing or NHEJ mutations in the cells of the early embryo, affecting the somatic tissues and fecundity of the adult mosquito.

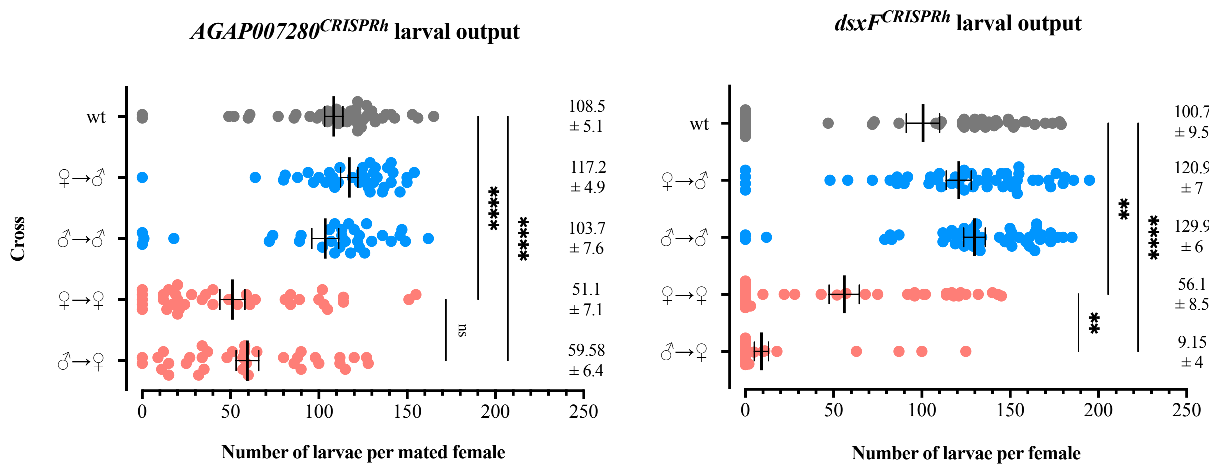


Figure 5.6 – Comparison of the fecundity of the *AGAP007280*^{CRISPRh} and *dsxF*^{CRISPRh} gene drive lines. In both gene drive lines, the fertility of heterozygous females is impaired when using the zpg promoter to target the female fertility locus. In the gene drive targeting the *AGAP007280* gene (left) we do not see evidence of parental effect since female offspring appear to have the same fertility whether they receive the gene drive allele maternally or paternally. On the contrary, females for the *doublesex* gene drive (right) that received the gene drive allele paternally (♂→♀) appear to have a substantially reduced fertility when compared with those that received the allele maternally (♀→♀), suggesting a paternal deposition of the nuclease in the early embryo that later on affects the phenotype of the adult females. This difference on the parental effect suggests that the genomic location of the target locus of a gene drive might also play a significant role to either the degree of nuclease deposition in the embryo or even to its own accessibility to the nuclease. Data for the *AGAP007280*^{CRISPRh} gene drive line were obtained from Hammond et al. 2019. Bars and error bars indicate the mean and standard error of the mean respectively (**, P <0.01; ****, P <0.0001; Kruskal- Wallis test).

Supporting this hypothesis of mosaicism due to early activity of the nuclease in the embryo, is the fact that among the progeny of male heterozygotes mated with wild type, and therefore containing only one copy of the gene drive allele, we retrieve female mosquitoes that have partially formed male characteristics (Figure 5.7). What is also noteworthy is that we never observe full conversion of tissues to the intersex phenotype but rather we detect a partial conversion, which is also random within the population. Since

the mosquitoes receive only one gene drive allele from their parent, this conversion is most likely the result of stochastic early activity of the nuclease in the embryo that converts the wild-type allele, through NHEJ or embryonic homing, to a null allele in a portion of cells.

Figure 5.7 – Partial conversion of heterozygous gene drive females to the intersex phenotype. Females heterozygous for the gene drive appear to carry partial characteristics of the intersex phenotype. Panels *A* and *B* represent $dsxF^{CRISPRh}$/+ females with an incomplete set of claspers (arrows), a male-specific characteristic, which are also present fully formed and dorsally rotated on the intersex phenotype (panel *C*). Even though females that are heterozygous for the $dsxF^{-}$ knockout do not carry any male-specific characteristics, a portion of $dsxF^{CRISPRh}$/+ females show evidence of some conversion of tissues to homozygosity, probably the result of nuclease activity in the early embryo, supporting the hypothesis of the parental effect. The lack of consistency between the samples also indicates strong evidence that the nuclease activity in the embryo is stochastic. Image reproduced with permission of the rights holder, Copyright © 2018, Springer Nature, Kyrou et al. 2018.

5.3.2 Heterozygous mosquitoes transmit the gene drive to more than half of the progeny

The larvae obtained through the fertility assays were subsequently screened for the presence of the RFP marker to assess the inheritance rate of the gene drive allele in the progeny of the $dsxF^{CRISPRh}$ line. From all four crosses, I retrieved transgenic rates between 94% and 100% in the progeny (Table 5.2), indicating that the homing process of the gene drive construct in exon 5 is highly efficient and able to bias its own inheritance in the offspring (Kyrou et al. 2018). In addition to homing, other repair events such as NHEJ and microhomology-mediated end-joining (MMEJ) are possible during the repair of a broken chromosome. Investigating the hypothesis that in those individuals that showed relatively low homing rates, these other pathways might be having a prominent effect, I looked at the genetic composition of offspring derived from 10 males and 1 female that showed transgenic rates below 90% (Figure 5.8). Sequence analysis of 3 RFP^{-} larvae from all 11 parents revealed 11 alleles out of 33 that were repaired using MMEJ, and 10 alleles that were repaired using partial homology-directed repair (pHDR). The rest 12 alleles were wild type that were possibly never cleaved by the nuclease or were repaired by a perfect NHEJ. The MMEJ repair was

powered by a pair of six-bp-long strings of nucleotides (GGTCAA) that are identical and surround the cut site. During MMEJ, both ends of the broken chromosome carrying the six-bp sequence recombine, resulting in the removal of any intervening nucleotides – in this example five - but also one of the two six-bp strings. This MMEJ behaviour results in an allele with an 11-bp deletion and an out-of-frame product that was predicted during the design of the gRNA and was one of the characteristics of gRNA-T1 that made it the final choice for this gene drive system (see Subsection 4.2.1).

		Larvae screened			
***dsxF^{CRISPRh}*/+ Parent**	***dsxF^{CRISPRh}*/+**	**RFP^{+}**	**RFP^{-}**	**Total**	**Transgenic rate**
Male	Males (44)	5,661	109	5,770	97.9%
	Females (9)	401	0	401	100.0%
Female	Males (43)	5,015	284	5,299	94.1%
	Females (24)	2,063	7	2070	99.2%

Table 5.2 – Screening results from the gene drive line targeting the *doublesex* gene. *dsxF^{CRISPRh}*/+ were separated based on the parent from which they received the CRISPR^{h} cassette and were crossed to wild type to determine the efficiency of the homing process of the line. All four groups presented high transgenic rates of up to 100%, indicating an efficient conversion of germline using the *zpg* promoter. The number in brackets indicate the total number of *dsxF^{CRISPRh}*/+ mosquitoes that produced progeny form the 50 that were mated to wild type.

Partial homology-directed repair (pHDR) result in a null allele of the target locus and an incomplete gene drive element and likely result as a consequence of the homology between the gRNA sequence located on the CRISPR^{h} cassette and the target sequence on the broken chromosome. Because the cleavage point on a gRNA of a Cas9 system is not located at the centre of its sequence but rather three bp upstream of the PAM sequence, one of the two ends of the broken chromosome results with more homology to the gRNA than the other. This difference, but also the alignment orientation of the gRNA in the genome, determines which part of the CRISPR^{h} cassette will be partially homed and incorporated in the broken allele. In our gene drive system, the forward alignment orientation of the gRNA to the genome results in the exclusion of the RFP marker and the *hCas9* from partially homed alleles (Figure 5.9). In any case, the presence of random synthetic elements in partially homed alleles interfere with the coding sequence of the target gene, producing inactive products.

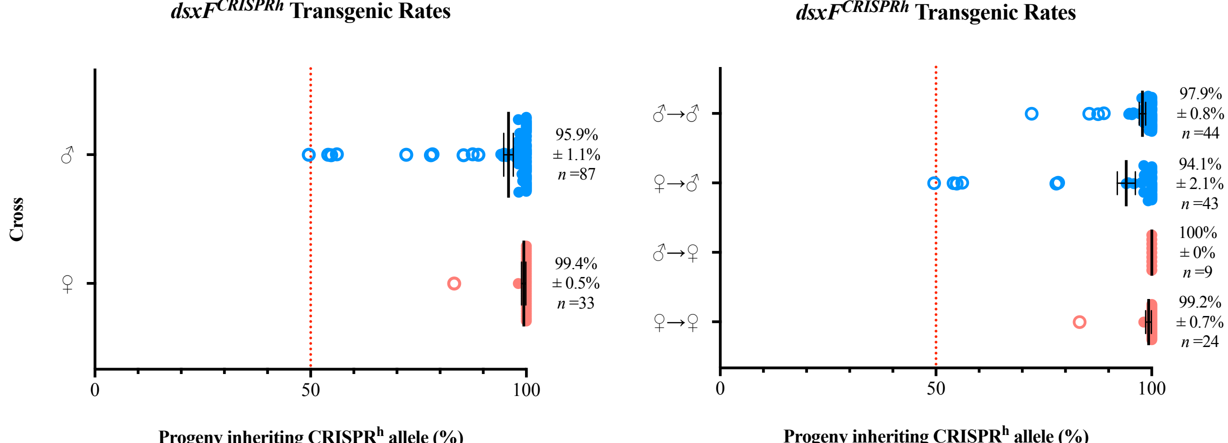


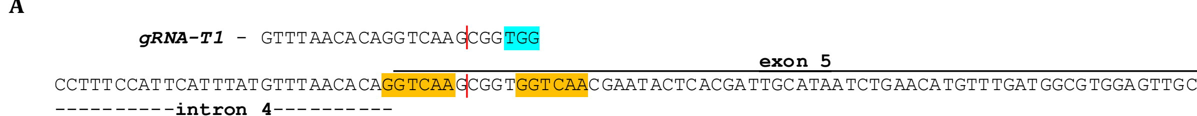
Figure 5.8 – Transgenic rates of heterozygous *dsxF*^{CRISPRh} individuals. Both male and female *dsxF*^{CRISPRh}/+ mosquitoes when mated with wild type, bypass Mendelian inheritance (red line) and pass the gene drive allele to more than 50% of the progeny, indicating an efficient homing process in the germline of the mosquito using the *zpg* promoter. The left panel shows the transgenic rates of *dsxF*^{CRISPRh}/+ males and females irrespective from which parent they received the CRISPR^{h} cassette. The right panel shows the same data grouped based on the parental inheritance of the CRISPR^{h} cassette (e.g. ♀→♂represents a heterozygous male that received the gene drive allele maternally). The empty points represent the 11 cases where RFP^{-} progeny was used to assess the meiotic NHEJ, MMEJ and pHDR. Bars and error bars represent the mean and the standard error of the mean respectively. Image reproduced with permission of the rights holder, Copyright © 2018, Springer Nature, (Kyrou et al. 2018) 2018.

5.4 Discussion

Gene drive systems are selfish genetic elements that bypass their own inheritance and increase in frequency among the progeny. The systems rely on the use of programmable nucleases that successfully digest specific DNA sequences in the genome. By placing the gene drive system within its own recognition sequence, the gene drive elements are copied between homologous chromosomes as a by process of HDR. If the genomic loci where the gene drive is placed is also essential for fertility, the presence of the gene drive mutates the locus and also spreads this mutation in the progeny using super-Mendelian inheritance. The use of gene drives for the control of mosquito borne diseases has been suggested more than a decade ago but the development of this technology has been recently boosted with the discovery of the CRISPR/Cas9 system. This easily programmable genetic toolbox was quickly adopted and pioneered to allow the rapid development of gene drive technology. The Cas9 nuclease has replaced previously used homing endonucleases due to its easy manipulation and fast reprogramming.

A gene drive system targeting the female-specific isoform of the *dsx* gene in *An. gambiae* was created using a simple cassette containing the *hCas9* gene under the expression of the *zpg* promoter, the gRNA-T1

A



B

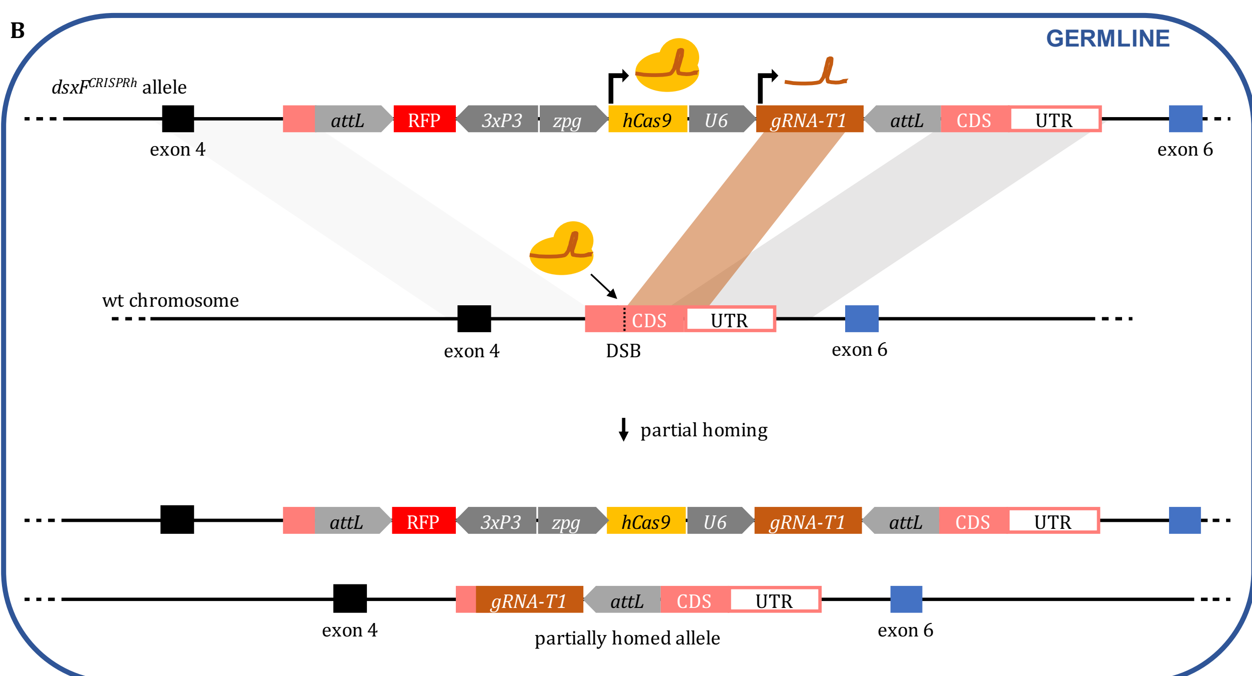


Figure 5.9 – Microhomology-mediated end-joining and partial homing at the target site. (A) Six-bp homologous strings (shaded in yellow) that surround the *dsxF*^{CRISPRh} target site (red dotted line) are used by the repair machinery to repair the broken chromosome. By recombining with each other, any intervening sequence is lost alongside with one of the two six-bp strings. In this example, the 11-nt sequence that is removed results in an out-of-frame product that does not restore the function of the locus and AgdsxF functionality. The PAM of the gRNA-T1 target site is shaded in blue. (B) The gRNA sequence located on the CRISPR^{h} cassette of the gene drive allele can in some cases interfere with the repair process of the broken wild-type (wt) locus in *dsxF*^{CRISPRh}/+ mosquitoes. The repair machinery sometimes cannot distinguish between the wild-type sequence of the gRNA located on the CRISPR^{h} cassette and the homology sequence flanking the cassette, using the former as a template to repair the broken allele. Based on the orientation of the gRNA coding sequence in the genome, the repair machinery uses either homology side to complete the repair. This results in pHDR and the generation of a partially homed allele. In this example, the repair machinery uses the gRNA sequence (brown) and right homology (dark grey) to repair the locus. The resulting allele includes parts of the CRISPR^{h} cassette that prevent it from producing a functional DSX protein.

under the expression of the *U6* promoter and an RFP marker for visualisation. The CRISPR^{h} cassette was inserted in exon 5 of the gene by using the *dsxF*$^{-}$ knockout line and RMCE. Homing individuals were able to pass the gene drive to the majority of the progeny, indicating high homing rates in the germline of the mosquito using the *zpg* promoter. Although a small reduction in fertility was expected in female gene drive individuals because of minor somatic leakiness of the *zpg* promoter, the *dsxF*^{CRISPRh} gene drive females presented additional fertility issues. The extra fertility cost observed in females was not distributed evenly among groups that inherited the gene drive allele maternally or paternally. A significant reduction of fertility was observed in females that received the gene drive paternally, indicating a locus- and sex-depended deposition of the nuclease in the early embryo. An alternative explanation may be the degree of NHEJ and MMEJ or embryonic homing in the early embryo that could be also locus-dependent. Despite all that, the *zpg* promoter is still more suitable than the previously used *vas2* promoter, and is currently the best germline-specific promoter we have to drive the homing process in the germline of the mosquito. The *zpg* promoter appears to be more germline restrictive, prohibiting the conversion of somatic tissues surrounding the germline to homozygous null mutants that affect the fertility of the female gene drive individuals. This is an important factor to consider when designing gene drives for population control, since the transgenic mosquitoes will have to compete with wild types in the field.

The gene drive targeting the female-specific exon of the *dsx* gene has all the necessary characteristics that allow it to be evaluated in caged mosquito populations. The target exon of the gene is haplosufficient and only homozygous females have abnormal sex development and are completely sterile. This female-specificity allows the gene drive heterozygous and homozygous male mosquitoes to develop normally and be fully fertile. Nonetheless, for a successful gene drive of this nature, the contribution of both carrier males and females is equally important for the spread of the gene drive. That means that even with fully fertile heterozygous and homozygous male mosquitoes, the gene drive cannot be sustained in the population if both the homozygous and heterozygous females are completely sterile. Even though the *dsxF*^{CRISPRh} heterozygous females are not fully fertile, they are still able to produce enough progeny to sustain a continuous release of gene drive alleles in the population. In the next chapter, I test the ability of the *dsxF*^{CRISPRh} gene drive to successfully spread in caged *An. gambiae* mosquito populations. I assess the efficiency and spread dynamics of the drive and compare them to previous systems that failed to increase in frequency and spread in the populations because of resistance.

Chapter 6

Testing the *doublesex* gene drive system in caged mosquitoes

6.1 Gene drive systems for population control

The ultimate goal for designing and building a gene drive system targeting the reproductive capacity of a major malaria mosquito vector, is to use it to complement current interventions that aim to reduce the burden of the disease. Testing a gene drive system in a lab setting is the first step out of a series of long and thorough checkpoints that ensure the safe and controlled deployment of a gene drive in the wild. Surprisingly, the creation of a gene drive and its assessment in caged mosquito populations are only a small section in the deployment process of a gene drive system (James et al. 2018). Because of their self-sustaining nature, gene drive systems have to pass through vigorous validation processes and regulatory checkpoints that ensure the safe introduction of such systems in nature. In the lab we can mimic a real-life release scenario of a gene drive system in caged mosquito populations, which can give us an idea of how the system might perform in the wild. Of course, the lab and nature are vastly different environments and at this point we can only speculate how the lab setting performance might be reflected in the wild. In this chapter, I describe the process I followed to test the $dsxF^{CRISPRh}$ gene drive system in small caged mosquito populations in the lab. I assess the dynamics of the system and compare them with previous attempts that failed to result in population suppression because of resistance at the target site. Finally, I reveal the benefits of using highly conserved genes as targets for such systems and how they affect the dynamics of the gene drive when tested in larger caged mosquito populations.

6.2 Assessment of the *dsxF*^{CRISPRh} gene drive system in caged *An. gambiae*

In a real-life release scenario, gene drives against malaria are most likely to be deployed through the release of male mosquitoes rather than females, since females contribute to the transmission of the disease. In order to decide on an optimal population size and release frequency for this experiment, a power calculation was performed to minimise the stochastic loss of the gene drive allele from the populations using a male-only release scenario in a laboratory setup. The power analysis revealed that in a population of 600 individuals, the stochastic loss of the gene drive allele was dramatically reduced almost to zero, even when as low as 20 *dsxF*^{CRISPRh} heterozygous males were used to seed the gene drive allele in the populations (Figure 6.1). To minimise the length of the experiment, we decided to seed the populations using a 12.5% gene drive allele frequency, which translates to 150 *dsxF*^{CRISPRh}/+ males in a total population of 600 individuals. Although higher than the minimum frequency suggested by the power analysis, this gene drive allele frequency was lower than the 50% used in the first generation gene drive targeting the *AGAP007280* gene (Hammond et al. 2016).

6.2.1 The gene drive imposes a reproductive load on the populations until they collapse

Initially, a total of two caged populations were used to test the dynamics of the gene drive when released at 12.5% allele frequency using only *dsxF*^{CRISPRh}/+ males. The caged populations, named cage 1 and 2, were monitored every generation for their clutch size and gene drive frequency within the populations, as indicated by the number of transgenic mosquitoes screened every generation. In both cages the frequency of transgenic mosquitoes increased from 25% to ≈69% within the first three generations but diverged after generation 4 (Figure 6.2). Both cages reached 100% transgenic frequency by generation 11 and 7 respectively and failed to produce any eggs in the next generation. The egg output of the cages was gradually being reduced with every generation, possibly as the result of the presence of more *dsxF*^{CRISPRh} female mosquitoes in the cages. The dynamics of the spread of the gene drive in both cages followed a mathematical model, which was created to predict the outcome of 20 stochastic simulations. The model was calculated using the properties of the *dsxF*^{CRISPRh} gene drive line, such as the inheritance bias of the gene drive allele, the parental effect and fertility rates of the line, but also the phenotype of the intersex individuals (Kyrou et al. 2018).

After the successful collapse of both cage 1 and cage 2 populations, I repeated the experiment using two more caged populations (cage 3 and 4). The cages were set up using the exact same conditions and monitored every generation for their clutch size and transgenic frequency. In both cages 3 and 4, the transgenic frequency increased from 25% to ≈50% within one generation, similarly to cages 1 and 2, but

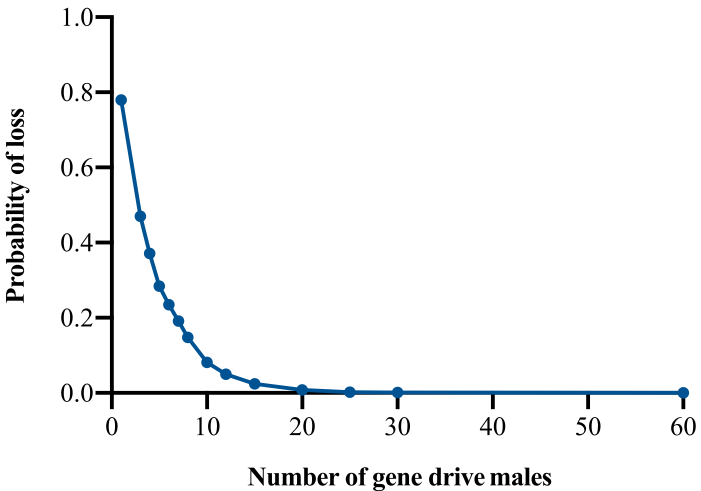


Figure 6.1 – Probability of the stochastic loss of the gene drive from the caged populations in a male-only release scenario. A power calculation was used to assess the probability of the stochastic loss of the gene drive from the caged populations based on the number of *dsxF*^{CRISPRh}/+ males used to introduce the gene drive in the cages. The model takes into consideration the size of the caged populations but most importantly the properties of the *dsxF*^{CRISPRh} gene drive line, such as fecundity and homing efficiency. In a population of 600 mosquitoes with equal male-to-female sex ratios, the probability to lose the gene drive allele by chance is dramatically reduced when as low as 15 *dsxF*^{CRISPRh}/+ males (5% of total males) are used to seed the gene drive into the populations. Each data point on the graph represents 1000 individual simulations of stochastic cage models as described in Kyrou et al. 2018. Image reproduced with permission of the rights holder, Copyright © 2018, Springer Nature, Kyrou et al. 2018.

later the two cages diverged, reaching 100% frequency at generations 7 and 17 respectively (Figure 6.2). Similarly with cages 1 and 2, cage 3 remained within the model prediction boundaries of the experiment, whereas cage 4 diverged from the predictions during the later generations. Interestingly, I observed a similarity between the dynamics of cages 1 and 4 and cages 2 and 3, prompting me to test the impact of the location of the cages in the insectary on the gene drive dynamics. The idea behind this was to test if changes in humidity or temperature had a noticeable effect in the function of the efficiency of the gene drive. For that, I set up two different sub-populations from cage 4 at generation 7 and reared them side by side in the insectary. Both populations (cage 4m and 4r) followed a very different transgenic frequency trajectory, indicating that there is no strong correlation between the location of the cages in the insectary and the dynamics of the gene drive in the populations. The successful suppression of the mosquito populations was the first of its kind and hinted at the advantage of using highly conserved genes like *doublesex* as targets when designing such interventions.

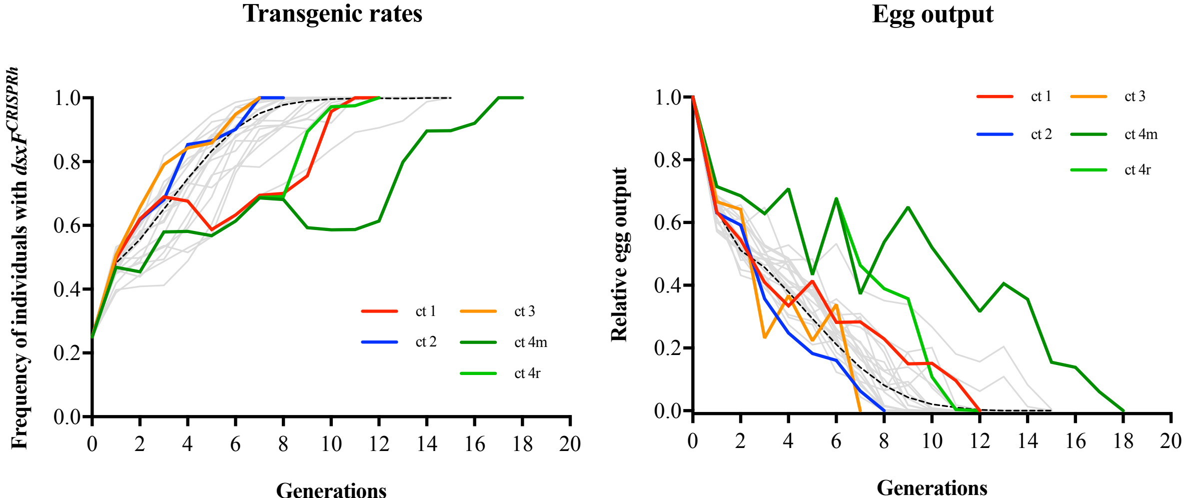


Figure 6.2 – Dynamics of the *dsxF*^{CRISPRh} gene drive in five caged mosquito populations. In total, five cages have been used to test the efficiency of the *dsxF*^{CRISPRh} gene drive to suppress mosquito populations in the lab. Both the transgenic frequency (left) and relative egg output (righ) were monitored in every generation. The gene drive targeting exon 5 gradually increased in frequency reaching 100% transgenic rate in all cages, leading to a complete population suppression. Cage 2 and 3 reached 100% transgenic frequency at generation 7, whereas cage 1, 4r and 4m reached 100% prevalence at generations 11, 12 and 17 respectively. The accumulation of *dsxF*^{CRISPRh} females in the populations but also the increasing production of intersex individuals towards the later generations imposed a dramatic reproductive load on the populations that eventually failed to produce eggs and sustain their numbers. The grey lines indicate 20 stochastic simulations that were calculated assuming a population size of 650 mosquitoes, whereas the dashed black line follows a deterministic model. Image reproduced with permission of the rights holder, Copyright © 2018, Springer Nature, Kyrou et al. 2018.

6.2.2 Target site resistance on the *doublesex* locus was not detected within the populations

To characterise the repertoire of indel mutations accumulated at the target site, but also their frequency within the cages, I have sequenced the mosquito populations at different generations using deep amplicon sequencing. The method, was designed to capture all alleles in the populations that were created as the result of MMEJ or NHEJ and ergo did not carry a gene drive cassette. Partially homed events were not able to be detected using this method because of limitation on the amplicon size of the method. The amplicon sequencing revealed the existence of indel mutations of different sizes in all cage trials, but none have managed to increase in frequenciy above 2.5% within the population of the cages. Specifically, the same 11-bp deletion that was detected during the assessment of the gene drive line (Figure 5.9) was found to reach a maximum of 1.45% frequency among the non-homed alleles within later generations of some

populations. The majority of other indels, including in-frame mutations that could theoretically restore the functionality of the AgdsF isoform were kept below 0.5% (Figure 6.3). A 4-bp out-of-frame deletion reached the maximum of 2.15% allele frequency in cage trial 2 at generation 7, most probably because all mosquitoes of the population were either heterozygous or homozygous for the gene drive allele by that generation. Nevertheless, in all instances, the most prominent allele detected using this method was the wild type, which was never found to be below 94%, indicating that none of the other alleles produced were able to restore the female phenotype and fertility of the *dsxF*$^{-}$ knockout and confer resistance.

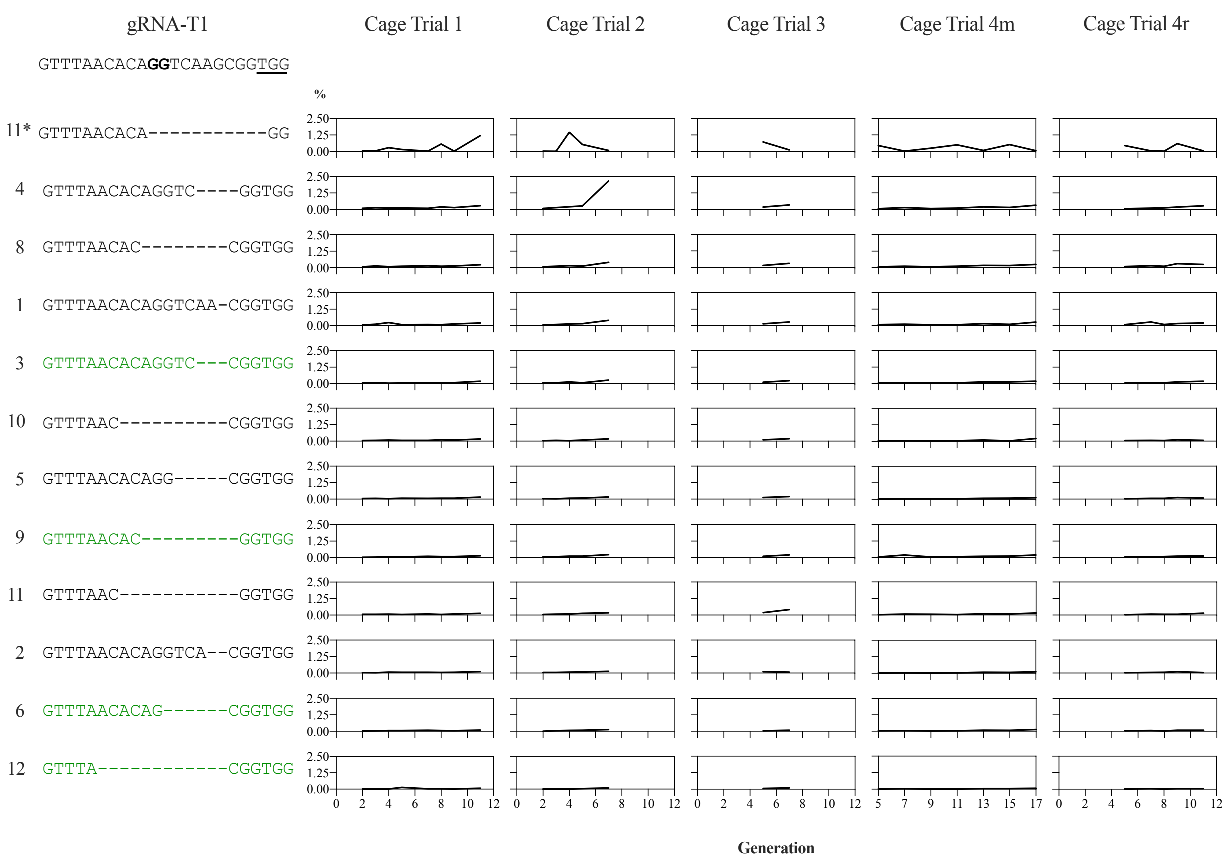


Figure 6.3 – Dynamics of the most prevalent indels detected in the cage trial populations. The deep amplicon sequencing revealed a repertoire of in-frame (green) and out-of-frame (black) mutations that were shared across the different caged populations. The frequency of the mutated alleles was recorded every two generations where possible. The 11-bp deletion detected during the assessment of the homing rate of the gene drive line (*) reached a maximum of 1.45% in cage trial 2 at generation 4, whereas a 4-bp deletion reached the maximum frequency among the non-homed alleles at 2.15% in cage trial 2 at generation 7. All other indels, including in-frame deletions, failed to reach frequencies above 0.5%, indicating a strong negative selection against them. It is possible that indels that were extending in the splice site between intron 4 and exon 5 (Gs in bold) might have impaired the alternative splicing of the female-specific exon. The PAM sequence of the gRNA-T1 is underlined.

Deletions that could theoretically restore the coding sequence of the AgdsxF isoform have been detected in all cage trials. Specifically, the detected three-bp deletion, which in theory could result in the removal of a glutamine residue and the replacement of an alanine residue with a proline (Figure 6.4A), has reached a maximum frequency of 0.3% in the populations. Three alternative in-frame deletions (six, nine and 12 bp) were also recorded in the populations but they were either adjacent to the AG-G splice site of the intron/exon boundary or were deleting the splice site altogether. The six-bp deletion results in the removal of the exonic -G residue that is necessary for exon splicing, but does not interfere with the AG- site that is found at the 3' site of introns and is required for splicing. Assuming this deletion does not interfere with the splicing of the female-specific exon, it results in the removal of both a glycine and glutamine residues at the start of the exon. On the contrary, the nine- and 12-bp deletions remove the splice site altogether, resulting most probably in an incorrect splicing of the exon. In any case, all four deletions were unable to restore the function of the splice site and the exon, minimising the formation of in-frame mutations. This again signifies the highly conserved character of the coding sequence of the female-specific exon of the *dsx* gene.

The results obtained through amplicon sequencing were in contrast with the results obtained through the same analysis when performed on the gene drive targeting the non-conserved *AGAP007280* gene (Hammond et al. 2017, 2020a). In those experiments, in-frame mutations have increased in frequency towards later generations by up to 95%, stalling the increasing frequency of the gene drive and restoring female fertility in the populations. To determine the difference in the conservation between the two target sites, I investigated the frequency of single nucleotide polymorphisms (SNPs) using genomic data from species of the Anophelinae subfamily (Neafsey et al. 2015) and genomic sequences from wild-caught *Anopheles* mosquitoes obtained from sub-Saharan Africa (Miles et al. 2017). The investigation revealed a great level of divergence between the two target sites that explains the difference observed among the frequencies of in-frame alleles in the populations of the two experiments. The target site in *AGAP007280* proved to be greatly polymorphic with no signs of conservation between species (Figure 6.5A). The 21-bp target site sequence of the gene drive has nine SNPs spread across the whole length of the sequence with frequencies in the population varying from 0.07% up to 31.2% (Figure 6.5B). On the contrary, when the *dsx* locus was investigated using the same datasets, there was clearly a dramatic difference in the number of SNPs and variation within the Anophelinae subfamily. The $dsxF^{CRISPRh}$ target site on the intron 4 – exon 5 boundary is much more conserved within the Anophelinae subfamily and even more conserved within wild-type *An. gambiae* populations in Africa, with only one silent SNP being present at 2.9% allele frequency. This indicates that the locus is highly conserved and under strong selective pressure that does not allow the accumulation of sequence variation at that site.

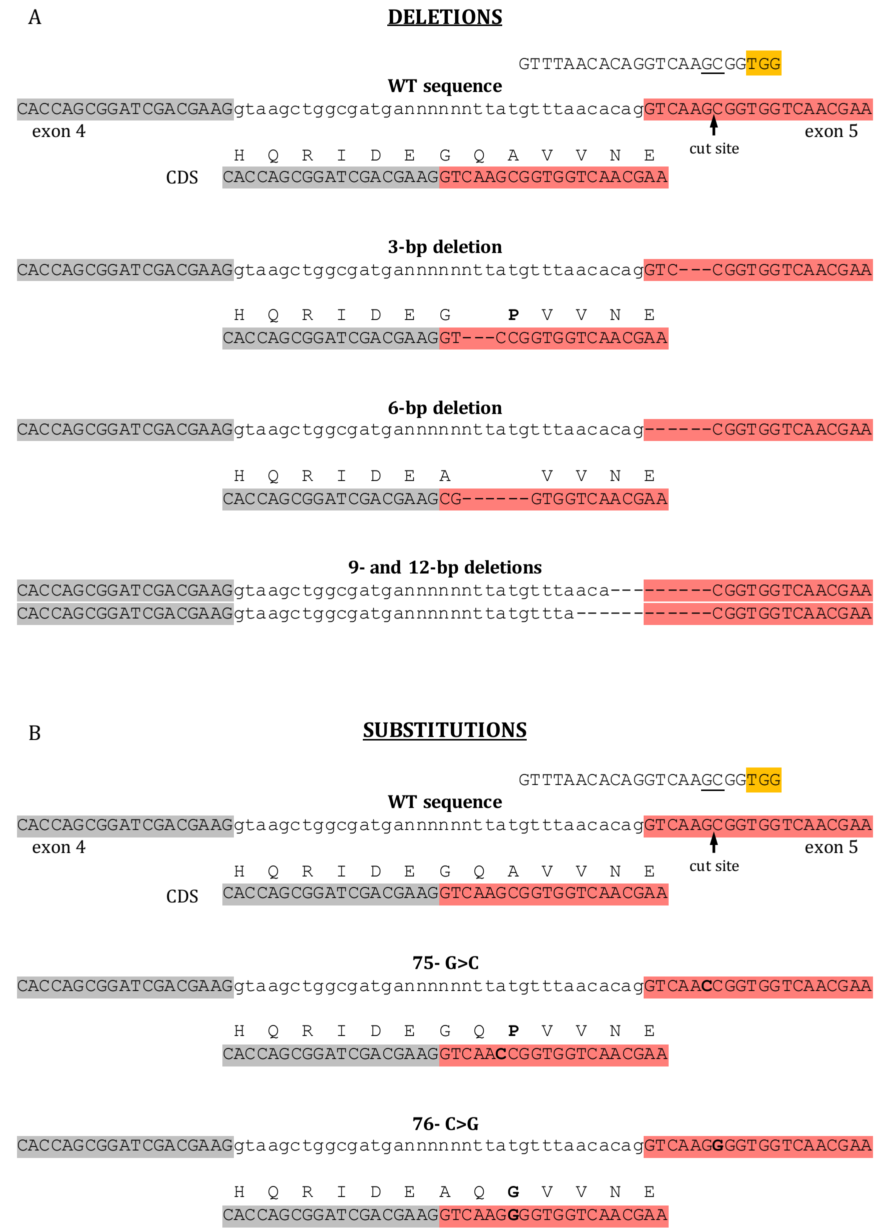


Figure 6.4 – In-frame deletions and substitutions detected within the caged populations and their impact on the coding sequence of exon 5. (A) The three-bp deletion detected in the populations of the cage trials could in theory result in the removal of a glutamine residue (Q) from the coding sequence of the exon and the replacement of an alanine (A) with proline (P). The six-bp deletion results in the removal of the nucleotides adjacent to the intron – exon boundary and the removal of both glutamine and alanine residues from the CDS of the exon, assuming the splicing of the exon is not affected. Both the nine- and 12-bp deletion extend in the splice site of the intron - exon boundary, resulting presumably to the disruption of the splicing of the female-specific exon. (B) The two substitutions that were detected with frequencies above 0.5% at positions 75 and 76 result in the plausible replacement of an alanine residue with proline (P) and glycine (G) respectively. Exon 4 and 5 are shaded with grey and red respectively. 15-bp sections from both the 5' and 3' sides of intron 4 are shown in lower case, whereas the intervening sequence is not drawn to scale. The gRNA-T1 target site is also annotated with the PAM highlighted in yellow and the nucleotides surrounding the cut site (arrow) underlined.

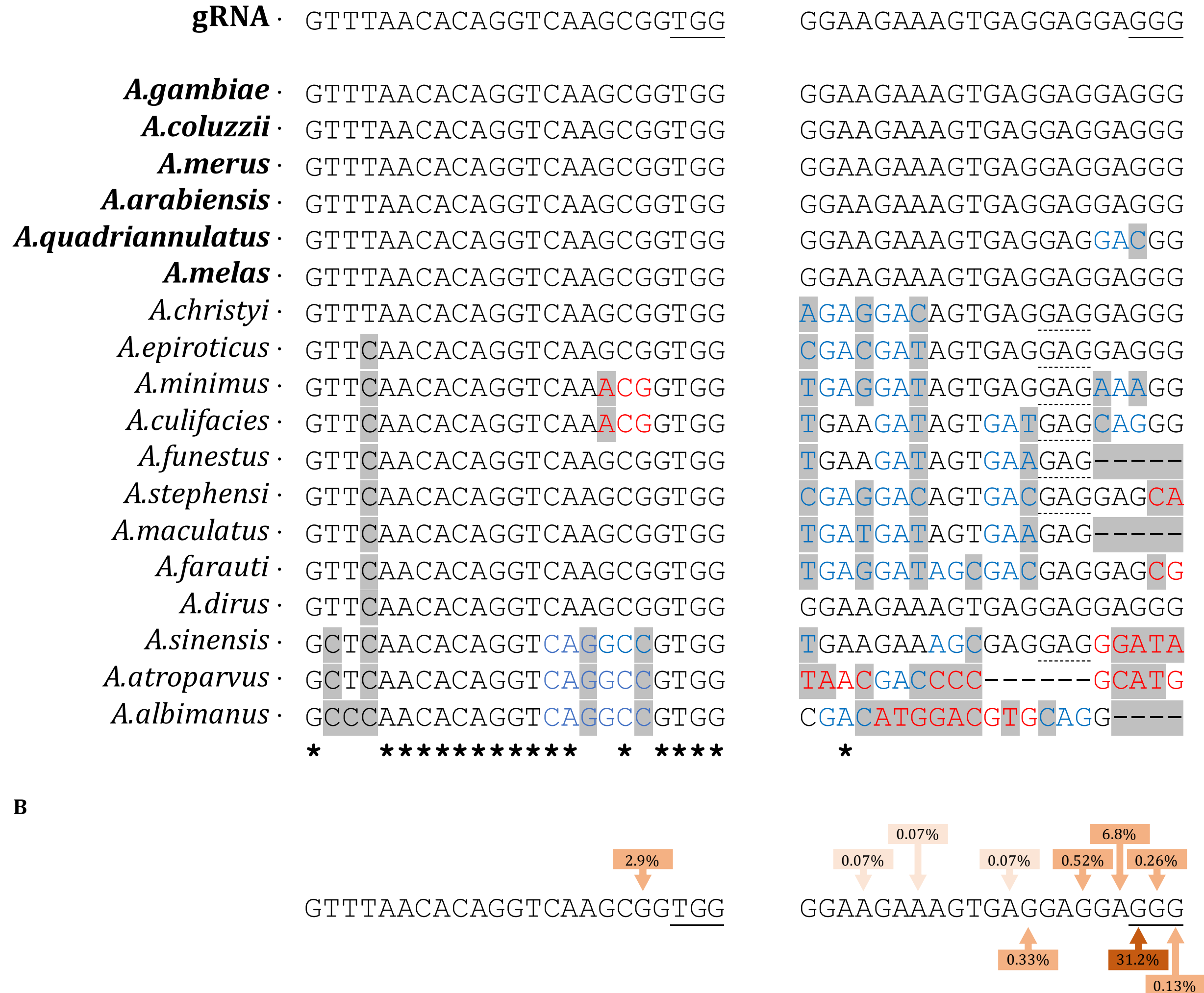


Figure 6.5 – Conservation comparison between the *dsx* and *nudel* gene drive target sites. (A) The *dsx* locus is identical within members of the *An. gambiae* species complex (bold) but also highly conserved within species of the Anophelinae subfamily. In contrast, the *nudel* locus is diverse and highly variable within the subfamily, indicating a less stringent conservation pressure. Shaded nucleotides indicate alterations from the *An. gambiae* wild-type sequence, whereas blue and red indicate silent and missense codon variations respectively. Underlined nucleotides indicate the PAM sequence in the gRNA, whereas the presence of insertions in the locus is underlined using a dashed line. Asterisks at the bottom indicate conserved nucleotides throughout the subfamily. (B) Genomic data obtained from 765 wild-type *An. gambiae* and *An. coluzzii* mosquitoes from regions from sub-Saharan Africa reveal a dramatic difference in the presence of SNPs in the two loci. The *dsx* locus contains only one SNP at 2.9% frequency, whereas the *nudel* locus contains nine SNPs varying from 0.07% up to 31.2% frequency in the populations. This is an indication of the different degree of conservation that describes the two loci. The PAM sequence of the gRNA is underlined, and the SNPs are coloured in relation to their frequency.

From the deep amplicon sequence analysis performed on the populations of the $dsxF^{CRISPRh}$ gene drive, I was also able to analyse single nucleotide variant (SNV) frequency at the target site as part of the Cas9 activity on the locus. Analysing the SNV frequencies and mapping them across the background signal of SNVs created during the amplification process, I was not able to detect any significant difference between the frequencies of SNVs at the cut site and those on the rest of the amplicon in the majority of the populations (Figure 6.6). Across the length of the 250-bp amplicon, the frequency of the SNVs rarely surpassed the 0.5% threshold that we set at the start of the experiment. With the exception of the error-prone ends of the amplicon, only two SNVs were detected to be present in more than 0.5% of alleles examined in two distinct populations. In cage trial 3 and cage trial 4r, at generations five and nine respectively, a G>C substitution at location 75 and a C>G substitution at location 76 were found to be present within close proximity of the cut site. The two SNVs were present at a frequency of 0.57% and 0.73% respectively. When these substitutions are mapped to the coding sequence of the exon they result in the plausible replacement of an alanine residue with proline and glycine respectively (Figure 6.4B). It is unclear if these missense mutations did not confer resistance to the gene drive either because they could be cleaved by the nuclease or because they did not restore the phenotype of the female-specific exon.

6.2.3 SNP found in wild-type populations can be cleaved by the Cas9 *in vitro*

The *doublesex* target site chosen for this gene drive is apparently under high functional constraint. Within the laboratory colony of mosquitoes in which it was tested, the target site on exon 5 appears to be invariant. However, when I investigated genomic data obtained from wild-caught *An. gambiae* and *An.coluzzii* mosquitoes from sub-Saharan Africa, I detected a natural SNP variant located two nucleotides upstream of the PAM site of the sequence (Figure 6.5B). The C>A SNP that was detected is present at a 2.9% total allele frequency, and when analysed *in-silico* it results to a missense mutation. To investigate the geographic distribution of the SNP in the continent, I leveraged the Ag1000G database, which contains 765 genomes from wild-caught female *An. gambiae* and *An.coluzzii* mosquitoes from sub-Saharan Africa (Miles et al. 2017). A detailed analysis of the data revealed that the SNP can reach frequencies up to 25% in local *An. coluzzii* populations and up to 12.5% in *An. gambiae* populations. The SNP was also not evenly distributed across the 8 countries where the samples originated from and it was only found in populations from western equatorial and southern countries (Figure 6.7). It is also likely that this particular SNP can restore the female phenotype and fertility to some degree, since it has been found in homozygosity in four female *An. coluzzii* out of 60 tested in one of these populations (Table 6.1). Since at the time I could not perform an *in vivo* assay to test the efficiency of our gene drive against this particular SNP variant, I used an *in vitro* cleavage assay to preliminary reveal the degree of resistance the SNP could impose to putative

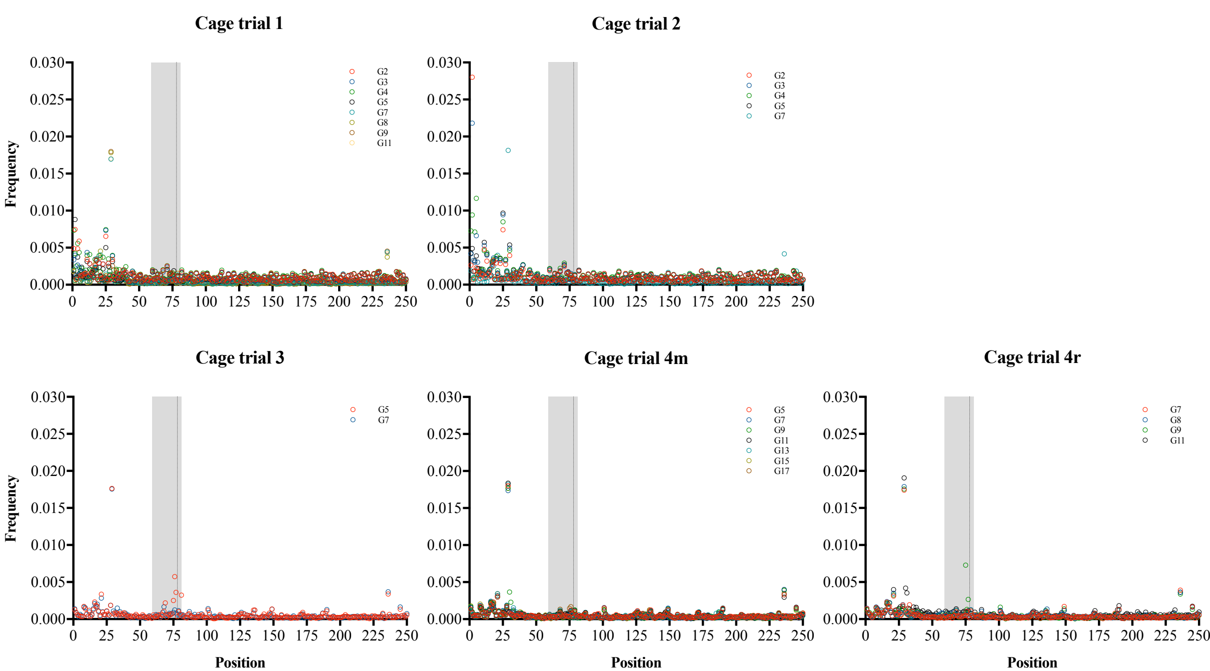


Figure 6.6 – Single nucleotide variance analysis across the cut site of the *dsxF*^{CRISPRh} gene drive. Deep amplicon sequencing performed across the gRNA-T1 (grey) and the cut site (dashed line) revealed the absence of significant variation caused by the *dsxF*^{CRISPRh} gene drive. In only two instances, in cage trial 3 and cage trial 4r, at generations five and nine respectively, two SNVs were detected at frequencies above 0.5%. The respective 75-G>C and 76-C>G substitutions were present at 0.57% and 0.73% frequency. The fact that these SNVs are located in close proximity with the cut site hints to their creation during the caged experiment and not during the amplification process. In any case, they maintained in low frequencies and failed to confer resistance either because they could not restore the function of the *doublesex* gene or because they could be recognised and cleaved by the nuclease.

releases of the gene drive in the continent. I used a combined method of restriction digestion enzymes and ribonucleoprotein complexes of Cas9 with gRNA-T1 to digest plasmids carrying the sequences of interest. The assay revealed that the target site carrying the G>A SNP can be cleaved as efficiently as the wild-type sequence (Figure 6.8), indicating that this variation found in populations in Africa might not pose a resistance barrier to our gene drive technology, allowing the spread of the gene drive allele and the eventual suppression of local populations. Of course, this hypothesis needs to be thoroughly investigated experimentally *in vivo* to unambiguously reveal the true impact of the SNP on our gene drive system.

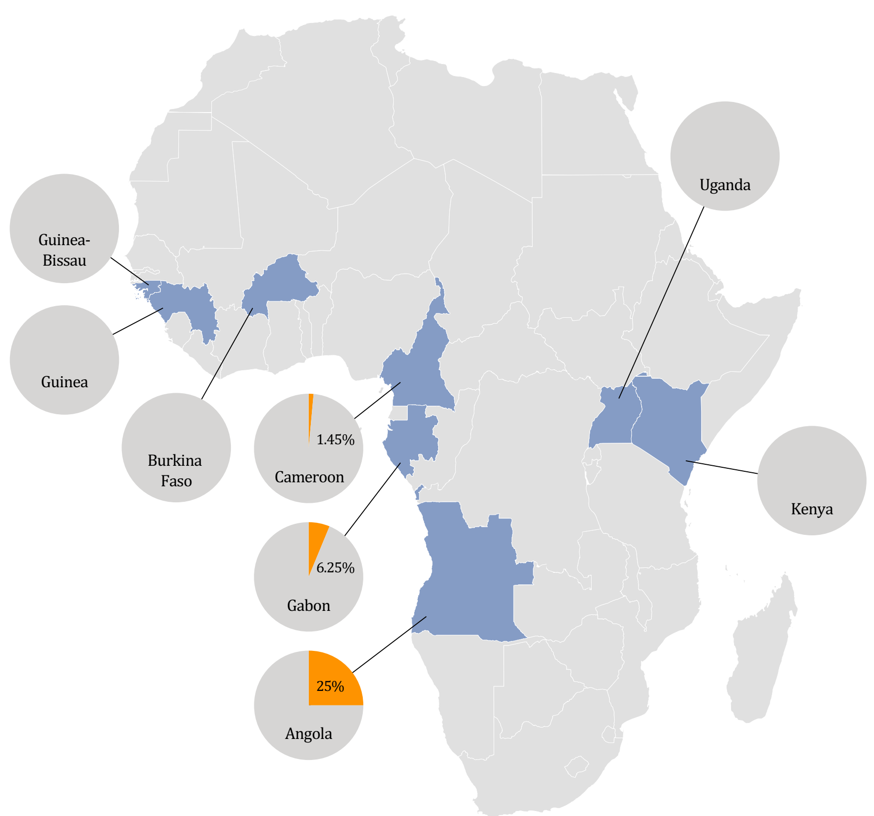


Figure 6.7 – Distribution of the G>A SNP found in the gRNA-T1 sequence in mosquitoes from sub-Saharan Africa. The particular SNP is not evenly distributed across populations tested from 8 different countries. The SNP is mainly present in western equatorial and southern African countries and reaches allele frequencies of up to 25% in populations from Angola.

		Total tested			Heterozygous		Homozygous		
Region	Country	*Agam*	*Acol*	Mix	*Agam*	*Acol*	*Agam*	*Acol*	SNP frequency
Luanda	Angola	0	60	0	0	22 (36.7%)	0	4 (6.7%)	25%
Bana	Burkina Faso	18	37	0	0	0	0	0	0%
Pala	Burkina Faso	39	10	0	0	0	0	0	0%
Sourukoudinga	Burkina Faso	24	22	0	0	0	0	0	0%
Daiguene	Cameroon	83	0	0	1 (1.2%)	0	0	0	0.6%
Gado-Badzere	Cameroon	70	0	0	2 (2.9%)	0	0	0	1.43%
Mayos	Cameroon	101	0	0	5 (4.9%)	0	0	0	2.48%
Zembe-Borongo	Cameroon	21	0	0	0	0	0	0	0%
Libreville	Gabon	56	0	0	7 (12.5%)	0	0	0	6.25%
Koundara	Guinea	15	0	1	0	0	0	0	0%
Antula	Guinea-Bissau	15	3	0	0	0	0	0	0%
Kilifi-Junju	Kenya	30	0	13	0	0	0	0	0%
Kilifi-Mbogolo	Kenya	31	0	0	0	0	0	0	0%
Tororo	Uganda	103	0	0	0	0	0	0	0%
Total		619	132	14	15 (2.4%)	22 (16.7%)	0	4 (3.0%)	2.9%

Table 6.1 – Distribution of the G>A SNP found in the gRNA-T1 sequence in mosquitoes from sub-Saharan Africa. The SNP is not evenly distributed throughout the continent but rather is present in scattered populations mostly in western equatorial and southern African countries. The SNP reaches frequencies up to 25% in *An. coluzzii* populations from Luanda, Angola, where homozygous females for the SNP have been also detected. The presence of females homozygous for the SNP indicates that this particular missense mutation can restore the female phenotype and fertility of the species to some degree. The percentages in brackets indicate the frequency of the respective mosquitoes in the local population. The 'Mix' column contains mosquitoes in which the simple *gambiae*/*coluzzii* dichotomy was not adequate using the current molecular markers at the time of the experiment. Data obtained from The *Anopheles gambiae* Ag1000G Genomes Consortium, Miles et al. 2017.

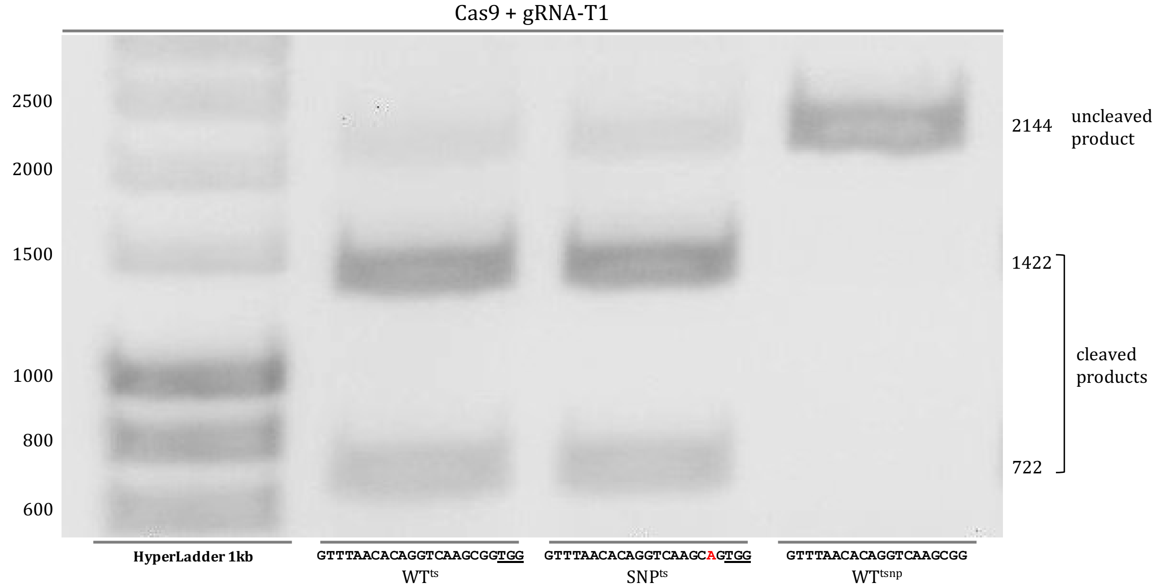


Figure 6.8 – *In vitro* cleavage assay testing the efficiency of the *dsxF^{CRISPRh}* gene drive against the SNP found in mosquito populations from Africa. Both the wild-type target site (WT^{ts}) and the target site carrying the SNP (SNP^{ts}) are cleaved with similar efficiency using *in vitro* ribonucleoprotein particles assembled using Cas9 protein and the gRNA-T1. A target site lacking the PAM sequence (WT^{tsnp}) served as a control. The SNP variant is indicated in red. Image reproduced with permission of the rights holder, Copyright © 2018, Springer Nature, Kyrou et al. 2018.

6.3 Controlled releases in large scale set ups successfully suppress mosquito populations

The success of the *dsxF^{CRISPRh}* gene drive to crash 5 caged mosquito populations in the lab is the first of its kind. This small success in the insectary paved the way for the construct to be tested in larger populations housed in bigger cages built to provide a more ecologically challenging environment for the mosquitoes. The big cages measuring 4.7 m^3 in volume, allow for more complex behaviours such as male swarming, oviposition site preference and other aspects of mosquito biology that cannot be recreated in small cages in the lab (Hammond et al. 2020b). For comparison, the small cages I used for the population experiments measure only 0.045 m^3 in volume. Six cages were built inside a climatic chamber designed to maintain the humidity and temperature but also mimic more accurately the light conditions of a full day cycle in the tropics. The cages were used to establish and maintain a mean population size of 574 wild-type *An. gambiae* mosquitoes for at least 74 days before the release of any gene drive (Hammond et al. 2020b). The *dsxF^{CRISPRh}* gene drive was released in four of the cages in duplicates using a 'Low' (12.5%) and an 'Intermediate' (25%) allelic frequency, whereas two cages without the gene drive were kept as controls. In contrast with the small cage experiment, which was were based on discrete generations, the setup of

the big cage experiment was relying on overlapping generations. By allowing the older populations to remain in the cage and overlap with the new generations, we reveal differences in life-span, mating and fecundity over time that cannot be captured in a discrete generation setup (Facchinelli et al. 2019). Large cage populations experiments with an overlapping generation setup are considered now to be an essential bridge between laboratory and outdoor or open release field testing for gene drive technology (James et al. 2018).

The *dsxF*^{CRISPRh} gene drive was introduced in the cages using transgenic heterozygous males to mimic a male-only release scenario. Due to the overlapping generation setup, the number of transgenic males required for each release was split into two small consecutive releases during the course of one week. For the 'Low' release, a total of 143 transgenics were introduced in the cages, whereas for the 'Intermediate' release a total of 287 transgenics were introduced, equivalent to 25% and 50% transgenic frequency respectively of the estimated mean adult population of 574 (Hammond et al. 2020b). Upon release, for the first 125 days, the gene drive allele spread slowly in the cages, fluctuating around the release allele frequency, as indicated by the presence of the gene drive marker among the offspring of the populations (Figure 6.9A). This initiation period was followed by a rapid invasion phase that lasted approximately 100 days and resulted in the gene drive reaching allele frequencies up to 90%. Towards the end of the experiment, the gene drive entered a final fixation period of approximately 50 days that ensured the complete transformation of the mosquito populations. The full prevalence of the gene drive in the populations resulted in a reduced egg output in all cages that started to be in effect between 200 and 250 days post release (Figure 6.9B). The reproductive load imposed on the populations during this period was due to the noticeable decline of fertile females and the increasing numbers of sterile intersex. Eventually, the populations were unable to sustain the numbers needed to seed subsequent generations, leading to a complete suppression in all four cages within an average of 275 days post release.

Similarly with the small caged experiments, no resistance against the gene drive has been detected within the large caged population during the one year experiment. Deep amplicon sequencing revealed the formation of indels at the cut site that were also found during the same analysis of the small caged populations. These mutations were found to reach only a maximum of 0.3% allele frequency in the populations, with in-frame deletions reaching only 0.19%. Two novel mutations, one in-frame and one out-of-frame, reached the maximum of 3.28% and 1.56% allele frequency respectively among the non-homed alleles. The in-frame mutation was the result of both an insertion and a substitution of three and one nucleotides respectively, resulting in the replacement of an alanine residue with a glycine and glutamine (Figure 6.10). The inability of the respective allele to increase in frequency and counteract the spread of the gene drive is an indication that the allele was under negative selective pressure, possibly because it

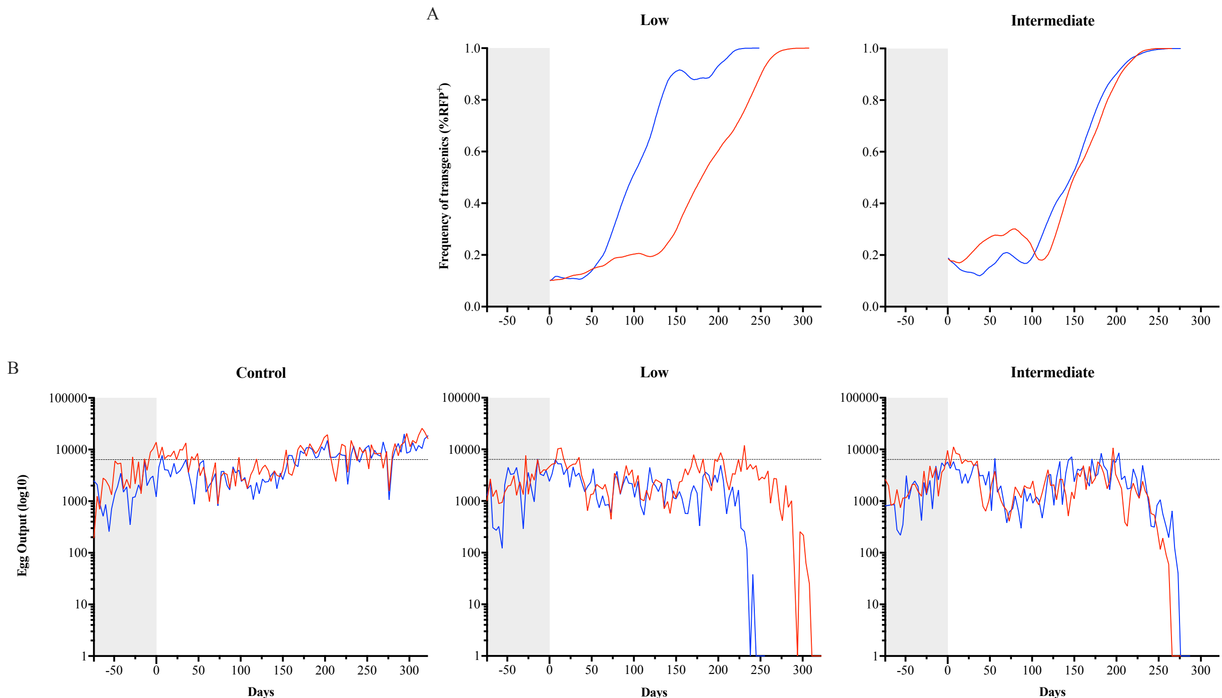


Figure 6.9 – Gene drive dynamics and egg output in large cage experiments. The gene drive targeting *dsx* was tested in four large cages housing a mean population of 574 mosquitoes. Wild-type populations were established in all cages for 74 days (grey) before the release of the gene drive in the cages. (A) The gene drive was released using transgenic males in duplicate cages (red and blue) using an allele frequency of 12.5% (Low) and 25% (Intermediate). In all cages the gene drive followed a sigmoidal pattern of invasion, fluctuating around the release allele frequency within the first 125 days, before spreading rapidly to fixation between 200 and 250 days. (B) The gene drive imposed a reproductive load on the populations leading to a dramatic decrease on egg output after 220 days post release when the gene drive reached full prevalence in the populations. In contrast with the control populations, all four cages housing the gene drive were unable to produce eggs and sustain subsequent generations to maintain the populations. The dotted line indicates the mean egg output of the control cages.

did not encode for a functional DSX protein. Nevertheless, the overall absence of resistance against the gene drive targeting the *dsx* gene in the large cage setup is a strong validation of the high conservation that characterises the gene. The *dsx* gene, and specifically the female-specific isoform, appears to be so far resilient to sequence variations and proves to be highly constrained even when tested in larger populations. This successful population suppression using a gene drive system is a landmark achievement and the next step towards the deployment and use of a gene drive system against vector populations (James et al. 2018). The $dsxF^{CRISPRh}$ gene drive targeting *doublesex* can now be shipped and integrated in local populations in Africa and tested in an outdoor field setup to enter the next phase towards the deployment in the wild.

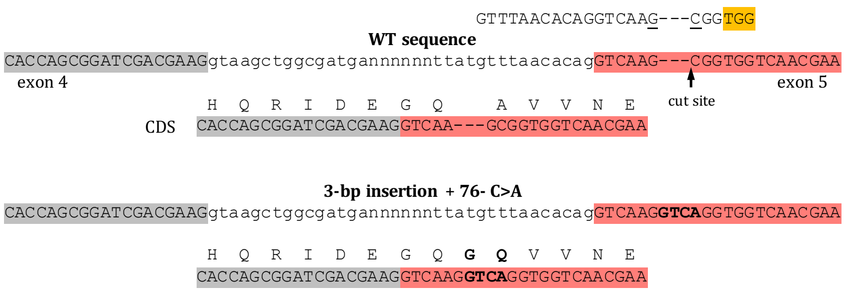


Figure 6.10 – In-frame indel detected in the large cages and its impact on the coding sequence of exon 5. Deep amplicon sequencing revealed the presence of a novel indel mutation consisting of a three-bp insertion and a substitution. The in-frame indel was detected at 3.28% frequency among the non-homed alleles and when mapped across the female-specific exon results in the insertion of a glycine residue (G) and the replacement of an alanine residue (A) with glutamine (Q). Exon 4 and 5 are shaded with grey and red respectively. 15-bp sections from both the 5' and 3' sides of intron 4 are shown in lower case, whereas the intervening sequence is not drawn to scale. The gRNA-T1 target site is also annotated with the PAM highlighted in yellow and the nucleotides surrounding the cut site (arrow) underlined.

6.4 Discussion

Designing and building a gene drive system in the lab is only a small part in a long process that is required for the controlled and successful release of such systems in nature. Because of their self-sustaining and invasive character, gene drive systems have to pass rigorous validation checkpoints during the deployment process before they are allowed to be released in the wild. From the design phase to the final release and post-release surveillance, a gene drive system is thoroughly examined in every step of the process for its efficiency, safety and ability to achieve population modification. To date, no gene drive targeting population suppression has managed to reach the second step in the deployment pathway, the release and testing in large indoor cages. Numerous gene drive systems have been created and tested in the lab for their efficiency and invasiveness but failed to show the necessary properties required to be tested on a larger scale. Many of these failed attempts were due to the rise of resistant alleles against the gene drive that accumulated in the tested populations through positive selection, counteracting the function of the gene drive system. The use of highly conserved genes as gene drive targets is one of the solutions implemented to minimise the creation and selection of such resistant alleles.

A gene drive targeting the highly conserved female-specific exon of the *dsx* gene has proven to be resilient to the formation of resistant alleles when tested in small caged populations in the lab. Using a single release of a small gene drive allele frequency, the gene drive has managed to reach full prevalence within the populations starting at generation 7. Alleles that were formed at the *dsx* target site as a result

of MMEJ and NHEJ were unable to restore the function of the *dsx* gene and block the efficiency of the gene drive. This has allowed the system to spread with high efficiency in all the mosquitoes of the cages, resulting in the complete transformation of the populations. As a result, the continuous depletion and replacement of fertile females with sterile homozygous intersex imposed a dramatic reproductive load on the populations affecting the egg output of the cages. Unable to produce enough progeny to sustain their numbers, five caged populations eventually collapsed one generation after the complete transformation of the cages. The successful collapse of the mosquito populations in the lab was the first of its kind, making this a landmark achievement for gene drive technology.

Following the success in the lab, the gene drive system targeting the *dsx* gene was quickly set to be tested in large indoor cages in accordance with the next phase of the gene drive deployment process. The cages were housed in specialised facilities with climatic chambers able to maintain the environmental conditions observed in sub-Saharan Africa and thus providing an ecologically challenging environment for the mosquitoes. Populations housed in larger cages and in such conditions are more likely to behave like in nature, revealing factors and parameters that cannot be recreated and tested in small cages the lab. Swarming behaviour, mating preference but also oviposition site preference are some of these factors that can have a tremendous impact on the efficiency of a gene drive system. Following a successful release in the cages using male transgenics, the gene drive targeting the female-specific exon of the *dsx* gene was able to reach full prevalence in all four cages, resulting in a complete population suppression within an average of 275 days. This accomplishment proved that the *dsx* gene is highly constrained and does not allow the formation of resistant alleles that could have blocked the gene drive system from achieving population suppression. This successful assessment of the gene drive system during the second phase of the deployment pathway has tremendous benefits for gene drive technology. The gene drive system targeting the highly conserved *dsx* gene can now be assessed for its efficiency and robustness in outdoor cages located in Africa, bringing the technology closer to the field and hopefully to be used in the future against the deadliest animal on earth.

Discussion and Future Directions

The development of a gene drive system capable of manipulating and suppressing mosquito populations is a long-sought scientific achievement for vector control strategies and malaria in particular. To date, gene drive technologies had failed to overcome limitations and challenges that arose during proof-of-principle experiments when using this technology to control vector populations in the lab. The gene drive system developed in this thesis overcomes the basic problem of resistance and successfully invades caged populations of *Anopheles gambiae* mosquitoes. The simple switch of target site to a more highly conserved locus such as exon 5 of *doublesex*, has dramatically reduced the creation of resistant alleles that could restore the function of the locus and counteract the spread of the gene drive. The successful population suppression observed in small cages in the lab but also during large semi-field testing is an indication that the degree of conservation observed in the *doublesex* locus so far is sufficient to allow the system to spread in the population, gradually imposing a reproductive load that leads to suppression. Evaluation of a gene drive in large confined indoor spaces is considered an essential intermediate and a bridge between the lab setting and large outdoor field trials for the safe and controlled deployment of a gene drive system in the wild.

The next step for the gene drive targeting *dsx* is the assessment of the system in outdoor cage trials in Africa in order to study and evaluate the efficiency of the system when integrated into local mosquito populations. Like the large indoor population experiments, the outdoor cage field trials provide an even more ecologically challenging environment for the mosquitoes, but this time under the native conditions of sub-Saharan Africa. The native environment and scale of the experiment allow for the accurate recreation of complex behaviours of mosquito biology such as male swarming and oviposition site preferences that can only be stimulated to some degree in large indoor semi-field experiments. The evaluation in native populations from Africa is also important in order to reveal any potential differences in life-span, mating, and fecundity of the carrier mosquitoes that are crucial parameters that shape the efficiency of a gene drive system. These parameters are important to know prior to a potential deployment phase because they will help determine an efficient mosquito capacity and shape the strategy that would need to be implemented in a potential release scenario in native populations.

Introgression of the *dsx* gene drive system to the native genetic background can also reveal the effects of natural genetic variation on the efficiency and prevalence of the gene drive. Even though highly conserved, the target site in exon 5 appears to carry small variations that are present in populations in Africa and could pose a threat to the efficiency of the gene drive system. Although *in vitro* experiments hinted that the gene drive might not be affected by current variation at the target site, the situation might be different when tested *in vivo*. Current experiments are planned to incorporate the variant target site in lab colonies and test the efficiency and the dynamics of the system when these alleles are present. Finally, natural variation that has not been detected yet in local populations in Africa could also be revealed through the incorporation of the gene drive to local populations. Early detection of these variants during the assessment process is beneficial for the final deployment of the gene drive system.

Although resistance has not been detected in the small and large cage populations experiments, a number of measures have already been proposed to reduce the impact of potential resistance that could arise later during the deployment process. The use of alternative nucleases like the Cas12a (Paul & Montoya 2020) has been suggested as a solution to overcome target site resistance. These nucleases have separate recognition and cleavage domains that could potentially solve the resistance issue since variations at the target site do not affect the efficiency of the nuclease. Unfortunately, experiments using the Cas12a in *An. gambiae* mosquitoes in the lab revealed the low efficiency of the system to function and home efficiently in the germline of the mosquito, possibly due to temperature sensitivity (unpublished with permission of Akash Sharma). Other approaches to overcome resistance rely on the use of multiple gRNAs to target multiple loci in close proximity on the *dsx* locus. Multiplexing gRNAs (Kurata et al. 2018) ensures that the homing process could still occur even if resistant alleles are formed in some of the targeted loci. This is because in theory only one gRNA is needed to cleave the DNA and activate the repair machinery and promote HDR. The use of multiple gRNAs ensures the gene drive system has enough time to invade the population and cause a significant reduction before all the target sites confer resistance. Gene drive systems that use two gRNAs to target *dsx* are currently under development and will soon be ready to be tested in small cage experiments.

Finally, the availability of genomic sequences from other major vectors of the Anophelinae subfamily allows the easy transfer of this technology to those vectors. Experiments are currently underway to replace - where different - the native *An. gambiae* exon 5 sequence of the *dsx* gene with synthetic alleles homologous to the target site sequences found in those species. These experiments aim to test the efficiency and robustness of the gene drive targeting *dsx* when found against these alleles and the dynamics we expect when the system is transferred to these populations. Currently, the gene drive targeting *Agdsx* is in the process of being introgressed in the genome of *Anopheles arabiensis* mosquitoes, and results about its

efficiency in the vector are imminent. Small cage population experiments using this vector are of priority to vigorously assess and evaluate the dynamics of the system when tested in the species.

Nevertheless, developing and testing a gene drive system for population suppression against malaria in Africa could be proved futile if the technology does not get approved to be released in the continent. Gene drives have no prior use-history and because of their self-sustaining nature are considered by a minority of people as a 'dangerous' concept, and attempts to pass moratoriums on the research and use of the technology with immediate effect have been put forward multiple times in the recent years. The argument brought forward by the groups opposing the use of gene drives in the natural world is the lack of research surrounding any effects the technology might have in both the ecosystem and human health. The groups argue, among other things, that all species are important and contribute equally to their ecosystem, and depletion of such an abundant species could have tremendous effects that could lead to the collapse of food chains in areas where the species is present. Such a valid argument is also shared predominantly among people that are no experts on the subject, prompting for the scientific community to provide a thorough risk assessment of the technology (Teem et al. 2019).

It is important to mention that *Anopheles gambiae* became a species of importance not because of its contributions to ecosystem food webs but because of its ability to transmit the malaria parasite, putting at risk almost half of the world's population. Because of its highly anthropophilic nature, the species populations have been always closely linked with the ever-increasing human population, multiplying the burden of the disease. It is now suggested that without new interventions that tackle one of the most important problems - the scale of the disease - we might not be able to defeat malaria at all. Interventions that aim to genetically reduce vector populations are now considered a complementary method that could help humanity eliminate malaria once and for all. Consequently, because these interventions are destined to alter the diversity of ecosystems, scientists have studied in parallel from the early beginning possible implications these interventions could have on natural ecosystems. To date, multiple studies have suggested that major changes in the population size or even depletion of vector species, such as *An. gambiae*, are unlikely to harm the biodiversity of its native ecosystems (Roberts et al. 2017). An extensive analysis on the predation of *An. gambiae* throughout its lifetime has emphasised the absence of any tight links between this particular species and predators (reviewed in Collins et al. 2019). Apart from a spider species that predominantly feeds on blood-fed *Anopheles* mosquitoes (Jackson & Nelson 2012) and lives in a restricted distribution near the shores of Lake Victoria, in other cases, the species has been described as a low-value, low-volume and disaggregated resource in the ecosystem food web (Collins et al. 2019).

Another infamous question that accompanies gene drives is the potential horizontal transfer in closely related species and the hazard of the 'leak' of the technology in an ecosystem. Although gene drives are

designed to be species-specific and destined to pass to next generations by vertical transmission, in theory, the horizontal transfer in other species is also possible if the two species can mate and create viable offspring. Hybrids of *An. gambiae* with its closely related species *An. arabiensis*, although rare, can occur in nature. In this context, the flow of the gene drive in the sibling species is considered beneficial in further reducing the transmission of the disease, since alongside *An. gambiae* , *An. arabiensis* is also considered a major malaria vector and a threat to humanity. It is important to mention that removing both species from an ecosystem would not create an opportunistic niche for other mosquito species to become major vectors of the disease. Past examples of similar scenarios in which selected mosquito species have been reduced dramatically by either using desalination of larval habitats (Missiroli 1938), IRS (Gillies & Smith 1960) or ITNs (Bayoh et al. 2010), have shown that a reduction in the population size of an anthropophilic vector in an ecosystem is quickly followed by the increase in the biomass of zoophilic species instead. This is not surprising if we consider that it requires a substantial amount of evolutionary time for a species to change its feeding preferences and adapt to feed predominantly on humans but also to be highly susceptible to the malaria parasite.

Another argument that is brought forth by opposing teams is the bio-ethics surrounding the use of gene drives in nature (Pugh 2016). The notion of 'playing God' with nature has fuelled an enormous discussion on whether it is ethically acceptable or not to alter the genome of an entire species using gene editing technologies. 'Interfering with nature' is considered by opposing groups an unacceptable method of resolving humanity's problems and the key to a Pandora's box that could have catastrophic implications for both nature and mankind. On the other hand, scientists argue that gene editing technologies, when used properly, can confer tremendous benefits to humanity and ultimately help us defeat diseases that have devastated this planet for years. Malaria alone claims the lives of almost 400,000 people annually so a big ethical question is also raised here; if gene editing and gene drive technology can help us eliminate this disease and save 270.000 children annually, why is it ethical not to use them?

With all being said, gene editing technologies and gene drives in particular, although new, proved to be valuable tools for science and promise to provide solutions to world problems that seemed impossible to solve until recently. The novelty of the technology and the lack of use-history have raised ethical conundrums and scepticism among scientists, regulatory bodies and the public, and strengthen the need for the development of a step-by-step assessment pathway for the safe deployment of any future gene drive technology. Vigorous evaluation and risk assessment of the technology will provide useful insights on its function and safety, and most importantly, help convince regulatory bodies, governments and the communities who will live alongside the technology to allow for gene drives to be used as an intervention method against wide-spread diseases such as malaria.

Bibliography

Akbari, O. S., Bellen, H. J., Bier, E., Bullock, S. L., Burt, A., Church, G. M., Cook, K. R., Duchek, P., Edwards, O. R., Esvelt, K. M., Gantz, V. M., Golic, K. G., Gratz, S. J., Harrison, M. M., Hayes, K. R., James, A. A., Kaufman, T. C., Knoblich, J., Malik, H. S., Matthews, K. A., O'Connor-Giles, K. M., Parks, A. L., Perrimon, N., Port, F., Russell, S., Ueda, R. & Wildonger, J. (2015), 'Safeguarding gene drive experiments in the laboratory', *Science.* **349**(6251), 927–929.
URL: *https://www.sciencemag.org/lookup/doi/10.1126/science.aac7932*

Akbari, O. S., Chen, C. H., Marshall, J. M., Huang, H., Antoshechkin, I. & Hay, B. A. (2014), 'Novel synthetic Medea selfish genetic elements drive population replacement in *Drosophila*; A theoretical exploration of medea -dependent population suppression', *ACS Synth. Biol.* **3**(12), 915–928.
URL: *https://pubs.acs.org/doi/10.1021/sb300079h*

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', *Curr. Biol.* **23**(8), 671–677.
URL: *http://dx.doi.org/10.1016/j.cub.2013.02.059*

Aliota, M. T., Walker, E. C., Uribe Yepes, A., Dario Velez, I., Christensen, B. M. & Osorio, J. E. (2016), 'The wMel strain of *Wolbachia* reduces transmission of chikungunya virus in *Aedes aegypti*', *PLoS Negl. Trop. Dis.* **10**(4), 1–13.
URL: *https://journals.plos.org/plosntds/article?id=10.1371/journal.pntd.0004677*

Alphey, L. (2002), 'Re-engineering the sterile insect technique.', *Insect Biochem. Mol. Biol.* **32**(10), 1243–7.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/12225915*

Alphey, L. (2014), 'Genetic control of mosquitoes', *Annu. Rev. Entomol.* **59**(1), 205–224.
URL: *https://www.annualreviews.org/doi/abs/10.1146/annurev-ento-011613-162002*

Alphey, L., Benedict, M., Bellini, R., Clark, G. G., Dame, D. A., Service, M. W. & Dobson, S. L. (2010), 'Sterile-insect methods for control of mosquito-borne diseases: An analysis', *Vector-Borne Zoonotic Dis.* **10**(3), 295–311.
URL: *https://www.ncbi.nlm.nih.gov/pubmed/19725763*

An, W., Cho, S., Ishii, H. & Wensink, P. C. (1996), 'Sex-specific and non-sex-specific oligomerization domains in both of the *doublesex* transcription factors from *Drosophila melanogaster*', *Mol. Cell. Biol.* **16**(6), 3106–3111.
URL: *https://europepmc.org/article/med/8649422*

Anderson, J. L., Marí, A., Braasch, I., Amores, A., Hohenlohe, P., Batzel, P. & Postlethwait, J. H. (2012), 'Multiple sex-associated regions and a putative sex chromosome in zebrafish revealed by RAD mapping and population genomics', *PLoS One* **7**(7).
URL: *https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0040701*

Ariey, F., Witkowski, B., Amaratunga, C., Beghain, J., Langlois, A.-C., Khim, N., Kim, S., Duru, V., Bouchier, C., Ma, L., Lim, P., Leang, R., Duong, S., Sreng, S., Suon, S., Chuor, C. M., Bout, D. M., Ménard, S., Rogers, W. O., Genton, B., Fandeur, T., Miotto, O., Ringwald, P., Le Bras, J., Berry, A., Barale, J.-C., Fairhurst, R. M., Benoit-Vical, F., Mercereau-Puijalon, O. & Ménard, D. (2014), 'A molecular marker of artemisinin-resistant *Plasmodium falciparum* malaria', *Nature* **505**(7481), 50–55.

URL: *http://www.nature.com/articles/nature12876*

Aryal, R. & Ming, R. (2014), 'Sex determination in flowering plants: Papaya as a model system', *Plant Sci.* **217-218**, 56–62.
URL: *http://dx.doi.org/10.1016/j.plantsci.2013.10.018*

Aryan, A., Anderson, M. A. E., Biedler, J. K., Qi, Y., Overcash, J. M., Naumenko, A. N., Sharakhova, M. V., Mao, C., Adelman, Z. N. & Tu, Z. (2020), *Proc. Natl. Acad. Sci.* p. 202001132.
URL: *http://www.pnas.org/lookup/doi/10.1073/pnas.2001132117*

Ashley, E. A., Dhorda, M., Fairhurst, R. M., Amaratunga, C., Lim, P., Suon, S., Sreng, S., Anderson, J. M., Mao, S., Sam, B., Sopha, C., Chuor, C. M., Nguon, C., Sovannaroth, S., Pukrittayakamee, S., Jittamala, P., Chotivanich, K., Chutasmit, K., Suchatsoonthorn, C., Runcharoen, R., Hien, T. T., Thuy-Nhien, N. T., Thanh, N. V., Phu, N. H., Htut, Y., Han, K. T., Aye, K. H., Mokuolu, O. A., Olaosebikan, R. R., Folaranmi, O. O., Mayxay, M., Khanthavong, M., Hongvanthong, B., Newton, P. N., Onyamboko, M. A., Fanello, C. I., Tshefu, A. K., Mishra, N., Valecha, N., Phyo, A. P., Nosten, F., Yi, P., Tripura, R., Borrmann, S., Bashraheil, M., Peshu, J., Faiz, M. A., Ghose, A., Hossain, M. A., Samad, R., Rahman, M. R., Hasan, M. M., Islam, A., Miotto, O., Amato, R., MacInnis, B., Stalker, J., Kwiatkowski, D. P., Bozdech, Z., Jeeyapant, A., Cheah, P. Y., Sakulthaew, T., Chalk, J., Intharabut, B., Silamut, K., Lee, S. J., Vihokhern, B., Kunasol, C., Imwong, M., Tarning, J., Taylor, W. J., Yeung, S., Woodrow, C. J., Flegg, J. A., Das, D., Smith, J., Venkatesan, M., Plowe, C. V., Stepniewska, K., Guerin, P. J., Dondorp, A. M., Day, N. P. & White, N. J. (2014), 'Spread of artemisinin resistance in *Plasmodium falciparum* malaria', *N. Engl. J. Med.* **371**(5), 411–423.
URL: *http://www.nejm.org/doi/10.1056/NEJMoa1314981*

Ashman, T. L., Bachtrog, D., Blackmon, H., Goldberg, E. E., Hahn, M. W., Kirkpatrick, M., Kitano, J., Mank, J. E., Mayrose, I., Ming, R., Otto, S. P., Peichel, C. L., Pennell, M. W., Perrin, N., Ross, L., Valenzuela, N. & Vamosi, J. C. (2014), 'Tree of Sex: A database of sexual systems', *Sci. Data* **1**, 1–8.
URL: *https://www.nature.com/articles/sdata201415*

Bachtrog, D. (2013), 'Y-chromosome evolution: Emerging insights into processes of Y-chromosome degeneration', *Nat. Rev. Genet.* **14**(2), 113–124.
URL: *https://www.nature.com/articles/nrg3366*

Bachtrog, D., Mank, J. E., Peichel, C. L., Kirkpatrick, M., Otto, S. P., Ashman, T. L., Hahn, M. W., Kitano, J., Mayrose, I., Ming, R., Perrin, N., Ross, L., Valenzuela, N., Vamosi, J. C., Mank, J. E., Peichel, C. L., Ashman, T. L., Blackmon, H., Goldberg, E. E., Hahn, M. W., Kirkpatrick, M., Kitano, J., Mayrose, I., Ming, R., Pennell, M. W., Perrin, N., Valenzuela, N. & Vamosi, J. C. (2014), 'Sex Determination: Why so many ways of doing it?', *PLoS Biol.* **12**(7), 1–13.
URL: *https://www.ncbi.nlm.nih.gov/pubmed/24983465*

Baker, B. S. & Wolfner, M. F. (1988), 'A molecular analysis of *doublesex*, a bifunctional gene that controls both male and female sexual differentiation in *Drosophila* melanogaster', *Genes Dev.* **2**(4), 477–489.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/2453394*

Baker, D. A., Nolan, T., Fischer, B., Pinder, A., Crisanti, A. & Russell, S. (2011), 'A comprehensive gene expression atlas of sex- and tissue-specificity in the malaria vector, *Anopheles gambiae*', *BMC Genomics* **12**(1), 296.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/21649883*

Baker, R. H. (1984), 'Chromosome rearrangements in the control of mosquitoes', *Prev. Vet. Med.* **2**(1-4), 529–540.
URL: *https://linkinghub.elsevier.com/retrieve/pii/0167587784900990*

Baker, R. H., Sakai, R. K. & Saifuddin, U. T. (1978), 'Genetic sexing technique for a mosquito sterile male release', *Nature* **274**(5668), 253–255.
URL: *https://www.nature.com/articles/274253a0*

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', *Nat. Commun.* **5**(1), 3985.
URL: *http://www.nature.com/articles/ncomms4985*

Baragaña, B., Hallyburton, I., Lee, M. C., Norcross, N. R., Grimaldi, R., Otto, T. D., Proto, W. R., Blagborough, A. M., Meister, S., Wirjanata, G., Ruecker, A., Upton, L. M., Abraham, T. S., Almeida, M. J., Pradhan, A., Porzelle, A., Martínez, M. S., Bolscher, J. M., Woodland, A., Norval, S., Zuccotto, F., Thomas, J., Simeons, F., Stojanovski, L., Osuna-Cabello, M., Brock, P. M., Churcher, T. S., Sala, K. A., Zakutansky, S. E., Jiménez-Díaz, M. B., Sanz, L. M., Riley, J., Basak, R., Campbell, M., Avery, V. M., Sauerwein, R. W., Dechering, K. J., Noviyanti, R., Campo, B., Frearson, J. A., Angulo-Barturen, I., Ferrer-Bazaga, S., Gamo, F. J., Wyatt, P. G., Leroy, D., Siegl, P., Delves, M. J., Kyle, D. E., Wittlin, S., Marfurt, J., Price, R. N., Sinden, R. E., Winzeler, E. A., Charman, S. A., Bebrevska, L., Gray, D. W., Campbell, S., Fairlamb, A. H., Willis, P. A., Rayner, J. C., Fidock, D. A., Read, K. D. & Gilbert, I. H. (2015), 'A novel multiple-stage antimalarial agent that inhibits protein synthesis', *Nature* **522**(7556), 315–320.
URL: *https://www.nature.com/articles/nature14451*

Bateman, J. R., Lee, A. M. & Wu, C.-t. (2006), 'Site-Specific Transformation of *Drosophila* via ϕC31 Integrase-Mediated Cassette Exchange', *Genetics* **173**(2), 769–777.
URL: *http://www.genetics.org/lookup/doi/10.1534/genetics.106.056945*

Bayoh, M. N., Mathias, D. K., Odiere, M. R., Mutuku, F. M., Kamau, L., Gimnig, J. E., Vulule, J. M., Hawley, W. A., Hamel, M. J. & Walker, E. D. (2010), '*Anopheles gambiae*: Historical population decline associated with regional distribution of insecticide-treated bed nets in western Nyanza Province, Kenya', *Malar. J.* **9**(1).
URL: *https://doi.org/10.1186/1475-2875-9-62*

Becker, N., Petrić, D., Zgomba, M., Boase, C., Madon, M., Dahl, C. & Kaiser, A. (2010), Morphology of Mosquitoes, *in* 'Mosquitoes Their Control', Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 63–87.
URL: *http://dx.doi.org/10.1007/978-3-540-92874-4{_}5*

Beeman, R. W., Friesen, K. S. & Denell, R. E. (1992), 'Maternal-effect selfish genes in flour beetles', *Science.* **256**(5053), 89–92.
URL: *https://science.sciencemag.org/content/256/5053/89/tab-article-info*

Bell, L. R., Horabin, J. I., Schedl, P. & Cline, T. W. (1991), 'Positive autoregulation of sex-lethal by alternative splicing maintains the female determined state in *Drosophila*', *Cell* **65**(2), 229–239.
URL: *https://www.ncbi.nlm.nih.gov/pubmed/2015624*

Bell, L. R., Maine, E. M., Schedl, P. & Cline, T. W. (1988), 'Sex-lethal, a *Drosophila* sex determination switch gene, exhibits sex-specific RNA splicing and sequence similarity to RNA binding proteins', *Cell* **55**(6), 1037–1046.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/3144435*

Belote, J. M., McKeown, M., Boggs, R. T., Ohkawa, R. & Sosnowski, B. A. (1989), 'Molecular genetics of *transformer*, a genetic switch controlling sexual differentiation in *Drosophila*', *Dev. Genet.* **10**(3), 143–154.
URL: *https://www.ncbi.nlm.nih.gov/pubmed/2472240*

Benedict, M. Q. & Robinson, A. S. (2003), 'The first releases of transgenic mosquitoes: An argument for the sterile insect technique', *Trends Parasitol.* **19**(8), 349–355.
URL: *https://www.meta.org/papers/the-first-releases-of-transgenic-mosquitoes-an/12901936*

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', *Proc Natl Acad Sci U S A* **111**(21), 7600–7605.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/24821795*

Bertin, R. I. & Newman, C. M. (1993), 'Dichogamy in angiosperms', *Bot. Rev.* **59**, 112–152.
URL: *https://link.springer.com/article/10.1007/BF02856676*

Beukeboom, L. W., Kamping, A. & van de Zande, L. (2007), 'Sex determination in the haplodiploid wasp

Nasonia vitripennis (Hymenoptera: Chalcidoidea): A critical consideration of models and evidence', *Semin. Cell Dev. Biol.* **18**(3), 371–378.
URL: *https://europepmc.org/article/med/17292644*

Beye, M. (2004), 'The dice of fate: The *csd* gene and how its allelic composition regulates sexual development in the honey bee, *Apis mellifera*', *BioEssays* **26**(10), 1131–1139.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1002/bies.20098*

Beye, M., Hasselmann, M., Fondrk, M. K., Page, R. E. & Omholt, S. W. (2003), 'The gene *csd* is the primary signal for sexual development in the honeybee and encodes an SR-type protein', *Cell* **114**(4), 419–429.
URL: *https://www.cell.com/fulltext/S0092-8674(03)00606-8*

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* **526**(7572), 207–211.
URL: *https://www.nature.com/articles/nature15535*

Bhaya, D., Davison, M. & Barrangou, R. (2011), 'CRISPR-Cas systems in bacteria and archaea: Versatile small RNAs for adaptive defense and regulation', *Annu. Rev. Genet.* **45**(1), 273–297.
URL: *https://www.annualreviews.org/doi/abs/10.1146/annurev-genet-110410-132430*

Bian, G., Joshi, D., Dong, Y., Lu, P., Zhou, G., Pan, X., Xu, Y., Dimopoulos, G. & Xi, Z. (2013), '*Wolbachia* invades *Anopheles stephensi* populations and induces refractoriness to *Plasmodium* infection', *Science.* **340**(6133), 748–751.
URL: *https://www.sciencemag.org/lookup/doi/10.1126/science.1236192*

Biedler, J. K. & Tu, Z. (2016), Chapter Two - Sex Determination in Mosquitoes, *in* A. S. Raikhel, ed., 'Progress in Mosquito Research', Vol. 51 of *Advances in Insect Physiology*, Academic Press, pp. 37 – 66.
URL: *https://www.sciencedirect.com/science/article/pii/S0065280616300273*

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', *Curr Biol* **16**(17), R766–76.
URL: *https://www.ncbi.nlm.nih.gov/pubmed/16950103*

Billeter, J. C., Villella, A., Allendorfer, J. B., Dornan, A. J., Richardson, M., Gailey, D. A. & Goodwin, S. F. (2006a), 'Isoform-specific control of male neuronal differentiation and behavior in *Drosophila* by the *fruitless* gene', *Curr. Biol.* **16**(11), 1063–1076.
URL: *https://www.cell.com/current-biology/fulltext/S0960-9822(06)01500-4*

Blackmon, H., Ross, L. & Bachtrog, D. (2017), 'Sex Determination, sex chromosomes, and karyotype evolution in insects', *J Hered* **108**(1), 78–93.
URL: *https://www.ncbi.nlm.nih.gov/pubmed/27543823*

Blagborough, A. M., Churcher, T. S., Upton, L. M., Ghani, A. C., Gething, P. W. & Sinden, R. E. (2013), 'Transmission-blocking interventions eliminate malaria from laboratory populations', *Nat. Commun.* **4**(May), 1–7.
URL: *https://www.nature.com/articles/ncomms2840*

Blasco, B., Leroy, D. & Fidock, D. A. (2017), 'Antimalarial drug resistance: linking *Plasmodium falciparum* parasite biology to the clinic', *Nat. Med.* **23**(8), 917–928.
URL: *http://www.nature.com/articles/nm.4381*

Boggs, R. T., Gregor, P., Idriss, S., Belote, J. M. & McKeown, M. (1987), 'Regulation of sexual differentiation in *D. melanogaster* via alternative splicing of RNA from the *transformer* gene', *Cell* **50**(5), 739–747.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/2441872*

Bopp, D., Bell, L. R., Cline, T. W. & Schedl, P. (1991), 'Developmental distribution of female-specific Sex-lethal proteins in *Drosophila* melanogaster', *Genes Dev* **5**(3), 403–415.

URL: *https://www.ncbi.nlm.nih.gov/pubmed/1900493*

Bopp, D., Calhoun, G., Horabin, J. I., Samuels, M. & Schedl, P. (1996), 'Sex-specific control of *Sex-lethal* is a conserved mechanism for sex determination in the genus *Drosophila*', *Development* **122**(3), 971–982.
URL: *https://www.zora.uzh.ch/id/eprint/487/*

Bopp, D., Saccone, G. & Beye, M. (2014), 'Sex determination in insects: Variations on a common theme', *Sex. Dev.* **8**(1-3), 20–28.
URL: *https://www.ncbi.nlm.nih.gov/pubmed/24335049*

Bradley, K. M., Breyer, J. P., Melville, D. B., Broman, K. W., Knapik, E. W. & Smith, J. R. (2011), 'An SNP-based linkage map for zebrafish reveals sex determination loci', *G3 Genes, Genomes, Genet.* **1**(1), 3–9.
URL: *https://www.g3journal.org/content/1/1/3*

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*', *Proc. Natl. Acad. Sci. U. S. A.* **115**(18), 4725–4730.
URL: *https://www.pnas.org/content/115/18/4725*

Burt, A. (2003), 'Site-specific selfish genes as tools for the control and genetic engineering of natural populations', *Proc Biol Sci* **270**(1518), 921–928.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/12803906*

Burtis, K. C. & Baker, B. S. (1989), '*Drosophila doublesex* gene controls somatic sexual differentiation by producing alternatively spliced mRNAs encoding related sex-specific polypeptides', *Cell* **56**(6), 997–1010.
URL: *https://www.ncbi.nlm.nih.gov/pubmed/2493994*

Burtis, K. C., Coschigano, K. T., Baker, B. S. & Wensink, P. C. (1991), 'The *doublesex* proteins of *Drosophila melanogaster* bind directly to a sex-specific *yolk* protein gene enhancer', *EMBO J* **10**(9), 2577–2582.
URL: *https://www.ncbi.nlm.nih.gov/pubmed/1907913*

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 Negl. Trop. Dis.* **9**(7), e0003864.
URL: *https://dx.plos.org/10.1371/journal.pntd.0003864*

Castillo, E. R., Marti, D. A. & Bidau, C. J. (2010), 'Sex and neo-sex chromosomes in orthoptera: A review', *J. Orthoptera Res.* **19**(2), 213–231.
URL: *https://doi.org/10.1665/034.019.0207*

Catteruccia, F., Benton, J. P. & Crisanti, A. (2005), 'An *Anopheles* transgenic sexing strain for vector control', *Nat. Biotechnol.* **23**(11), 1414–1417.
URL: *https://www.nature.com/articles/nbt1152*

CDC (2015), 'Ross and the discovery that mosquitoes transmit malaria parasites'. [Online; accessed 28-July-2020].
URL: *https://www.cdc.gov/malaria/about/history/ross.html*

Champer, J., Buchman, A. & Akbari, O. S. (2016), 'Cheating evolution: Engineering gene drives to manipulate the fate of wild populations', *Nat. Rev. Genet.* **17**(3), 146–159.
URL: *http://dx.doi.org/10.1038/nrg.2015.34*

Chan, Y. S., Naujoks, D. A., Huen, D. S. & Russell, S. (2011), 'Insect population control by homing endonuclease-based gene drive: an evaluation in *Drosophila melanogaster*', *Genetics* **188**(1), 33–44.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/21368273*

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).
URL: *https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0074254*

Charlesworth, B. & Charlesworth, D. (1978), ‘A model for the evolution of dioecy and gynodioecy’, *Am. Nat.* **6**(3), 14–15.
URL: *https://www.jstor.org/stable/2460344?seq=1*

Charlesworth, B. & Charlesworth, D. (2000), ‘The degeneration of Y chromosomes’, *Philos. Trans. R. Soc. B Biol. Sci.* **355**(1403), 1563–1572.
URL: *https://royalsocietypublishing.org/doi/10.1098/rstb.2000.0717*

Charlesworth, D. (2015), ‘Plant contributions to our understanding of sex chromosome evolution’, *New Phytol.* **208**(1), 52–65.
URL: *https://nph.onlinelibrary.wiley.com/doi/full/10.1111/nph.13497*

Charlesworth, D. & Willis, J. H. (2009), ‘The genetics of inbreeding depression’, *Nat. Rev. Genet.* **10**(11), 783–796.
URL: *https://www.nature.com/articles/nrg2664*

Chen, C.-h. 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.* **597**(2007), 597–600.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/17395794*

Chen, S. & Corces, V. G. (2001), ‘The gypsy insulator of *Drosophila* affects chromatin structure in a directional manner’, *Genetics* **159**(4), 1649–1658.
URL: *https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1461898/*

Cho, S. & Wensink, P. C. (1997), ‘DNA binding by the male and female *doublesex* proteins of *Drosophila melanogaster*’, *J Biol Chem* **272**(6), 3185–3189.
URL: *https://www.ncbi.nlm.nih.gov/pubmed/9013552*

Ciofi, C. & Swingland, I. R. (1997), ‘Environmental sex determination in reptiles’, *Appl. Anim. Behav. Sci.* **51**(3-4), 251–265.
URL: *https://www.sciencedirect.com/science/article/pii/S0168159196011082*

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’, *Nat. Biotechnol.* **37**(3), 224–226.
URL: *https://www.nature.com/articles/s41587-019-0032-3*

Cline, T. W. (1988), ‘Evidence that *sisterless-a* and *sisterless-b* are two of several discrete "numerator elements" of the X/A sex determination signal in *Drosophila* that switch *Sxl* between two alternative stable expression states.’, *Genetics* **119**(4), 829–862.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/3137120*

Cline, T. W. (1993), ‘The *Drosophila* sex determination signal - How do flies count to two?’, *Trends Genet.* **9**(11), 385–390.
URL: *https://www.cell.com/trends/genetics/fulltext/0168-9525(93)90138-8*

Coetzee, M., Craig, M. & Le Sueur, D. (2000), ‘Distribution of African malaria mosquitoes belonging to the *Anopheles gambiae* complex’, *Parasitol. Today* **16**(2), 74–77.
URL: *https://www.sciencedirect.com/science/article/pii/S016947589901563X*

Coetzee, M., Hunt, R. H., Wilkerson, R., Torre, A. D., Coulibaly, M. B. & Besansky, N. J. (2013), ‘*Anopheles coluzzii* and *Anopheles amharicus*, new members of the *Anopheles gambiae* complex’, *Zootaxa* **3619**(3), 246–274.
URL: *http://dx.doi.org/10.11646/zootaxa.3619.3.2*

Collins, C. M., Bonds, J. A., Quinlan, M. M. & Mumford, J. D. (2019), ‘Effects of the removal or reduction in density of the malaria mosquito, *Anopheles gambiae s.l.*, on interacting predators and competitors in local ecosystems’.
URL: *https://onlinelibrary.wiley.com/doi/full/10.1111/mve.12327*

Conway, D. J. (2015), 'Paths to a malaria vaccine illuminated by parasite genomics', *Trends Genet.* **31**(2), 97–107.
URL: *http://dx.doi.org/10.1016/j.tig.2014.12.005*

Cook, J. (1993), 'Sex determination in the Hymenoptera: a review of models and evidence', *Heredity (Edinb).* **71**(March), 421–435.
URL: *https://www.nature.com/articles/hdy1993157*

Coschigano, K. T. & Wensink, P. C. (1993), 'Sex-specific transcriptional regulation by the male and female *doublesex* proteins of *Drosophila*', *Genes Dev* **7**(1), 42–54.
URL: *https://www.ncbi.nlm.nih.gov/pubmed/8422987*

Craig, G. B., Hickey, W. A. & VandeHey, R. C. (1960), 'An iInherited male-producing factor in *Aedes aegypti*', *Science.* **132**(3443), 1887–1889.
URL: *https://www.sciencemag.org/lookup/doi/10.1126/science.132.3443.1887*

Criscione, F., Qi, Y., Saunders, R., Hall, B. & Tu, Z. (2013), 'A unique Y gene in the Asian malaria mosquito *Anopheles stephensi* encodes a small lysine-rich protein and is transcribed at the onset of embryonic development', *Insect Mol. Biol.* **22**(4), 433–441.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1111/imb.12034*

Cui, X., Wang, Q., Yin, W., Xu, H., Wilson, Z. A., Wei, C., Pan, S. & Zhang, D. (2012), 'PMRD: A curated database for genes and mutants involved in plant male reproduction', *BMC Plant Biol.* **12**.
URL: *https://bmcplantbiol.biomedcentral.com/articles/10.1186/1471-2229-12-215*

Curtis, C. F. (1968), 'Possible use of translocations to fix desirable genes in insect pest populations', *Nature* **218**(5139), 368–369.
URL: *http://www.nature.com/articles/218368a0*

Dalton, J. E., Fear, J. M., Knott, S., Baker, B. S., McIntyre, L. M. & Arbeitman, M. N. (2013), 'Male-specific *fruitless* isoforms have different regulatory roles conferred by distinct zinc finger DNA binding domains', *BMC Genomics* **14**(1), 1.
URL: *https://bmcgenomics.biomedcentral.com/articles/10.1186/1471-2164-14-659*

Dame, D. A., Benedict, M. Q., Robinson, A. S. & Knols, B. G. (2009), 'Historical applications of induced sterilisation in field populations of mosquitoes', *Malar. J.* **8**(SUPPL. 2), 1–10.
URL: *https://malariajournal.biomedcentral.com/articles/10.1186/1475-2875-8-S2-S2*

Davis, S., Bax, N. & Grewe, P. (2001), 'Engineered underdominance allows efficient and economical introgression of traits into pest populations', *J. Theor. Biol.* **212**(1), 83–98.
URL: *https://www.sciencedirect.com/science/article/pii/S0022519301923574*

Demir, E. & Dickson, B. J. (2005), '*Fruitless* splicing specifies male courtship behavior in *Drosophila*', *Cell* **121**(5), 785–794.
URL: *https://www.ncbi.nlm.nih.gov/pubmed/15935764*

Deredec, A., Burt, A. & Godfray, H. C. J. (2008), 'The population genetics of using homing endonuclease genes in vector and pest management', *Genetics* **179**(4), 2013–2026.
URL: *https://www.genetics.org/content/179/4/2013.figures-only*

Deredec, A., Godfray, H. C. & Burt, A. (2011), 'Requirements for effective malaria control with homing endonuclease genes', *Proc Natl Acad Sci U S A* **108**(43), E874–80.
URL: *https://www.ncbi.nlm.nih.gov/pubmed/21976487*

Dicarlo, J. E., Norville, J. E., Mali, P., Rios, X., Aach, J. & Church, G. M. (2013), 'Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems', *Nucleic Acids Res.* **41**(7), 4336–4343.
URL: *https://academic.oup.com/nar/article/41/7/4336/1075252*

Diggle, P. K., Di Stilio, V. S., Gschwend, A. R., Golenberg, E. M., Moore, R. C., Russell, J. R. & Sinclair, J. P. (2011), 'Multiple developmental processes underlie sex differentiation in angiosperms', *Trends Genet.* **27**(9), 368–376.

URL: *https://www.cell.com/trends/genetics/fulltext/S0168-9525(11)00064-3*

Doench, J. G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E. W., Donovan, K. F., Smith, I., Tothova, Z., Wilen, C., Orchard, R., Virgin, H. W., Listgarten, J. & Root, D. E. (2016), 'Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9', *Nat. Biotechnol.* **34**(2), 184–191.
URL: *https://www.nature.com/articles/nbt.3437*

Dondorp, A. M., Nosten, F., Yi, P., Das, D., Phyo, A. P., Tarning, J., Lwin, K. M., Ariey, F., Hanpithakpong, W., Lee, S. J., Ringwald, P., Silamut, K., Imwong, M., Chotivanich, K., Lim, P., Herdman, T., An, S. S., Yeung, S., Singhasivanon, P., Day, N. P., Lindegardh, N., Socheat, D. & White, N. J. (2009), 'Artemisinin resistance in *Plasmodium falciparum* malaria', *N. Engl. J. Med.* **361**(5), 455–467.
URL: *http://www.nejm.org/doi/abs/10.1056/NEJMoa0808859*

Duan, J., Xu, H., Ma, S., Guo, H., Wang, F., Zhang, L., Zha, X., Zhao, P. & Xia, Q. (2014), 'Ectopic expression of the male BmDSX affects formation of the chitin plate in female *Bombyx mori*', *Mol. Reprod. Dev.* **81**(3), 240–247.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1002/mrd.22290*

Dübendorfer, A., Hediger, M., Burghardt, G. & Bopp, D. (2002), '*Musca domestica*, a window on the evolution of sex-determining mechanisms in insects', *Int. J. Dev. Biol.* **46**(1), 75–79.
URL: *http://www.ijdb.ehu.es/web/paper.php?doi=11902690*

Duffy, J. B. & Gergen, J. P. (1991), 'The *Drosophila* segmentation gene *runt* acts as a position-specific numerator element necessary for the uniform expression of the sex-determining gene *sex-lethal*', *Genes Dev.* **5**(12a), 2176–2187.
URL: *https://europepmc.org/article/med/1748277*

Eggers, S. & Sinclair, A. (2012), 'Mammalian sex determination-insights from humans and mice', *Chromosom. Res.* **20**(1), 215–238.
URL: *https://link.springer.com/article/10.1007/s10577-012-9274-3*

Engler, C., Gruetzner, R., Kandzia, R. & Marillonnet, S. (2009), 'Golden gate shuffling: A one-pot DNA shuffling method based on type ils restriction enzymes', *PLoS One* **4**(5).
URL: *https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0005553*

Engler, C., Kandzia, R. & Marillonnet, S. (2008), 'A one pot, one step, precision cloning method with high throughput capability', *PLoS One* **3**(11).
URL: *https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0003647*

Eppley, S. M. & Jesson, L. K. (2008), 'Moving to mate: The evolution of separate and combined sexes in multicellular organisms', *J. Evol. Biol.* **21**(3), 727–736.
URL: *https://onlinelibrary.wiley.com/doi/10.1111/j.1420-9101.2008.01524.x*

Erdman, S. E. & Burtis, K. C. (1993), 'The *Drosophila doublesex* proteins share a novel zinc finger related DNA binding domain', *EMBO J* **12**(2), 527–535.
URL: *https://www.ncbi.nlm.nih.gov/pubmed/8440242*

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://www.ncbi.nlm.nih.gov/pubmed/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 Biol.* **5**(12), 2821–2830.
URL: *https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.0050332*

Ezaz, T., Stiglec, R., Veyrunes, F. & Graves, M. A. J. (2006), 'Relationships between vertebrate ZW and XY sex chromosome systems', *Curr. Biol.* **16**(17).
URL: *http://www.cell.com/current-biology/retrieve/pii/S0960982206019968*

Facchinelli, L., North, A. R., Collins, C. M., Menichelli, M., Persampieri, T., Bucci, A., Spaccapelo, R., Crisanti, A. & Benedict, M. Q. (2019), 'Large-cage assessment of a transgenic sex-ratio distortion strain on

populations of an African malaria vector', *Parasit. Vectors* **12**(1), 70.
URL: *https://dx.doi.org/10.1186/s13071-019-3289-y*

Flannery, E. L., Chatterjee, A. K. & Winzeler, E. A. (2013), 'Antimalarial drug discovery-approaches and progress towards new medicines', *Nat. Rev. Microbiol.* **11**(12), 849–862.
URL: *https://www.nature.com/articles/nrmicro3138*

Foronda, D., Martín, P. & Sánchez-Herrero, E. (2012), '*Drosophila Hox* and sex-determination genes control segment elimination through EGFR and *extramacrochetae* activity', *PLoS Genet.* **8**(8).
URL: *https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1002874*

Fox, R. M. & Weiser, J. (1959), 'A Microsporidian parasite of *Anopheles gambiae* in Liberia', *J. Parasitol.* **45**(1), 21–30.
URL: *https://www.cabdirect.org/cabdirect/abstract/19601000190*

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

Gailey, D. A., Billeter, J. C., Liu, J. H., Bauzon, F., Allendorfer, J. B. & Goodwin, S. F. (2006), 'Functional conservation of the *fruitless* male sex-determination gene across 250 Myr of insect evolution', *Mol Biol Evol* **23**(3), 633–643.
URL: *https://www.ncbi.nlm.nih.gov/pubmed/16319090*

Gajanana, A., Tewari, S. C., Reuben, R. & Rajagopalan, P. K. (1979), 'Partial suppression of malaria parasites in *Aedes aegypti* and *Anopheles stephensi* doubly infected with *Nosema algerae* and *Plasmodium*', *Indian J. Med. Res.* **70**, 417–423.
URL: *https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2555140/*

Galizi, R., Doyle, L. A., Menichelli, M., Bernardini, F., Deredec, A., Burt, A., Stoddard, B. L., Windbichler, N. & Crisanti, A. (2014), 'A synthetic sex ratio distortion system for the control of the human malaria mosquito', *Nat Commun* **5**, 3977.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/24915045*

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', *Sci. Rep.* **6**(1), 31139.
URL: *http://www.nature.com/articles/srep31139*

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://science.sciencemag.org/content/348/6233/442*

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*', *Proc. Natl. Acad. Sci.* **112**(49), E6736–E6743.
URL: *https://www.ncbi.nlm.nih.gov/pubmed/26598698*

Gempe, T. & Beye, M. (2011), 'Function and evolution of sex determination mechanisms, genes and pathways in insects', *BioEssays* **33**(1), 52–60.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1002/bies.201000043*

Gempe, T., Hasselmann, M., Schiøtt, M., Hause, G., Otte, M. & Beye, M. (2009), 'Sex Determination in honeybees: Two separate mechanisms induce and maintain the female pathway', *PLoS Biol.* **7**(10), e1000222.
URL: *https://dx.plos.org/10.1371/journal.pbio.1000222*

Geuverink, E. & Beukeboom, L. W. (2014), 'Phylogenetic distribution and evolutionary dynamics of the sex determination genes *doublesex* and *transformer* in insects', *Sex. Dev.* **8**(1-3), 38–49.
URL: *https://www.karger.com/Article/FullText/357056*

Ghorbal, M., Gorman, M., Macpherson, C. R., Martins, R. M., Scherf, A. & Lopez-Rubio, J.-J. (2014), 'Genome editing in the human malaria parasite *Plasmodium falciparum* using the CRISPR-Cas9 system', *Nat. Biotechnol.* **32**(8), 819–821.
URL: *http://www.nature.com/articles/nbt.2925*

Gibson, D. G., Young, L., Chuang, R. Y., Venter, J. C., Hutchison, C. A. & Smith, H. O. (2009), 'Enzymatic assembly of DNA molecules up to several hundred kilobases', *Nat. Methods* **6**(5), 343–345.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/19363495*

Gillies, M. T. & Smith, A. (1960), 'The effect of a residual house-spraying campaign in East Africa on species balance in the *Anopheles funestus* group. The replacement of *A. funestus* Giles by *A. rivulorum* Leeson', *Bull. Entomol. Res.* **51**(2), 243–252.
URL: *https://doi.org/10.1017/S0007485300057953*

Giraldo-Calderón, G. I., Emrich, S. J., MacCallum, R. M., Maslen, G., Dialynas, E., Topalis, P., Ho, N., Gesing, S., Madey, G., Collins, F. H. & Lawson, D. (2015), 'VectorBase: an updated bioinformatics resource for invertebrate vectors and other organisms related with human diseases', *Nucleic Acids Res.* **43**(D1), D707–D713.
URL: *https://www.ncbi.nlm.nih.gov/pubmed/25510499*

Gomes, F. M. & Barillas-Mury, C. (2018), 'Infection of Anopheline mosquitoes with *Wolbachia*: Implications for malaria control', *PLoS Pathog.* **14**(11), 1–6.
URL: *https://journals.plos.org/plospathogens/article?id=10.1371/journal.ppat.1007333*

Gomes, F. M., Hixson, B. L., Tyner, M. D., Ramirez, J. L., Canepa, G. E., Alves e Silva, T. L., Molina-Cruz, A., Keita, M., Kane, F., Traoré, B., Sogoba, N. & Barillas-Mury, C. (2017), 'Effect of naturally occurring *Wolbachia* in *Anopheles gambiae s.l.* mosquitoes from Mali on *Plasmodium falciparum* malaria transmission', *Proc. Natl. Acad. Sci. U. S. A.* **114**(47), 12566–12571.
URL: *https://www.pnas.org/content/114/47/12566*

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', *Nat. Biotechnol.* **23**(4), 453–456.
URL: *http://www.nature.com/articles/nbt1071*

Goodwin, S. F., Taylor, B. J., Villella, A., Foss, M., Ryner, L. C., Baker, B. S. & Hall, J. C. (2000), 'Aberrant splicing and altered spatial expression patterns in *fruitless* mutants of *Drosophila melanogaster*', *Genetics* **154**(2), 725–745.
URL: *https://www.ncbi.nlm.nih.gov/pubmed/10655225*

Gotoh, H., Miyakawa, H., Ishikawa, A., Ishikawa, Y., Sugime, Y., Emlen, D. J., Lavine, L. C. & Miura, T. (2014), 'Developmental link between sex and nutrition; *doublesex* regulates sex-specific mandible growth via juvenile hormone signaling in Stag beetles', *PLoS Genet.* **10**(1), 1–9.
URL: *https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1004098*

Griffin, J. T., Bhatt, S., Sinka, M. E., Gething, P. W., Lynch, M., Patouillard, E., Shutes, E., Newman, R. D., Alonso, P., Cibulskis, R. E. & Ghani, A. C. (2016), *Lancet Infect. Dis.* **16**(4), 465–472.
URL: *http://dx.doi.org/10.1016/S1473-3099(15)00423-5*

Guler, Y., Short, S., Kile, P. & Ford, A. T. (2012), 'Integrating field and laboratory evidence for environmental sex determination in the amphipod, *Echinogammarus marinus*', *Mar. Biol.* **159**(12), 2885–2890.
URL: *https://link.springer.com/article/10.1007/s00227-012-2042-2*

Gullan, P. & Cranston, P. (2010), *The Insects: An outline of entomology 4th Edition*, Wiley-Blackwell.
URL: *https://www.wiley.com/en-gb/The+Insects%3A+An+Outline+of+Entomology%2C+4th+Edition-p-9781444317671*

Guy Reeves, R., Bryk, J., Altrock, P. M., Denton, J. A. & Reed, F. A. (2014), 'First steps towards underdominant genetic transformation of insect populations', *PLoS One* **9**(5).
URL: *https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0097557*

Hafez, M. & Hausner, G. (2012), 'Homing endonucleases: DNA scissors on a mission', *Genome* **55**(8), 553–569.
URL: *https://www.nrcresearchpress.com/doi/full/10.1139/g2012-049#.Xyffbi2ZPNA*

Hall, A. B., Basu, S., Jiang, X., Qi, Y., Timoshevskiy, V. A., Biedler, J. K., Sharakhova, M. V., Elahi, R., Anderson, M. A., Chen, X. G., Sharakhov, I. V., Adelman, Z. N. & Tu, Z. (2015), 'A male-determining factor in the mosquito *Aedes aegypti*', *Science.* **348**(6240), 1268–1270.
URL: *https://www.ncbi.nlm.nih.gov/pubmed/25999371*

Hall, A. B., Papathanos, P. A., Sharma, A., Cheng, C., Akbari, O. S., Assour, L., Bergman, N. H., Cagnetti, A., Crisanti, A., Dottorini, T., Fiorentini, E., Galizi, R., Hnath, J., Jiang, X., Koren, S., Nolan, T., Radune, D., Sharakhova, M. V., Steele, A., Timoshevskiy, V. A., Windbichler, N., Zhang, S., Hahn, M. W., Phillippy, A. M., Emrich, S. J., Sharakhov, I. V., Tu, Z. J. & Besansky, N. J. (2016), 'Radical remodeling of the Y chromosome in a recent radiation of malaria mosquitoes', *Proc. Natl. Acad. Sci. U. S. A.* **113**(15), E2114–E2123.
URL: *https://www.pnas.org/content/113/15/E2114*

Hall, A. B., Qi, Y., Timoshevskiy, V., Sharakhova, M. V., Sharakhov, I. V. & Tu, Z. (2013), 'Six novel Y chromosome genes in *Anopheles* mosquitoes discovered by independently sequencing males and females', *BMC Genomics* **14**(1).
URL: *https://bmcgenomics.biomedcentral.com/articles/10.1186/1471-2164-14-273*

Hamilton, W. D. (1967), 'Extraordinary sex ratios', *Science.* **156**(3774), 477–488.
URL: *https://www.sciencemag.org/lookup/doi/10.1126/science.156.3774.477*

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*', *Nat. Biotechnol.* **34**(1), 78–83.
URL: *http://www.nature.com/articles/nbt.3439*

Hammond, A., Karlsson, X., Morianou, I., Kyrou, K., Beaghton, A., Gribble, M., Kranjc, N., Galizi, R., Burt, A., Crisanti, A. & Nolan, T. (2020a), 'Regulation of gene drive expression increases invasive potential and mitigates resistance', *bioRxiv* .
URL: *https://www.biorxiv.org/content/10.1101/360339v2*

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 Genet.* **13**(10), e1007039.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/28976972*

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. (2020b), *Successful suppression of age-structured large cage populations of the malaria vector using the Ag(QFS)1 suppression gene drive.* Manuscript in preparation.

Han, B. & Weiss, L. M. (2017), 'Microsporidia: obligate intracellular pathogens within the fungal kingdom', *Microbiol. Spectr.* **5**(2).
URL: *https://www.asmscience.org/content/journal/microbiolspec/10.1128/microbiolspec.FUNK-0018-2016*

Haq, N., Reisen, W. K. & Aslamkhan, M. (1981), 'The effects of *Nosema algerae* on the horizontal life table attributes of *Anopheles stephensi* under laboratory conditions', *J. Invertebr. Pathol.* **37**(3), 236–242.
URL: *https://www.sciencedirect.com/science/article/abs/pii/0022201181900811*

Harris, A. F., McKemey, A. R., Nimmo, D., Curtis, Z., Black, I., Morgan, S. A., Oviedo, M. N., Lacroix, R., Naish, N., Morrison, N. I., Collado, A., Stevenson, J., Scaife, S., Dafa'alla, T., Fu, G., Phillips, C., Miles, A., Raduan, N., Kelly, N., Beech, C., Donnelly, C. A., Petrie, W. D. & Alphey, L. (2012), 'Successful suppression of a field mosquito population by sustained release of engineered male mosquitoes', *Nat. Biotechnol.*

30(9), 828–830.
URL: *https://www.nature.com/articles/nbt.2350*

Hasselmann, M., Gempe, T., Schiøtt, M., Nunes-Silva, C. G., Otte, M. & Beye, M. (2008), 'Evidence for the evolutionary nascence of a novel sex determination pathway in honeybees', *Nature* **454**(7203), 519–522.
URL: *https://www.nature.com/articles/nature07052*

Hay, S. I., Sinka, M. E., Okara, R. M., Kabaria, C. W., Mbithi, P. M., Tago, C. C., Benz, D., Gething, P. W., Howes, R. E., Patil, A. P., Temperley, W. H., Bangs, M. J., Chareonviriyaphap, T., Elyazar, I. R. F., Harbach, R. E., Hemingway, J., Manguin, S., Mbogo, C. M., Rubio-Palis, Y. & Godfray, H. C. J. (2010), 'Developing global maps of the dominant *Anopheles* vectors of human malaria', *PLoS Med.* **7**(2), e1000209.
URL: *https://dx.plos.org/10.1371/journal.pmed.1000209*

Hediger, M., Henggeler, C., Meier, N., Perez, R., Saccone, G. & Bopp, D. (2010), 'Molecular characterization of the key switch F provides a basis for understanding the rapid divergence of the sex-determining pathway in the housefly', *Genetics* **184**(1), 155–170.
URL: *https://europepmc.org/article/pmc/pmc2815913*

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. Natl. Acad. Sci. U. S. A.* **97**(15), 8229–8232.
URL: *https://www.pnas.org/content/97/15/8229*

Herren, J. K., Mbaisi, L., Mararo, E., Makhulu, E. E., Mobegi, V. A., Butungi, H., Mancini, M. V., Oundo, J. W., Teal, E. T., Pinaud, S., Lawniczak, M. K. N., Jabara, J., Nattoh, G. & Sinkins, S. P. (2020), 'A microsporidian impairs *Plasmodium falciparum* transmission in *Anopheles arabiensis* mosquitoes', *Nat. Commun.* **11**(1), 2187.
URL: *https://www.nature.com/articles/s41467-020-16121-y*

Hertig, M. & Wolbach, S. B. (1924), 'Studies on *Rickettsia*-like micro-organisms in insects.', *J. Med. Res.* **44**(3), 329–374.7.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/19972605*

Hickey, W. A. & Craig Jr., G. B. (1966), 'Distortion of sex ratio in populations of *Aedes aegypti*', *Can. J. Genet. Cytol.* **8**(2), 260–278.
URL: *http://www.nrcresearchpress.com/doi/10.1139/g66-033*

Hildreth, P. E. (1965), '*Doublesex*, a recessive gene that transforms both males and females of *Drosophila* into intersexes', *Genetics* **51**(4), 659–678.
URL: *http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1224459/*

Hoffman, S. L., Rogers, W. O., Carucci, D. J. & Venter, J. C. (1998), 'From genomics to vaccines: Malaria as a model system', *Nat. Med.* **4**(12), 1351–1353.
URL: *https://doi.org/10.1038/3934*

Hoffmann, A. A., Iturbe-Ormaetxe, I., Callahan, A. G., Phillips, B. L., Billington, K., Axford, J. K., Montgomery, B., Turley, A. P. & O'Neill, S. L. (2014), 'Stability of the wMel *Wolbachia* infection following invasion into *Aedes aegypti* populations', *PLoS Negl. Trop. Dis.* **8**(9).
URL: *https://journals.plos.org/plosntds/article?id=10.1371/journal.pntd.0003115*

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* **476**(7361), 454–459.
URL: *https://www.nature.com/articles/nature10356*

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://science.sciencemag.org/content/298/5591/129.long*

Hong, C. S., Park, B. Y. & Saint-Jeannet, J. P. (2007), 'The function of *Dmrt* genes in vertebrate development: It is not just about sex', *Dev. Biol.* **310**(1), 1–9.
URL: *https://www.ncbi.nlm.nih.gov/pubmed/17720152*

Hughes, G. L., Dodson, B. L., Johnson, R. M., Murdock, C. C., Tsujimoto, H., Suzuki, Y., Patt, A. A., Cui, L., Nossa, C. W., Barry, R. M., Sakamoto, J. M., Hornett, E. A. & Rasgon, J. L. (2014), 'Native microbiome impedes vertical transmission of *Wolbachia* in *Anopheles* mosquitoes', *Proc. Natl. Acad. Sci. U. S. A.* **111**(34), 12498–12503.
URL: *https://www.pnas.org/content/early/2014/08/06/1408888111*

Hughes, G. L., Koga, R., Xue, P., Fukatsu, T. & Rasgon, J. L. (2011), '*Wolbachia* infections are virulent and inhibit the human malaria parasite *Plasmodium falciparum* in *Anopheles gambiae*', *PLoS Pathog.* **7**(5), 3–10.
URL: *https://journals.plos.org/plospathogens/article?id=10.1371/journal.ppat.1002043*

Hulme, P. E. (2006), 'Beyond control: Wider implications for the management of biological invasions', *J. Appl. Ecol.* **43**(5), 835–847.
URL: *https://besjournals.onlinelibrary.wiley.com/doi/full/10.1111/j.1365-2664.2006.01227.x*

Hwang, W. Y., Fu, Y. F., Reyon, D., Maeder, M. L., Tsai, S. Q., Sander, J. D., Peterson, R. T., Yeh, J. R. J. & Joung, J. K. (2013), 'Efficient genome editing in zebrafish using a CRISPR-Cas system', *Nat. Biotechnol.* **31**(3), 227–229.
URL: *https://www.nature.com/articles/nbt.2501*

Ito, H., Sato, K., Koganezawa, M., Ote, M., Matsumoto, K., Hama, C. & Yamamoto, D. (2012), '*Fruitless* recruits two antagonistic chromatin factors to establish single-neuron sexual dimorphism', *Cell* **149**(6), 1327–1338.
URL: *http://dx.doi.org/10.1016/j.cell.2012.04.025*

Ito, Y., Harigai, A., Nakata, M., Hosoya, T., Araya, K., Oba, Y., Ito, A., Ohde, T., Yaginuma, T. & Niimi, T. (2013), 'The role of *doublesex* in the evolution of exaggerated horns in the Japanese rhinoceros beetle', *EMBO Rep* **14**(6), 561–567.
URL: *https://www.ncbi.nlm.nih.gov/pubmed/23609854*

Jackson, R. R. & Nelson, X. J. (2012), '*Evarcha culicivora* chooses blood-fed *Anopheles* mosquitoes but other East African jumping spiders do not', *Med. Vet. Entomol.* **26**(2), 233–235.
URL: *https://doi.org/10.1111/j.1365-2915.2011.00986.x*

James, S., Collins, F. H., Welkhoff, P. A., Emerson, C., Godfray, H. C. J., Gottlieb, M., Greenwood, B., Lindsay, S. W., Mbogo, C. M., Okumu, F. O., Quemada, H., Savadogo, M., Singh, J. A., Tountas, K. H. & Touré, Y. T. (2018), 'Pathway to deployment of gene drive mosquitoes as a potential biocontrol tool for elimination of malaria in sub-Saharan Africa: Recommendations of a scientific working group', *Am. J. Trop. Med. Hyg.* **98**(6_Suppl), 1–49.

URL: *http://www.ajtmh.org/content/journals/10.4269/ajtmh.18-0083*

Jarne, P. & Auld, J. R. (2006), 'Animals mix it up too: The distribution of self-fertilization among hermaphroditic animals', *Evolution (N. Y).* **60**(9), 1816.
URL: *https://doi.org/10.1554/06-246.1*

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 Res.* **3**, 113.
URL: *https://wellcomeopenresearch.org/articles/3-113/v2*

Jensen, M. P., Allen, C. D., Eguchi, T., Bell, I. P., LaCasella, E. L., Hilton, W. A., Hof, C. A. & Dutton, P. H. (2018), 'Environmental warming and feminization of one of the largest sea turtle populations in the world', *Curr. Biol.* **28**(1), 154–159.e4.
URL: *https://doi.org/10.1016/j.cub.2017.11.057*

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**(6096), 816–821.
URL: *https://www.sciencemag.org/lookup/doi/10.1126/science.1225829*

Kato, Y., Kobayashi, K., Oda, S., Tatarazako, N., Watanabe, H. & Iguchi, T. (2010), 'Sequence divergence and expression of a *transformer* gene in the branchiopod crustacean, *Daphnia magna*', *Genomics* **95**(3), 160–165.
URL: *https://www.ncbi.nlm.nih.gov/pubmed/20060040*

Kato, Y., Kobayashi, K., Watanabe, H. & Iguchi, T. (2011), 'Environmental sex determination in the branchiopod crustacean *Daphnia magna*: Deep conservation of a *doublesex* gene in the sex-determining pathway', *PLoS Genet.* **7**(3).
URL: *https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1001345*

Kijimoto, T., Moczek, A. P. & Andrews, J. (2012), 'Diversification of *doublesex* function underlies morph-, sex-, and species-specific development of beetle horns', *Proc Natl Acad Sci U S A* **109**(50), 20526–20531.
URL: *https://www.ncbi.nlm.nih.gov/pubmed/23184999*

Killeen, G. F. (2013), 'Modulating malaria with *Wolbachia*', *Nat. Med.* **19**(8), 974–975.
URL: *http://www.nature.com/articles/nm.3298*

Kim, S. S., Seawright, J. A. & Kaiser, P. E. (1987), 'A genetic sexing strain of *Anopheles quadrimaculatus, species A.*', *J. Am. Mosq. Control Assoc.* **3**(1), 50–53.
URL: *https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6304763/*

Kim, Y. G., Cha, J. & Chandrasegaran, S. (1996), 'Hybrid restriction enzymes: Zinc finger fusions to Fok I cleavage domain', *Proc. Natl. Acad. Sci. U. S. A.* **93**(3), 1156–1160.
URL: *https://www.pnas.org/content/93/3/1156*

Kimura, K., Ote, M., Tazawa, T. & Yamamoto, D. (2005), '*Fruitless* specifies sexually dimorphic neural circuitry in the *Drosophila* brain', *Nature* **438**(7065), 229–233.
URL: *https://www.ncbi.nlm.nih.gov/pubmed/16281036*

Kitron, U. & Spielman, A. (1989), 'Suppression of transmission of malaria through source reduction: Antianopheline measures applied in Israel, the United States, and Italy', *Rev. Infect. Dis.* **11**(3), 391–406.
URL: *https://academic.oup.com/cid/article/11/3/391/396856*

Kiuchi, T., Koga, H., Kawamoto, M., Shoji, K., Sakai, H., Arai, Y., Ishihara, G., Kawaoka, S., Sugano, S., Shimada, T., Suzuki, Y., Suzuki, M. G. & Katsuma, S. (2014), 'A single female-specific piRNA is the primary determiner of sex in the silkworm', *Nature* **509**(7502), 633–+.
URL: *https://www.nature.com/articles/nature13315*

Klein, S. L., Shann, F., Moss, W. J., Benn, C. S. & Aabye, P. (2016), 'RTS,S malaria vaccine and increased

mortality in girls', *MBio* **7**(2), 2–3.
URL: *https://mbio.asm.org/content/7/2/e00514-16*

Knipling, E. F. (2015), 'Possibilities of insect control or eradication through the use of sexually sterile males', *J. Econ. Entomol.* **48**(4), 459–462.
URL: *https://academic.oup.com/jee/article-abstract/48/4/459/2205947?redirectedFrom=fulltext*

Kobayashi, Y., Nagahama, Y. & Nakamura, M. (2013), 'Diversity and plasticity of sex determination and differentiation in fishes', *Sex. Dev.* **7**(1-3), 115–125.
URL: *https://www.karger.com/Article/FullText/342009*

Konet, D. S., Anderson, J., Piper, J., Akkina, R., Suchman, E. & Carlson, J. (2007), 'Short-hairpin RNA expressed from polymerase III promoters mediates RNA interference in mosquito cells', *Insect Mol Biol* **16**(2), 199–206.
URL: *https://www.ncbi.nlm.nih.gov/pubmed/17298556*

Kopp, A. (2012), '*Dmrt* genes in the development and evolution of sexual dimorphism', *Trends Genet.* **28**(4), 175–184.
URL: *http://dx.doi.org/10.1016/j.tig.2012.02.002*

Kopp, A., Duncan, I. & Carroll, S. B. (2000), 'Genetic control and evolution of sexually dimorphic characters in *Drosophilia*', *Nature* **408**(6812), 553–559.
URL: *https://www.nature.com/articles/35046017*

Krafsur, E. S. (1998), 'Sterile insect technique for suppressing and eradicating insect population: 55 years and counting', *J. Agric. Urban Entomol.* **15**(4), 303–317.
URL: *https://lib.dr.iastate.edu/ent_pubs/417/*

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: *http://www.sciencemag.org/lookup/doi/10.1126/science.aaf5605*

Kumar, S., Stecher, G., Suleski, M. & Hedges, S. B. (2017), 'TimeTree: A Resource for Timelines, Timetrees, and Divergence Times', *Mol. Biol. Evol.* **34**(7), 1812–1819.
URL: *https://academic.oup.com/mbe/article/34/7/1812/3091705*

Kunte, K., Zhang, W., Tenger-Trolander, A., Palmer, D. H., Martin, A., Reed, R. D., Mullen, S. P. & Kronforst, M. R. (2014), '*Doublesex* is a mimicry supergene', *Nature* **507**(7491), 229–232.
URL: *https://www.nature.com/articles/nature13112*

Kurata, M., Wolf, N. K., Lahr, W. S., Weg, M. T., Kluesner, M. G., Lee, S., Hui, K., Shiraiwa, M., Webber, B. R. & Moriarity, B. S. (2018), 'Highly multiplexed genome engineering using CRISPR/Cas9 gRNA arrays', *PLoS One* **13**(9), e0198714.
URL: *https://dx.plos.org/10.1371/journal.pone.0198714*

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', *Nat. Biotechnol.* **36**(11), 1062–1066.
URL: *http://www.nature.com/articles/nbt.4245*

Labun, K., Montague, T. G., Gagnon, J. A., Thyme, S. B. & Valen, E. (2016), 'CHOPCHOP v2: a web tool for the next generation of CRISPR genome engineering', *Nucleic Acids Res* **44**(W1), W272–W276.
URL: *https://academic.oup.com/nar/article/44/W1/W272/2499370*

Lacroix, R., McKemey, A. R., Raduan, N., Kwee Wee, L., Hong Ming, W., Guat Ney, T., Siti Rahidah, A. A., Salman, S., Subramaniam, S., Nordin, O., Norhaida Hanum, A. T., Angamuthu, C., Marlina Mansor, S., Lees, R. S., Naish, N., Scaife, S., Gray, P., Labbé, G., Beech, C., Nimmo, D., Alphey, L., Vasan, S. S., Han Lim, L., Wasi A., N. & Murad, S. (2012), 'Open field release of genetically engineered sterile male *Aedes aegypti* in Malaysia', *PLoS One* **7**(8), 1–9.
URL: *https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0042771*

Lagos, D., Koukidou, M., Savakis, C. & Komitopoulou, K. (2007), 'The *transformer* gene in *Bactrocera oleae*: The genetic switch that determines its sex fate', *Insect Mol. Biol.* **16**(2), 221–230.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2583.2006.00717.x*

Lagos, D., Ruiz, M. F., Sánchez, L. & Komitopoulou, K. (2005), 'Isolation and characterization of the *Bactrocera oleae* genes orthologous to the sex determining *Sex-lethal* and *doublesex* genes of *Drosophila melanogaster*', *Gene* **28**(348), 111–121.
URL: *https://linkinghub.elsevier.com/retrieve/pii/S0378111904007942*

Lahn, B. T. & Page, C. (1999), 'Four evolutionary strata on the human X chromosome', *Science.* **286**(5441), 964–967.
URL: *https://www.sciencemag.org/lookup/doi/10.1126/science.286.5441.964*

Laven, H. (1956), 'Cytoplasmic inheritance in *Culex*', *Nature* **177**(4499), 141–142.
URL: *http://www.nature.com/articles/177141a0*

Laven, H., Cousserans, J. & Guille, G. (1972), 'Eradicating mosquitoes using translocations: a first field experiment', *Nature* **236**(5348), 456–457.
URL: *http://www.nature.com/articles/236456a0*

Lee, G., Foss, M., Goodwin, S. F., Carlo, T., Taylor, B. J. & Hall, J. C. (2000), 'Spatial, temporal, and sexually dimorphic expression patterns of the *fruitless* gene in the *Drosophila* central nervous system', *J. Neurobiol.* **43**(4), 404–426.
URL: *https://pubmed.ncbi.nlm.nih.gov/10861565/*

Liew, W. C., Bartfai, R., Lim, Z., Sreenivasan, R., Siegfried, K. R. & Orban, L. (2012), 'Polygenic sex determination system in zebrafish', *PLoS One* **7**(4).
URL: *https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0034397*

Lindquist, D. A., Abusowa, M. & Hall, M. J. (1992), 'The New World screwworm fly in Libya: a review of its introduction and eradication', *Med. Vet. Entomol.* **6**(1), 2–8.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2915.1992.tb00027.x*

Lindsay, S. W., Adiamah, J. H., Miller, J. E. & Armstrong, J. R. (1991), 'Pyrethroid-treated bednet effects on mosquitoes of the *Anopheles gambiae* complex in The Gambia', *Med. Vet. Entomol.* **5**(4), 477–483.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2915.1991.tb00576.x*

Lines, J. D., Myamba, J. & Curtis, C. F. (1987), 'Experimental hut trials of permethrin-impregnated mosquito nets and eave curtains against malaria vectors in Tanzania', *Med. Vet. Entomol.* **1**(1), 37–51.
URL: *http://doi.wiley.com/10.1111/j.1365-2915.1987.tb00321.x*

Loehlin, D. W., Oliveira, D. C., Edwards, R., Giebel, J. D., Clark, M. E., Cattani, M. V., Van De Zande, L., Verhulst, E. C., Beukeboom, L. W., Muñoz-Torres, M. & Werren, J. H. (2010), 'Non-coding changes cause sex-specific wing size differences between closely related species of *Nasonia*', *PLoS Genet.* **6**(1).
URL: *https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1000821*

Lucchesi, J. (1978), 'Gene dosage compensation and the evolution of sex chromosomes', *Science.* **202**(4369), 711–716.
URL: *https://www.sciencemag.org/lookup/doi/10.1126/science.715437*

Lukhtanov, V. A. (2000), 'Sex chromatin and sex chromosome systems in nonditrysian Lepidoptera (Insecta)', *J. Zool. Syst. Evol. Res.* **38**(2), 73–79.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1046/j.1439-0469.2000.382130.x*

Luo, S. D., Shi, G. W. & Baker, B. S. (2011), 'Direct targets of the *D. melanogaster* DSXF protein and the evolution of sexual development', *Development* **138**(13), 2761–2771.
URL: *https://www.ncbi.nlm.nih.gov/pubmed/21652649*

Lynch, K. W. & Maniatis, T. (1996), 'Assembly of specific SR protein complexes on distinct regulatory elements of the *Drosophila doublesex* splicing enhancer', *Genes Dev.* **10**(16), 2089–2101.
URL: *http://genesdev.cshlp.org/content/10/16/2089*

Magnusson, K., Mendes, A. M., Windbichler, N., Papathanos, P. A., Nolan, T., Dottorini, T., Rizzi, E., Christophides, G. K. & Crisanti, A. (2011), 'Transcription regulation of sex-biased genes during ontogeny in the malaria vector *Anopheles gambiae*', *PLoS One* **6**(6).
URL: *https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0021572*

Mahmoudi, S. & Keshavarz, H. (2017), 'Efficacy of phase 3 trial of RTS, S/AS01 malaria vaccine: The need for an alternative development plan', *Hum. Vaccines Immunother.* **13**(9), 2098–2101.
URL: *https://www.tandfonline.com/doi/full/10.1080/21645515.2017.1295906*

Maïga, H., Damiens, D., Niang, A., Sawadogo, S. P., Fatherhaman, O., Lees, R. S., Roux, O., Dabiré, R. K., Ouédraogo, G. A., Tripet, F., Diabaté, A. & Gilles, J. R. (2014), 'Mating competitiveness of sterile male Anopheles coluzzii in large cages', *Malar. J.* **13**(1), 460.
URL: *https://malariajournal.biomedcentral.com/articles/10.1186/1475-2875-13-460*

Majori, G. (2012), 'Short history of malaria and its eradication in Italy with short notes on the fight against the infection in the Mediterranean basin', *Mediterr. J. Hematol. Infect. Dis.* **4**(1).
URL: *http://www.mjhid.org/index.php/mjhid/article/view/2012.016*

Manoli, D. S., Foss, M., Villella, A., Taylor, B. J., Hall, J. C. & Baker, B. S. (2005), 'Male-specific *fruitless* specifies the neural substrates of *Drosophila* courtship behaviour', *Nature* **436**(7049), 395–400.
URL: *https://www.nature.com/articles/nature03859*

Marec, F. & Novák, K. (1998), 'Absence of sex chromatin corresponds with a sex-chromosome univalent in females of Trichoptera', *Eur. J. Entomol.* **95**(2), 197–209.
URL: *https://www.eje.cz/artkey/eje-199802-0002_Absence_of_sex_chromatin_corresponds_with_a_sex-chromosome_univalent_in_females_of_Trichoptera.php*

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', *Malar. J.* **11**(1), 302.
URL: *http://malariajournal.biomedcentral.com/articles/10.1186/1475-2875-11-302*

Marshall, J. M. & Hay, B. A. (2012), 'General principles of single-construct chromosomal gene drive', *Evolution (N. Y).* **66**(7), 2150–2166.
URL: *https://onlinelibrary.wiley.com/doi/full/10.1111/j.1558-5646.2012.01582.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.* **323**(5910), 141–144.
URL: *https://www.sciencemag.org/lookup/doi/10.1126/science.1165326*

Meise, M., Hilfiker-Kleiner, D., Dubendorfer, A., Brunner, C., Nothiger, R. & Bopp, D. (1998), '*Sex-lethal*, the master sex-determining gene in *Drosophila*, is not sex-specifically regulated in *Musca domestica*', *Development* **125**(8), 1487–1494.
URL: *https://europepmc.org/article/med/9502729*

Mellert, D. J., Robinett, C. C. & Baker, B. S. (2012), '*Doublesex* functions early and late in gustatory sense organ development', *PLoS One* **7**(12).
URL: *https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0051489*

Merchant-Larios, H. & Díaz-Hernández, V. (2013), 'Environmental sex determination mechanisms in reptiles', *Sex. Dev.* **7**(1-3), 95–105.
URL: *https://www.karger.com/Article/Abstract/341936*

Metz, C. W. (1938), 'Chromosome behaviour, inheritance and sex determination in *Sciara*', *Am. Nat.* **72**(743), 485–520.
URL: *https://www.journals.uchicago.edu/doi/10.1086/280803*

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: *http://www.nature.com/articles/nature24995*

Miller, J. C., Tan, S., Qiao, G., Barlow, K. A., Wang, J., Xia, D. F., Meng, X., Paschon, D. E., Leung, E., Hinkley, S. J., Dulay, G. P., Hua, K. L., Ankoudinova, I., Cost, G. J., Urnov, F. D., Zhang, H. S., Holmes, M. C., Zhang, L., Gregory, P. D. & Rebar, E. J. (2011), 'A TALE nuclease architecture for efficient genome editing', *Nat. Biotechnol.* **29**(2), 143–150.
URL: *https://www.nature.com/articles/nbt.1755*

Miller, J. E., Lindsay, S. W. & Armstrong, J. R. (1991), 'Experimental hut trials of bednets impregnated with synthetic pyrethroid or organophosphate insecticide for mosquito control in The Gambia', *Med. Vet. Entomol.* **5**(4), 465–476.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2915.1991.tb00575.x*

Ming, R., Bendahmane, A. & Renner, S. S. (2011), 'Sex chromosomes in land plants', *Annu. Rev. Plant Biol.* **62**(1), 485–514.
URL: *https://www.annualreviews.org/doi/abs/10.1146/annurev-arplant-042110-103914*

Misof, B., Liu, S., Meusemann, K., Peters, R. S., Donath, A., Mayer, C., Frandsen, P. B., Ware, J., Flouri, T., Beutel, R. G., Niehuis, O., Petersen, M., Izquierdo-Carrasco, F., Wappler, T., Rust, J., Aberer, A. J., Aspock, U., Aspock, H., Bartel, D., Blanke, A., Berger, S., Bohm, A., Buckley, T. R., Calcott, B., Chen, J., Friedrich, F., Fukui, M., Fujita, M., Greve, C., Grobe, P., Gu, S., Huang, Y., Jermiin, L. S., Kawahara, A. Y., Krogmann, L., Kubiak, M., Lanfear, R., Letsch, H., Li, Y., Li, Z., Li, J., Lu, H., Machida, R., Mashimo, Y., Kapli, P., McKenna, D. D., Meng, G., Nakagaki, Y., Navarrete-Heredia, J. L., Ott, M., Ou, Y., Pass, G., Podsiadlowski, L., Pohl, H., von Reumont, B. M., Schutte, K., Sekiya, K., Shimizu, S., Slipinski, A., Stamatakis, A., Song, W., Su, X., Szucsich, N. U., Tan, M., Tan, X., Tang, M., Tang, J., Timelthaler, G., Tomizuka, S., Trautwein, M., Tong, X., Uchifune, T., Walzl, M. G., Wiegmann, B. M., Wilbrandt, J., Wipfler, B., Wong, T. K. F., Wu, Q., Wu, G., Xie, Y., Yang, S., Yang, Q., Yeates, D. K., Yoshizawa, K., Zhang, Q., Zhang, R., Zhang, W., Zhang, Y., Zhao, J., Zhou, C., Zhou, L., Ziesmann, T., Zou, S., Li, Y., Xu, X., Zhang, Y., Yang, H., Wang, J., Wang, J., Kjer, K. M. & Zhou, X. (2014), 'Phylogenomics resolves the timing and pattern of insect evolution', *Science.* **346**(6210), 763–767.
URL: *https://www.sciencemag.org/lookup/doi/10.1126/science.1257570*

Missiroli, A. (1938), 'Varieties of *Anopheles maculipennis* and the malaria problem in italy', *Transactions int. Congr. Ent.* **7**(2).
URL: *https://www.cabdirect.org/cabdirect/abstract/19391000099*

Mita, K., Kasahara, M., Sasaki, S., Nagayasu, Y., Yamada, T., Kanamori, H., Namiki, N., Kitagawa, M., Yamashita, H., Yasukochi, Y., Kadono-Okuda, K., Yamamoto, K., Ajimura, M., Ravikumar, G., Shimomura, M., Nagamura, Y., Shin-I, T., Abe, H., Shimada, T., Morishita, S. & Sasaki, T. (2004), 'The genome sequence of silkworm, *Bombyx mori*', *DNA Res.* **11**(1), 27–35.
URL: *https://academic.oup.com/dnaresearch/article/11/1/27/342818*

Monnat, R. J., Hackmann, A. F. & Cantrell, M. A. (1999), 'Generation of highly site-specific DNA double-strand breaks in human cells by the homing endonucleases I-PpoI and I-CreI', *Biochem. Biophys. Res. Commun.* **255**(1), 88–93.
URL: *https://www.sciencedirect.com/science/article/pii/S0006291X99901523*

Mora, C., Tittensor, D. P., Adl, S., Simpson, A. G. & Worm, B. (2011), 'How many species are there on earth and in the ocean?', *PLoS Biol.* **9**(8), 1–8.
URL: *https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1001127*

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://www.cell.com/fulltext/S0092-8674(09)01500-1*

Morrison, N. I., Simmons, G. S., Fu, G., O'Connell, S., Walker, A. S., Dafa'alla, T., Walters, M., Claus, J., Tang, G., Jin, L., Marubbi, T., Epton, M. J., Harris, C. L., Staten, R. T., Miller, E., Miller, T. A. & Alphey, L. (2012), 'Engineered repressible lethality for controlling the pink bollworm, a lepidopteran pest of cotton', *PLoS One* **7**(12), 1–10.
URL: *https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0050922*

Mukherjee, A., Bopp, S., Magistrado, P., Wong, W., Daniels, R., Demas, A., Schaffner, S., Amaratunga, C., Lim, P., Dhorda, M., Miotto, O., Woodrow, C., Ashley, E. A., Dondorp, A. M., White, N. J., Wirth, D., Fairhurst, R. & Volkman, S. K. (2017), 'Artemisinin resistance without *pfkelch13* mutations in *Plasmodium falciparum* isolates from Cambodia', *Malar. J.* **16**(1), 1–12.
URL: *https://malariajournal.biomedcentral.com/articles/10.1186/s12936-017-1845-5*

Müller-Holtkamp, F. (1995), 'The *sex-lethal* gene homologue in *Chrysomya rufifacies* is highly conserved in sequence and exon-intron organization', *J. Mol. Evol.* **41**(4), 467–477.
URL: *https://agris.fao.org/agris-search/search.do?recordID=US9602581*

Munday, P. L., Buston, P. M. & Warner, R. R. (2006), 'Diversity and flexibility of sex-change strategies in animals', *Trends Ecol. Evol.* **21**(2), 89–95.
URL: *https://www.sciencedirect.com/science/article/pii/S0169534705003599*

Munga, S., Minakawa, N., Zhou, G., Barrack, O.-O. J., Githeko, A. K. & Yan, G. (2006), 'Oviposition site preference and egg hatchability of *Anopheles gambiae*: Effects of land cover types', *J. Med. Entomol.* **42**(6), 993–997.
URL: *https://europepmc.org/article/med/16465740*

Muscarella, D. E., Ellison, E. L., Ruoff, B. M. & Vogt, V. M. (1990), 'Characterization of I-Ppo, an intron-encoded endonuclease that mediates homing of a group I intron in the ribosomal DNA of *Physarum polycephalum*', *Mol. Cell. Biol.* **10**(7), 3386–3396.
URL: *https://mcb.asm.org/content/10/7/3386*

Muscarella, D. E. & Vogt, V. M. (1989), 'A mobile group I intron in the nuclear rDNA of *Physarum polycephalum*', *Cell* **56**(3), 443–454.
URL: *https://www.cell.com/fulltext/0092-8674(89)90247-X*

Muyle, A., Shearn, R. & Marais, G. A. (2017), 'The evolution of sex chromosomes and dosage compensation in plants', *Genome Biol. Evol.* **9**(3), 627–645.
URL: *https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5629387/*

Navarro-Martín, L., Viñas, J., Ribas, L., Díaz, N., Gutiérrez, A., Di Croce, L. & Piferrer, F. (2011), 'DNA methylation of the gonadal aromatase *(cyp19a)* promoter is involved in temperature-dependent sex ratio shifts in the European sea bass', *PLoS Genet.* **7**(12).
URL: *https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1002447*

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. M., Deligianni, E., Dottorini, T., Dritsou, V., Gabriel, S. B., Guelbeogo, W. M., Hall, A. B., Han, M. V., Hlaing, T., Hughes, D. S. T., 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. N., 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. C., Tojo, M., Topalis, P., Tubio, J. M. C., 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. T., 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), 1258522.
URL: *https://www.sciencemag.org/lookup/doi/10.1126/science.1258522*

Neville, M. C., Nojima, T., Ashley, E., Parker, D. J., Walker, J., Southall, T., Van De Sande, B., Marques, A. C., Fischer, B., Brand, A. H., Russell, S., Ritchie, M. G., Aerts, S. & Goodwin, S. F. (2014), 'Male-specific *fruitless* isoforms target neurodevelopmental genes to specify a sexually dimorphic nervous system', *Curr. Biol.* **24**(3), 229–241.
URL: *http://dx.doi.org/10.1016/j.cub.2013.11.035*

Ng'Habi, K. R., Knols, B. G., Lee, Y., Ferguson, H. M. & Lanzaro, G. C. (2011), 'Population genetic structure of *Anopheles arabiensis* and *Anopheles gambiae* in a malaria endemic region of southern Tanzania', *Malar. J.* **10**, 1–10.
URL: *https://malariajournal.biomedcentral.com/articles/10.1186/1475-2875-10-289*

Nigro, R. G., Campos, M. C. C. & Perondini, A. L. P. (2007), 'Temperature and the progeny sex-ratio in *Sciara ocellaris* (Diptera, Sciaridae)', *Genet. Mol. Biol.* **30**(1), 152–158.
URL: *https://www.scielo.br/scielo.php?script=sci_arttext&pid=S1415-47572007000100026*

Nissen, I., Müller, M. & Beye, M. (2012), 'The Am-tra2 gene is an essential regulator of female splice regulation at two levels of the sex determination hierarchy of the honeybee', *Genetics* **192**(3), 1015–1026.
URL: *http://www.genetics.org/lookup/doi/10.1534/genetics.112.143925*

Noedl, H., Se, Y., Schaecher, K., Smith, B. L., Socheat, D. & Fukuda, M. M. (2008), 'Evidence of artemisinin-resistant malaria in Western Cambodia', *N. Engl. J. Med.* **359**(24), 2619–2620.
URL: *http://www.nejm.org/doi/abs/10.1056/NEJMc0805011*

Normark, B. B. (2003), 'The evolution of alternative genetic systems in insects', *Annu. Rev. Entomol.* **48**(1), 397–423.
URL: *http://www.annualreviews.org/doi/10.1146/annurev.ento.48.091801.112703*

Ogura, T. & Hiraga, S. (1983), 'Mini-F plasmid genes that couple host cell division to plasmid proliferation.', *Proc. Natl. Acad. Sci.* **80**(15), 4784–4788.
URL: *http://www.pnas.org/cgi/doi/10.1073/pnas.80.15.4784*

Pane, A., Salvemini, M., Bovi, P. D., Polito, C. & Saccone, G. (2002), 'The *transformer* gene in *Ceratitis capitata* provides a genetic basis for selecting and remembering the sexual fate', *Development* **129**(15), 3715–3725.
URL: *https://europepmc.org/article/med/12117820*

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 Mol. Biol.* **10**, 1–13.
URL: *https://bmcmolbiol.biomedcentral.com/articles/10.1186/1471-2199-10-65*

Pates, H. V., Takken, W., Curtis, C. F. & Jamet, H. (2006), 'Zoophilic *Anopheles quadriannulatus* species B found in a human habitation in Ethiopia', *Ann. Trop. Med. Parasitol.* **100**(2), 177–179.
URL: *https://www.tandfonline.com/doi/abs/10.1179/136485906X86374*

Paul, B. & Montoya, G. (2020), 'CRISPR-Cas12a: Functional overview and applications', *Biomed. J.* **43**(1), 8–17.
URL: *https://doi.org/10.1016/j.bj.2019.10.005*

Pen, I., Uller, T., Feldmeyer, B., Harts, A., While, G. M. & Wapstra, E. (2010), 'Climate-driven population divergence in sex-determining systems', *Nature* **468**(7322), 436–438.
URL: *http://dx.doi.org/10.1038/nature09512*

Penalva, L. O., Sakamoto, H., Navarro-Sabaté, A., Sakashita, E., Granadino, B., Segarra, C. & Sánchez, L. (1996), 'Regulation of the gene *Sex-lethal*: A comparative analysis of *Drosophila melanogaster* and *Drosophila subobscura*', *Genetics* **144**(4), 1653–1664.
URL: *https://www.genetics.org/content/144/4/1653.long*

Phillips-Howard, P. A., Nahlen, B. L., Alaii, J. A., Ter Kuile, F. O., Gimnig, J. E., Terlouw, D. J., Kachur, S. P., Hightower, A. W., Lal, A. A., Schoute, E., Oloo, A. J. & Hawley, W. A. (2003), 'The efficacy of permethrin-treated bed nets on child mortality and morbidity in western Kenya I. Development of infrastructure and description of study site', *Am. J. Trop. Med. Hyg.* **68**(4 SUPPL.), 3–9.
URL: *http://www.ajtmh.org/content/journals/10.4269/ajtmh.2003.68.3?crawler=true*

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 Biol.* **5**, 1–11.
URL: *https://bmcbiol.biomedcentral.com/articles/10.1186/1741-7007-5-11*

Pomerantz, A. F., Hoy, M. A. & Kawahara, A. Y. (2015), 'Molecular characterization and evolutionary insights into potential sex-determination genes in the western orchard predatory mite *Metaseiulus occidentalis* (Chelicerata: Arachnida: Acari: Phytoseiidae)', *J. Biomol. Struct. Dyn.* **33**(6), 1239–1253.
URL: *http://www.tandfonline.com/doi/abs/10.1080/07391102.2014.941402*

Port, F., Chen, H.-M., Lee, T. & Bullock, S. L. (2014), 'Optimized CRISPR/Cas tools for efficient germline and somatic genome engineering in *Drosophila*', *Proc. Natl. Acad. Sci.* **111**(29), E2967–E2976.
URL: *http://www.pnas.org/cgi/doi/10.1073/pnas.1405500111*

Price, D. C., Egizi, A. & Fonseca, D. M. (2015), 'The ubiquity and ancestry of insect *doublesex*', *Sci Rep* **5**, 13068.
URL: *https://www.ncbi.nlm.nih.gov/pubmed/26278009*

Price, D. C., Egizi, A. & Fonseca, D. M. (2015a), 'Characterization of the *doublesex* gene within the *Culex pipiens* complex suggests regulatory plasticity at the base of the mosquito sex determination cascade', *Bmc Evol. Biol.* **15**(108).
URL: *https://bmcevolbiol.biomedcentral.com/articles/10.1186/s12862-015-0386-1*

Pugh, J. (2016), 'Driven to extinction? The ethics of eradicating mosquitoes with gene-drive technologies', *J. Med. Ethics* **42**(9), 578–581.
URL: *https://jme.bmj.com/content/42/9/578*

Quinn, A. E., Georges, A., Sarre, S. D., Guarino, F., Ezaz, T. & Marshall Graves, J. A. (2007), 'Temperature sex reversal implies sex gene dosage in a reptile', *Science.* **316**(5823), 411.
URL: *https://science.sciencemag.org/content/316/5823/411.long*

Raymond, C. S., Kettlewell, J. R., Hirsch, B., Bardwell, V. J. & Zarkower, D. (1999a), 'Expression of *Dmrt1* in the genital ridge of mouse and chicken embryos suggests a role in vertebrate sexual development', *Dev. Biol.* **215**(2), 208–220.
URL: *https://www.ncbi.nlm.nih.gov/pubmed/10545231*

Raymond, C. S., Parker, E. D., Kettlewell, J. R., Brown, L. G., Page, D. C., Kusz, K., Jaruzelska, J., Reinberg, Y., Flejter, W. L., Bardwell, V. J., Hirsch, B. & Zarkower, D. (1999), 'A region of human chromosome 9p required for testis development contains two genes related to known sexual regulators', *Hum. Mol. Genet.* **8**(6), 989–996.
URL: *https://academic.oup.com/hmg/article/8/6/989/601886*

Raymond, C. S., Shamu, C. E., Shen, M. M., Seifert, K. J., Hirsch, B., Hodgkin, J. & Zarkower, D. (1998), 'Evidence for evolutionary conservation of sex-determining genes', *Nature* **391**(6668), 691–695.
URL: *https://www.nature.com/articles/35618*

Redman, M., King, A., Watson, C. & King, D. (2016), 'What is CRISPR/Cas9?', *Arch. Dis. Child. Educ. Pract. Ed.* **101**(4), 213–215.
URL: *https://ep.bmj.com/content/101/4/213*

Reidenbach, K. R., Cook, S., Bertone, M. A., Harbach, R. E., Wiegmann, B. M. & Besansky, N. J. (2009), 'Phylogenetic analysis and temporal diversification of mosquitoes (Diptera: Culicidae) based on nuclear genes and morphology', *BMC Evol Biol* **9**, 298.
URL: *https://www.ncbi.nlm.nih.gov/pubmed/20028549*

Renner, S. S. & Ricklefs, R. E. (1995), 'Dioecy and its correlates in the flowering plants', *Am. J. Bot.* **82**(5), 596–606.
URL: *https://bsapubs.onlinelibrary.wiley.com/doi/abs/10.1002/j.1537-2197.1995.tb11504.x*

Roberts, A., De Andrade, P. P., Okumu, F., Quemada, H., Savadogo, M., Singh, J. A. & James, S. (2017), 'Results from the workshop "problem formulation for the use of gene drive in mosquitoes"'.
URL: *https://www.ajtmh.org/view/journals/tpmd/96/3/article-p530.xml*

Ross, M. T., Grafham, D. V., Coffey, A. J., Scherer, S., McLay, K., Muzny, D., Platzer, M., Howell, G. R., Burrows, C., Bird, C. P., Frankish, A., Lovell, F. L., Howe, K. L., Ashurst, J. L., Fulton, R. S., Sudbrak, R., Wen, G., Jones, M. C., Hurles, M. E., Andrews, T. D., Scott, C. E., Searle, S., Ramser, J., Whittaker, A., Deadman, R., Carter, N. P., Hunt, S. E., Chen, R., Cree, A., Gunaratne, P., Havlak, P., Hodgson, A., Metzker, M. L., Richards, S., Scott, G., Steffen, D., Sodergren, E., Wheeler, D. A., Worley, K. C., Ainscough, R., Ambrose, K. D., Ansari-Lari, M. A., Aradhya, S., Ashwell, R. I. S., Babbage, A. K., Bagguley, C. L., Ballabio, A., Banerjee, R., Barker, G. E., Barlow, K. F., Barrett, I. P., Bates, K. N., Beare, D. M., Beasley, H., Beasley, O., Beck, A., Bethel, G., Blechschmidt, K., Brady, N., Bray-Allen, S., Bridgeman, A. M., Brown, A. J., Brown, M. J., Bonnin, D., Bruford, E. A., Buhay, C., Burch, P., Burford, D., Burgess, J., Burrill, W., Burton, J., Bye, J. M., Carder, C., Carrel, L., Chako, J., Chapman, J. C., Chavez, D., Chen, E., Chen, G., Chen, Y., Chen, Z., Chinault, C., Ciccodicola, A., Clark, S. Y., Clarke, G., Clee, C. M., Clegg, S., Clerc-Blankenburg, K., Clifford, K., Cobley, V., Cole, C. G., Conquer, J. S., Corby, N., Connor, R. E., David, R., Davies, J., Davis, C., Davis, J., Delgado, O., DeShazo, D., Dhami, P., Ding, Y., Dinh, H., Dodsworth, S., Draper, H., Dugan-Rocha, S., Dunham, A., Dunn, M., Durbin, K. J., Dutta, I., Eades, T., Ellwood, M., Emery-Cohen, A., Errington, H., Evans, K. L., Faulkner, L., Francis, F., Frankland, J., Fraser, A. E., Galgoczy, P., Gilbert, J., Gill, R., Glöckner, G., Gregory, S. G., Gribble, S., Griffiths, C., Grocock, R., Gu, Y., Gwilliam, R., Hamilton, C., Hart, E. A., Hawes, A., Heath, P. D., Heitmann, K., Hennig, S., Hernandez, J., Hinzmann, B., Ho, S., Hoffs, M., Howden, P. J., Huckle, E. J., Hume, J., Hunt, P. J., Hunt, A. R., Isherwood, J., Jacob, L., Johnson, D., Jones, S., de Jong, P. J., Joseph, S. S., Keenan, S., Kelly, S., Kershaw, J. K., Khan, Z., Kioschis, P., Klages, S., Knights, A. J., Kosiura, A., Kovar-Smith, C., Laird, G. K., Langford, C., Lawlor, S., Leversha, M., Lewis, L., Liu, W., Lloyd, C., Lloyd, D. M., Loulseged, H., Loveland, J. E., Lovell, J. D., Lozado, R., Lu, J., Lyne, R., Ma, J., Maheshwari, M., Matthews, L. H., McDowall, J., McLaren, S., McMurray, A., Meidl, P., Meitinger, T., Milne, S., Miner, G., Mistry, S. L., Morgan, M., Morris, S., Müller, I., Mullikin, J. C., Nguyen, N., Nordsiek, G., Nyakatura, G., O'Dell, C. N., Okwuonu, G., Palmer, S., Pandian, R., Parker, D., Parrish, J., Pasternak, S., Patel, D., Pearce, A. V., Pearson, D. M., Pelan, S. E., Perez, L., Porter, K. M., Ramsey, Y., Reichwald, K., Rhodes, S., Ridler, K. A., Schlessinger, D., Schueler, M. G., Sehra, H. K., Shaw-Smith, C., Shen, H., Sheridan, E. M., Shownkeen, R., Skuce, C. D., Smith, M. L., Sotheran, E. C., Steingruber, H. E., Steward, C. A., Storey, R., Swann, R. M., Swarbreck, D., Tabor, P. E., Taudien, S., Taylor, T., Teague, B., Thomas, K., Thorpe, A., Timms, K., Tracey, A., Trevanion, S., Tromans, A. C., D'Urso, M., Verduzco, D., Villasana, D., Waldron, L., Wall, M., Wang, Q., Warren, J., Warry, G. L., Wei, X., West, A., Whitehead, S. L., Whiteley, M. N., Wilkinson, J. E., Willey, D. L., Williams, G., Williams, L., Williamson, A., Williamson, H., Wilming, L., Woodmansey, R. L., Wray, P. W., Yen, J., Zhang, J., Zhou, J., Zoghbi, H., Zorilla, S., Buck, D., Reinhardt, R., Poustka, A., Rosenthal, A., Lehrach, H., Meindl, A., Minx, P. J., Hillier, L. W., Willard, H. F., Wilson, R. K., Waterston, R. H., Rice, C. M., Vaudin, M., Coulson, A., Nelson, D. L., Weinstock, G., Sulston, J. E., Durbin, R., Hubbard, T., Gibbs, R. A., Beck, S., Rogers, J. & Bentley, D. R. (2005), 'The DNA sequence of the human X chromosome', *Nature* **434**(7031), 325–337.
URL: *http://www.nature.com/articles/nature03440*

RTS, S. C. T. P. (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* **386**(9988), 31–45.
URL: *http://dx.doi.org/10.1016/S0140-6736(15)60721-8*

Ruiz, M. F., Milano, A., Salvemini, M., Eirín-López, J. M., Perondini, A. L., Selivon, D., Polito, C., Saccone, G. & Sánchez, L. (2007), 'The gene *transformer* of *Anastrepha* fruit flies (Diptera, Tephritidae) and its evolution in insects', *PLoS One* **2**(11).
URL: *https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0001239*

Ryner, L. C., Goodwin, S. F., Castrillon, D. H., Anand, A., Villella, A., Baker, B. S., Hall, J. C., Taylor, B. J. & Wasserman, S. A. (1996), *Cell* **87**(6), 1079–1089.
URL: *https://www.ncbi.nlm.nih.gov/pubmed/8978612*

Saccone, G., Peluso, I., Artiaco, D., Giordano, E., Bopp, D. & Polito, L. C. (1998), 'The *Ceratitis capitata* homologue of the *Drosophila* sex-determining gene *Sex-lethal* is structurally conserved, but not sex-specifically regulated', *Development* **125**(8), 1495–1500.
URL: *https://www.zora.uzh.ch/id/eprint/489/*

Sahara, K., Yoshido, A. & Traut, W. (2012), 'Sex chromosome evolution in moths and butterflies', *Chromosom. Res.* **20**(1), 83–94.
URL: *https://link.springer.com/article/10.1007/s10577-011-9262-z*

Salvemini, M., D'Amato, R., Petrella, V., Aceto, S., Nimmo, D., Neira, M., Alphey, L., Polito, L. C. & Saccone, G. (2013), 'The orthologue of the fruitfly sex behaviour gene *fruitless* in the mosquito *Aedes aegypti*: Evolution of genomic organisation and alternative splicing', *PLoS One* **8**(2), e48554.
URL: *https://www.ncbi.nlm.nih.gov/pubmed/23418412*

Salvemini, M., Mauro, U., Lombardo, F., Milano, A., Zazzaro, V., Arca, B., Polito, L. C. & Saccone, G. (2011), 'Genomic organization and splicing evolution of the *doublesex* gene, a *Drosophila* regulator of sexual differentiation, in the dengue and yellow fever mosquito *Aedes aegypti*', *Bmc Evol. Biol.* **11**(41).
URL: *https://bmcevolbiol.biomedcentral.com/articles/10.1186/1471-2148-11-41*

Salz, H. K., Maine, E. M., Keyes, L. N., Samuels, M. E., Cline, T. W. & Schedl, P. (1989), 'The *Drosophila* female-specific sex-determination gene, *Sex-lethal*, has stage-, tissue-, and sex-specific RNAs suggesting multiple modes of regulation', *Genes Dev.* **3**(5), 708–719.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/2473007*

Samuels, M. E., Schedl, P. & Cline, T. W. (1991), 'The complex set of late transcripts from the *Drosophila* sex determination gene *sex-lethal* encodes multiple related polypeptides', *Mol. Cell. Biol.* **11**(7), 3584–3602.
URL: *https://www.ncbi.nlm.nih.gov/pubmed/1710769*

Sánchez, L. (2008), 'Sex-determining mechanisms in insects', *Int. J. Dev. Biol.* **52**(7), 837–856.
URL: *https://www.ncbi.nlm.nih.gov/pubmed/18956315*

Scali, C., Catteruccia, F., Li, Q. & Crisanti, A. (2005), 'Identification of sex-specific transcripts of the *Anopheles gambiae doublesex* gene', *J. Exp. Biol.* **208**(Pt 19), 3701–3709.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/16169947*

Scott, T. W. (2005), 'Containment of arthropod disease vectors', *ILAR J.* **46**(1), 53–61.
URL: *https://academic.oup.com/ilarjournal/article-lookup/doi/10.1093/ilar.46.1.53*

Sefton, L., Timmer, J. R., Zhang, Y., Beranger, F. & Cline, T. W. (2000), 'An extracellular activator of the *Drosophila* JAK/STAT pathway is a sex-determination signal element', *Nature* **405**(6789), 970–973.
URL: *https://pubmed.ncbi.nlm.nih.gov/10879541/*

Serebrovsky, A. S. (1940), 'On the possibility of a new method for the control of insect pests', *Zool. Zhurnal* **19**, 618–630.
URL: *https://www.cabdirect.org/cabdirect/abstract/19701000720*

Shao, C., Li, Q., Chen, S., Zhang, P., Lian, J., Hu, Q., Sun, B., Jin, L., Liu, S., Wang, Z., Zhao, H., Jin, Z., Liang, Z., Li, Y., Zheng, Q., Zhang, Y., Wang, J. & Zhang, G. (2014), 'Epigenetic modification and inheritance in

sexual reversal of fish', *Genome Res.* **24**(4), 604–615.
URL: *https://genome.cshlp.org/content/24/4/604.abstract*

Shaw, W. R., Marcenac, P., Childs, L. M., Buckee, C. O., Baldini, F., Sawadogo, S. P., Dabiré, R. K., Diabaté, A. & Catteruccia, F. (2016), '*Wolbachia* infections in natural *Anopheles* populations affect egg laying and negatively correlate with *Plasmodium* development', *Nat. Commun.* **7**(11772).
URL: *https://www.nature.com/articles/ncomms11772*

Shelley, H. & Aziz, M. (1949), '*Anopheles* eradication in Cyprus', *Br. Med. J.* **1**(4608), 767–768.
URL: *https://www.bmj.com/content/1/4608/767/related*

Shen, M. M. & Hodgkin, J. (1988), '*mab-3*, a gene required for sex-specific *yolk* protein expression and a male-specific lineage in *C. elegans*', *Cell* **54**(7), 1019–1031.
URL: *https://www.ncbi.nlm.nih.gov/pubmed/3046751*

Shirangi, T. R., Dufour, H. D., Williams, T. M. & Carroll, S. B. (2009), 'Rapid evolution of sex pheromone-producing enzyme expression in *Drosophila*', *PLoS Biol* **7**(8), e1000168.
URL: *https://www.ncbi.nlm.nih.gov/pubmed/19652700*

Shukla, J. N. & Palli, S. R. (2012*a*), '*Doublesex* target genes in the red flour beetle, *Tribolium castaneum*', *Sci. Rep.* **2**, 948.
URL: *https://www.ncbi.nlm.nih.gov/pubmed/23230513*

Shukla, J. N. & Palli, S. R. (2012*b*), 'Sex determination in beetles: Production of all male progeny by parental RNAi knockdown of *transformer*', *Sci. Rep.* **2**(602).
URL: *https://www.nature.com/articles/srep00602*

Sievert, V., Kuhn, S., Paululat, A. & Traut, W. (2000), 'Sequence conservation and expression of the *Sex-lethal* homologue in the fly *Megaselia scalaris*', *Genome* **43**(2), 382–390.
URL: *https://www.nrcresearchpress.com/doi/abs/10.1139/g99-132#.XynF-S2ZM1I*

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 Res.* **42**(11), 7461–7472.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/24803674*

Sinka, M. E., Bangs, M. J., Manguin, S., Coetzee, M., Mbogo, C. M., Hemingway, J., Patil, A. P., Temperley, W. H., Gething, P. W., Kabaria, C. W., Okara, R. M., Van Boeckel, T., Godfray, H. C. J., Harbach, R. E. & Hay, S. I. (2010), 'The dominant *Anopheles* vectors of human malaria in Africa, Europe and the Middle East: occurrence data, distribution maps and bionomic précis', *Parasit. Vectors* **3**(1), 117.
URL: *http://link.springer.com/10.1007/978-4-431-54883-6{_}12*

Sinkins, S. P. (2013), '*Wolbachia* and arbovirus inhibition in mosquitoes', *Future Microbiol.* **8**(10), 1249–1256.
URL: *https://www.futuremedicine.com/doi/10.2217/fmb.13.95*

Smidler, A. L., Terenzi, O., Soichot, J., Levashina, E. A. & Marois, E. (2013), 'Targeted mutagenesis in the malaria mosquito using TALE nucleases', *PLoS One* **8**(8), e74511.
URL: *https://dx.plos.org/10.1371/journal.pone.0074511*

Smith, R. C., Walter, M. F., Hice, R. H., O'Brochta, D. A. & Atkinson, P. W. (2007), 'Testis-specific expression of the β2 tubulin promoter of *Aedes aegypti* and its application as a genetic sex-separation marker', *Insect Mol. Biol.* **16**(1), 61–71.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2583.2006.00701.x*

Steinemann, S. & Steinemann, M. (2005), 'Y chromosomes: Born to be destroyed', *BioEssays* **27**(10), 1076–1083.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1002/bies.20288*

Stevens, N. M. (1911), 'Further studies on heterochromosomes in mosquitoes', *Biol. Bull.* **20**(2), 109–120.
URL: *http://www.jstor.org/stable/1536039 https://www.journals.uchicago.edu/doi/10.2307/1536039*

Stoute, J. A., Slaoui, M., Heppner, D. G., Momin, P., Kester, K. E., Desmons, P., Wellde, B. T., Garçon, N., Krzych, U., Marchand, M., Ballou, W. R. & Cohen, J. D. (1997), 'A preliminary evaluation of a recombinant circumsporozoite protein vaccine against *Plasmodium falciparum* malaria', *N. Engl. J. Med.* **336**(2), 86–91.
URL: *https://www.nejm.org/doi/full/10.1056/NEJM199701093360202*

Straimer, J., Gnädig, N. F., Witkowski, B., Amaratunga, C., Duru, V., Ramadani, A. P., Dacheux, M., Khim, N., Zhang, L., Lam, S., Gregory, P. D., Urnov, F. D., Mercereau-Puijalon, O., Benoit-Vical, F., Fairhurst, R. M., Ménard, D. & Fidock, D. A. (2015), 'K13-propeller mutations confer artemisinin resistance in *Plasmodium falciparum* clinical isolates', *Science.* **347**(6220), 428–431.
URL: *https://www.sciencemag.org/lookup/doi/10.1126/science.1260867*

Stuart, J. J. & Hatchett, J. H. (1991), 'Genetics of sex determination in the hessian fly *Mayetiola destructor*', *J. Hered.* **82**(1), 43–52.
URL: *https://academic.oup.com/jhered/article/82/1/43/971943*

Suzuki, M. G., Funaguma, S., Kanda, T., Tamura, T. & Shimada, T. (2003), 'Analysis of the biological functions of a *doublesex* homologue in *Bombyx mori*', *Dev. Genes Evol.* **213**(7), 345–354.
URL: *https://link.springer.com/article/10.1007%2Fs00427-003-0334-8*

Suzuki, M. G., Imanishi, S., Dohmae, N., Asanuma, M. & Matsumoto, S. (2010), 'Identification of a male-specific RNA binding protein that regulates sex-specific splicing of *Bmdsx* by Increasing RNA Binding activity of *BmPSI*', *Mol. Cell. Biol.* **30**(24), 5776–5786.
URL: *https://mcb.asm.org/content/30/24/5776*

Sweeny, T. & Barr, A. R. (1978), 'Sex ratio distortion caused by meiotic drive in a mosquito, *Culex pipiens L.*', *Genetics* **88**(3), 427–446.
URL: *https://pubmed.ncbi.nlm.nih.gov/17248804/*

Tacoli, C., Gai, P. P., Bayingana, C., Sifft, K., Geus, D., Ndoli, J., Sendegeya, A., Gahutu, J. B. & Mockenhaupt, F. P. (2016), 'Artemisinin resistance-associated K13 polymorphisms of *Plasmodium falciparum* in Southern Rwanda, 2010-2015', *Am. J. Trop. Med. Hyg.* **95**(5), 1090–1093.
URL: *http://www.ajtmh.org/content/journals/10.4269/ajtmh.16-0483*

Tazuke, S. I., Schulz, C., Gilboa, L., Fogarty, M., Mahowald, A. P., Guichet, A., Ephrussi, A., Wood, C. G., Lehmann, R. & Fuller, M. T. (2002), 'A germline-specific gap junction protein required for survival of differentiating early germ cells.', *Development* **129**(10), 2529–39.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/11973283*

Teem, J. L., Ambali, A., Glover, B., Ouedraogo, J., Makinde, D. & Roberts, A. (2019), 'Problem formulation for gene drive mosquitoes designed to reduce malaria transmission in Africa: Results from four regional consultations 2016-2018', *Malar. J.* **18**(1).
URL: *https://doi.org/10.1186/s12936-019-2978-5*

Tennessen, J. A., Ingham, V. A., Hyacinthe Toé, K., Moussa Guelbéogo, W., Sagnon, N., Kuzma, R., Ranson, H. & Neafsey, D. E. (2020), 'A population genomic unveiling of a new cryptic mosquito taxon within the malaria-transmitting *Anopheles gambiae* complex', *bioaRxiv* .
URL: *https://doi.org/10.1101/2020.05.26.116988*

Thomas, D. D., Donnelly, C. A., Wood, R. J. & Alphey, L. S. (2000), 'Insect population control using a dominant, repressible, lethal genetic system', *Science.* **287**(5462), 2474–2476.
URL: *https://science.sciencemag.org/content/287/5462/2474/tab-article-info*

Thyme, S. B., Boissel, S. J., Arshiya Quadri, S., Nolan, T., Baker, D. A., Park, R. U., Kusak, L., Ashworth, J. & Baker, D. (2014), 'Reprogramming homing endonuclease specificity through computational design and directed evolution', *Nucleic Acids Res.* **42**(4), 2564–2576.
URL: *https://academic.oup.com/nar/article/42/4/2564/2437382*

Turner, J. M. (2007), 'Meiotic sex chromosome inactivation', *Development* **134**(10), 1823–1831.
URL: *https://dev.biologists.org/content/134/10/1823*

Undeen, A. H. & Alger, N. E. (1975), ‘The effect of the microsporidan, *Nosema algerae*, on *Anopheles stephensi*’, *J. Invertebr. Pathol.* **25**(1), 19–24.
URL: *https://www.sciencedirect.com/science/article/abs/pii/0022201175902815*

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

Usui-Aoki, K., Ito, H., Ui-Tei, K., Takahashi, K., Lukacsovich, T., Awano, W., Nakata, H., Piao, Z. F., Nilsson, E. E., Tomida, J. & Yamamoto, D. (2000), ‘Formation of the male-specific muscle in female *Drosophila* by ectopic *fruitless* expression’, *Nat Cell Biol* **2**(8), 500–506.
URL: *https://www.ncbi.nlm.nih.gov/pubmed/10934470*

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’, *Nat. Med.* .
URL: *http://www.nature.com/articles/s41591-020-1005-2*

Valcarcel, J., Singh, R., Zamore, P. D. & Green, M. R. (1993), ‘The protein *Sex-lethal* antagonizes the splicing factor U2AF to regulate alternative splicing of *transformer* pre-mRNA’, *Nature* **362**(6416), 171–175.
URL: *https://www.ncbi.nlm.nih.gov/pubmed/7680770*

Valenzuela, N., Literman, R., Neuwald, J. L., Mizoguchi, B., Iverson, J. B., Riley, J. L. & Litzgus, J. D. (2019), ‘Extreme thermal fluctuations from climate change unexpectedly accelerate demographic collapse of vertebrates with temperature-dependent sex determination’, *Sci. Rep.* **9**(1), 1–11.
URL: *http://dx.doi.org/10.1038/s41598-019-40597-4*

Van den Hurk, A. F., Hall-Mendelin, S., Pyke, A. T., Frentiu, F. D., McElroy, K., Day, A., Higgs, S. & O’Neill, S. L. (2012), ‘Impact of *Wolbachia* on infection with chikungunya and yellow fever viruses in the mosquito vector *Aedes aegypti*’, *PLoS Negl. Trop. Dis.* **6**(11).
URL: *https://journals.plos.org/plosntds/article?id=10.1371/journal.pntd.0001892*

Venkatraman, N., Tiono, A., Bowyer, G., Powlson, J., Collins, K., Coulibaly, S., Datoo, M., Silman, D., Ouedraogo, A., Nebie, I., Imoukhuede, E., Brod, F., Folegatti, P., Dickinson, E., Jamieson, S., Bougouma, E., Wright, D., Bellamy, D., Diarra, A., Bliss, C., Morter, R., Glenn, G., Fries, L., Reimer, J., Lovgren-Bengtsson, K., Baker, M., Poulton, I., Moyle, S., Berrie, E., Green, N., Mukhopadhyay, E., Viebig, N., Angus, B., Lawrie, A., Roberts, R., Gilbert, S., Lewis, D., Sirima, S., Ewer, K. & Hill, A. (2019), ‘Phase I assessments of first-in-human administration of a novel malaria anti-sporozoite vaccine candidate, R21 in matrix-M adjuvant, in UK and Burkinabe volunteers’, *medRxiv* .
URL: *https://www.medrxiv.org/content/10.1101/19009282v1*

Verhulst, E. C., Beukeboom, L. W. & van de Zande, L. (2010), ‘Maternal control of haplodiploid sex determination in the wasp *Nasonia*’, *Science.* **328**(5978), 620–623.
URL: *https://www.sciencemag.org/lookup/doi/10.1126/science.1185805*

Verhulst, E. C. & van de Zande, L. (2015), ‘Double nexus - *Doublesex* is the connecting element in sex determination’, *Brief. Funct. Genomics* **14**(6), 396–406.
URL: *https://www.ncbi.nlm.nih.gov/pubmed/25797692*

Villarreal, S. M., Pitcher, S., Helinski, M. E., Johnson, L., Wolfner, M. F. & Harrington, L. C. (2018), ‘Male contributions during mating increase female survival in the disease vector mosquito *Aedes aegypti*’, *J. Insect Physiol.* **108**, 1–9.
URL: *https://doi.org/10.1016/j.jinsphys.2018.05.001*

Vizioli, J., Bulet, P., Hoffmann, J. A., Kafatos, F. C., Müller, H. M. & Dimopoulos, G. (2001), ‘Gambicin: A novel immune responsive antimicrobial peptide from the malaria vector *Anopheles gambiae*’, *Proc. Natl. Acad. Sci. U. S. A.* **98**(22), 12630–12635.

URL: *https://www.pnas.org/content/98/22/12630/tab-article-info*

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* **5**(6), 1151–1163.
URL: *https://www.ncbi.nlm.nih.gov/pubmed/25869647*

Vossbrinck, C. R. & Debrunner-Vossbrinck, B. A. (2005), 'Molecular phylogeny of the Microsporidia: Ecological, ultrastructural and taxonomic considerations', *Folia Parasitol.* **52**(1-2), 131–142.
URL: *https://folia.paru.cas.cz/artkey/fol-200501-0017_molecular_phylogeny_of_the_microsporidia_ecological_ultrastructural_and_taxonomic_considerations.php*

Vreysen, M. J. B., Saleh, K. M., Ali, M. Y., Abdulla, A. M., Zhu, Z.-R., Juma, K. G., Dyck, V. A., Msangi, A. R., Mkonyi, P. A. & Feldmann, H. U. (2000), '*Glossina austeni* (Diptera: Glossinidae) eradicated on the island of Unguja, Zanzibar, using the Sterile Insect Technique', *J. Econ. Entomol.* **93**(1), 123–135.
URL: *https://academic.oup.com/jee/article-lookup/doi/10.1603/0022-0493-93.1.123*

Walker, G. (2005), 'Sex determination in the larvae of the parasitic barnacle *Heterosaccus lunatus*: An experimental approach', *J. Exp. Mar. Bio. Ecol.* **318**(1), 31–38.
URL: *https://espace.library.uq.edu.au/view/UQ:76590*

Walker, T., Johnson, P. H., Moreira, L. A., Iturbe-Ormaetxe, I., Frentiu, F. D., McMeniman, C. J., Leong, Y. S., Dong, Y., Axford, J., Kriesner, P., Lloyd, A. L., Ritchie, S. A., O'Neill, S. L. & Hoffmann, A. A. (2011), 'The wMel *Wolbachia* strain blocks dengue and invades caged *Aedes aegypti* populations', *Nature* **476**(7361), 450–455.
URL: *https://www.nature.com/articles/nature10355*

Wang, J., Na, J. K., Yu, Q., Gschwend, A. R., Han, J., Zeng, F., Aryal, R., VanBuren, R., Murray, J. E., Zhang, W., Navajas-Pérez, R., Feltus, F. A., Lemke, C., Tong, E. J., Chen, C., Wai, C. M., Singh, R., Wang, M. L., Min, X. J., Alam, M., Charlesworth, D., Moore, P. H., Jiang, J., Paterson, A. H. & Ming, R. (2012), 'Sequencing papaya X and Y^h chromosomes reveals molecular basis of incipient sex chromosome evolution', *Proc. Natl. Acad. Sci. U. S. A.* **109**(34), 13710–13715.
URL: *https://www.pnas.org/content/early/2012/08/01/1207833109*

Wang, S., Dos-Santos, A. L. A., Huang, W., Liu, K. C., Oshaghi, M. A., Wei, G., Agre, P. & Jacobs-Lorena, M. (2017), 'Driving mosquito refractoriness to *Plasmodium falciparum* with engineered symbiotic bacteria', *Science.* **357**(6358), 1399–1402.
URL: *https://www.sciencemag.org/lookup/doi/10.1126/science.aan5478*

Warner, R. R., Fitch, D. L. & Standish, J. D. (1996), 'Social control of sex change in the shelf limpet, *Crepidula norrisiarum*: Size-specific responses to local group composition', *J. Exp. Mar. Bio. Ecol.* **204**(1-2), 155–167.
URL: *https://europepmc.org/article/pmc/pmc6046201*

Warrell, D. A. & Gilles, H. M., eds (2002), *Essential malariology*, 4th edn, CRC press, London.
URL: *https://doi.org/10.1016/S0035-9203(03)90118-0*

Werren, J. H., Baldo, L. & Clark, M. E. (2008), '*Wolbachia*: Master manipulators of invertebrate biology', *Nat. Rev. Microbiol.* **6**(10), 741–751.
URL: *https://www.nature.com/articles/nrmicro1969*

White, B. J., Collins, F. H. & Besansky, N. J. (2011), 'Evolution of *Anopheles gambiae* in relation to humans and malaria', *Annu. Rev. Ecol. Evol. Syst.* **42**(1), 111–132.
URL: *https://www.annualreviews.org/doi/abs/10.1146/annurev-ecolsys-102710-145028*

WHO (2015*a*), *Global technical strategy for malaria 2016-2030*, World Health Organization.
URL: *https://www.who.int/malaria/areas/global_technical_strategy/en/*

WHO (2015*b*), *Indoor residual spraying: An operational manual for indoor residual spraying (IRS) for malaria transmission control and elimination, 2nd ed.*, 2nd ed edn, World Health Organization, Geneva PP

- Geneva.
URL: *https://apps.who.int/iris/handle/10665/177242*

WHO (2019), *World malaria report 2019*, World Health Organization.
URL: *https://www.who.int/malaria/publications/world-malaria-report-2019/en/*

WHO (2020), '*Vector-borne diseases; Key facts*'. [Online; accessed 26-April-2020].
URL: *https://www.who.int/en/news-room/fact-sheets/detail/vector-borne-diseases*

Wiedenheft, B., Sternberg, S. H. & Doudna, J. A. (2012), 'RNA-guided genetic silencing systems in bacteria and archaea', *Nature* **482**(7385), 331–338.
URL: *https://www.nature.com/articles/nature10886*

Williams, T. M., Selegue, J. E., Werner, T., Gompel, N., Kopp, A. & Carroll, S. B. (2008), 'The regulation and evolution of a genetic switch controlling sexually dimorphic traits in *Drosophila*', *Cell* **134**(4), 610–623.
URL: *https://www.cell.com/cell/fulltext/S0092-8674(08)00840-4*

Windbichler, N., Menichelli, M., Papathanos, P. A., Thyme, S. B., Li, H., Ulge, U. Y., Hovde, B. T., Baker, D., Monnat Jr., R. J., Burt, A. & Crisanti, A. (2011), 'A synthetic homing endonuclease-based gene drive system in the human malaria mosquito', *Nature* **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 Res.* **35**(17), 5922–5933.
URL: *https://academic.oup.com/nar/article/35/17/5922/2402736*

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 Genet.* **4**(12).
URL: *https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1000291*

Wise de Valdez, M. R., Nimmo, D., Betz, J., Gong, H.-F., James, A. A., Alphey, L. & Black, W. C. (2011), 'Genetic elimination of dengue vector mosquitoes', *Proc. Natl. Acad. Sci.* **108**(12), 4772–4775.
URL: *http://www.pnas.org/cgi/doi/10.1073/pnas.1019295108*

Wyss, J. H. (2006), 'Screwworm eradication in the Americas', *Ann. N. Y. Acad. Sci.* **916**(1), 186–193.
URL: *http://doi.wiley.com/10.1111/j.1749-6632.2000.tb05289.x*

Yen, J. H. & Barr, A. R. (1971), 'New Hypothesis of the cause of Cytoplasmic Incompatibility in *Culex pipiens L.*', *Nature* **232**(5313), 657–658.
URL: *http://www.nature.com/articles/232657a0*

Yukich, J. O., Lengeler, C., Tediosi, F., Brown, N., Mulligan, J. A., Chavasse, D., Stevens, W., Justino, J., Conteh, L., Maharaj, R., Erskine, M., Mueller, D. H., Wiseman, V., Ghebremeskel, T., Zerom, M., Goodman, C., McGuire, D., Urrutia, J. M., Sakho, F., Hanson, K. & Sharp, B. (2008), 'Costs and consequences of large-scale vector control for malaria', *Malar. J.* **7**, 1–12.
URL: *https://malariajournal.biomedcentral.com/articles/10.1186/1475-2875-7-258*

Zhu, L. Y., Wilken, J., Phillips, N. B., Narendra, U., Chan, G., Stratton, S. M., Kent, S. B. & Weiss, M. A. (2000), 'Sexual dimorphism in diverse metazoans is regulated by a novel class of intertwined zinc fingers', *Genes Dev.* **14**(14), 1750–1764.
URL: *https://www.ncbi.nlm.nih.gov/pmc/articles/PMC316782/*

Appendix A

Population Experiment Data for *dsx* Gene Drive

Gen.	Larvae								Pupae		Eggs	
	RFP$^+$ (*n*)	RFP$^+$ (%)	RFP$^-$ (*n*)	RFP$^-$ (%)	Total (*n*)	Reduction (%)	Hatching Rate (%)	Mortality (%)	Emerged (*n*)	Reduction (%)	Total (*n*)	Reprod. Load (%)
0	150	25.00	450	75.00	600				600		27462	0.00
1	286	49.65	290	50.35	576	4.00	88.62	8.51	527	12.17	17405	36.62
2	302	62.01	185	37.99	487	18.83	74.92	13.35	422	29.67	14957	45.54
3	344	68.94	155	31.06	499	16.83	76.77	7.41	462	23.00	11249	59.04
4	316	67.67	151	32.33	467	22.17	71.85	23.13	359	40.17	9170	66.61
5	264	58.67	186	41.33	450	25.00	69.23	7.56	416	30.67	11364	58.62
6	288	63.30	167	36.70	455	24.17	70.00	25.05	341	43.17	7727	71.86
7	355	69.47	156	30.53	511	14.83	78.62	9.78	461	23.17	7785	71.65
8	323	70.07	138	29.93	461	23.17	70.92	14.32	395	34.17	5643	79.45
9	325	75.58	105	24.42	430	28.33	66.15	7.44	398	33.67	4107	85.04
10	357	95.71	16	4.29	373	37.83	57.38	9.38	338	43.67	4146	84.90
11	374	100.00	0	0.00	374	37.67	57.54	2.41	365	39.17	2645	90.37
12	253	100.00	0	0.00	253	57.83	38.92	7.11	235	60.83	0	100.00
13	0		0		0	100.00			0	100.00	0	

Table A.1 – Population experiment - Cage trial 1

Gen.	Larvae								Pupae		Eggs	
	RFP+ (*n*)	RFP+ (%)	RFP- (*n*)	RFP- (%)	Total (*n*)	Reduction (%)	Hatching Rate (%)	Mortality (%)	Emerged (*n*)	Reduction (%)	Total (*n*)	Reprod. Load (%)
0	150	25.00	450	75.00	600				600		26277	0.00
1	280	50.00	280	50.00	560	6.67	86.15	18.21	458	23.67	16578	36.91
2	325	61.79	201	38.21	526	12.33	80.92	11.98	463	22.83	15565	40.77
3	328	68.05	154	31.95	482	19.67	74.15	17.22	399	33.50	9376	64.32
4	398	85.41	68	14.59	466	22.33	71.69	14.38	399	33.50	6514	75.21
5	346	86.50	54	13.50	400	33.33	61.54	7.75	369	38.50	4805	81.71
6	309	90.09	34	9.91	343	42.83	52.77	9.33	311	48.17	4210	83.98
7	363	100.00	0	0.00	363	39.50	55.85	6.89	338	43.67	1668	93.65
8	278	100.00	0	0.00	278	53.67	42.77	20.14	222	63.00	0	100.00
9	0		0		0	100.00			0	100.00	0	

Table A.2 – Population experiment - Cage trial 2

Gen.	Larvae								Pupae		Eggs	
	RFP+ (*n*)	RFP+ (%)	RFP- (*n*)	RFP- (%)	Total (*n*)	Reduction (%)	Hatching Rate (%)	Mortality (%)	Emerged (*n*)	Reduction (%)	Total (*n*)	Reprod. Load (%)
0	150	25.00	450	75.00	600				600		18872	0.00
1	281	50.72	273	49.28	554	7.67	85.23	23.10	426	29.00	12575	33.37
2	332	65.74	173	34.26	505	15.83	77.69	12.87	440	26.67	12116	35.80
3	268	79.06	71	20.94	339	43.50	52.15	2.06	332	44.67	4381	76.79
4	348	84.26	65	15.74	413	31.17	63.54	10.41	370	38.33	6934	63.26
5	388	85.84	64	14.16	452	24.67	69.54	13.27	392	34.67	4209	77.70
6	351	94.86	19	5.14	370	38.33	56.92	4.86	352	41.33	6382	66.18
7	341	100.00	0	0.00	341	43.17	52.46	11.14	303	49.50	0	100.00
8	0		0		0	100.00			0	100.00	0	

Table A.3 – Population experiment - Cage trial 3

Gen.	Larvae								Pupae		Eggs	
	RFP+ (*n*)	RFP+ (%)	RFP- (*n*)	RFP- (%)	Total (*n*)	Reduction (%)	Hatching Rate (%)	Mortality (%)	Emerged (*n*)	Reduction (%)	Total (*n*)	Reprod. Load (%)
0	150	25.00	450	75.00	600				600		21247	0.00
1	289	46.92	327	53.08	616	0.00	94.77	15.91	518	13.67	15181	28.55
2	237	45.49	284	54.51	521	13.17	80.15	11.32	462	23.00	14557	31.49
3	278	57.92	202	42.08	480	20.00	73.85	7.92	442	26.33	13346	37.19
4	305	58.10	220	41.90	525	12.50	80.77	16.19	440	26.67	15038	29.22
5	283	56.83	215	43.17	498	17.00	76.62	21.69	390	35.00	9210	56.65
6	261	61.41	164	38.59	425	29.17	65.38	4.71	405	32.50	14380	32.32
7	378	68.60	173	31.40	551	8.17	84.77	15.43	466	22.33	7910	62.77
8	333	68.10	156	31.90	489	18.50	75.23	7.36	453	24.50	11428	46.21
9	256	59.26	176	40.74	432	28.00	66.46	10.88	385	35.83	13794	35.08
10	288	58.66	203	41.34	491	18.17	75.54	17.92	403	32.83	11058	47.96
11	249	58.73	175	41.27	424	29.33	65.23	6.37	397	33.83	8894	58.14
12	261	61.41	164	38.59	425	29.17	65.38	8.94	387	35.50	6726	68.34
13	314	79.90	79	20.10	393	34.50	60.46	9.41	356	40.67	8614	59.46
14	415	89.63	48	10.37	463	22.83	71.23	8.21	425	29.17	7550	64.47
15	418	89.70	48	10.30	466	22.33	71.69	17.60	384	36.00	3282	84.55
16	299	92.00	26	8.00	325	45.83	50.00	5.54	307	48.83	2943	86.15
17	434	100.00	0	0.00	434	27.67	66.77	17.97	356	40.67	1519	93.94
18	333	100.00	0	0.00	333	44.50	51.23	10.51	298	50.33	0	100.00
19	0		0		0	100.00			0	100.00	0	

Table A.4 – Population experiment - Cage trial 4m

Gen.	Larvae								Pupae		Eggs	
	RFP+ (*n*)	RFP+ (%)	RFP- (*n*)	RFP- (%)	Total (*n*)	Reduction (%)	Hatching Rate (%)	Mortality (%)	Emerged (*n*)	Reduction (%)	Total (*n*)	Reprod. Load (%)
7	392	68.77	178	31.23	570	5.00	87.69	23.16	438	27.00	9852	53.63
8	334	69.01	150	30.99	484	19.33	74.46	15.08	411	31.50	8272	61.07
9	429	89.38	51	10.63	480	20.00	73.85	15.42	406	32.33	7589	64.28
10	241	97.18	7	2.82	248	58.67	38.15	15.32	210	65.00	2304	89.16
11	239	97.55	6	2.45	245	59.17	37.69	11.43	217	63.83	75	99.65
12	37	100.00	0	0.00	37	93.83	5.69	13.51	32	94.67	0	100.00
13	0		0		0	100.00			0	100.00	0	

Table A.5 – Population experiment - Cage trial 4r

Appendix B

Licences and Permissions

Page	Name of work	Source of work	Copyright holder	Requested	Permission note
30	FIG. 2.3. Map of malaria case incidence rate (cases per 1000 population at risk) by country, 2018	World malaria report 2019. Geneva World Health Organization; 2019 Licence: CC BY-NC-SA 3.0 IGO	© World Health Organization 2019 www.who.int/malaria infogmp@who.int	09.05.20	Creative Commons Attribution NonCommercial-ShareAlike 3.0 IGO licence (CC BY-NC-SA 3.0 IGO)
32	Figure 3: Projected geographical distribution of *Plasmodium falciparum* malaria under the Accelerate 2 scenario between 2015 and 2030	Potential for reduction of burden and local elimination of malaria by reducing *Plasmodium falciparum* malaria transmission: a mathematical modelling study. The Lancet Infectious Diseases vol.16(4), pp.465-472 (2016)	© 2016, Elsevier Griffin et al., 2016 Open Access article distributed under the terms of CC BY. Published by Elsevier Ltd.	09.05.20	Creative Commons Attribution (CC BY 4.0)
36	File:*Culex* mosquito life cycle nol text.svg	Wikimedia Commons https://commons.wikimedia.org/wiki/File:Culex_mosquito_life_cycle_en.svg	Credit: Mariana Ruiz Villarreal LadyofHats MRV@LadyofHats.com	09.05.20	Creative Commons Public Domain
53	Figure 1. The CRISPR/Cas9 system	What is CRISPR/Cas9? Archives of Disease in Childhood - Education and Practice vol.101, pp.213-215 (2016)	© 2016, BMJ Publishing Group Ltd. Redman et al., 2016 Open Access article distributed under the terms of CC BY.	10.05.20	Creative Commons Attribution (CC BY 4.0)
58	Figure 3. Diversity of sex determination systems for representative plant and animal clades	Sex Determination: Why So Many Ways of Doing It? PLOS Biology vol.12(7) (2014)	© 2014, Bachtrog et al. Open Access article distributed under the terms of CC BY. Published by PLOS	14.05.20	Creative Commons Attribution (CC BY 4.0)
61 67	Figure 3: Distribution and sample of vertebrate data from the Tree of Sex Database. Figure 2: Distribution and sample of invertebrate data from the Tree of Sex Database.	Tree of Sex: A database of sexual systems Scientific Data vol.1:140015 (2014)	© 2015, Springer Nature The Tree of Sex Consortium Open Access article distributed under the terms of CC BY. Published by Springer Nature	14.05.20	Creative Commons Attribution (CC BY 4.0)

Table B.1 – Licences and permissions for materials and figures used in this report.

<table>
<tr><th>Page</th><th>Name of work</th><th>Source of work</th><th>Copyright holder</th><th>Requested</th><th>Permission note</th></tr>
<tr><td>98</td><td>Figure 1. Targeting the female-specific isoform of doublesex.</td><td rowspan="7">A CRISPR–Cas9 gene drive targeting doublesex causes complete population suppression in caged Anopheles gambiae mosquitoes.
Nature Biotechnology vol.13(10) pp.1062-1066 (2018)</td><td rowspan="7">© 2018, Springer Nature
Kyrou et al., 2018
Open Access article distributed under the terms of CC BY.
Published by Springer Nature</td><td rowspan="7">15.06.20</td><td rowspan="7">Creative Commons Attribution
(CC BY 4.0)</td></tr>
<tr><td>109</td><td>Figure 2. Morphological analysis of homozygous $dsxF^{-/-}$ mutants.</td></tr>
<tr><td>112</td><td>Figure 3. Reproductive phenotype of dsxF mutants.</td></tr>
<tr><td>132</td><td>Figure 5. Dynamics of the spread of the $dsxF^{CRISPRh}$ allele and effect on population reproductive capacity</td></tr>
<tr><td>111</td><td>Figure S2. Morphology of the $dsxF^{-}$ internal reproductive organs.</td></tr>
<tr><td>131</td><td>Figure S6. Probability of stochastic loss of the drive as a function of initial number of male drive heterozygotes.</td></tr>
<tr><td>141</td><td>Figure S9. In vitro cleavage assay testing the efficiency of the gRNA in the $dsxF^{CRISPRh}$ gene drive to cleave the dsx exon 5 target site with the SNP found in wild populations in Africa.</td></tr>
</table>

Table B.2 – Licences and permissions for materials and figures from authored publications.

Page	Name of work	Source of work	Copyright holder	Requested	Granted?	Permission note
39	Figure 2: Changing endemicity and effect of interventions 2000–2015	The effect of malaria control on *Plasmodium falciparum* in Africa between 2000 and 2015. Nature vol.526, pp.207-211 (2015)	© 2015, Springer Nature Bhatt et al., 2016	05.06.20	yes	Written permission
59	Figure 1. Possible scenario of animal Y chromsome evolution	Y chromosomes: borne to be destroyed. BioEssays vol.27(10) (2005)	© 2005, Wiley Periodicals, Inc Steinemann & Steinemann, 2005	16.07.20	yes	Written permission
65	Figure 1. Dated phylogenetic tree of insect relationships	Phylogenomics resolves the timing and pattern of insect evolution. Science vol.346(6210) (2014)	© 2016, American Association for the Advancement of Science Misof et al., 2014	30.05.20	yes	Written permission
23	Figure 1. Common sex determination systems in insects	Sex Determination, Sex Chromosomes, and Karyotype Evolution in Insects. Journal of Heredity vol.108(1) (2017)	© 2016, Oxford University Press Blackmon et al., 2017	05.06.20	yes	Written permission

Table B.3 – Written permissions for materials and figures used in this report.

SPRINGER NATURE LICENSE
TERMS AND CONDITIONS

Jun 05, 2020

This Agreement between Mr. Kyros Kyrou ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	4842591381206
License date	Jun 05, 2020
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Nature
Licensed Content Title	The effect of malaria control on Plasmodium falciparum in Africa between 2000 and 2015
Licensed Content Author	S. Bhatt et al
Licensed Content Date	Sep 16, 2015
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	print and electronic
Portion	figures/tables/illustrations

Number of figures/tables/illustrations	1
High-res required	no
Will you be translating?	no
Circulation/distribution	1 - 29
Author of this Springer Nature content	no
Title	Targeting sex determination for genetic control of the malaria mosquito
Institution name	Imperial College London
Expected presentation date	Aug 2020
Portions	Figure 2: Changing endemicity and effect of interventions 2000–2015.
Requestor Location	Mr. Kyros Kyrou Flat 6, Melina Court 7 Gipsy Lane London, SW15 5RG United Kingdom Attn: Mr. Kyros Kyrou
Total	0.00 GBP
Terms and Conditions	

Springer Nature Customer Service Centre GmbH
Terms and Conditions

This agreement sets out the terms and conditions of the licence (the **Licence**) between you and **Springer Nature Customer Service Centre GmbH** (the **Licensor**). By clicking

'accept' and completing the transaction for the material (**Licensed Material**), you also confirm your acceptance of these terms and conditions.

1. Grant of License

1. 1. The Licensor grants you a personal, non-exclusive, non-transferable, world-wide licence to reproduce the Licensed Material for the purpose specified in your order only. Licences are granted for the specific use requested in the order and for no other use, subject to the conditions below.

1. 2. The Licensor warrants that it has, to the best of its knowledge, the rights to license reuse of the Licensed Material. However, you should ensure that the material you are requesting is original to the Licensor and does not carry the copyright of another entity (as credited in the published version).

1. 3. If the credit line on any part of the material you have requested indicates that it was reprinted or adapted with permission from another source, then you should also seek permission from that source to reuse the material.

2. Scope of Licence

2. 1. You may only use the Licensed Content in the manner and to the extent permitted by these Ts&Cs and any applicable laws.

2. 2. A separate licence may be required for any additional use of the Licensed Material, e.g. where a licence has been purchased for print only use, separate permission must be obtained for electronic re-use. Similarly, a licence is only valid in the language selected and does not apply for editions in other languages unless additional translation rights have been granted separately in the licence. Any content owned by third parties are expressly excluded from the licence.

2. 3. Similarly, rights for additional components such as custom editions and derivatives require additional permission and may be subject to an additional fee. Please apply to Journalpermissions@springernature.com/bookpermissions@springernature.com for these rights.

2. 4. Where permission has been granted **free of charge** for material in print, permission may also be granted for any electronic version of that work, provided that the material is incidental to your work as a whole and that the electronic version is essentially equivalent to, or substitutes for, the print version.

2. 5. An alternative scope of licence may apply to signatories of the STM Permissions Guidelines, as amended from time to time.

3. Duration of Licence

3. 1. A licence for is valid from the date of purchase ('Licence Date') at the end of the relevant period in the below table:

Scope of Licence	Duration of Licence
Post on a website	12 months
Presentations	12 months
Books and journals	Lifetime of the edition in the language purchased

4. Acknowledgement

4. 1. The Licensor's permission must be acknowledged next to the Licenced Material in print. In electronic form, this acknowledgement must be visible at the same time as the figures/tables/illustrations or abstract, and must be hyperlinked to the journal/book's homepage. Our required acknowledgement format is in the Appendix below.

5. Restrictions on use

5. 1. Use of the Licensed Material may be permitted for incidental promotional use and minor editing privileges e.g. minor adaptations of single figures, changes of format, colour and/or style where the adaptation is credited as set out in Appendix 1 below. Any other changes including but not limited to, cropping, adapting, omitting material that affect the meaning, intention or moral rights of the author are strictly prohibited.

5. 2. You must not use any Licensed Material as part of any design or trademark.

5. 3. Licensed Material may be used in Open Access Publications (OAP) before publication by Springer Nature, but any Licensed Material must be removed from OAP sites prior to final publication.

6. Ownership of Rights

6. 1. Licensed Material remains the property of either Licensor or the relevant third party and any rights not explicitly granted herein are expressly reserved.

7. Warranty

IN NO EVENT SHALL LICENSOR BE LIABLE TO YOU OR ANY OTHER PARTY OR ANY OTHER PERSON OR FOR ANY SPECIAL, CONSEQUENTIAL, INCIDENTAL OR INDIRECT DAMAGES, HOWEVER CAUSED, ARISING OUT OF OR IN CONNECTION WITH THE DOWNLOADING, VIEWING OR USE OF THE MATERIALS REGARDLESS OF THE FORM OF ACTION, WHETHER FOR BREACH OF CONTRACT, BREACH OF WARRANTY, TORT, NEGLIGENCE, INFRINGEMENT OR OTHERWISE (INCLUDING, WITHOUT LIMITATION, DAMAGES BASED ON LOSS OF PROFITS, DATA, FILES, USE, BUSINESS OPPORTUNITY OR CLAIMS OF

THIRD PARTIES), AND
WHETHER OR NOT THE PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. THIS LIMITATION SHALL APPLY NOTWITHSTANDING ANY FAILURE OF ESSENTIAL PURPOSE OF ANY LIMITED REMEDY PROVIDED HEREIN.

8. Limitations

8. 1. ***BOOKS ONLY:***Where **'reuse in a dissertation/thesis'** has been selected the following terms apply: Print rights of the final author's accepted manuscript (for clarity, NOT the published version) for up to 100 copies, electronic rights for use only on a personal website or institutional repository as defined by the Sherpa guideline (www.sherpa.ac.uk/romeo/).

9. Termination and Cancellation

9. 1. Licences will expire after the period shown in Clause 3 (above).

9. 2. Licensee reserves the right to terminate the Licence in the event that payment is not received in full or if there has been a breach of this agreement by you.

Appendix 1 — Acknowledgements:

For Journal Content:
Reprinted by permission from [**the Licensor**]: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication)

For **Advance Online Publication papers:**
Reprinted by permission from [**the Licensor**]: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication), advance online publication, day month year (doi: 10.1038/sj.[JOURNAL ACRONYM].)

For Adaptations/Translations:
Adapted/Translated by permission from [**the Licensor**]: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication)

Note: For any republication from the British Journal of Cancer, the following credit line style applies:

Reprinted/adapted/translated by permission from [**the Licensor**]: on behalf of Cancer Research UK: : [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT]** (year of publication)

For **Advance Online Publication** papers:
Reprinted by permission from The [**the Licensor**]: on behalf of Cancer Research UK: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication), advance online publication, day month year (doi: 10.1038/sj.[JOURNAL ACRONYM])

For Book content:
Reprinted/adapted by permission from [**the Licensor**]: [**Book Publisher** (e.g. Palgrave Macmillan, Springer etc) [**Book Title**] by [**Book author**(s)] [**COPYRIGHT**] (year of publication)

Other Conditions:

Version 1.2

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.

JOHN WILEY AND SONS LICENSE
TERMS AND CONDITIONS

Jul 16, 2020

This Agreement between Mr. Kyros Kyrou ("You") and John Wiley and Sons ("John Wiley and Sons") consists of your license details and the terms and conditions provided by John Wiley and Sons and Copyright Clearance Center.

License Number	4870760726745
License date	Jul 16, 2020
Licensed Content Publisher	John Wiley and Sons
Licensed Content Publication	BioEssays
Licensed Content Title	Y chromosomes: born to be destroyed
Licensed Content Author	Manfred Steinemann, Sigrid Steinemann
Licensed Content Date	Sep 14, 2005
Licensed Content Volume	27
Licensed Content Issue	10
Licensed Content Pages	8
Type of use	Dissertation/Thesis

Requestor type	University/Academic
Format	Print and electronic
Portion	Figure/table
Number of figures/tables	1
Will you be translating?	No
Title	Targeting sex determination for genetic control of the malaria mosquito
Institution name	Imperial College London
Expected presentation date	Aug 2020
Portions	Figure 1. Possible scenario of animal Y chromosome evolution.
Requestor Location	Mr. Kyros Kyrou Flat 6, Melina Court 7 Gipsy Lane London, SW15 5RG United Kingdom Attn: Mr. Kyros Kyrou
Publisher Tax ID	EU826007151
Total	0.00 GBP

Terms and Conditions

TERMS AND CONDITIONS

This copyrighted material is owned by or exclusively licensed to John Wiley & Sons, Inc. or one of its group companies (each a"Wiley Company") or handled on behalf of a society with which a Wiley Company has exclusive publishing rights in relation to a particular work

(collectively "WILEY"). By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the billing and payment terms and conditions established by the Copyright Clearance Center Inc., ("CCC's Billing and Payment terms and conditions"), at the time that you opened your RightsLink account (these are available at any time at http://myaccount.copyright.com).

Terms and Conditions

- The materials you have requested permission to reproduce or reuse (the "Wiley Materials") are protected by copyright.

- You are hereby granted a personal, non-exclusive, non-sub licensable (on a stand-alone basis), non-transferable, worldwide, limited license to reproduce the Wiley Materials for the purpose specified in the licensing process. This license, **and any CONTENT (PDF or image file) purchased as part of your order,** is for a one-time use only and limited to any maximum distribution number specified in the license. The first instance of republication or reuse granted by this license must be completed within two years of the date of the grant of this license (although copies prepared before the end date may be distributed thereafter). The Wiley Materials shall not be used in any other manner or for any other purpose, beyond what is granted in the license. Permission is granted subject to an appropriate acknowledgement given to the author, title of the material/book/journal and the publisher. You shall also duplicate the copyright notice that appears in the Wiley publication in your use of the Wiley Material. Permission is also granted on the understanding that nowhere in the text is a previously published source acknowledged for all or part of this Wiley Material. Any third party content is expressly excluded from this permission.

- With respect to the Wiley Materials, all rights are reserved. Except as expressly granted by the terms of the license, no part of the Wiley Materials may be copied, modified, adapted (except for minor reformatting required by the new Publication), translated, reproduced, transferred or distributed, in any form or by any means, and no derivative works may be made based on the Wiley Materials without the prior permission of the respective copyright owner.**For STM Signatory Publishers clearing permission under the terms of the STM Permissions Guidelines only, the terms of the license are extended to include subsequent editions and for editions in other languages, provided such editions are for the work as a whole in situ and does not involve the separate exploitation of the permitted figures or extracts,** You may not alter, remove or suppress in any manner any copyright, trademark or other notices displayed by the Wiley Materials. You may not license, rent, sell, loan, lease, pledge, offer as security, transfer or assign the Wiley Materials on a stand-alone basis, or any of the rights granted to you hereunder to any other person.

- The Wiley Materials and all of the intellectual property rights therein shall at all times remain the exclusive property of John Wiley & Sons Inc, the Wiley Companies, or their respective licensors, and your interest therein is only that of having possession of and the right to reproduce the Wiley Materials pursuant to Section 2 herein during the continuance of this Agreement. You agree that you own no right, title or interest in or to the Wiley Materials or any of the intellectual property rights therein. You shall have no rights hereunder other than the license as provided for above in Section 2. No right,

license or interest to any trademark, trade name, service mark or other branding ("Marks") of WILEY or its licensors is granted hereunder, and you agree that you shall not assert any such right, license or interest with respect thereto

- NEITHER WILEY NOR ITS LICENSORS MAKES ANY WARRANTY OR REPRESENTATION OF ANY KIND TO YOU OR ANY THIRD PARTY, EXPRESS, IMPLIED OR STATUTORY, WITH RESPECT TO THE MATERIALS OR THE ACCURACY OF ANY INFORMATION CONTAINED IN THE MATERIALS, INCLUDING, WITHOUT LIMITATION, ANY IMPLIED WARRANTY OF MERCHANTABILITY, ACCURACY, SATISFACTORY QUALITY, FITNESS FOR A PARTICULAR PURPOSE, USABILITY, INTEGRATION OR NON-INFRINGEMENT AND ALL SUCH WARRANTIES ARE HEREBY EXCLUDED BY WILEY AND ITS LICENSORS AND WAIVED BY YOU.

- WILEY shall have the right to terminate this Agreement immediately upon breach of this Agreement by you.

- You shall indemnify, defend and hold harmless WILEY, its Licensors and their respective directors, officers, agents and employees, from and against any actual or threatened claims, demands, causes of action or proceedings arising from any breach of this Agreement by you.

- IN NO EVENT SHALL WILEY OR ITS LICENSORS BE LIABLE TO YOU OR ANY OTHER PARTY OR ANY OTHER PERSON OR ENTITY FOR ANY SPECIAL, CONSEQUENTIAL, INCIDENTAL, INDIRECT, EXEMPLARY OR PUNITIVE DAMAGES, HOWEVER CAUSED, ARISING OUT OF OR IN CONNECTION WITH THE DOWNLOADING, PROVISIONING, VIEWING OR USE OF THE MATERIALS REGARDLESS OF THE FORM OF ACTION, WHETHER FOR BREACH OF CONTRACT, BREACH OF WARRANTY, TORT, NEGLIGENCE, INFRINGEMENT OR OTHERWISE (INCLUDING, WITHOUT LIMITATION, DAMAGES BASED ON LOSS OF PROFITS, DATA, FILES, USE, BUSINESS OPPORTUNITY OR CLAIMS OF THIRD PARTIES), AND WHETHER OR NOT THE PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. THIS LIMITATION SHALL APPLY NOTWITHSTANDING ANY FAILURE OF ESSENTIAL PURPOSE OF ANY LIMITED REMEDY PROVIDED HEREIN.

- Should any provision of this Agreement be held by a court of competent jurisdiction to be illegal, invalid, or unenforceable, that provision shall be deemed amended to achieve as nearly as possible the same economic effect as the original provision, and the legality, validity and enforceability of the remaining provisions of this Agreement shall not be affected or impaired thereby.

- The failure of either party to enforce any term or condition of this Agreement shall not constitute a waiver of either party's right to enforce each and every term and condition of this Agreement. No breach under this agreement shall be deemed waived or excused by either party unless such waiver or consent is in writing signed by the party granting such waiver or consent. The waiver by or consent of a party to a breach of any provision of this Agreement shall not operate or be construed as a waiver of or

consent to any other or subsequent breach by such other party.

- This Agreement may not be assigned (including by operation of law or otherwise) by you without WILEY's prior written consent.

- Any fee required for this permission shall be non-refundable after thirty (30) days from receipt by the CCC.

- These terms and conditions together with CCC's Billing and Payment terms and conditions (which are incorporated herein) form the entire agreement between you and WILEY concerning this licensing transaction and (in the absence of fraud) supersedes all prior agreements and representations of the parties, oral or written. This Agreement may not be amended except in writing signed by both parties. This Agreement shall be binding upon and inure to the benefit of the parties' successors, legal representatives, and authorized assigns.

- In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall prevail.

- WILEY expressly reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.

- This Agreement will be void if the Type of Use, Format, Circulation, or Requestor Type was misrepresented during the licensing process.

- This Agreement shall be governed by and construed in accordance with the laws of the State of New York, USA, without regards to such state's conflict of law rules. Any legal action, suit or proceeding arising out of or relating to these Terms and Conditions or the breach thereof shall be instituted in a court of competent jurisdiction in New York County in the State of New York in the United States of America and each party hereby consents and submits to the personal jurisdiction of such court, waives any objection to venue in such court and consents to service of process by registered or certified mail, return receipt requested, at the last known address of such party.

WILEY OPEN ACCESS TERMS AND CONDITIONS

Wiley Publishes Open Access Articles in fully Open Access Journals and in Subscription journals offering Online Open. Although most of the fully Open Access journals publish open access articles under the terms of the Creative Commons Attribution (CC BY) License only, the subscription journals and a few of the Open Access Journals offer a choice of Creative Commons Licenses. The license type is clearly identified on the article.

The Creative Commons Attribution License

The Creative Commons Attribution License (CC-BY) allows users to copy, distribute and transmit an article, adapt the article and make commercial use of the article. The CC-BY license permits commercial and non-

Creative Commons Attribution Non-Commercial License

The Creative Commons Attribution Non-Commercial (CC-BY-NC)License permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.(see below)

Creative Commons Attribution-Non-Commercial-NoDerivs License

The Creative Commons Attribution Non-Commercial-NoDerivs License (CC-BY-NC-ND) permits use, distribution and reproduction in any medium, provided the original work is properly cited, is not used for commercial purposes and no modifications or adaptations are made. (see below)

Use by commercial "for-profit" organizations

Use of Wiley Open Access articles for commercial, promotional, or marketing purposes requires further explicit permission from Wiley and will be subject to a fee.

Further details can be found on Wiley Online Library
http://olabout.wiley.com/WileyCDA/Section/id-410895.html

Other Terms and Conditions:

v1.10 Last updated September 2015

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.

THE AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE LICENSE TERMS AND CONDITIONS

May 30, 2020

This Agreement between Mr. Kyros Kyrou ("You") and The American Association for the Advancement of Science ("The American Association for the Advancement of Science") consists of your license details and the terms and conditions provided by The American Association for the Advancement of Science and Copyright Clearance Center.

License Number	4838890764899
License date	May 30, 2020
Licensed Content Publisher	The American Association for the Advancement of Science
Licensed Content Publication	Science
Licensed Content Title	Phylogenomics resolves the timing and pattern of insect evolution
Licensed Content Author	Bernhard Misof,Shanlin Liu,Karen Meusemann,Ralph S. Peters,Alexander Donath,Christoph Mayer,Paul B. Frandsen,Jessica Ware,Tomáš Flouri,Rolf G. Beutel,Oliver Niehuis,Malte Petersen,Fernando Izquierdo-Carrasco,Torsten Wappler,Jes Rust,Andre J. Aberer,Ulrike Aspöck,Horst Aspöck,Daniela Bartel,Alexander Blanke,Simon Berger,Alexander Böhm,Thomas R. Buckley,Brett Calcott,Junqing Chen,Frank Friedrich,Makiko Fukui,Mari Fujita,Carola Greve,Peter Grobe,Shengchang Gu,Ying Huang,Lars S. Jermiin,Akito Y. et al.
Licensed Content Date	Nov 7, 2014

Licensed Content Volume	346
Licensed Content Issue	6210
Volume number	346
Issue number	6210
Type of Use	Thesis / Dissertation
Requestor type	Scientist/individual at a research institution
Format	Print and electronic
Portion	Figure
Number of figures/tables	1
Title	Targeting sex determination for genetic control of the malaria mosquito
Institution name	Imperial College London
Expected presentation date	Aug 2020
Portions	Fig. 1 Dated phylogenetic tree of insect relationships.

Requestor Location
Mr. Kyros Kyrou
Flat 6, Melina Court
7 Gipsy Lane

London, SW15 5RG
United Kingdom
Attn: Mr. Kyros Kyrou

Total 0.00 GBP

Terms and Conditions

American Association for the Advancement of Science TERMS AND CONDITIONS

Regarding your request, we are pleased to grant you non-exclusive, non-transferable permission, to republish the AAAS material identified above in your work identified above, subject to the terms and conditions herein. We must be contacted for permission for any uses other than those specifically identified in your request above.

The following credit line must be printed along with the AAAS material: "From [Full Reference Citation]. Reprinted with permission from AAAS."

All required credit lines and notices must be visible any time a user accesses any part of the AAAS material and must appear on any printed copies and authorized user might make.

This permission does not apply to figures / photos / artwork or any other content or materials included in your work that are credited to non-AAAS sources. If the requested material is sourced to or references non-AAAS sources, you must obtain authorization from that source as well before using that material. You agree to hold harmless and indemnify AAAS against any claims arising from your use of any content in your work that is credited to non-AAAS sources.

If the AAAS material covered by this permission was published in Science during the years 1974 - 1994, you must also obtain permission from the author, who may grant or withhold permission, and who may or may not charge a fee if permission is granted. See original article for author's address. This condition does not apply to news articles.

The AAAS material may not be modified or altered except that figures and tables may be modified with permission from the author. Author permission for any such changes must be secured prior to your use.

Whenever possible, we ask that electronic uses of the AAAS material permitted herein include a hyperlink to the original work on AAAS's website (hyperlink may be embedded in the reference citation).

AAAS material reproduced in your work identified herein must not account for more than 30% of the total contents of that work.

AAAS must publish the full paper prior to use of any text.

AAAS material must not imply any endorsement by the American Association for the Advancement of Science.

This permission is not valid for the use of the AAAS and/or Science logos.

AAAS makes no representations or warranties as to the accuracy of any information contained in the AAAS material covered by this permission, including any warranties of

merchantability or fitness for a particular purpose.

If permission fees for this use are waived, please note that AAAS reserves the right to charge for reproduction of this material in the future.

Permission is not valid unless payment is received within sixty (60) days of the issuance of this permission. If payment is not received within this time period then all rights granted herein shall be revoked and this permission will be considered null and void.

In the event of breach of any of the terms and conditions herein or any of CCC's Billing and Payment terms and conditions, all rights granted herein shall be revoked and this permission will be considered null and void.

AAAS reserves the right to terminate this permission and all rights granted herein at its discretion, for any purpose, at any time. In the event that AAAS elects to terminate this permission, you will have no further right to publish, publicly perform, publicly display, distribute or otherwise use any matter in which the AAAS content had been included, and all fees paid hereunder shall be fully refunded to you. Notification of termination will be sent to the contact information as supplied by you during the request process and termination shall be immediate upon sending the notice. Neither AAAS nor CCC shall be liable for any costs, expenses, or damages you may incur as a result of the termination of this permission, beyond the refund noted above.

This Permission may not be amended except by written document signed by both parties.

The terms above are applicable to all permissions granted for the use of AAAS material. Below you will find additional conditions that apply to your particular type of use.

FOR A THESIS OR DISSERTATION
If you are using figure(s)/table(s), permission is granted for use in print and electronic versions of your dissertation or thesis. A full text article may be used in print versions only of a dissertation or thesis.

Permission covers the distribution of your dissertation or thesis on demand by ProQuest / UMI, provided the AAAS material covered by this permission remains in situ.

If you are an Original Author on the AAAS article being reproduced, please refer to your License to Publish for rules on reproducing your paper in a dissertation or thesis.

FOR JOURNALS:
Permission covers both print and electronic versions of your journal article, however the AAAS material may not be used in any manner other than within the context of your article.

FOR BOOKS/TEXTBOOKS:
If this license is to reuse figures/tables, then permission is granted for non-exclusive world rights in all languages in both print and electronic formats (electronic formats are defined below).

If this license is to reuse a text excerpt or a full text article, then permission is granted for non-exclusive world rights in English only. You have the option of securing either print or electronic rights or both, but electronic rights are not automatically granted and do garner additional fees. Permission for translations of text excerpts or full text articles into other languages must be obtained separately.

Licenses granted for use of AAAS material in electronic format books/textbooks are valid only in cases where the electronic version is equivalent to or substitutes for the print version of the book/textbook. The AAAS material reproduced as permitted herein must remain in situ and must not be exploited separately (for example, if permission covers the use of a full text article, the article may not be offered for access or for purchase as a stand-alone unit), except in the case of permitted textbook companions as noted below.

You must include the following notice in any electronic versions, either adjacent to the reprinted AAAS material or in the terms and conditions for use of your electronic products: "Readers may view, browse, and/or download material for temporary copying purposes only, provided these uses are for noncommercial personal purposes. Except as provided by law, this material may not be further reproduced, distributed, transmitted, modified, adapted, performed, displayed, published, or sold in whole or in part, without prior written permission from the publisher."

If your book is an academic textbook, permission covers the following companions to your textbook, provided such companions are distributed only in conjunction with your textbook at no additional cost to the user:

- Password-protected website
- Instructor's image CD/DVD and/or PowerPoint resource
- Student CD/DVD

All companions must contain instructions to users that the AAAS material may be used for non-commercial, classroom purposes only. Any other uses require the prior written permission from AAAS.

If your license is for the use of AAAS Figures/Tables, then the electronic rights granted herein permit use of the Licensed Material in any Custom Databases that you distribute the electronic versions of your textbook through, so long as the Licensed Material remains within the context of a chapter of the title identified in your request and cannot be downloaded by a user as an independent image file.

Rights also extend to copies/files of your Work (as described above) that you are required to provide for use by the visually and/or print disabled in compliance with state and federal laws.

This permission only covers a single edition of your work as identified in your request.

FOR NEWSLETTERS:

Permission covers print and/or electronic versions, provided the AAAS material reproduced as permitted herein remains in situ and is not exploited separately (for example, if permission covers the use of a full text article, the article may not be offered for access or for purchase as a stand-alone unit)

FOR ANNUAL REPORTS:
Permission covers print and electronic versions provided the AAAS material reproduced as permitted herein remains in situ and is not exploited separately (for example, if permission covers the use of a full text article, the article may not be offered for access or for purchase as a stand-alone unit)

FOR PROMOTIONAL/MARKETING USES:
Permission covers the use of AAAS material in promotional or marketing pieces such as information packets, media kits, product slide kits, brochures, or flyers limited to a single print run. The AAAS Material may not be used in any manner which implies endorsement or promotion by the American Association for the Advancement of Science (AAAS) or Science of any product or service. AAAS does not permit the reproduction of its name, logo or text on promotional literature.

If permission to use a full text article is permitted, The Science article covered by this permission must not be altered in any way. No additional printing may be set onto an article copy other than the copyright credit line required above. Any alterations must be approved in advance and in writing by AAAS. This includes, but is not limited to, the placement of sponsorship identifiers, trademarks, logos, rubber stamping or self-adhesive stickers onto the article copies.

Additionally, article copies must be a freestanding part of any information package (i.e. media kit) into which they are inserted. They may not be physically attached to anything, such as an advertising insert, or have anything attached to them, such as a sample product. Article copies must be easily removable from any kits or informational packages in which they are used. The only exception is that article copies may be inserted into three-ring binders.

FOR CORPORATE INTERNAL USE:
The AAAS material covered by this permission may not be altered in any way. No additional printing may be set onto an article copy other than the required credit line. Any alterations must be approved in advance and in writing by AAAS. This includes, but is not limited to the placement of sponsorship identifiers, trademarks, logos, rubber stamping or self-adhesive stickers onto article copies.

If you are making article copies, copies are restricted to the number indicated in your request and must be distributed only to internal employees for internal use.

If you are using AAAS Material in Presentation Slides, the required credit line must be visible on the slide where the AAAS material will be reprinted

If you are using AAAS Material on a CD, DVD, Flash Drive, or the World Wide Web, you must include the following notice in any electronic versions, either adjacent to the reprinted AAAS material or in the terms and conditions for use of your electronic products: "Readers may view, browse, and/or download material for temporary copying purposes only, provided these uses are for noncommercial personal purposes. Except as provided by law, this

material may not be further reproduced, distributed, transmitted, modified, adapted, performed, displayed, published, or sold in whole or in part, without prior written permission from the publisher." Access to any such CD, DVD, Flash Drive or Web page must be restricted to your organization's employees only.

FOR CME COURSE and SCIENTIFIC SOCIETY MEETINGS:
Permission is restricted to the particular Course, Seminar, Conference, or Meeting indicated in your request. If this license covers a text excerpt or a Full Text Article, access to the reprinted AAAS material must be restricted to attendees of your event only (if you have been granted electronic rights for use of a full text article on your website, your website must be password protected, or access restricted so that only attendees can access the content on your site).

If you are using AAAS Material on a CD, DVD, Flash Drive, or the World Wide Web, you must include the following notice in any electronic versions, either adjacent to the reprinted AAAS material or in the terms and conditions for use of your electronic products: "Readers may view, browse, and/or download material for temporary copying purposes only, provided these uses are for noncommercial personal purposes. Except as provided by law, this material may not be further reproduced, distributed, transmitted, modified, adapted, performed, displayed, published, or sold in whole or in part, without prior written permission from the publisher."

FOR POLICY REPORTS:
These rights are granted only to non-profit organizations and/or government agencies. Permission covers print and electronic versions of a report, provided the required credit line appears in both versions and provided the AAAS material reproduced as permitted herein remains in situ and is not exploited separately.

FOR CLASSROOM PHOTOCOPIES:
Permission covers distribution in print copy format only. Article copies must be freestanding and not part of a course pack. They may not be physically attached to anything or have anything attached to them.

FOR COURSEPACKS OR COURSE WEBSITES:
These rights cover use of the AAAS material in one class at one institution. Permission is valid only for a single semester after which the AAAS material must be removed from the Electronic Course website, unless new permission is obtained for an additional semester. If the material is to be distributed online, access must be restricted to students and instructors enrolled in that particular course by some means of password or access control.

FOR WEBSITES:
You must include the following notice in any electronic versions, either adjacent to the reprinted AAAS material or in the terms and conditions for use of your electronic products: "Readers may view, browse, and/or download material for temporary copying purposes only, provided these uses are for noncommercial personal purposes. Except as provided by law, this material may not be further reproduced, distributed, transmitted, modified, adapted, performed, displayed, published, or sold in whole or in part, without prior written permission from the publisher."

Permissions for the use of Full Text articles on third party websites are granted on a case by case basis and only in cases where access to the AAAS Material is restricted by some means

of password or access control. Alternately, an E-Print may be purchased through our reprints department (brocheleau@rockwaterinc.com).

REGARDING FULL TEXT ARTICLE USE ON THE WORLD WIDE WEB IF YOU ARE AN 'ORIGINAL AUTHOR' OF A SCIENCE PAPER

If you chose "Original Author" as the Requestor Type, you are warranting that you are one of authors listed on the License Agreement as a "Licensed content author" or that you are acting on that author's behalf to use the Licensed content in a new work that one of the authors listed on the License Agreement as a "Licensed content author" has written.

Original Authors may post the 'Accepted Version' of their full text article on their personal or on their University website and not on any other website. The 'Accepted Version' is the version of the paper accepted for publication by AAAS including changes resulting from peer review but prior to AAAS's copy editing and production (in other words not the AAAS published version).

FOR MOVIES / FILM / TELEVISION:
Permission is granted to use, record, film, photograph, and/or tape the AAAS material in connection with your program/film and in any medium your program/film may be shown or heard, including but not limited to broadcast and cable television, radio, print, world wide web, and videocassette.

The required credit line should run in the program/film's end credits.

FOR MUSEUM EXHIBITIONS:
Permission is granted to use the AAAS material as part of a single exhibition for the duration of that exhibit. Permission for use of the material in promotional materials for the exhibit must be cleared separately with AAAS (please contact us at permissions@aaas.org).

FOR TRANSLATIONS:
Translation rights apply only to the language identified in your request summary above.

The following disclaimer must appear with your translation, on the first page of the article, after the credit line: "This translation is not an official translation by AAAS staff, nor is it endorsed by AAAS as accurate. In crucial matters, please refer to the official English-language version originally published by AAAS."

FOR USE ON A COVER:
Permission is granted to use the AAAS material on the cover of a journal issue, newsletter issue, book, textbook, or annual report in print and electronic formats provided the AAAS material reproduced as permitted herein remains in situ and is not exploited separately

By using the AAAS Material identified in your request, you agree to abide by all the terms and conditions herein.

Questions about these terms can be directed to the AAAS Permissions department permissions@aaas.org.

Other Terms and Conditions:

v 2

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.

OXFORD UNIVERSITY PRESS LICENSE
TERMS AND CONDITIONS

Jun 05, 2020

This Agreement between Mr. Kyros Kyrou ("You") and Oxford University Press ("Oxford University Press") consists of your license details and the terms and conditions provided by Oxford University Press and Copyright Clearance Center.

License Number	4842590950015
License date	Jun 05, 2020
Licensed content publisher	Oxford University Press
Licensed content publication	The Journal of Heredity
Licensed content title	Sex Determination, Sex Chromosomes, and Karyotype Evolution in Insects
Licensed content author	Blackmon, Heath; Ross, Laura
Licensed content date	Aug 20, 2016
Type of Use	Thesis/Dissertation
Institution name	
Title of your work	Targeting sex determination for genetic control of the malaria mosquito
Publisher of your work	Imperial College London

Expected publication date	Aug 2020
Permissions cost	0.00 GBP
Value added tax	0.00 GBP
Total	0.00 GBP
Title	Targeting sex determination for genetic control of the malaria mosquito
Institution name	Imperial College London
Expected presentation date	Aug 2020
Portions	Figure 1. Common sex determination systems in insects.
Requestor Location	Mr. Kyros Kyrou Flat 6, Melina Court 7 Gipsy Lane London, SW15 5RG United Kingdom Attn: Mr. Kyros Kyrou
Publisher Tax ID	GB125506730
Total	0.00 GBP
Terms and Conditions	

STANDARD TERMS AND CONDITIONS FOR REPRODUCTION OF MATERIAL FROM AN OXFORD UNIVERSITY PRESS JOURNAL

1. Use of the material is restricted to the type of use specified in your order details.

2. This permission covers the use of the material in the English language in the following territory: world. If you have requested additional permission to translate this material, the terms and conditions of this reuse will be set out in clause 12.

3. This permission is limited to the particular use authorized in (1) above and does not allow you to sanction its use elsewhere in any other format other than specified above, nor does it apply to quotations, images, artistic works etc that have been reproduced from other sources which may be part of the material to be used.

4. No alteration, omission or addition is made to the material without our written consent. Permission must be re-cleared with Oxford University Press if/when you decide to reprint.

5. The following credit line appears wherever the material is used: author, title, journal, year, volume, issue number, pagination, by permission of Oxford University Press or the sponsoring society if the journal is a society journal. Where a journal is being published on behalf of a learned society, the details of that society must be included in the credit line.

6. For the reproduction of a full article from an Oxford University Press journal for whatever purpose, the corresponding author of the material concerned should be informed of the proposed use. Contact details for the corresponding authors of all Oxford University Press journal contact can be found alongside either the abstract or full text of the article concerned, accessible from www.oxfordjournals.org Should there be a problem clearing these rights, please contact journals.permissions@oup.com

7. If the credit line or acknowledgement in our publication indicates that any of the figures, images or photos was reproduced, drawn or modified from an earlier source it will be necessary for you to clear this permission with the original publisher as well. If this permission has not been obtained, please note that this material cannot be included in your publication/photocopies.

8. While you may exercise the rights licensed immediately upon issuance of the license at the end of the licensing process for the transaction, provided that you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (either by Oxford University Press or by Copyright Clearance Center (CCC)) as provided in CCC's Billing and Payment terms and conditions. If full payment is not received on a timely basis, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC's Billing and Payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and Oxford University Press reserves the right to take any and all action to protect its copyright in the materials.

9. This license is personal to you and may not be sublicensed, assigned or transferred by you to any other person without Oxford University Press's written permission.

10. Oxford University Press reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing

transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.

11. You hereby indemnify and agree to hold harmless Oxford University Press and CCC, and their respective officers, directors, employs and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.

12. Other Terms and Conditions:

v1.4

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.

Appendix C

Publications

The following two publication are attached to the next pages.

Kyrou, K., Hammond, A.M., Galizi, R., Kranjc, N., Burt, A., Beaghton, A.K., Nolan, T. and Crisanti, A. A CRISPR-Cas9 gene drive targeting *doublesex* causes complete population suppression in caged *Anopheles gambiae* mosquitoes. *Nat. Biotechnol.*, **36**(11):1062–1066, 2018. ISSN 15461696. doi: 10.1038/nbt.4245

Hammond, A.M., Kyrou, K., Bruttini, M., North, A., Galizi, R., Karlsson, X., Carpi, F.M., D'Aurizio, R. Crisanti, A. and Nolan, T. The creation and selection of mutations resistant to a gene drive over multiple generations in the malaria mosquito. *PLoS Genet.*, **13**(10):e1007039, 2017. ISSN 15537404. doi: 10.1371/journal.pgen.1007039

OPEN

A CRISPR–Cas9 gene drive targeting *doublesex* causes complete population suppression in caged *Anopheles gambiae* mosquitoes

Kyros Kyrou[1,2], Andrew M Hammond[1,2], Roberto Galizi[1], Nace Kranjc[1], Austin Burt[1], Andrea K Beaghton[1], Tony Nolan[1] & Andrea Crisanti[1]

In the human malaria vector *Anopheles gambiae*, the gene *doublesex* (*Agdsx*) encodes two alternatively spliced transcripts, *dsx-female* (*AgdsxF*) and *dsx-male* (*AgdsxM*), that control differentiation of the two sexes. The female transcript, unlike the male, contains an exon (exon 5) whose sequence is highly conserved in all *Anopheles* mosquitoes so far analyzed. We found that CRISPR–Cas9-targeted disruption of the intron 4–exon 5 boundary aimed at blocking the formation of functional AgdsxF did not affect male development or fertility, whereas females homozygous for the disrupted allele showed an intersex phenotype and complete sterility. A CRISPR–Cas9 gene drive construct targeting this same sequence spread rapidly in caged mosquitoes, reaching 100% prevalence within 7–11 generations while progressively reducing egg production to the point of total population collapse. Owing to functional constraint of the target sequence, no selection of alleles resistant to the gene drive occurred in these laboratory experiments. Cas9-resistant variants arose in each generation at the target site but did not block the spread of the drive.

CRISPR–Cas9 nucleases have been applied in gene drive constructs to target endogenous sequences of the human malaria vectors*A. gambiae* and *A. stephensi* with the objective of vector control[1,2]. These proof-of-principle experiments translated a hypothesis into a genetic tool able to suppress the reproductive capability of the mosquito population. According to mathematical modeling, suppression of *A. gambiae* mosquito reproductive capability can be achieved using gene drive systems targeting haplosufficient female fertility genes[3,4] or by introducing a sex distorter on the Y chromosome in the form of a nuclease designed to shred the X chromosome during meiosis, an approach known as Y-drive[4–6]. Both strategies could cause a progressive decrease in the number of fertile females that would eventually collapse the population.

However, several technical and scientific issues remain before these proof-of-principle demonstrations are advanced to effect vector population suppression. The development of a Y-drive has so far proven difficult because of the complete transcriptional shut down of the sex chromosomes during meiosis, which prevents the expression of a Y-linked sex distorter during gamete formation[6,7]. A gene drive designed to disrupt the *A. gambiae* fertility gene *AGAP007280* initially increased in frequency, but the selection of nuclease-resistant, functional variants that could be detected as early as generation 2 completely blocked the spread of the drive[2]. Resistant variants comprised small insertions or deletions (indels) of differing length generated by nonhomologous end joining repair following nuclease activity at the target site. The development of resistance to any nuclease-based gene drive was predicted[3] and is regarded as the main technical obstacle for the use of gene drives for vector control[8–12] (**Supplementary Table 1**). Gene drive targets with functional or structural constraints that might prevent the development of resistant variants could offer a route to successful population control. With this in mind, we evaluated the potential for disruption of the sex determination pathway in *A. gambiae* mosquitoes to selectively block the formation of the female splice transcript of the gene *doublesex* (*dsx*).

RESULTS

doublesex and sex differentiation in *A. gambiae*

Sex differentiation in insects follows a common pattern in which a primary signal activates a central gene that induces a cascade of molecular mechanisms that control alternative splicing of the *doublesex* (*dsx*) gene[13,14]. Although the molecular mechanisms and the genes involved in regulating sex differentiation in *A. gambiae* are not well understood, except that *Yob1* functions as a Y-linked male determining factor[15], available data indicated an important role of *dsx* in determining sexual dimorphism in this mosquito species[16]. In *A. gambiae*, *dsx* (*Agdsx*) consists of seven exons, distributed over an 85-kb region on chromosome 2R, a gene structure similar to that of *Drosophila melanogaster dsx* (*Dmdsx*) and other insect orthologs, and is alternately spliced to produce the female and male transcripts *AgdsxF* and *AgdsxM*, respectively. The female transcript consists of a 5′ segment common with that of males, a highly conserved female-specific exon (exon 5) and a 3′ common region, while the male transcript comprises only the

[1]Department of Life Sciences, Imperial College London, UK. [2]These authors contributed equally to this work. Correspondence should be addressed to A.C. (acrs@imperial.ac.uk).

Received 6 April; accepted 3 August; published online 24 September 2018; doi:10.1038/nbt.4245

5′ and 3′ common segments. The male-specific isoform contains an additional domain at the C terminus that is transcribed as a noncoding 3′ untranslated region in females (**Fig. 1a**).

To investigate whether *dsx* is a suitable target for a gene drive to suppress population reproductive capacity, we disrupted the intron 4–exon 5 boundary of *dsx* (**Fig. 1b**) to prevent the formation of functional AgdsxF while leaving the AgdsxM transcript unaffected. We injected *A. gambiae* embryos with a source of Cas9 and a single-guide RNA (sgRNA) designed to recognize and cleave a sequence overlapping the intron 4–exon 5 boundary, in combination with a template for homology-directed repair (HDR) to insert an eGFP transcription unit (**Fig. 1c**). Transformed individuals were intercrossed to generate homozygous and heterozygous mutants among the progeny. HDR-mediated integration was confirmed with diagnostic PCR using primers that spanned the insertion site: a large amplicon for the HDR event and a smaller amplicon for the wild-type allele enabled facile confirmation of genotypes (**Fig. 1d**). The knock-in of eGFP resulted in the complete disruption of the exon 5 (*dsxF*⁻) coding sequence and was confirmed by PCR and genomic sequencing of the chromosomal integration (**Supplementary Fig. 1**). Crosses of heterozygous individuals produced wild-type, heterozygous and homozygous individuals for the *dsxF*⁻ allele at the expected Mendelian ratio 1:2:1, indicating that there was no obvious lethality associated with the mutation during development (**Supplementary Table 2**). Larvae heterozygous for the exon 5 disruption developed into adult male and female mosquitoes with a sex ratio close to 1:1. However, half of $dsxF^{-/-}$ individuals developed into normal males whereas the other half had both male and female morphological features, as well as a number of developmental anomalies in the internal and external reproductive organs (intersex phenotype). To establish the sex genotype of these $dsxF^{-/-}$ intersex mosquitoes, we introgressed the mutation into a line containing a Y-linked visible marker (RFP) and used the presence of this marker to unambiguously assign sex genotype among individuals heterozygous and homozygous for the null mutation. This approach revealed that the intersex phenotype was observed only in females that were homozygous for the null mutation. We saw no phenotype in heterozygous mutants, suggesting that the female-specific isoform of *dsx* is haplosufficient. Examination of external sexually dimorphic structures in $dsxF^{-/-}$ genotypic females ($n > 50$) showed several phenotypic abnormalities, including the development of dorsally rotated male claspers (and absence of female cerci) and longer flagellomeres associated with male-like plumose antennae (**Fig. 2** and **Supplementary Table 3**). Analyses of the internal reproductive organs of the same set of insects revealed the absence of fully developed ovaries and spermathecae; instead these were replaced with male accessory glands and in some cases (~20%) by rudimentary pear-shaped organs resembling unstructured testes (**Supplementary Fig. 2**). Males carrying the *dsxF*⁻ null mutation in heterozygosity or homozygosity showed wild-type levels of fertility as measured by clutch size and larval hatching per mated female, as did heterozygous *dsxF*⁻ female mosquitoes. Intersex XX $dsxF^{-/-}$ female mosquitoes, although attracted to anesthetized mice, were unable to take a blood meal and failed to produce any eggs (**Fig. 3**). The drastic phenotype of $dsxF^{-/-}$ in females indicates that exon 5 of *dsx* has a fundamental role in the previously poorly understood sex differentiation pathway of *A. gambiae* mosquitoes and suggested that its sequence might represent a suitable target for gene drives designed for population suppression.

Building a gene drive to target *dsx*

We used recombinase-mediated cassette exchange to replace the *3xP3::GFP* transcription unit with a $dsxF^{CRISPRh}$ gene drive construct

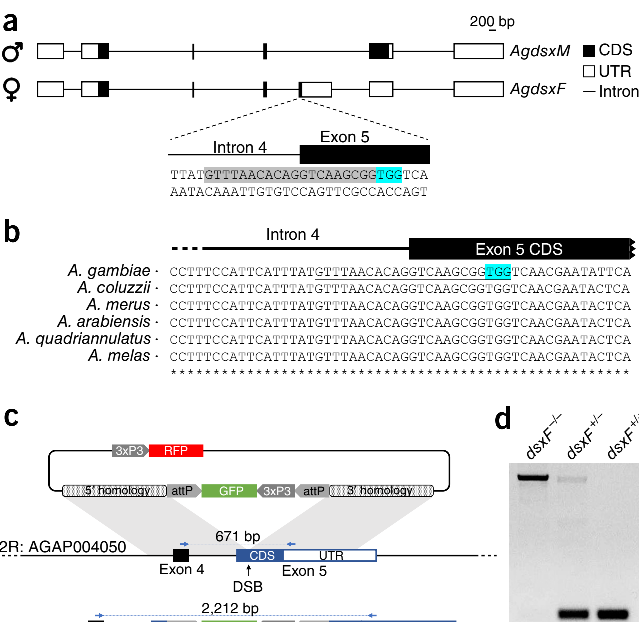


Figure 1 Targeting the female-specific isoform of *doublesex*. (**a**) Schematic representation of the male- and female-specific *dsx* transcripts and the gRNA sequence used to target the gene (shaded in gray). The gRNA spans the intron 4–exon 5 boundary. The protospacer-adjacent motif (PAM) of the gRNA is highlighted in blue. Scale bar, 200 bp. Coding regions of exons (CDS) are shaded in black, noncoding regions in white. Introns are not drawn to scale. UTR, untranslated region. (**b**) Sequence alignment of the *dsx* intron 4–exon 5 boundary in six of the species from the *A. gambiae* complex. The sequence is highly conserved within the complex suggesting tight functional constraint at this region of the *dsx* gene. The gRNA used to target the gene is underlined and the protospacer-adjacent motif is highlighted in blue. (**c**) Schematic representation of the HDR knockout construct specifically recognizing exon 5 and the corresponding target locus. DSB, double-strand break. (**d**) Diagnostic PCR using a primer set (blue arrows in **c**) to discriminate between the wild-type and *dsxF* allele in homozygous ($dsxF^{-/-}$), heterozygous ($dsxF^{+/-}$) and wild-type ($dsxF^{+/+}$) individuals.

that comprised an RFP marker gene, a transcription unit to express the guide RNA (gRNA) targeting *dsxF*, and *cas9* under the control of the germline promoter of *zero population growth* (*zpg*) and its terminator sequence (**Fig. 4a** and **Supplementary Fig. 3**). The *zpg* promoter has improved germline restriction of expression, resulting in increased female fertility compared with the *vasa* promoter used in previous gene drive constructs[17] (**Supplementary Fig. 4**). Successful cassette exchange events that incorporated $dsxF^{CRISPRh}$ into the target locus were confirmed in those individuals that had swapped the GFP for the RFP marker ($n = 17$ G_1 individuals) (**Supplementary Fig. 3**). During meiosis the Cas9–gRNA complex cleaves the wild-type allele at the target sequence and the $dsxF^{CRISPRh}$ cassette is copied into the wild-type locus by HDR ('homing'), disrupting exon 5 in the process. The ability of the $dsxF^{CRISPRh}$ construct to home and bypass Mendelian inheritance was analyzed by scoring the rates of RFP inheritance in the progeny of heterozygous parents (referred to as $dsxF^{CRISPRh}/+$ hereafter) crossed to wild-type mosquitoes. High $dsxF^{CRISPRh}$ transmission rates were observed in the progeny of both heterozygous $dsxF^{CRISPRh}/+$ male (95.9% ± 1.1% s.e.m.; $n = 87$) and female mosquitoes (99.4% ± 0.5%; $n = 33$) (**Fig. 4b**). The fertility of the $dsxF^{CRISPRh}$ line was also assessed to unravel potential negative effects due to ectopic expression of the nuclease in somatic cells and/or parental deposition of the nuclease into the newly fertilized embryos (**Fig. 4c**). These

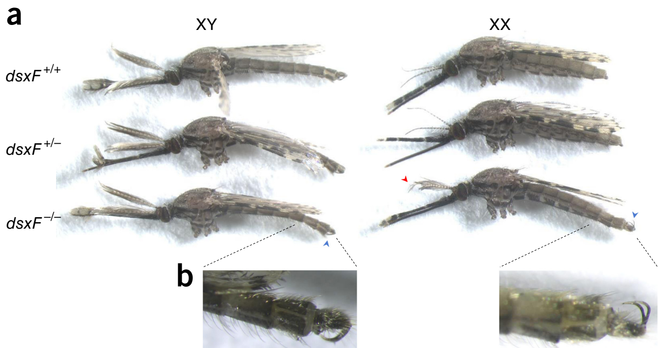


Figure 2 Morphological analysis of homozygous *dsxF*$^{-/-}$ mutants. (**a**) Morphological appearance of genetic males and females heterozygous (*dsxF*$^{+/-}$) or homozygous (*dsxF*$^{-/-}$) for the exon 5 null allele. This assay was performed in a strain containing a dominant RFP marker linked to the Y chromosome, whose presence permits unambiguous determination of male or female genotype. Anomalies in sexual morphology were observed only in *dsxF*$^{-/-}$ genetic female mosquitoes. This group of XX individuals showed male-specific traits, including a plumose antenna (red arrowhead) and claspers (blue arrowheads). This group also showed anomalies in the proboscis and accordingly they could not bite and feed on blood. Representative samples of each genotype are shown. (**b**) Magnification of the external genitalia. All *dsxF*$^{-/-}$ females carried claspers, a male-specific characteristic. The claspers were dorsally rotated rather than in the normal ventral position.

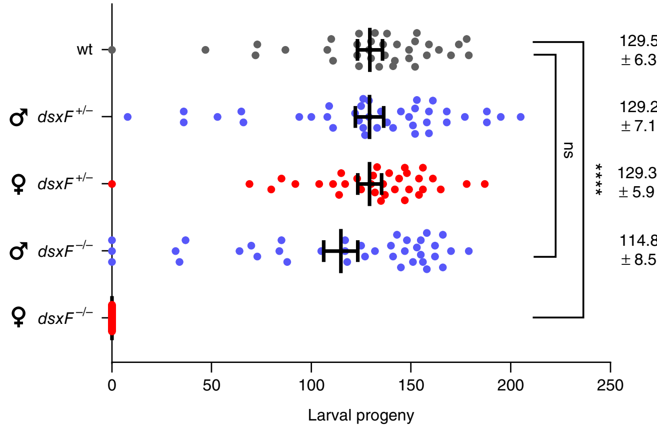


Figure 3 Reproductive phenotype of *dsxF* mutants. Male and female *dsxF*$^{-/-}$ and *dsxF*$^{+/-}$ individuals were mated with the corresponding wild-type sexes. Females were given access to a blood meal and subsequently allowed to lay individually. Fecundity was investigated by counting the number of larval progeny per lay ($n \geq 43$). Using wild type (wt) as a comparator, we saw no significant differences ('ns') in any genotype other than *dsxF*$^{-/-}$ females, which were unable to feed on blood and therefore failed to produce a single egg (****$P < 0.0001$; Kruskal–Wallis test). Vertical bars indicate the mean and the s.e.m. Blue and red indicate the crosses of male or female *dsxF* mutants, respectively, to wild type, whereas the gray dots represent wild-type-only crosses.

experiments showed that while heterozygous *dsxF*^{CRISPRh}/+ males showed a fecundity rate (assessed as larval progeny per fertilized female) that did not differ from that of wild-type males, heterozygous *dsxF*^{CRISPRh}/+ females had reduced fecundity overall (mean fecundity 49.8% ± 6.3% s.e.m., $P < 0.0001$). We noticed a greater reduction in the fertility of heterozygous females when the drive allele was inherited from the father (mean fecundity 21.7% ± 8.6%; $P < 0.0001$) ($n = 15$) rather than the mother (64.9% ± 6.9%; $P < 0.001$) ($n = 28$) (**Supplementary Fig. 5**). This could be explained by assuming a paternal deposition of active Cas9 nuclease into the newly fertilized zygote that stochastically induces conversion of *dsx* to *dsxF*$^{-}$, either through end-joining or HDR, in a substantial number of embryonic cells, which in females results in a reduced fertility. Consistent with this hypothesis, some heterozygous females (9 of 31 examined) receiving a paternal *dsxF*^{CRISPRh} allele showed a somatic mosaic phenotype that included, with varying penetrance, the absence of spermatheca and/or the formation of an incomplete clasper set (**Supplementary Fig. 2c**).

Assessment of *dsx* gene drive in caged insects

Using a mathematical model that includes the inheritance bias of the construct, the fecundity of heterozygous individuals, the phenotype of intersex, and the effect of the paternal deposition of the nuclease on female fertility (Online Methods), we found that the *dsxF*^{CRISPRh} had the potential to reach 100% frequency in caged population in 9–13 generations considering a starting allele frequency of 12.5% and stochasticity (**Fig. 5a**). To test this hypothesis, we mixed caged wild-type mosquito populations with heterozygous individuals carrying the *dsxF*^{CRISPRh} allele and monitored progeny at each generation to assess the spread of the drive and to quantify effect(s) on reproductive output. We started the experiment in two replicate cages, each with an initial drive allele frequency of 12.5% (300 wild-type female mosquitoes with 150 wild-type male mosquitoes and 150 *dsxF*^{CRISPRh}/+ male individuals). The initial drive allele frequency that we selected minimizes the stochastic loss of the drive (**Supplementary Fig. 6**) and represents a realistic field release scenario, being severalfold lower than that used in non-invasive genetic control strategies[18]. All of the eggs produced by the entire cage population were counted, and then 650 eggs were randomly selected to seed the next generations. The larvae that hatched from the eggs were counted and screened for the presence of the RFP marker to score the number of the progeny containing the *dsxF*^{CRISPRh} allele in each generation.

During the first three generations we observed an increase of the drive allele from 25% to ~69% in both caged populations, but at generation 4 the outcomes in the two cages diverged. In cage 2 the drive reached 100% frequency by generation 7; in generation 8, no eggs were produced and the population collapsed. In cage 1 the drive allele reached 100% frequency at generation 11 after remaining at around 65–70% for generations 4 through 8. In generation 12 the cage 1 population also failed to produce eggs (**Fig. 5b**). While the dynamics of spread of the gene drive in the two caged populations was different, both sets of finding fall within the prediction range of our mathematical model (**Fig. 5**).

Potential for resistance to *dsx* gene drive

We monitored the occurrence of mutations at the drive target site in generations 2, 3, 4 and 5 to identify the occurrence of nuclease-resistant, functional variants. Amplicon sequencing of the target sequence from pooled samples containing a minimum of 359 mosquitoes, which were collected in generations 2–5, revealed several low-frequency indels present at the target site (up to 1.16% frequency among nondrive alleles), none of which appeared to encode a functional *AgdsxF* transcript (**Supplementary Fig. 7**). In addition, none of the variants identified showed any signs of positive selection, which would be expected to cause them to increase in frequency as the drive progressively increased in frequency over generations, suggesting that the selected target sequence has rigid functional or structural constraints. This hypothesis is supported by the exceptionally high conservation of exon 5 in *A. gambiae* mosquitoes[19,20] and the presence of a strictly regulated splice site that is crucial in mosquito

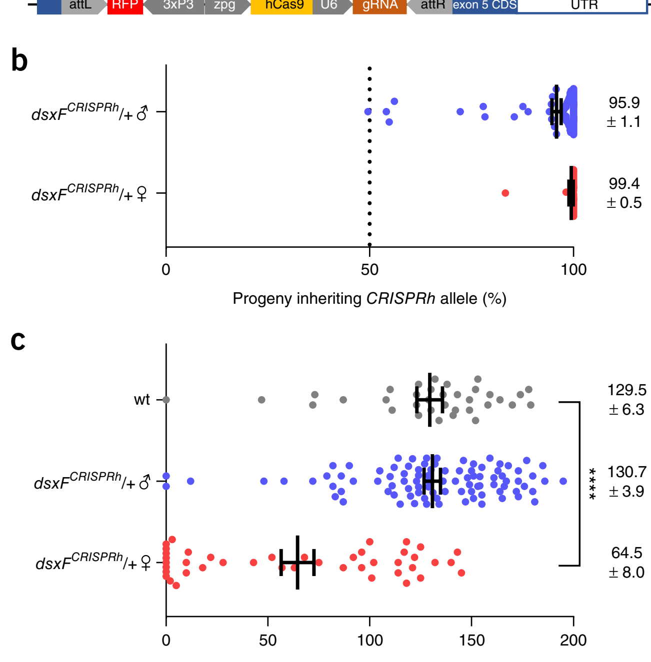
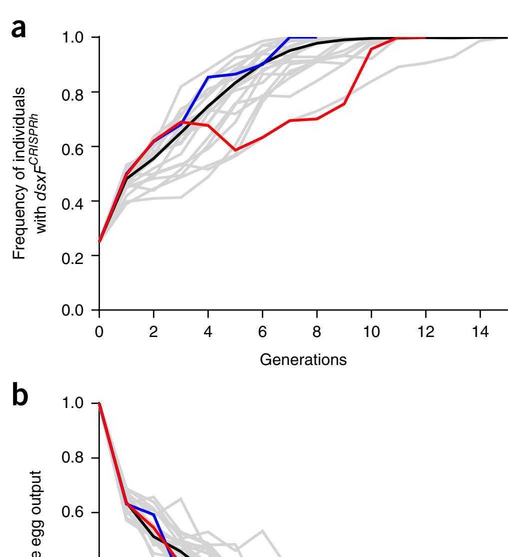


Figure 4 Transmission rate of the *dsxF*^{CRISPRh} driving allele and fecundity analysis of heterozygous male and female mosquitoes. (**a**–**c**) Male and female mosquitoes heterozygous for the *dsxF*^{CRISPRh} allele (**a**) were analyzed in crosses with wild-type mosquitoes to assess the inheritance bias of the *dsxF*^{CRISPRh} drive construct (**b**) and for the effect of the construct on their reproductive phenotype (**c**). (**b**) Scatter plot of the transgenic rate observed in the progeny of *dsxF*^{CRISPRh}/+ female or male mosquitoes that gave progeny when crossed to wild-type individuals ($n \geq 33$). Each dot represents the progeny derived from a single female. Both male and female *dsxF*^{CRISPRh}/+ showed a high transmission rate of up to 100% of the *dsxF*^{CRISPRh} allele to the progeny. The transmission rate was determined by visually scoring offspring for the RFP marker that is linked to the *dsxF*^{CRISPRh} allele. The dotted line indicates the expected Mendelian inheritance. Mean transmission rate (± s.e.m.) is shown. (**c**) Scatter plot showing the number of larvae produced by single females ($n \geq 35$) from crosses of *dsxF*^{CRISPRh}/+ mosquitoes with wild-type individuals after one blood meal. Mean progeny count (± s.e.m.) is shown (****$P < 0.0001$; Kruskal–Wallis test).



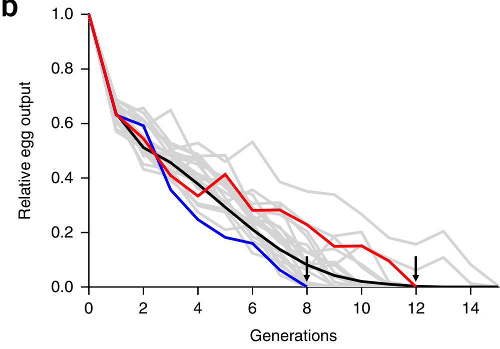


Figure 5 Dynamics of the spread of the *dsxF*^{CRISPRh} allele and effect on population reproductive capacity. Two cages were set up with a starting population of 300 wild-type females, 150 wild-type males and 150 *dsxF*^{CRISPRh}/+ males, seeding each cage with a *dsxF*^{CRISPRh} allele frequency of 12.5%. (**a**) The frequency of *dsxF*^{CRISPRh} mosquitoes was scored for each generation. The drive allele reached 100% prevalence in both cage 2 (blue) and cage 1 (red) at generation 7 and 11, respectively, in agreement with a deterministic model (black line) that takes into account the parameter values retrieved from the fecundity assays. Twenty stochastic simulations were run (gray lines) assuming a maximum population size of 650 individuals. (**b**) Total egg output deriving from each generation of the cage was measured and normalized relative to the output from the starting generation. Suppression of the reproductive output of each cage led the population to collapse completely (black arrows) by generation 8 (cage 2) or generation 12 (cage 1). Parameter estimates included in the model are provided in **Supplementary Table 5**.

reproductive biology. Furthermore, large-scale resequencing of 765 wild-caught mosquitoes from eight sub-Saharan African countries[20] revealed only a single rare SNP within the drive target site, present at 2.9% frequency (**Supplementary Fig. 8**). This naturally occurring variant could block the spread of the drive. To investigate this hypothesis, we tested whether this SNP variant was as susceptible to cleavage *in vitro* by Cas9 as the wild-type sequence, using the sgRNA from our gene drive construct. We found that the gRNA in our gene drive construct efficiently cleaved both the wild-type and the SNP sequence variant, which may indicate that our gene drive would be able to spread even if this conserved SNP was present (**Supplementary Fig. 9**). However, it is important to note that we cannot state that our drive target site is 'resistance-proof', since at scale, and over time, it is possible that nuclease-induced mutations could be produced that do restore sufficient function to the gene to be positively selected. This notwithstanding, targeting gene drives to functionally constrained sequences is clearly advantageous, as evidenced by the population collapse effected by this gene drive in both caged mosquito populations. Distinct, highly conserved sequences may have varying levels of functional constraint, and the relative strength of selection for maintaining sequence conservation versus the strength of selection imposed by the gene drive will ultimately determine their suitability as targets for gene drives.

Our data not only provide important functional insights into the role of *dsx* in *A. gambiae* sex determination, but also represent a substantial step toward the development of effective gene drive vector-control measures that aim to suppress insect populations. The intersex phenotype of *dsxF*$^{-/-}$ genetic females shows that exon 5 is crucial for the production of a functional female transcript, as was initially hypothesized on the basis of the expression profile of the *dsx* splice variants in the two sexes[16]. Furthermore, the observation that heterozygous *dsxF*^{CRISPRh}/+ females are fertile and produce almost 100% inheritance of the drive might indicate that most of the germ cells in these females are homozygous and, unlike somatic cells, do not undergo autonomous *dsx*-mediated sex commitment[21].

DISCUSSION

The development of a gene drive capable of collapsing a human malaria vector population to levels that cannot support malaria transmission is a long-sought scientific and technical goal[22]. The gene

drive *dsxF*^{CRISPRh} targeting exon 5 of *dsx* has several features that make it suitable for future field testing. Specifically, this drive has high inheritance bias, heterozygous individuals are fully fertile, homozygous females are sterile and unable to bite, and we found no evidence for nuclease-resistant functional variants at the drive target site. We note that these proof-of-principle experiments cannot conclude that this drive is resistance proof. This is in contrast to a recent study in *Drosophila* that targeted the *transformer* gene, upstream of *doublesex*. Invasion of the drive in *transformer* was rapidly compromised by the accumulation of large numbers of functional and nonfunctional resistant alleles[23].

Our *doublesex* gene drive now needs to be rigorously evaluated in large confined spaces that more closely mimic native ecological conditions, in accordance with the recommendations of the US National Academy of Sciences[24]. Under such conditions, competition for resources or mating success may disproportionately affect individuals harboring the gene drive, resulting in invasion dynamics substantially different from those observed in insectary cage experiments. Indeed, previous work with other genetically manipulated insects would suggest that in the less ideal conditions present in field cages and natural landscapes (competition for food, presence of predators and environmental stressors), heterozygous female mosquitoes carrying the drive allele might have a further reduction in fitness as result of the combined effect of the genetic background of the laboratory strain and the presence of the drive construct itself (**Supplementary Table 1**)[25–27]. To mimic less ideal conditions, we modeled varying levels of additional reduction in fitness (over the experimentally observed value of reproduction rate) associated with the heterozygous gene drive and evaluated the effects on penetrance of the *doublesex* gene drive (**Supplementary Fig. 10**). An additional reduction in fitness (over the experimentally observed value) of up to 40% would still allow the drive to reach 100% frequency and cause population suppression, albeit more slowly. Further reductions in fitness would result in different equilibrium frequencies that might still cause a large reproductive load on the population.

Our results may have implications beyond malaria vector control. The role of *doublesex* in sex determination in all insect species so far analyzed, and the high degree of *doublesex* sequence conservation among members of the same species (in gene regions involved in sex-specific splicing), suggests that these sequences might be an Achilles heel present in many insect pests that could be targeted with gene editing approaches.

METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We are grateful to L. Calisse, Y.W. Leong and F. Drury for technical assistance on the project and to D. Meacci and M. Gribble for insect husbandry. This work was funded by grants from the Bill & Melinda Gates Foundation and additionally from the BBSRC (BB/GCRF-IAA/10 and BB/GCRF-IAA/17/10) and DARPA (HR0011-17-2-0042).

AUTHOR CONTRIBUTIONS

A.C. conceived the overall experimental approach; K.K., A.M.H., T.N. and A.C. designed the research; K.K., A.M.H. and R.G. performed research; N.K. performed bioinformatic analysis; A.K.B. and A.B. performed modeling; K.K., A.M.H., A.K.B., T.N. and A.C. analyzed data; K.K., T.N. and A.C. wrote the paper, with input from all authors.

COMPETING INTERESTS

The authors declare no competing interests.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This work is licensed under a Creative Commons Attribution 4.0 International licence. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons licence, users will need to obtain permission from the licence holder to reproduce the material. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

1. Gantz, V.M. *et al.* Highly efficient Cas9-mediated gene drive for population modification of the malaria vector mosquito *Anopheles stephensi*. *Proc. Natl. Acad. Sci. USA* **112**, E6736–E6743 (2015).
2. Hammond, A. *et al.* A CRISPR-Cas9 gene drive system targeting female reproduction in the malaria mosquito vector *Anopheles gambiae*. *Nat. Biotechnol.* **34**, 78–83 (2016).
3. Burt, A. Site-specific selfish genes as tools for the control and genetic engineering of natural populations. *Proc. Biol. Sci.* **270**, 921–928 (2003).
4. Deredec, A., Godfray, H.C. & Burt, A. Requirements for effective malaria control with homing endonuclease genes. *Proc. Natl. Acad. Sci. USA* **108**, E874–E880 (2011).
5. Hamilton, W.D. Extraordinary sex ratios. A sex-ratio theory for sex linkage and inbreeding has new implications in cytogenetics and entomology. *Science* **156**, 477–488 (1967).
6. Galizi, R. *et al.* A synthetic sex ratio distortion system for the control of the human malaria mosquito. *Nat. Commun.* **5**, 3977 (2014).
7. Magnusson, K. *et al.* Demasculinization of the *Anopheles gambiae* X chromosome. *BMC Evol. Biol.* **12**, 69 (2012).
8. Champer, J. *et al.* Novel CRISPR/Cas9 gene drive constructs reveal insights into mechanisms of resistance allele formation and drive efficiency in genetically diverse populations. *PLoS Genet.* **13**, e1006796 (2017).
9. Hammond, A.M. *et al.* The creation and selection of mutations resistant to a gene drive over multiple generations in the malaria mosquito. *PLoS Genet.* **13**, e1007039 (2017).
10. Marshall, J.M., Buchman, A., Sánchez C, H.M. & Akbari, O.S. Overcoming evolved resistance to population-suppressing homing-based gene drives. *Sci. Rep.* **7**, 3776 (2017).
11. Unckless, R.L., Clark, A.G. & Messer, P.W. Evolution of resistance against CRISPR/Cas9 gene drive. *Genetics* **205**, 827–841 (2017).
12. Champer, J. *et al.* Reducing resistance allele formation in CRISPR gene drive. *Proc. Natl. Acad. Sci. USA* **115**, 5522–5527 (2018).
13. Burtis, K.C. & Baker, B.S. *Drosophila* doublesex gene controls somatic sexual differentiation by producing alternatively spliced mRNAs encoding related sex-specific polypeptides. *Cell* **56**, 997–1010 (1989).
14. Graham, P., Penn, J.K. & Schedl, P. Masters change, slaves remain. *Bioessays* **25**, 1–4 (2003).
15. Krzywinska, E., Dennison, N.J., Lycett, G.J. & Krzywinski, J. A maleness gene in the malaria mosquito *Anopheles gambiae*. *Science* **353**, 67–69 (2016).
16. Scali, C., Catteruccia, F., Li, Q. & Crisanti, A. Identification of sex-specific transcripts of the *Anopheles gambiae doublesex* gene. *J. Exp. Biol.* **208**, 3701–3709 (2005).
17. Hammond, A.M. *et al.* Improved CRISPR-based suppression gene drives mitigate resistance and impose a large reproductive load on laboratory-contained mosquito populations. Preprint at. *bioRxiv* https://doi.org/10.1101/360339 (2018).
18. Harris, A.F. *et al.* Field performance of engineered male mosquitoes. *Nat. Biotechnol.* **29**, 1034–1037 (2011).
19. Neafsey, D.E. *et al.* Highly evolvable malaria vectors: the genomes of 16 *Anopheles* mosquitoes. *Science* **347**, 1258522 (2015).
20. The *Anopheles gambiae* 1000 Genomes Consortium. Genetic diversity of the African malaria vector *Anopheles gambiae*. *Nature* **552**, 96–100 (2017).
21. Murray, S.M., Yang, S.Y. & Van Doren, M. Germ cell sex determination: a collaboration between soma and germline. *Curr. Opin. Cell Biol.* **22**, 722–729 (2010).
22. Curtis, C.F. Possible use of translocations to fix desirable genes in insect pest populations. *Nature* **218**, 368–369 (1968).
23. KaramiNejadRanjbar, M. *et al.* Consequences of resistance evolution in a Cas9-based sex conversion-suppression gene drive for insect pest management. *Proc. Natl. Acad. Sci. USA* **115**, 6189–6194 (2018).
24. National Academies of Sciences, Engineering & Medicine. Gene Drives on the Horizon: Advancing Science, Navigating Uncertainty, and Aligning Research with Public Values (National Academies Press, Washington, DC, 2016).
25. Carvalho, D.O. *et al.* Suppression of a field population of *Aedes aegypti* in Brazil by sustained release of transgenic male mosquitoes. *PLoS Neglect. Trop. Dis.* **9**, e0003864 (2015).
26. Facchinelli, L. *et al.* Field cage studies and progressive evaluation of genetically-engineered mosquitoes. *PLoS Neglect. Trop. Dis.* **7**, e2001 (2013).
27. Harris, A.F. *et al.* Successful suppression of a field mosquito population by sustained release of engineered male mosquitoes. *Nat. Biotechnol.* **30**, 828–830 (2012).

ONLINE METHODS

Choice of target site. To selectively disrupt the female-specific isoform of *dsx* we targeted the upstream intron–exon boundary of exon 5, which has been shown to be expressed only in the female mosquito[16]. This exon spans a region of 1,712 bp on chromosome 2R (48,712,937–48,714,648) and contains at its 5′ end 89 bp encoding the sequence-specific portion of the female *A. gambiae dsx* isoform (AgdsxF). We identified a potential gRNA target site that showed almost complete sequence conservation across 16 different anopheline species and complete conservation across the *A. gambiae* species complex[19] (viewed using http://people.csail.mit.edu/waterhouse/alnloc.cgi), with no nucleotide variation at 22 of the 23 targeted bases across 765 wild-caught *A. gambiae* collected as part of the *Anopheles gambiae* 1000 Genomes project[20]. A single nucleotide variant existing in the target site was represented at 2.9% allele frequency in the wild-caught mosquitoes (**Supplementary Fig. 8**). *In vitro* testing of this SNP variant revealed it to be as susceptible as the wild-type sequence to Cas9 cleavage directed by the gRNA used in our gene drive construct (**Supplementary Fig. 9**). The gRNA target and protospacer-adjacent motif (5′-GTTTAACACAGGTCAAGCGGTGG-3′) was also assessed *in silico* for off-target activity using the online-based tool ChopChop (http://chopchop.cbu.uib.no)[28,29].

Generation of CRISPR and donor constructs. We engineered available template plasmids to develop the CRISPR (p16510) and donor (pK101) constructs used to induce a double-strand break on the *dsx* target sequence and to provide template for homology-mediated repair, respectively. In practice a CRISPR construct[2] containing a U6::gRNA spacer cloning cassette was utilized, using Golden Gate cloning, to generate a PolIII transcription unit containing the *dsx*-specific gRNA. The plasmid also contained a human-codon-optimized *Cas9* coding sequence (*hCas9*) under the control of the *vasa2* promoter, which directs the expression of the Cas9 protein in the pole cells of the developing embryo. The donor plasmid was designed to contain a *GFP* transcription unit under the control of the *3xP3* promoter enclosed within two reversible φC31 *attP* recombination sequences flanked both 5′ and 3′ by 2 kb sequence immediately upstream and downstream, respectively, of the target site in *dsx* exon 5. The homology recombination regions flanking the *dsx* target site were amplified using primers adapted for Gibson assembly (dsxφ31L-F + dsxφ31L-R, dsxφ31R-F + dsxφ31R-R) (**Supplementary Table 4**), and the 3xP3::GFP cassette and backbone were excised using restriction enzymes from plasmid p163 (ref. 2). The final donor vector was named K101 (GenBank accession code MH541846) and was assembled using the standard Gibson assembly protocol[30].

Generation of the *dsxF* CRISPR homing allele (*dsxF^{CRISPRh}*). The *dsxF^{CRISPRh}* homing allele was generated *in vivo* by φC31 recombinase-mediated cassette exchange (RMCE)[31] using construct p17410, which encompassed the *hCas9* and the *dsx* gRNA transcription units, as well as reporter 3xP3::RFP cassette within two reversible φC31 *attB* recombination sequences. The gene drive construct targeting *dsxF* is identical in design to that described in Hammond *et al.*[2] except for the promoter and 3′ UTR surrounding the Cas9 gene: where previously these were from the ortholog of *vasa* (*AGAP008578*), in the current construct these are replaced by 1,074 bp upstream and 1,034 bp downstream of the germline-specific gene *AGAP006241*, the putative ortholog of *zero population growth* (*zpg*). A comparison of the fertility and homing rates in individuals heterozygous *vasa*- and *zpg*-driven gene CRISPR^h constructs at the exact same target locus (in *AGAP007280*, previously described by Hammond *et al.*[2]), showed improved fertility in the *zpg*-driven constructs[17] (summarized in **Supplementary Fig. 4**).

To make p17410 (GenBank accession code MH541847), we amplified both the promoter and terminator using primers carrying arms suitable for a subsequent Gibson assembly (**Supplementary Table 4**). The promoter, a 1,074-bp region upstream of the gene also containing the 5′ UTR, was amplified using primers zpgprCRISPR-F and zpgprCRISPR-R from the wild-type G3 mosquito strain. The terminator, a 1,037-bp region downstream also containing the 3′ UTR, was amplified using primers zpgteCRISPR-F and zpgteCRISPR-R. Using restriction enzymes, we removed the *hCas9* gene, backbone and gRNA cassette from p16510 and reassembled everything in a Gibson assembly reaction using the *zpg* promoter and terminator fragments.

Microinjection of embryos and selection of transformed mosquitoes. All mosquitoes were reared under standard conditions of 80% relative humidity and 28 °C. The mosquitoes were blood-fed on anesthetized mice, and freshly laid embryos were aligned and used for microinjections as described before[32]. We injected embryos with solution containing both p16510 and pK101 (each at 300 ng/μl) to generate mosquitoes (*dsxF*$^-$) in which the splicing junction of *dsx* exon 5 had been disrupted by the insertion of the eGFP φC31 acceptor construct. To generate the *dsxF* CRISPR homing allele, embryos from the *dsxF*$^-$ knock-in line were injected with solution containing p17410 and a plasmid-based source of φC31 integrase[2]. All the surviving G_0 larvae were crossed to wild-type mosquitoes and G_1 positive transformants were identified using a fluorescence microscope (Eclipse TE200) as GFP$^+$ larvae for the knock-in events and RFP$^+$ larvae for the RMCE events.

Containment of gene drive mosquitoes. All mosquitoes were housed at Imperial College London in an insectary that is compliant with Arthropod Containment Guidelines Level 2 (ref. 33). All GM work was performed under institutionally approved biosafety and GM protocols. In particular, GM mosquitoes containing constructs with the potential to show gene drive were housed in dedicated cubicles, separated by at least six doors from the external environment and requiring two levels of security card access. Moreover, because of its location in a city with a northern temperate climate, *A. gambiae* mosquitoes housed in the insectary are also ecologically contained. The physical and ecological containment of the insectary are compliant with guidelines set out in a recent commentary calling for safeguards in the study of synthetic gene drive technologies[34].

Molecular confirmation of gene targeting and cassette integration. Successful integration of *dsxF*$^-$ and *dsxF^{CRISPRh}* cassettes into *Agdsx* at exon 5 was confirmed by PCR using genomic DNA extracted using the Wizard Genomic DNA purification kit (Promega). Generation of the HDR-mediated *dsxF*$^-$ allele was confirmed using primers binding the integrated cassette (GFP-F and 3xP3-R) and the neighboring genomic integration site, external to the sequence included on the homology arms (dsxin3-F and dsxex6-R). *dsxF*$^-$ heterozygotes and homozygotes could be further distinguished by PCR using primers that bind either side of the inserted cassette (dsxex4-F and dsxex5-R), giving rise to a smaller and/or larger product corresponding to the empty wild-type locus or the predicted *dsxF*$^-$ allele, respectively.

RCME of the *dsxF^{CRISPRh}* construct into the *dsx* locus was confirmed using primers binding the drive cassette (hCas9-F and RFP-R) and the neighboring genomic integration site (dsxin4-F and dsxex5-R1). Primer sequences can be found in **Supplementary Table 4**.

Phenotypic characterization and microdissections. Microdissection and phenotypic characterization were carried out using Olympus SZX7 optical microscopes. Mosquitoes were collected in Falcon tubes and anesthetized on ice 5 min before dissection. For phenotypic comparison, the legs of the mosquitoes were removed to achieve the profile orientation. Pictures were taken using a HiChrome-SMII GXCAM digital mounted camera (GT Vision). Pictures of gonads were taken using the EVOS imaging system (Thermo-Fisher).

Phenotypic assays. Phenotypic assays designed to examine relative fecundity in mosquitoes carrying either *dsxF*$^-$ or *dsxF^{CRISPRh}* alleles were carried out essentially as described before[2]. Briefly, the offspring of intercrossed heterozygous *dsxF*$^{-/+}$ individuals were screened for heterozygous or homozygous knock-in on the basis of weak or strong GFP expression, respectively. Nonfluorescent progeny were kept as controls. Groups of 50 male and 50 female mosquitoes from each of the three classes were mated to an equal number of wild-type mosquitoes for 5 d, blood-fed, and a minimum of 45 females allowed to lay individually. The entire egg and larval progeny were counted for each lay and a minimum of 20 progeny investigated to confirm zygosity of the *dsxF*$^-$ allele in the parent. Females that failed to give progeny and had no evidence of sperm in their spermathecae were excluded from the analysis. Phenotypic assays for *dsxF^{CRISPRh}* individuals were performed essentially the same way with the exception that the entire larval progeny were screened for presence of DsRed, which is linked to the *dsxF^{CRISPRh}* allele. Statistical differences between genotypes were assessed using the Kruskal–Wallis test.

Cage trial assays. Two cage trials were initiated using 300 wild-type females, 150 wild-type males and 150 *dsxF*^{CRISPRh}/+ males. The wild-type and *dsxF*^{CRISPRh} lines were reared in parallel and kept under the same conditions. For the starting generation only, age-matched male and female pupae were allowed to emerge in separate cages and were mixed only when all the pupae had emerged. Both *dsxF*^{CRISPRh} and wild-type male pupae were screened for the presence of the RFP marker. Mosquitoes were left to mate for 5 days before they were blood fed on anesthetized mice. Two days after, the mosquitoes were set to lay in a 300-ml egg bowl filled with water and lined with filter paper. The eggs produced from the cage were photographed and counted using JMicroVision V1.27. Prior to counting, eggs were dispersed using gentle water spraying in the egg bowl to homogenize the population, and 650 eggs were randomly selected to seed the next generation. Larvae emerging from the 650 eggs were counted and screened for the presence of the RFP marker to score the transgenic rate of the progeny. The number of pupae used to seed the next generation was also recorded.

PCR of target site and deep sequencing analysis preparation. For the deep sequence analysis, a limiting PCR reaction was performed on 40 ng of genomic material extracted *en masse* using the Wizard Genomic DNA purification kit (Promega) from a minimum of 359 mosquitoes taken at G_2, G_3, G_4 and G_5 from both cage experiments. Using the KAPA HiFi HotStart Ready Mix PCR kit (Kapa Biosystems) and primers that carried the Illumina Nextera Transposase adapters (underlined), 4050-Illumina-F (TCGTCGGCAGCGTCAGATGTG TATAAGAGACAGACTTATCGGCATCAGTTGCG) and 4050-Illumina-R (GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGTGAATTCCGTC AGCCAGCA), we amplified a 358-bp locus containing the target site in 50-µl reactions. To maintain the proportion of the reads corresponding to particular alleles at the target site, the PCR reactions were performed under nonsaturating conditions; they were allowed to run for 20 cycles before 25 µl were removed and stored at −20 °C. The remnant 25 µl were run for another 20 cycles and used to verify the amplification on an agarose gel. Annealing time and temperature were adjusted to 68 °C for 20 s to minimize off-target amplification.

Libraries were prepared in accordance with the Illumina 16S Metagenomic Sequencing Library Preparation protocol and the Nextera XT Index Kit. AMPure XP beads were used to purify the amplicons. Dual indices and Illumina sequencing adapters were attached in a second PCR step using the Nextera XT Indexing Kit and purified with the AMPure XP beads. The resulting libraries were validated using an Agilent 2100 Bioanalyzer (DNA High Sensitivity kit, sample dilution 1:5) to determine size distribution and a Qubit 3.0 fluorometer to determine concentration of libraries. Indexed DNA libraries were normalized to 4 nM, pooled and loaded at a concentration of 9 pM onto an Illumina Flowcell v2 with 19% of φX control and sequenced using the Illumina MiSeq, 2 × 250 bp v2 paired end run.

Deep sequencing analysis. We ran CRISPResso[35] software v1.0.8 on raw sequencing data to detect mutations at the target site using parameter -q 30, setting the minimum average read quality score (phred33) to 30. Raw sequencing data was deposited in the NCBI BioProject database (accession code PRJNA476358). Resulting allele frequency tables were processed using *ad hoc* Python and R scripts to group, filter and visualize indels and substitutions in the amplicon. To visualize the frequency of the most abundant indels around the cut site in both cages over the four generations, we calculated the mean frequency of indels occurring within the target region, including 20 bp upstream and downstream of the target site. The top ten alleles with the highest mean frequency were then selected to show the change of frequency of each allele throughout four generations. To plot and show the distribution of indels and substitutions in the whole amplicon, we filtered out alleles with less than three reads.

Modeling. We use discrete-generation deterministic and stochastic models with random mating and males and females treated separately as in Hammond *et al.*[9], and incorporate different homing rates in males and females and a modified treatment of embryonic cleavage and repair from paternally and maternally derived nuclease, as observed (see "Population genetics model" below[9,36]). We include wild-type (W), driver (D), and nonfunctional nuclease-resistant (R) alleles. Cleavage followed by homing and repair occurs in the germline in heterozygous W/D females and males; otherwise inheritance is Mendelian. Gametes (W, D or R) from W/D females and W/D, D/R and D/D males carry nuclease that is transmitted to the zygote and acts in the embryo in somatic cells to reduce fitness if wild-type alleles are present, so that W/W, W/R and W/D females have fitness w10, w01, w11 or 1, depending on whether nuclease was derived from a transgenic mother, father, both or neither. All males are assumed to have fitness 1, and we assume no effects of parentally deposited nuclease in germline cells. In the stochastic version of the model, probabilities of mating, egg production, hatching and emergence from pupae are estimated from experiments (**Supplementary Table 5**) and random numbers for these events are taken from the appropriate multinomial distributions. To model the cage experiments, 300 females and 150 male wild-type adults along with 150 male drive heterozygotes (from transgenic fathers) are initially present. Females may fail to mate or may mate once randomly with a male of a given genotype according to its frequency in the male population. The number of eggs produced from each mated female is randomly chosen by sampling with replacement from experimental values in **Supplementary Table 6**. To start the next generation, 650 eggs are randomly selected, and these hatch with a probability that also depends upon on the genotype of the mother. The probability of subsequent survival to adulthood is assumed to be equal across genotypes.

Population genetics model. To model the results of the cage experiments, we use discrete-generation recursion equations for the genotype frequencies, treating males and females separately. $F_{ij}(t)$ and $M_{ij}(t)$ denote the frequency of females (or males) of genotype *i*/*j* in the total female (or male) population. We consider three alleles, W (wild-type), D (driver) and R (nonfunctional resistant), and therefore six genotypes.

Homing. Adults of genotype W/D produce gametes at meiosis in the ratio W:D:R as follows:

$$(1-d_\mathrm{f})(1-u_\mathrm{f}):d_\mathrm{f}:(1-d_\mathrm{f})u_\mathrm{f} \quad \text{in females}$$
$$(1-d_\mathrm{m})(1-u_\mathrm{m}):d_\mathrm{m}:(1-d_\mathrm{m})u_\mathrm{m} \quad \text{in males}$$

Here d_f and d_m are the rates of transmission of the driver allele in the two sexes and u_f and u_m are the fractions of nondrive gametes that are nonfunctional resistant (R alleles) from meiotic end-joining. In all other genotypes, inheritance is Mendelian.

Fitness. Let $w_{ij} \leq 1$ represent the fitness of genotype *i*/*j* relative to $w_{\mathrm{WW}} = 1$ for the wild-type homozygote. We assume no fitness effects in males. Fitness effects in females are manifested as differences in the relative ability of genotypes to participate in mating and reproduction. We assume the target gene is needed for female fertility, and thus D/D, D/R and R/R females are sterile; there is no reduction in fitness in females with only one copy of the target gene (W/D, W/R).

Parental effects. We consider that further cleavage of the W allele and repair can occur in the embryo if nuclease is present, due to one or both contributing gametes derived from a parent with one or two driver alleles. The presence of parental nuclease is assumed to affect somatic cells and therefore female fitness but has no effect in germline cells that would alter gene transmission. Previously, embryonic EJ effects (maternal only) were modeled as acting immediately in the zygote. Here, we consider that experimental measurements of female individuals of different genotypes and origins show a range of fitnesses, suggesting that individuals may be mosaics with intermediate phenotypes. We therefore model genotypes W/X (X = W, D, R) with parental nuclease as individuals with an intermediate reduced fitness w_{WX}^{10}, w_{WX}^{01} or w_{WX}^{11} depending on whether nuclease was derived from a transgenic mother, father or both. We assume that parental effects are the same whether the parent(s) had one or two drive alleles. For simplicity, a baseline reduced fitness of w_{10}, w_{01}, w_{11} is assigned to all genotypes W/X (X = W, D, R) with maternal, paternal and maternal/paternal effects, with fitness estimated as the product of mean egg production values and hatching rates relative to wild type in **Supplementary Table 5** in the deterministic model. In the stochastic version of the model, egg production from female individuals with different parentage is sampled with replacement from experimental values.

Recursion equations. We first consider the gamete contributions from each genotype, including parental effects on fitness. In addition to W and R gametes that are derived from parents that have no drive allele and therefore have no deposited nuclease, gametes from W/D females and W/D, D/R and D/D males carry nuclease that is transmitted to the zygote, and these are denoted W^*, D^* and R^*. The proportion

doi:10.1038/nbt.4245

e_i of type i alleles in eggs produced by females participating in reproduction are given in terms of male and female genotype frequencies below. Frequencies of mosaic individuals with parental effects (i.e., reduced fitness) due to nuclease from mothers, fathers or both are denoted by superscripts 10, 01 or 11.

$$e_{\mathrm{W}} = \left(\begin{array}{l} F_{\mathrm{WW}} + w^{10}_{\mathrm{WW}} F^{10}_{\mathrm{WW}} + w^{01}_{\mathrm{WW}} F^{01}_{\mathrm{WW}} + w^{11}_{\mathrm{WW}} F^{11}_{\mathrm{WW}} + \\ \left(F_{\mathrm{WR}} + w^{10}_{\mathrm{WR}} F^{10}_{\mathrm{WR}} + w^{01}_{\mathrm{WR}} F^{01}_{\mathrm{WR}} + w^{11}_{\mathrm{WR}} F^{11}_{\mathrm{WR}} \right) / 2 \end{array} \right) \Big/ \bar{w}_{\mathrm{f}}$$

$$e_{\mathrm{R}} = \frac{1}{2} \left(F_{\mathrm{WR}} + w^{10}_{\mathrm{WR}} F^{10}_{\mathrm{WR}} + w^{01}_{\mathrm{WR}} F^{01}_{\mathrm{WR}} + w^{11}_{\mathrm{WR}} F^{11}_{\mathrm{WR}} \right) \Big/ \bar{w}_{\mathrm{f}}$$

$$e^{*}_{\mathrm{W}} = (1-d_{\mathrm{f}})(1-u_{\mathrm{f}}) \left(w^{10}_{\mathrm{WD}} F^{10}_{\mathrm{WD}} + w^{01}_{\mathrm{WD}} F^{01}_{\mathrm{WD}} + w^{11}_{\mathrm{WD}} F^{11}_{\mathrm{WD}} \right) \Big/ \bar{w}_{\mathrm{f}}$$

$$e^{*}_{\mathrm{D}} = d_{\mathrm{f}} \left(w^{10}_{\mathrm{WD}} F^{10}_{\mathrm{WD}} + w^{01}_{\mathrm{WD}} F^{01}_{\mathrm{WD}} + w^{11}_{\mathrm{WD}} F^{11}_{\mathrm{WD}} \right) \Big/ \bar{w}_{\mathrm{f}}$$

$$e^{*}_{\mathrm{R}} = (1-d_{\mathrm{f}}) u_{\mathrm{f}} \left(w^{10}_{\mathrm{WD}} F^{10}_{\mathrm{WD}} + w^{01}_{\mathrm{WD}} F^{01}_{\mathrm{WD}} + w^{11}_{\mathrm{WD}} F^{11}_{\mathrm{WD}} \right) \Big/ \bar{w}_{\mathrm{f}}$$

The proportions s_i of type i alleles in sperm are

$$s_{\mathrm{W}} = \left(\begin{array}{l} M_{\mathrm{WW}} + M^{10}_{\mathrm{WW}} + M^{01}_{\mathrm{WW}} + M^{11}_{\mathrm{WW}} + \\ \left(M_{\mathrm{WR}} + M^{10}_{\mathrm{WR}} + M^{01}_{\mathrm{WR}} + M^{11}_{\mathrm{WR}} \right) / 2 \end{array} \right) \Big/ \bar{w}_{\mathrm{m}}$$

$$s_{\mathrm{R}} = \left(M_{\mathrm{RR}} + \left(M_{\mathrm{WR}} + M^{10}_{\mathrm{WR}} + M^{01}_{\mathrm{WR}} + M^{11}_{\mathrm{WR}} \right) / 2 \right) \Big/ \bar{w}_{\mathrm{m}}$$

$$s^{*}_{\mathrm{W}} = (1-d_{\mathrm{m}})(1-u_{\mathrm{m}}) \left(M^{10}_{\mathrm{WD}} + M^{01}_{\mathrm{WD}} + M^{11}_{\mathrm{WD}} \right) \Big/ \bar{w}_{\mathrm{m}}$$

$$s^{*}_{\mathrm{D}} = \left(M_{\mathrm{DD}} + M_{\mathrm{DR}}/2 + d_{\mathrm{m}} \left(M^{10}_{\mathrm{WD}} + M^{01}_{\mathrm{WD}} + M^{11}_{\mathrm{WD}} \right) \right) \Big/ \bar{w}_{\mathrm{m}}$$

$$s^{*}_{\mathrm{R}} = \left(M_{\mathrm{DR}}/2 + (1-d_{\mathrm{m}}) u_{\mathrm{m}} \left(M^{01}_{\mathrm{WD}} + M^{10}_{\mathrm{WD}} + M^{11}_{\mathrm{WD}} \right) \right) \Big/ \bar{w}_{\mathrm{m}}$$

Above, $\bar{w}_{\mathrm{f}}$ and $\bar{w}_{\mathrm{m}}$ are the average female and male fitness:

$$\begin{aligned} \bar{w}_{\mathrm{m}} &= M_{\mathrm{WW}} + M^{10}_{\mathrm{WW}} + M^{01}_{\mathrm{WW}} + M^{11}_{\mathrm{WW}} + M^{10}_{\mathrm{WD}} + M^{01}_{\mathrm{WD}} + M^{11}_{\mathrm{WD}} \\ &\quad + M_{\mathrm{WR}} + M^{10}_{\mathrm{WR}} + M^{01}_{\mathrm{WR}} + M^{11}_{\mathrm{WR}} + M_{\mathrm{DD}} + M_{\mathrm{DR}} + M_{\mathrm{RR}} \\ &= 1 \end{aligned}$$

To model cage experiments, we start with an equal number of males and females, with an initial frequency of wild-type females in the female population of $F_{\mathrm{WW}} = 1$, wild-type males in the male population of $M_{\mathrm{WW}} = 1/2$, and $M^{01}_{\mathrm{WD}} = 1/2$ heterozygote drive males that inherited the drive from their fathers. Assuming a 1:1 ratio of males and females in progeny, after the starting generation, genotype frequencies of type i/j in the next generation $(t + 1)$ are the same in males and females, $F_{ij}(t + 1) = M_{ij}(t + 1)$. Both are given by $G_{ij}(t + 1)$ in the following set of equations in terms of the gamete proportions in the previous generation, assuming random mating:

$$G_{\mathrm{WW}}(t+1) = e_{\mathrm{W}} s_{\mathrm{W}}$$

$$G^{10}_{\mathrm{WW}}(t+1) = e^{*}_{\mathrm{W}} s_{\mathrm{W}}$$

$$G^{01}_{\mathrm{WW}}(t+1) = e_{\mathrm{W}} s^{*}_{\mathrm{W}}$$

$$G^{11}_{\mathrm{WW}}(t+1) = e^{*}_{\mathrm{W}} s^{*}_{\mathrm{W}}$$

$$G^{10}_{\mathrm{WD}}(t+1) = e^{*}_{\mathrm{D}} s_{\mathrm{W}}$$

$$G^{01}_{\mathrm{WD}}(t+1) = e_{\mathrm{W}} s^{*}_{\mathrm{D}}$$

$$G^{11}_{\mathrm{WD}}(t+1) = e^{*}_{\mathrm{W}} s^{*}_{\mathrm{D}} + e^{*}_{\mathrm{D}} s^{*}_{\mathrm{W}}$$

$$G_{\mathrm{WR}}(t+1) = e_{\mathrm{W}} s_{\mathrm{R}} + e_{\mathrm{R}} s_{\mathrm{W}}$$

$$G^{10}_{\mathrm{WR}}(t+1) = e^{*}_{\mathrm{W}} s_{\mathrm{R}} + e^{*}_{\mathrm{R}} s_{\mathrm{W}}$$

$$G^{01}_{\mathrm{WR}}(t+1) = e_{\mathrm{W}} s^{*}_{\mathrm{R}} + e_{\mathrm{R}} s^{*}_{\mathrm{W}}$$

$$G^{11}_{\mathrm{WR}}(t+1) = e^{*}_{\mathrm{W}} s^{*}_{\mathrm{R}} + e^{*}_{\mathrm{R}} s^{*}_{\mathrm{W}}$$

$$G_{\mathrm{DD}}(t+1) = e^{*}_{\mathrm{D}} s^{*}_{\mathrm{D}}$$

$$G_{\mathrm{DR}}(t+1) = \left(e_{\mathrm{R}} + e^{*}_{\mathrm{R}} \right) s^{*}_{\mathrm{D}} + e^{*}_{\mathrm{D}} \left(s_{\mathrm{R}} + s^{*}_{\mathrm{R}} \right)$$

$$G_{\mathrm{RR}} = \left(e_{\mathrm{R}} + e^{*}_{\mathrm{R}} \right) \left(s_{\mathrm{R}} + s^{*}_{\mathrm{R}} \right)$$

The frequency of transgenic individuals can be compared with experiment: the fraction of RFP^{+} individuals is given by

$$\begin{aligned} f_{\mathrm{RFP}^{+}} &= F^{10}_{\mathrm{WD}} + F^{01}_{\mathrm{WD}} + F^{11}_{\mathrm{WD}} + F_{\mathrm{DD}} + F_{\mathrm{DR}} \\ &+ M^{10}_{\mathrm{WD}} + M^{01}_{\mathrm{WD}} + M^{11}_{\mathrm{WD}} + M_{\mathrm{DD}} + M_{\mathrm{DR}} \end{aligned}$$

All calculations are carried out using Wolfram Mathematica (Wolfram Research Inc.)

***In vitro* cleavage assay against wild-type and SNP variant target site.** We performed an *in vitro* cleavage assay to test the ability of the gRNA used in this study to cleave the target site that incorporates the SNP found in wild populations in Africa (**Supplementary Fig. 9**). Using Golden Gate cloning and primers modified to carry suitable overhangs, we introduced the two target sequences separately into a 2-kb plasmid. As a control, we also prepared a plasmid that carries a modified version of the *dsx* target site without the SNP that lacks the PAM sequence, necessary for Cas9 cleavage. All three vectors were linearized and verified on a gel before the cleavage assay. For the cleavage assay we used a ready-to-use sgRNA provided by Synthego (USA) and *S. pyogenes* Cas9 nuclease in the form of enzyme (NEB). To form ribonucleoprotein particles (RNPs), we mixed a 1:1 molar ration of the sgRNA and the Cas9 protein into a 40-μl reaction to a final concentration of 400 nM and left it to incubate at room temperature for 10 min. The linearized substrate was added to the reactions in a final concentration of 40 nM in a final volume of 50 μl and incubated at 37 °C for 30 min. Proteinase K was added to stop the reaction and 20 μl were verified on a gel. The primers used to create the three target sequences are outlined in **Supplementary Table 4**.

Ethics statement. All animal work was conducted according to UK Home Office Regulations and approved under Home Office License PPL 70/8914.

Life Sciences Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability. Raw sequencing data were deposited in the NCBI BioProject database under accession code PRJNA476358.

28. Montague, T.G., Cruz, J.M., Gagnon, J.A., Church, G.M. & Valen, E. CHOPCHOP: a CRISPR/Cas9 and TALEN web tool for genome editing. *Nucleic Acids Res.* **42**, W401–W407 (2014).
29. Labun, K., Montague, T.G., Gagnon, J.A., Thyme, S.B. & Valen, E. CHOPCHOP v2: a web tool for the next generation of CRISPR genome engineering. *Nucleic Acids Res.* **44**, W272–W276 (2016).
30. Gibson, D.G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).
31. Bateman, J.R., Lee, A.M. & Wu, C.T. Site-specific transformation of *Drosophila* via phiC31 integrase-mediated cassette exchange. *Genetics* **173**, 769–777 (2006).
32. Fuchs, S., Nolan, T. & Crisanti, A. Mosquito transgenic technologies to reduce *Plasmodium* transmission. *Methods Mol. Biol.* **923**, 601–622 (2013).
33. Scott, T.W. Containment of arthropod disease vectors. *ILAR J.* **46**, 53–61 (2005).
34. Akbari, O.S. *et al.* Safeguarding gene drive experiments in the laboratory. *Science* **349**, 927–929 (2015).
35. Pinello, L. *et al.* Analyzing CRISPR genome-editing experiments with CRISPResso. *Nat. Biotechnol.* **34**, 695–697 (2016).
36. Papathanos, P.A., Windbichler, N., Menichelli, M., Burt, A. & Crisanti, A. 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 Mol. Biol.* **10**, 65 (2009).

nature research

Corresponding author(s): Andrea Crisanti

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a	Confirmed	
☐	☒	The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's d, Pearson's r), indicating how they were calculated
☐	☒	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	No customised software was used.
Data analysis	No customised software was used.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All data generated or analysed during this study are already included in this published article (and its supplementary information files). Accession codes for sequencing data and plasmids are indicated in the text. Accession code PRJNA476358; Genbank accession code: MH541846; Genbank accession code MH541847

Nature Biotechnology: doi:10.1038/nbt.4245

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample size was chosen consistent with the previous literature reporting similar assays . Sample size was maximised within the feasibility of performing biological assays with live insects. Starting frequencies in population experiments were chosen in order to minimise the effect of stochastic noise
Data exclusions	No data exclusions
Replication	Biological replicates were performed as described in the main text
Randomization	Randomisation was not appropriate for this experimental approach
Blinding	Blinding was not performed in this study. Many genotypes had distinguishing physical characteristics.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☒	☐ Unique biological materials
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☐	☒ Animals and other organisms
☒	☐ Human research participants

Methods

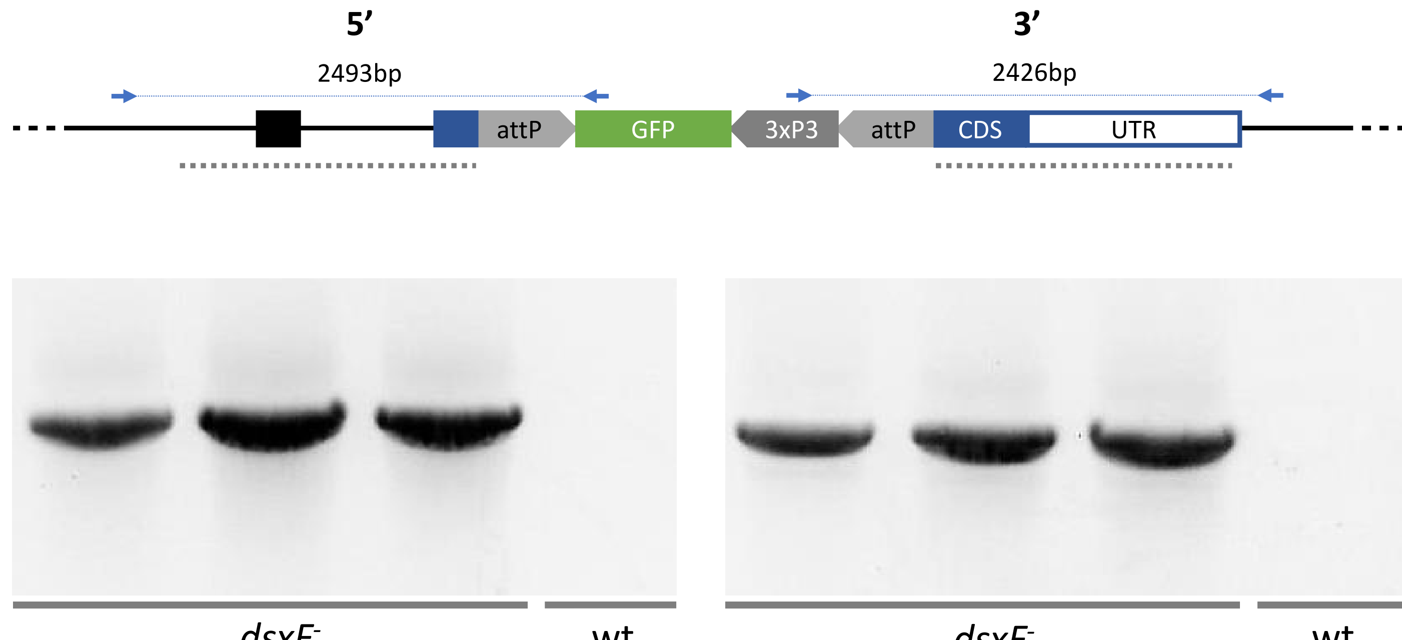
n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	For the purpose of giving mosquitoes a bloodmeal anesthetized mice (CD1 strain) were used. All animal work was conducted according to UK Home Office Regulations and approved under Home Office License PPL 70/8914.
Wild animals	The study did not involve wild animals.
Field-collected samples	The study did not involve field collected samples.

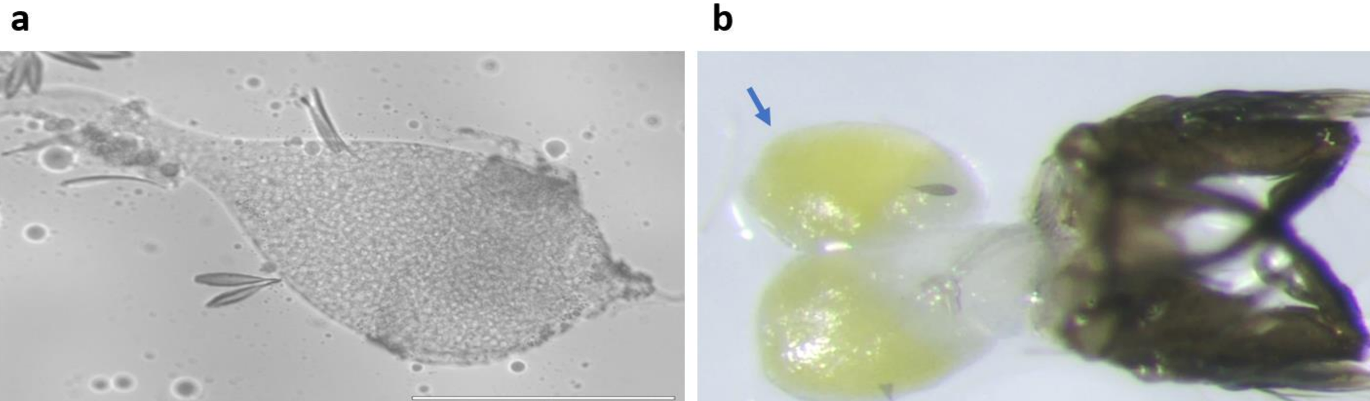
Nature Biotechnology: doi:10.1038/nbt.4245

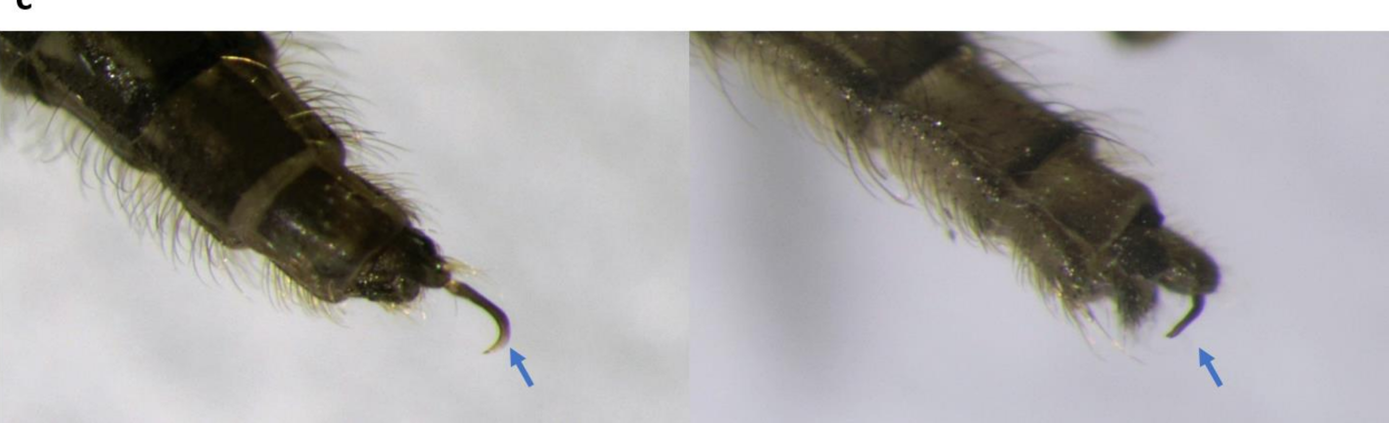


Supplementary Figure 1

Molecular confirmation of the correct integration of the HDR-mediated event to generate *dsxF*

PCRs were performed to verify the location of the *dsx* ϕC31 knock-in integration. Primers (blue arrows) were designed to bind internal of the ϕC31 construct and outside of the regions used for homology directed repair (HDR) (dotted grey lines) which were included in the Donor plasmid K101. Amplicons of the expected sizes should only be produced in the event of a correct HDR integration. The gel shows PCRs performed on the 5' (left) and 3' (right) of 3 individuals for the *dsx* ϕC31 knock-in line (*dsxF*$^{-}$) and wild type (wt) as a negative control.

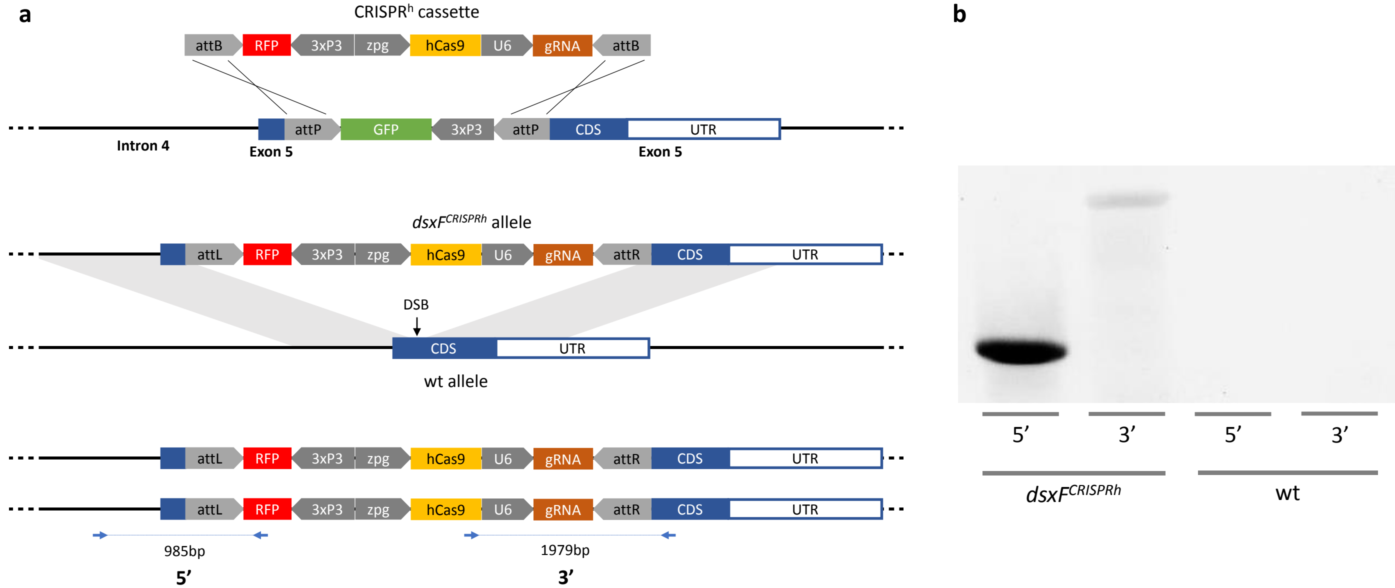




Supplementary Figure 2

Morphology of the *dsxF*$^{-/-}$ internal reproductive organs

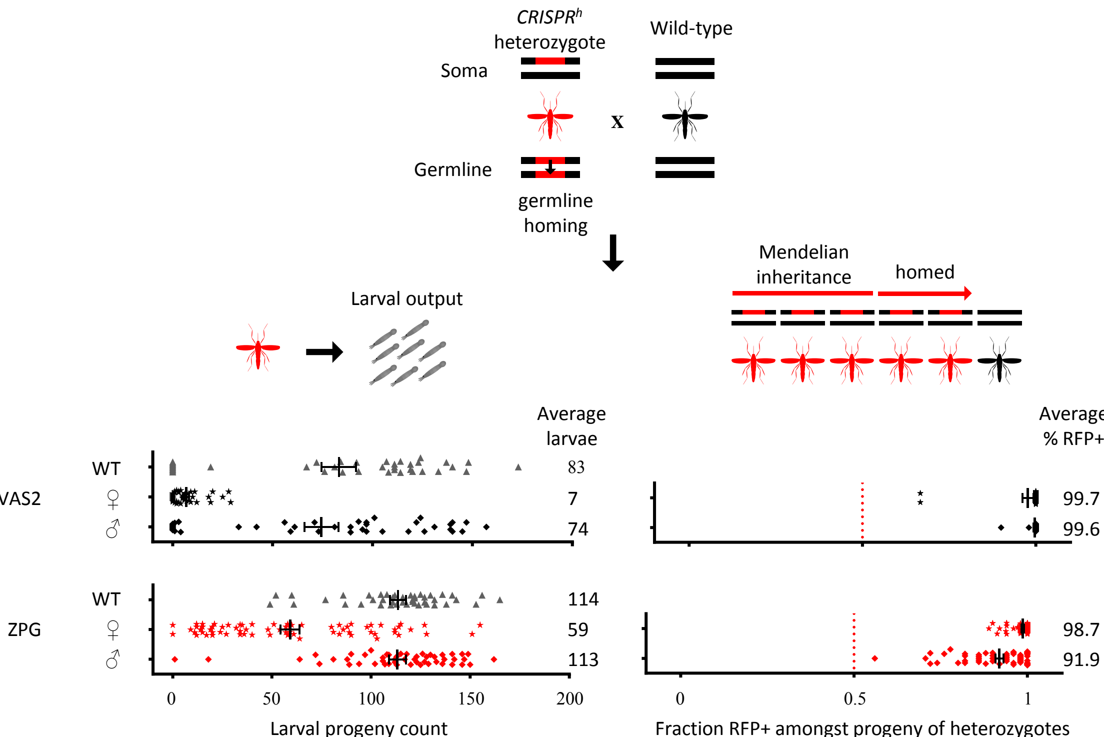
(**a**) Testis-like gonad from 3-days old female *dsxF*$^{-/-}$ individual. There was no layer division between the cells and there was no evidence of sperm. (**b**) Dissections performed on *dsxF*$^{-/-}$ genetic females revealed the presence of organs resembling accessory glands, a typical male internal reproductive organ. (**c**) somatic mosaicism of penetrance of *dsxF*$^{-/-}$ phenotype in *dsxF*^{CRISPRh}/+ females due to paternal deposition of nuclease, that can result in partial formation of clasper sets.



Supplementary Figure 3

Development of *dsxF*^{CRISPRh} drive construct and its predicted homing process and molecular confirmation of the locus

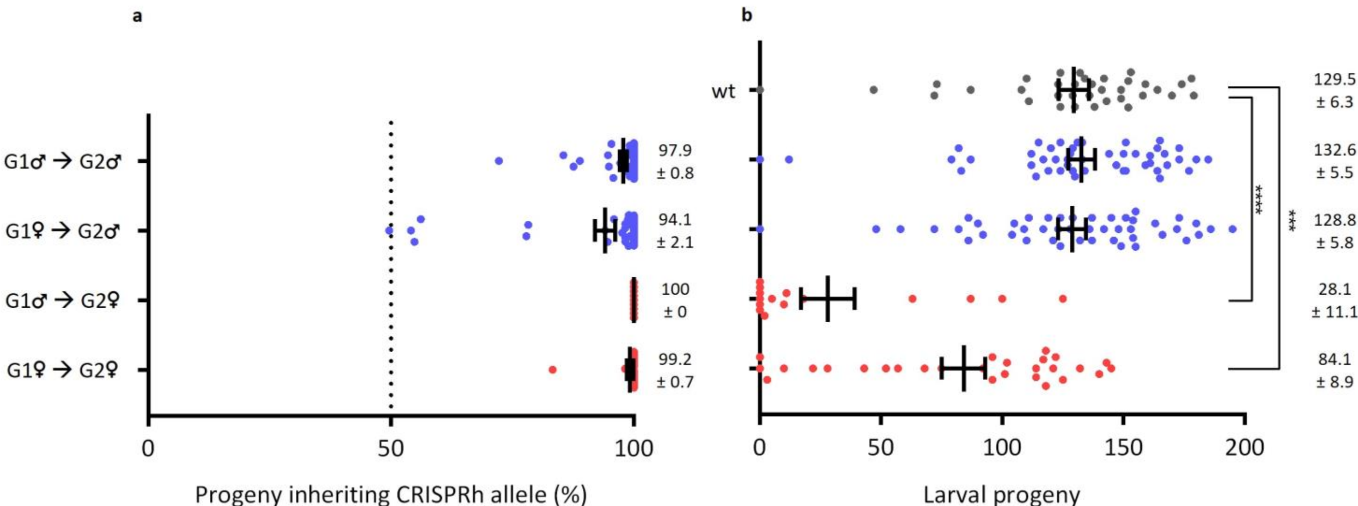
(**a**) The drive construct (*CRISPR^{h}* cassette) contained the transcription unit of a human codon-optimised *Cas9* controlled by the germline-restrictive *zpg* promoter, the RFP gene under the control of the neuronal *3xP3* promoter and the gRNA under the control of the constitutive *U6* promoter, all enclosed within two *attB* sequences. The cassette was inserted at the target locus using recombinase-mediated cassette exchange (RMCE) by injecting embryos with a plasmid containing the cassette and a plasmid containing a ϕC31 recombination transcription unit. During meiosis the Cas9/gRNA complex cleaves the wild-type allele at the target locus (DSB) and the construct is copied across to the wild-type allele via HDR (homing) disrupting exon 5 in the process. (**b**) Representative example of molecular confirmation of successful RMCE events. Primers (blue arrows) that bind components of the *CRISPR^{h}* cassette were combined with primers that bind the genomic region surrounding the construct. PCRs were performed on both sides of the *CRISPR^{h}* cassette (5’ and 3’) on many individuals as well as wild-type controls (wt).



Supplementary Figure 4

Gene drives designed to express Cas9 under regulation of the promoter and terminator regions of *zpg* show high rates of biased transmission and substantially improved fertility compared with the *vasa*2 promoter at a previously validated female fertility locus (*AGAP007280*)

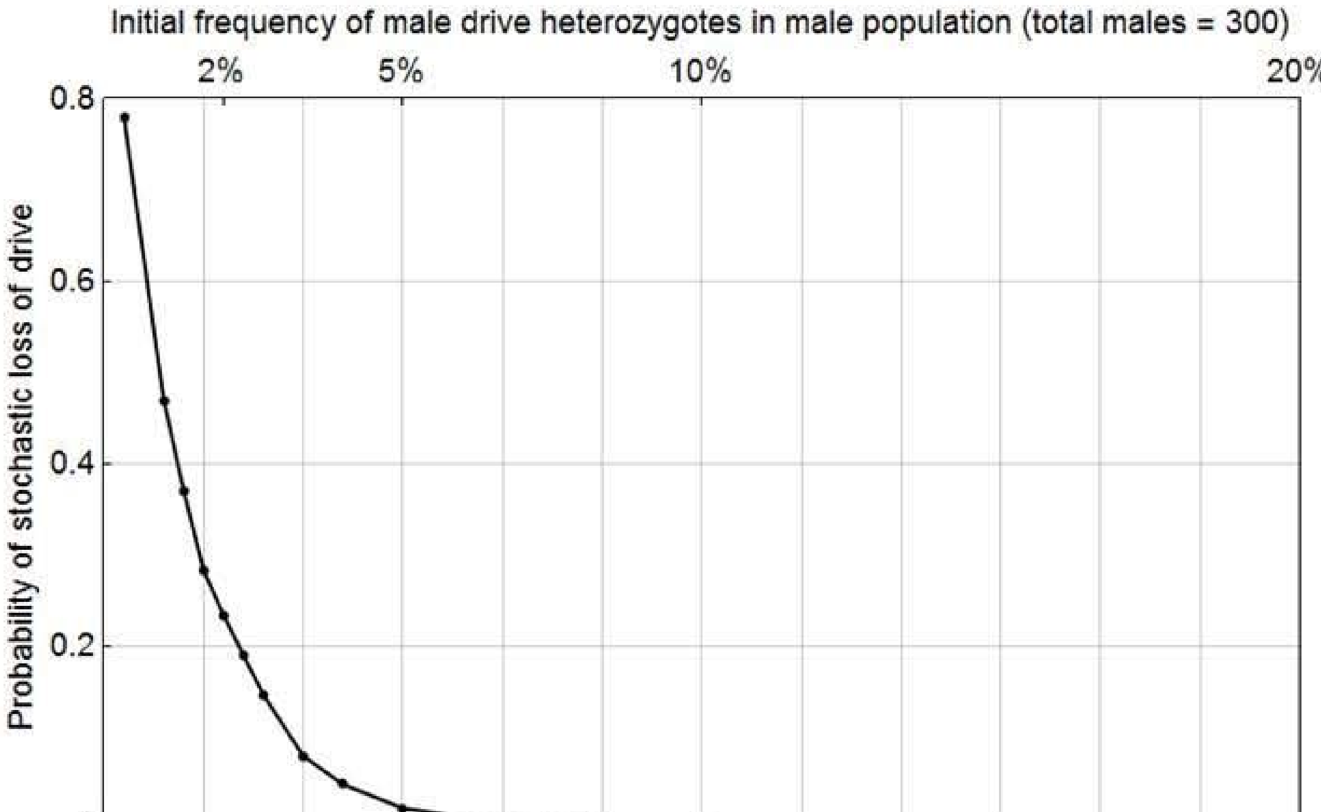
Phenotypic assays were performed to measure fertility and transmission rates for each gene drive based upon the *vasa* and *zpg* promoters. The data for the *vasa-CRISPR*^{h} is previously reported in Hammond et al. (2016). The *zpg-CRISPR*^{h} construct targeting AGAP007280 recognised exactly the same target site and was inserted in identical fashion to the *vasa-CRISPR*^{h}, through recombinase-mediated cassette exchange[9]. The larval output was determined for individual drive heterozygotes crossed to wild type (left), and their progeny scored for the presence of DsRed linked to the construct (right). The average progeny count and transmission rate is also shown (± s.e.m.).



Supplementary Figure 5

Maternal or paternal inheritance of the *dsxF^{CRISPRh}* driving allele affect fecundity and transmission bias in heterozygotes

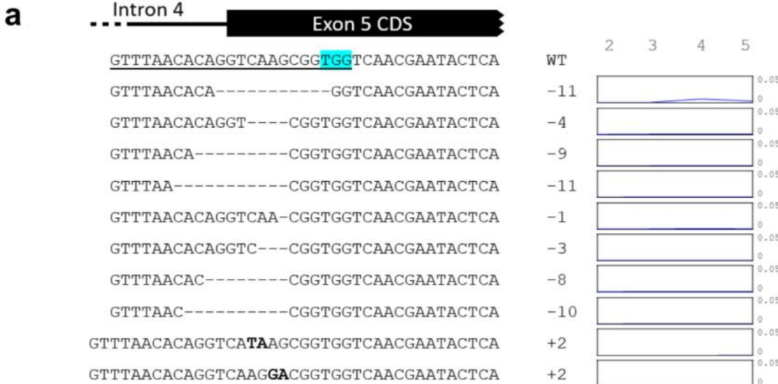
Male and female *dsxF^{CRISPRh}* heterozygotes (*dsxF^{CRISPRh}/+*) that had inherited a maternal or paternal copy of the driving allele were crossed to wild type and assessed for inheritance bias of the construct (**a**) and reproductive phenotype (**b**). (**a**) Progeny from single crosses (n≥15) were screened for the fraction that inherited DsRed marker gene linked to the *dsxF^{CRISPRh}* driving allele (e.g. G_1♂→G_2♀ represents a heterozygous female that received the drive allele from her father). Levels of homing were similarly high in males and females whether the allele had been inherited maternally or paternally. The dotted line indicates the expected Mendelian inheritance. Mean transmission rate (± s.e.m.) is shown. (**b**) Counts of hatched larvae for the individual crosses revealed a fertility cost in female *dsxF^{CRISPRh}* heterozygotes that was stronger when the allele was inherited paternally. Mean progeny count (± s.e.m.) is shown. (***, p<0.001;****, p<0.0001; Kruskal-Wallis test).

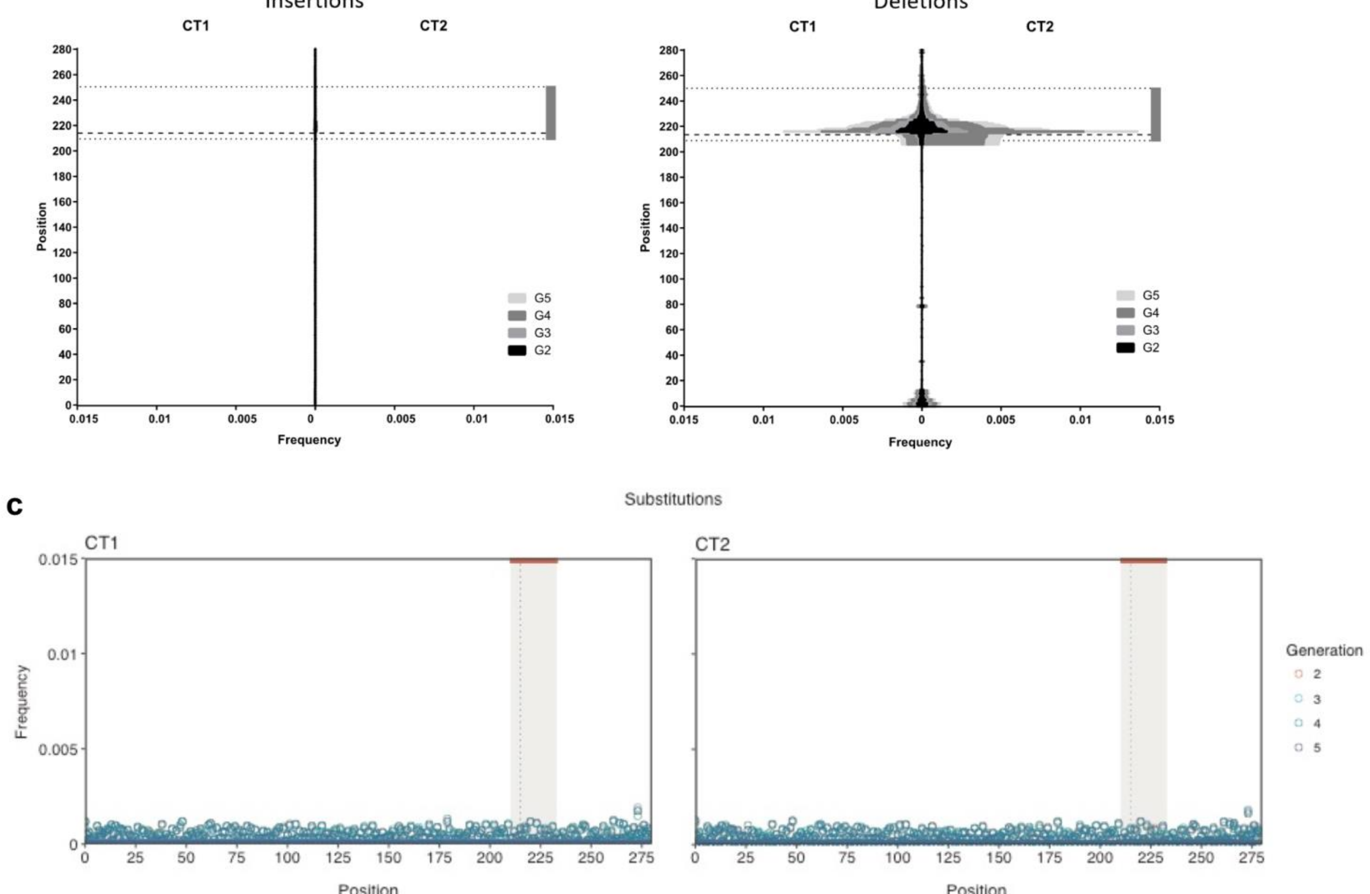


Supplementary Figure 6

Probability of stochastic loss of the drive as a function of initial number of male drive heterozygotes

To calculate the probability of stochastic loss of the drive in the cage experiment setup, for each initial number (h0) of male drive heterozygous individuals, out of 1000 simulations of the stochastic cage model (described in Supp Info), we recorded the number of times the drive was not present at 40 generations (and consequently population elimination did not occur). Each data point represents 1000 individual simulations of the stochastic cage model.



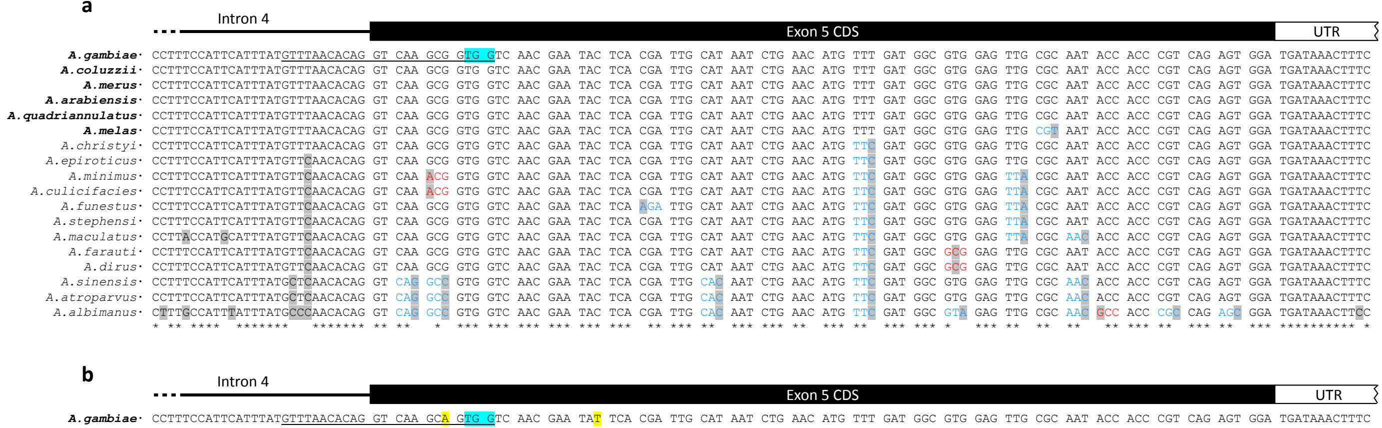


Supplementary Figure 7

Frequency plots of variants and indels in target sequence

Pooled amplicon sequencing of the target site from 4 generations of the cage experiment (generations 2, 3, 4 and 5) revealed a range of very low frequency indels at the target site (**a**), none of which showed any sign of positive selection. Insertion, deletion and substitution frequencies per nucleotide position were calculated, as a fraction of all non-drive

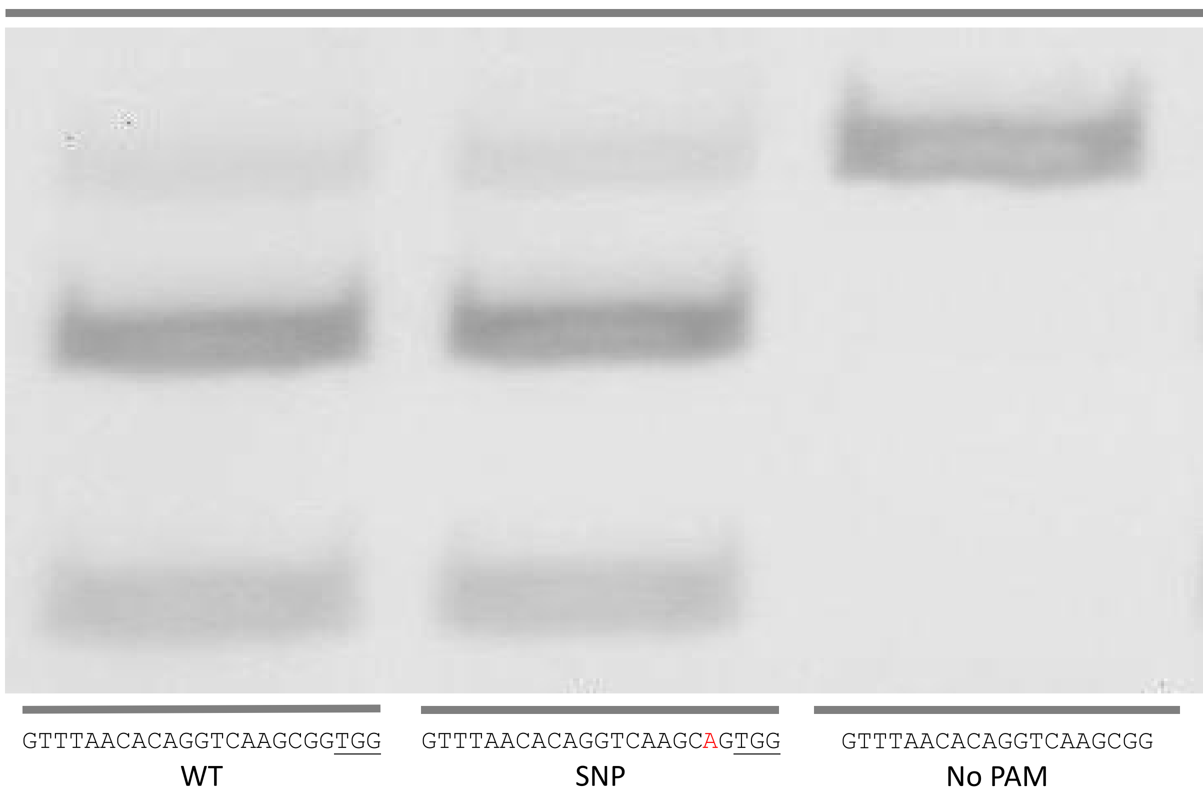
alleles, from the deep sequencing analysis for both cages. Distribution of insertions and deletions (**b**) in the amplicon is shown for each cage. Contribution of insertions and deletions arising from different generations is displayed with the frequency in each generation represented by a different colour. Significant change ($p<0.01$) in the overall indel frequency was observed in the region around the cut-site (dotted area ± 20 bp) for both cages. No significant changes were observed in the substitution frequency (**c**) around the cut-site (shaded area ± 20 bp) when compared with the rest of the amplicon, confirming that the gene drive did not generate any substitution activity at the target locus and that the laboratory colony is devoid of any standing variation in the form of SNPs within the entire amplicon.



Supplementary Figure 8

Sequence comparison of the *dsx* female-specific exon 5 across members of the *Anopheles* genus and SNP data obtained from *A. gambiae* mosquitoes in Africa.

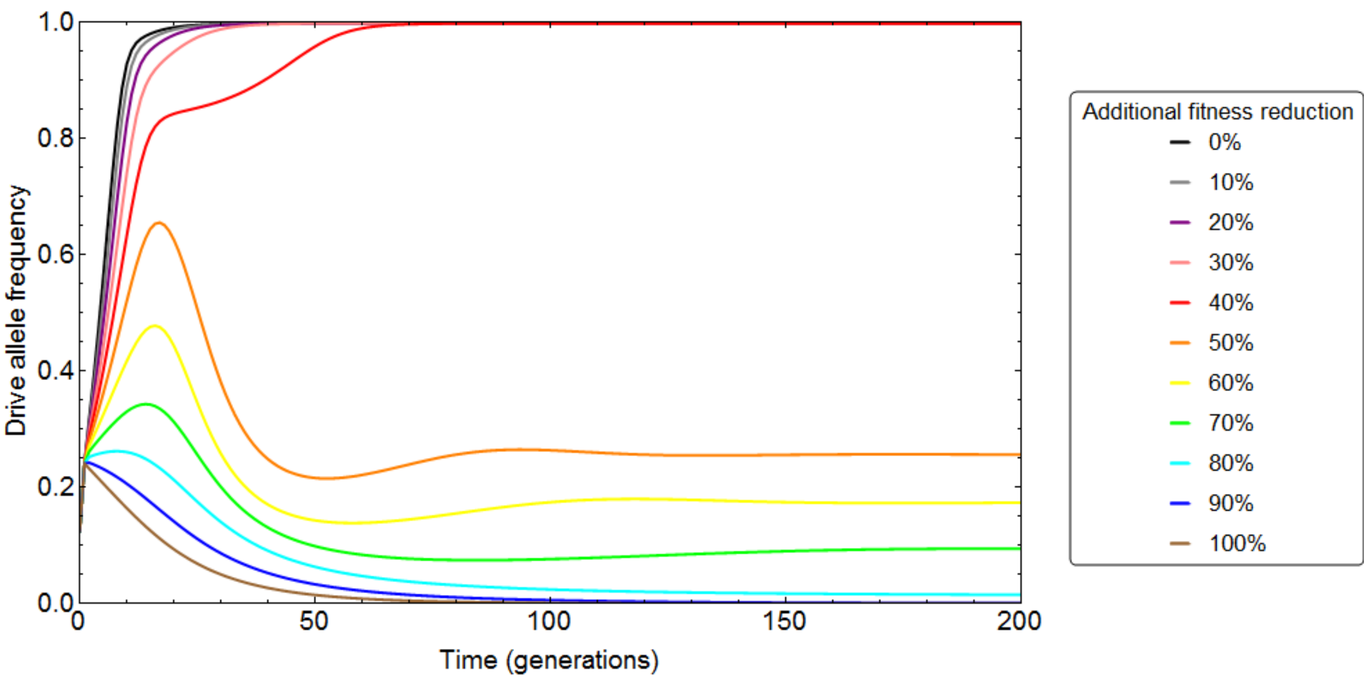
(**a**) Sequence comparison of the *dsx* intron 4-exon 5 boundary and the *dsx* female-specific exon 5 within the 16 anopheline species[16]. The sequence of the intron 4-exon 5 boundary is completely conserved within the six species that form the *Anopheles gambiae* species complex (noted in **bold**). The gRNA used to target the gene is underlined and the PAM is highlighted in blue. Changes in the DNA sequence are shaded grey and codon silent and missense substitutions are noted in blue and red respectively. (**b**) SNP frequencies obtained from 765 *Anopheles gambiae* mosquitoes captured across Africa[17]. Across the *dsx* female-specific Exon 5 there are only 2 SNP variants (noted in yellow) with frequencies of 2.9% (the SNP in the gRNA-complementary sequence) and 0.07%.



Supplementary Figure 9

***In vitro* cleavage assay testing the efficiency of the gRNA in the *dsxF*^{CRISPRh} gene drive to cleave the *dsx* exon 5 target site with the SNP found in wild populations in Africa**

An in vitro cleavage assay using an RNP complex of Cas9 enzyme and the gRNA used in this study was performed against linearised plasmids containing either wild-type (WT) target site in *dsx* exon 5 or the same site containing the single SNP found in wild caught populations (SNP). Products of the in vitro cleavage assay were purified and analysed on a gel. Both the WT and SNP-containing target sites are susceptible to the cleavage activity of the RNP complex as shown by the diminished high molecular band and the presence of the two cleavage products of the expected size. A *dsx* exon 5 target site containing the WT sequence complementary to the gRNA but without the PAM sequence was used as a control ('no PAM').



Supplementary Figure 10

Modeling the effect of unforeseen additional fitness reduction encountered by heterozygous gene drive females

Time dynamics of drive allele frequency as predicted by the deterministic model, with different coloured lines representing additional percentage reductions from zero to 100% in the baseline fertility of females, mimicking an ecologically more realistic scenario in which there were more severe fitness effects associated with the gene drive than in the laboratory. The reduction in fitness is assumed to affect the overall reproductive success (i.e mating success, longevity, fertility etc.). The baseline fertility values relative to the wild type are those reported in Supplementary Table S4 (0.65 for females with transgenic mothers, and 0.217 for females with transgenic fathers), that describes these and all other parameters estimated from experiment. Fitness reductions of up to 40% are predicted to crash the population. The spread of the drive is computed using the deterministic model in Figure 5.

SUPPLEMENTARY TABLES

Supp Table S1 Comparison of performance of gene drive and conventional genetic control approaches in terms of fitness and generation of resistance

Insect Species	Genetic modification	Intended control measure	Field/semi -field fitness	Homing rate	Resistance rate	Selection of resistance in cage	Population suppressio n	Study
Drosophila melanogaster	*tra^{nCHE}* targeting the *transformer* gene	Gene Drive	Not tested	56% males 0% females	Male germline: 28% r2, 14% r1	yes	no	KaramiNejadRanjbar et al. 2018
Drosophila melanogaster	*D-white(2-gRNA)* targeting the X-linked *white* gene, two separate gRNAs	Gene Drive	Not tested	76% males 0% females	Male germline: 23% r2, 0% r1; Embryonic 77% r2,0% r1	Not tested	Not tested (not designed for suppression)	Champer et al. 2018
Drosophila melanogaster	*D-cinnabar* targeting the *cinnabar* gene	Gene Drive	Not tested	38% males; 54% females	Male germline: 62% r2, 0% r1 Female germline: 46% r2, 0%r1 Embryonic: 100% r2	Not tested	Not tested (not designed for suppression)	Champer et al. 2018
Anopheles stephensi	*AsMCRkh2* targeting the *kynurenine hydroxylase* carrying a single chain antibody	Gene Drive	Not tested	97% males 99%females	Germline: not determined; embryonic: detected due to maternal deposition	Not tested	Not tested (not designed for suppression)	Gantz et al. 2015

Anopheles gambiae	*vasa-CRISPR*^{h} targeting the autosomal gene *nudel* (*AGAP007280*)	Gene Drive	Not tested	99% males 95% females	males: 0.28% r2, 0.14% r1 embryonic due to maternal deposition: 80%	yes	no	Hammond et al. 2016; Hammond et al. 2017
Anopheles gambiae	*zpg-CRISPR*^{h} targeting female *dsx* exon 5 (*AGAP004050*)	Gene Drive	Not tested	92% males 99% females	*male: 4.6% r2, 0% r1	no	yes	This study
Aedes aegypti	OX513A, a construct causing dominant lethality	RIDL SIT	0.56	n.a	n.a	n.a	n.a	AF Harris et al. 2011
Aedes aegypti	OX3604C a construct causing a female-specific flightless phenotype	RIDL SIT	0.03	n.a	n.a	n.a	n.a	Facchinelli et al. 2013

*Among the rare offspring of males that did not contain the drive allele (8 %) we sequenced 27 individuals, 12 of which had the wild-type allele and 15 of which had a putative non-functional resistant (r2) allele - either an out of frame 11bp deletion consistent with microhomology-mediated end joining, or a partial homing event.

Supp Table S2 | Ratio of larvae recovered by intercrossing heterozygous *dsx* ϕC31-knock-in mosquitoes			
GFP strong (*dsxF*$^{-/-}$)	**GFP weak (*dsxF*$^{-/+}$)**	**no GFP (*+/+*)**	**Total**
262 (24.9%)	**523 (49.7%)**	**268 (25.5%)**	**1053**

Supp Table S2 | Heterozygous and homozygous individuals for the *dsxF*$^{-}$ allele were separated based on the intensity of fluorescence afforded by the GFP transcription unit within the knockout allele. Homozygous mutants were distinguishable as recovered in the expected Mendelian ratio of 1:2:1 suggesting that the disruption of the female-specific isoform of *Agdsx* is not lethal at the L1 larval stage.

Nature Biotechnology: doi:10.1038/nbt.4245

Supp Table S3 | Genetic females homozygous for the insertion carry male-specific characteristics						
	Genetic Males			**Genetic Females**		
Characteristic	***dsxF$^{+/+}$***	***dsxF$^{+/-}$***	***dsxF$^{-/-}$***	***dsxF$^{+/+}$***	***dsxF$^{+/-}$***	***dsxF$^{-/-}$***
Pupal genital lobe	male	male	male	female	female	male
Claspers	✓	✓	✓	✗	✗	✓
Cercus	✗	✗	✗	✓	✓	✗
Spermatheca	✗	✗	✗	✓	✓	✗
MAGs	✓	✓	✓	✗	✗	✓
Feed on blood	✗	✗	✗	✓	✓	✗
Can lay eggs	✗	✗	✗	✓	✓	✗
Plumose antennae	✓	✓	✓	✗	✗	✓
Pilose antennae	✗	✗	✗	✓	✓	✗

Supp Table S3 Phenotypic characteristics observed on adult mosquitoes taken from the *dsxF^{-}* crosses. Female mosquitoes of the *dsxF$^{-/-}$* class present a profile of characteristics that matches the male sex rather than the female.

Supp Table S4 | Primers used in this study	
dsxgRNA-F	TGCTGTTTAACACAGGTCAAGCGG
dsxgRNA-R	AAACCCGCTTGACCTGTGTTAAAC
dsxϕ31L-F	GCTCGAATTAACCATTGTGGACCGGTCTTGTGTTTAGCAGGCAGGGGA
dsxϕ31L-R	CACCAAGACAGTTAACGTATCCGTTACCTTGACCTGTGTTAAACATAAAT
dsxϕ31R-F	GGTGGTAGTGCCACACAGAGAGCTTCGCGGTGGTCAACGAATACTCACG
dsxϕ31R-R	TCCACCTCACCCATGGGACCCACGCGTGGTGCGGGTCACCGAGATGTTC
zpgprCRISPR-F	GCTCGAATTAACCATTGTGGACCGGTCAGCGCTGGCGGTGGGGA
zpgprCRISPR-R	TCGTGGTCCTTATAGTCCATCTCGAGCTCGATGCTGTATTTGTTGT
zpgteCRISPR-F	AGGCAAAAAAGAAAAAGTAATTAATTAAGAGGACGGCGAGAAGTAATCAT
zpgteCRISPR-R	TTCAAGCGCACGCATACAAAGGCGCGCCTCGCATAATGAACGAACCAAAGG
dsxin3-F	GGCCCTTCAACCCGAAGAAT
GFP-F	GCCCTGAGCAAAGACCCCAA
dsxex4-F	GCACACCAGCGGATCGACGAAG
dsxex5-R	CCCACATACAAAGATACGGACAG
dsxex6-R	GAATTTGGTGTCAAGGTTCAGG
3xP3	TATACTCCGGCGGTCGAGGGTT
hCas9-F	CCAAGAGAGTGATCCTGGCCGA
dsxex5-R1	CTTATCGGCATCAGTTGCGCAC

dsxin4-F	GGTGTTATGCCACGTTCACTGA
RFP-R	CAAGTGGGAGCGCGTGATGAAC
† Dsx-original-target-F	TAGGGTTTAACACAGGTCAAGCGGTGG
† Dsx-original-target-R	AAACCCACCGCTTGACCTGTGTTAAAC
† Dsx-SNP-target-F	TAGGGTTTAACACAGGTCAAGCAGTGG
† Dsx-SNP-target-R	AAACCCACTGCTTGACCTGTGTTAAAC
† Dsx-noPAM-target-F	TAGGTTTAACACAGGTCAAGCGG
† Dsx-noPAM-target-R	AAACCCGCTTGACCTGTGTTAAA

Supp Table S4 Table listing the primers used in this study. Gibson assembly and Golden Gate cloning overhangs are underlined with a single and a double line respectively. † Primers used to create the target sequences for the *in vitro* RNP cleavage assay.

Supp Table S5 | Parameters for stochastic cage model		
Parameter	**Estimate**	**Method of estimation**
Mating probability	0.85 for heterozygotes; 0 for D/D, D/R and R/R homozygotes	Estimated from Hammond et al. 2017
Egg production from wild-type female (no parental nuclease)	Mean 137.4. Sampling with replacement of observed values (10, 61, 96, 98, 111, 111, 113, 127, 128, 129, 132, 132, 134, 135, 137, 138, 138, 139, 142, 142, 146, 146, 149, 152, 152, 152, 158, 160, 162, 164, 170, 179, 186, 189, 191)	From assays of mated females
Egg production from W/D heterozygote female (nuclease from ♀)	Mean 118.96. Sampling with replacement of observed values (12, 31, 76, 90, 96, 100, 106, 106, 107, 113, 117, 118, 119, 130, 133, 136, 136, 136, 137, 138, 139, 142, 143, 145, 146, 148, 157, 174)	From assays of mated females
Egg production from W/D heterozygote female (nuclease from ♂)	Mean 59.67. Sampling with replacement of observed values (0, 0, 0, 0, 0, 34, 47, 50, 65, 105, 113, 115, 115, 125, 126)	From assays of mated females
Hatching probability, wild-type female (no parental nuclease)	0.941	From assays of mated females
Hatching probability, W/D heterozygote female (nuclease from ♀)	0.707	From assays of mated females
Hatching probability,	0.47	From assays of mated females

W/D heterozygote female (nuclease from ♂)		
Probability of emergence from pupa (survival from larva)	0.8708	Average of observations over all generations and both cage experiments
Drive in W/D females	0.9985	Observed fraction transgenic from assays
Drive in W/D males	0.9635	Observed fraction transgenic from assays
Meiotic EJ parameter (fraction non-drive alleles that are resistant)	0.4685	Estimated from Hammond et al. 2016

Supp Table S5 **|** Parameters for stochastic cage model

We assume that parental effects on fitness (egg production and hatching rates) for non-drive (W/W, W/R) females with nuclease from one or both parents are the same as observed values for drive heterozygote (W/D) females with parental effects. For combined maternal and paternal effects (nuclease from both parents), the minimum of the observed values for maternal and paternal effect is assumed.

Supp Table S6 | Summary of values obtained from the cage trials								
	Cage Trial 1				Cage Trial 2			
Generation	Transgenic Rate (%)	Hatching Rate (%)	Egg Output (N)	Repr. Load (%)	Transgenic Rate (%)	Hatching Rate (%)	Egg Output (N)	Repr. Load (%)
G0	25 (150/600)	-	27462	-	25 (150/600)	-	26895	-
G1	49.65 (286/576)	88.62 (576/650)	17405	36.62	50 (280/560)	86.15 (560/650)	16578	38.36
G2	62.01 (302/487)	74.92 (487/650)	14957	45.54	61.79 (325/526)	80.92 (526/650)	15565	42.13
G3	68.94 (344/499)	76.77 (499/650)	11249	59.04	68.05 (328/482)	74.15 (482/650)	9376	65.14
G4	67.67 (316/467)	71.85 (467/650)	9170	66.61	85.41 (398/466)	71.69 (466/650)	6514	75.78
G5	58.67 (264/450)	69.23 (450/650)	11364	58.62	86.5 (346/400)	61.54 (400/650)	4805	82.13
G6	63.3 (288/455)	70 (455/650)	7727	71.86	90.09 (309/343)	52.77 (343/650)	4210	84.35
G7	69.47 (355/511)	78.62 (511/650)	7785	71.65	100 (363/363)	55.85 (363/650)	1668	93.8
G8	70.07 (323/461)	70.92 (461/650)	6293	77.08	100 (278/278)	42.77 (278/650)	0	100
G9	75.58 (325/430)	66.15 (430/650)	4107	85.04	-	-	-	-
G10	95.71 (357/373)	57.38 (373/650)	4146	84.90				
G11	100 (374/374)	57.54 (374/650)	2645	90.37				
G12	100 (253/253)	38.92 (253/650)	0	100				

Supp Table S6 | Summary of values obtained from the cage trials

Transgenic rate, hatching rate, egg output and reproductive load at each generation during the cage experiment. The reproductive load indicates the suppression of egg production at each generation compared to the first generation.

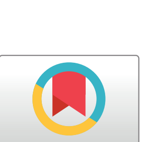
RESEARCH ARTICLE

The creation and selection of mutations resistant to a gene drive over multiple generations in the malaria mosquito

Andrew M. Hammond[1☯], Kyros Kyrou[1☯], Marco Bruttini[2], Ace North[3], Roberto Galizi[1], Xenia Karlsson[1], Nace Kranjc[1], Francesco M. Carpi[1], Romina D'Aurizio[4], Andrea Crisanti[1]*, Tony Nolan[1]*

1 Dept. of Life Sciences, Imperial College London, London, United Kingdom, **2** Polo d'Innovazione Genomica, Genetica e Biologia, Siena, Italy, **3** Department of Zoology, University of Oxford, Oxford, England, **4** Laboratory of Integrative Systems Medicine (LISM), Institute of Informatics and Telematics (IIT), CNR, Pisa, Italy

☯ These authors contributed equally to this work.
* acrs@imperial.ac.uk (AC); t.nolan@imperial.ac.uk (TN)



Citation: Hammond AM, Kyrou K, Bruttini M, North A, Galizi R, Karlsson X, et al. (2017) The creation and selection of mutations resistant to a gene drive over multiple generations in the malaria mosquito. PLoS Genet 13(10): e1007039. https://doi.org/10.1371/journal.pgen.1007039

Editor: Nick H. Barton, Institute of Science and Technology Austria (IST Austria), AUSTRIA

Received: August 8, 2017

Accepted: September 20, 2017

Published: October 4, 2017

Copyright: © 2017 Hammond et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files, except for sequencing data which is deposited on the NCBI Sequence Read Archive (SRA) under project accession number PRJNA397539.

Funding: This work was supported by a grant from the Foundation for the National Institutes of Health through the Vector-Based Control of Transmission: Discovery Research (VCTR) program of the Grand Challenges in Global Health initiative of the Bill & Melinda Gates Foundation. The funders had no role

Abstract

Gene drives have enormous potential for the control of insect populations of medical and agricultural relevance. By preferentially biasing their own inheritance, gene drives can rapidly introduce genetic traits even if these confer a negative fitness effect on the population. We have recently developed gene drives based on CRISPR nuclease constructs that are designed to disrupt key genes essential for female fertility in the malaria mosquito. The construct copies itself and the associated genetic disruption from one homologous chromosome to another during gamete formation, a process called homing that ensures the majority of offspring inherit the drive. Such drives have the potential to cause long-lasting, sustainable population suppression, though they are also expected to impose a large selection pressure for resistance in the mosquito. One of these population suppression gene drives showed rapid invasion of a caged population over 4 generations, establishing proof of principle for this technology. In order to assess the potential for the emergence of resistance to the gene drive in this population we allowed it to run for 25 generations and monitored the frequency of the gene drive over time. Following the initial increase of the gene drive we observed a gradual decrease in its frequency that was accompanied by the spread of small, nuclease-induced mutations at the target gene that are resistant to further cleavage and restore its functionality. Such mutations showed rates of increase consistent with positive selection in the face of the gene drive. Our findings represent the first documented example of selection for resistance to a synthetic gene drive and lead to important design recommendations and considerations in order to mitigate for resistance in future gene drive applications.

in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

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

Author summary

Gene drives are selfish genetic elements that are able to bias their own inheritance among offspring. Starting from very low frequencies they can rapidly invade a population in just a few generations, even when imposing a fitness cost. Gene drives based on the precise DNA cutting enzyme CRISPR have been shown recently to be highly efficient at copying themselves from one chromosome to the other during the process of gamete formation in mosquitoes, resulting in transmission to 99% of offspring instead of the 50% expected for a single gene copy. One proposed use for CRISPR-based gene drives is in the control of mosquitoes by designing the gene drive to target mosquito genes involved in fertility, thereby reducing their overall reproductive output and leading to population suppression. Like any intervention designed to suppress a population these gene drives are expected to select for mutations in the mosquito that are resistant to the drive and restore fertility to mosquitoes. We have analyzed the origin and selection of resistant alleles in caged populations of mosquitoes initiated with a gene drive construct targeting a female fertility gene. We find the selected alleles are in-frame insertions and deletions that are resistant to cleavage and restore female fertility. Our findings allow us to improve predictions on gene drive behaviour and to make concrete recommendations on how to improve future gene drive designs by decreasing the likelihood that they generate resistance.

Introduction

Naturally occurring gene drives—selfish genetic elements that are able to bias their own inheritance and rapidly invade a population, even starting from very low frequencies—have inspired proposals to harness their power to spread into a population of insect disease vectors traits that manipulate their biology in ways that could suppress or eliminate disease transmission [1–4]. In particular for malaria, transmitted exclusively by mosquitoes of the *Anopheles* genus, historical gains in reducing the disease burden have been largely achieved by the correct implementation of vector control measures (residual insecticides and bed nets) [5]. Though these measures have been instrumental in substantially reducing malaria transmission, they are insufficient by themselves to eradicate the disease in the near future at the current level of investment [6]. Gene drive technology could help in developing a self-sustaining, species-specific and affordable vector control measure much needed to achieve disease eradication in the future.

Gene drives based on the activity of DNA nucleases able to recognise specific target sequences were first proposed over a decade ago and have received much attention recently due to the advent of new, easily programmable nucleases such as CRISPR-Cas9 that have allowed us and others to build functioning gene drives that show rates of inheritance from a heterozygous parent close to 100%, compared to the expected Mendelian inheritance of 50% [1, 7–9]. The principle behind the technology is to re-program a nuclease to cleave a specific site of interest in the genome and to insert the nuclease within this recognition site. The gene drive is designed to be active in the germline, so that in diploid organisms heterozygous for the gene drive the nuclease causes a double stranded break (DSB) at the target site on the homologous chromosome not containing the gene drive. The DSB can be repaired either by simple end-joining (EJ) of the broken strands or via homology-directed repair (HDR) where the DSB is resected and the intact chromosome used as a template to synthesise the intervening sequence. In the case of a gene drive, repair via HDR thus leads to a copying of the drive from one chromosome to another and the conversion of a heterozygote into a homozygote. Hence

the force of gene drive is determined by a combination of the rate of cleavage of the nuclease in the germline, and the propensity for the cell machinery to repair the broken chromosome by HDR.

We and others have shown that in germline cells the rates of HDR following a nuclease-induced DSB can be almost two orders of magnitude greater than EJ, a fact which explains the extraordinarily high rates of gene drive inheritance observed [7, 8, 10, 11]. On the other hand EJ repair can lead to the creation of small insertions or deletions at the target site that, although occurring initially at low frequency, might be expected to be selected for in the target organism if they prevent the gene drive nuclease acting and there is a negative fitness cost associated with the gene drive [1, 7, 11–13]. This possibility has been recognised since the first proposal of this type of gene drive [1], with much theory being dedicated to it recently [13, 14] and recent empirical evidence of its occurrence in *Drosophila*[10]. To lower the likelihood of resistance arising there are several potential mitigation strategies including, but not limited to, the targeting of conserved sequences that are less tolerant of mutations and the targeting of multiple sequences, akin to combination therapy [1, 12].

We previously developed a gene drive designed to spread into a mosquito population and at the same time reduce its reproductive potential by disrupting a gene essential for female fertility, thus imposing a strong fitness load on the population [11]. To investigate the long term dynamics of the emergence of resistance to a gene drive imposing such a load we continued to monitor the frequency of this gene drive over generations and analysed the target locus for evidence of mutagenic activity that could lead to the development of resistant alleles that block gene drive activity and restore gene functionality. Our findings show that a range of different resistant alleles can be generated and some of these are subsequently selected for and show dynamics consistent with our modelling predictions. These results provide a quantitative framework for understanding the dynamics of resistance in a multi-generational setting and allow us to make recommendations for the improvement of future gene drive constructs that relate to choice of target site and regulation of nuclease expression in order to retard the emergence of resistance.

Results

Long term dynamics of spread of a population suppression gene drive in a caged mosquito population

A proof-of-principle CRISPR-based gene drive designed for population suppression was previously developed in our laboratory (Fig 1A). This gene drive disrupted a haplosufficient gene (AGAP007280, the putative mosquito ortholog of *nudel* [15]) required in the soma and essential for female fertility [11]. The gene drive also contained an RFP marker gene for the visual detection of individuals inheriting the drive. In our experiments individuals heterozygous for the gene drive transmitted the drive, regardless of their sex, to more than 99% of their offspring. We observed in these mosquitoes a marked reduction in fertility (~90%) in females heterozygous for the drive, due to ectopic expression of the nuclease under control of the germline *vasa2* promoter that resulted in conversion to the null phenotype in somatic cells. In spite of this fitness disadvantage experimental data showed that the gene drive could increase rapidly in frequency in a caged population due to the exceptionally high rates of inheritance bias. From a starting population (G_0) in two duplicate cages of 600 individuals with a 1:1 ratio of transgenic heterozygotes and wild type individuals, the gene drive progressively increased in frequency to 72–77% by G_4. This rate of increase was slightly higher than predicted by a deterministic model but within the limits of stochastic variation expected [11]. Due to a combination of the partial dominance of the sterility phenotype in heterozygous females and the

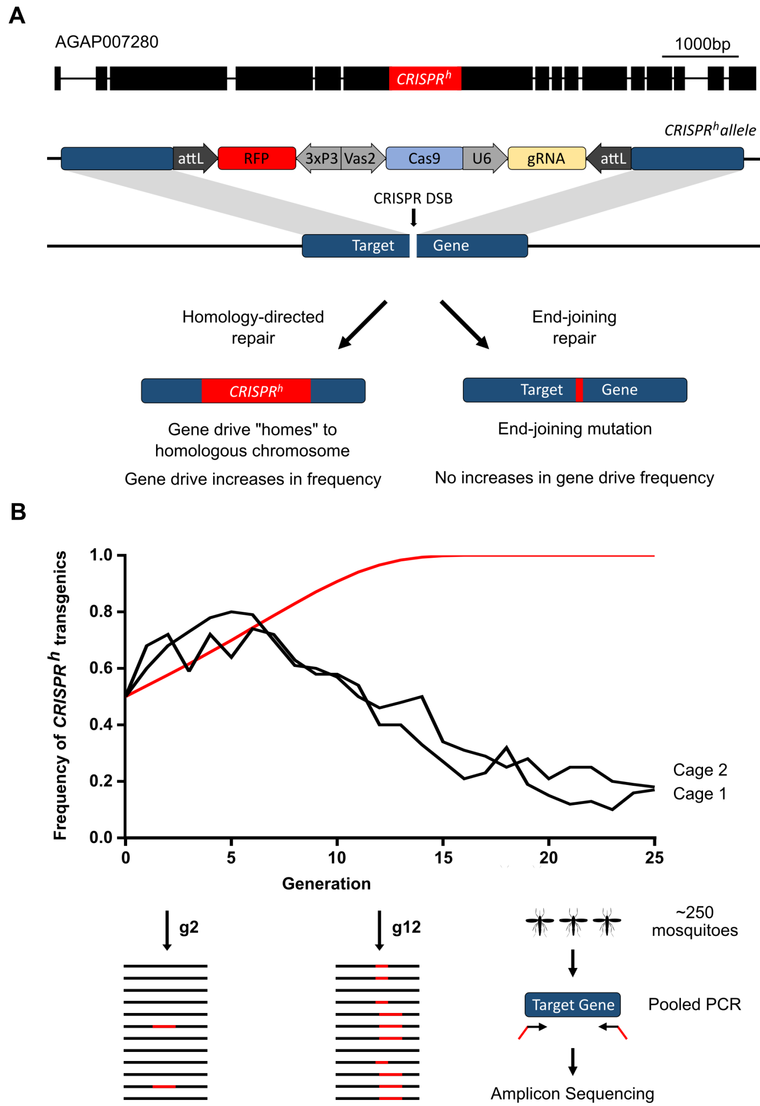


Fig 1. Dynamics of a population suppression gene drive construct over 25 generations. (A) Design of the CRISPR-based gene drive construct and the relevant position of its target site within AGAP007280; **(B)** The proportion of individuals containing at least one copy of the gene drive in two replicate cages, monitored each generation for 25 generations. Black lines represent the observed frequencies in each of the two cage trials (CT1 and CT2), red line represents the predicted frequency according to the previous deterministic model that did not take into account target site resistance [11]. Samples were taken for pooled PCR and sequencing analysis of the gene drive target site at G_2 and at G_{12}, whereby frequency of target site indels (shown figuratively as red bars) were revealed by their relative representation among the sequenced reads.

https://doi.org/10.1371/journal.pgen.1007039.g001

previously documented generation of target site mutations conferring resistance to the gene drive [11], this first gene drive was not expected to maintain high levels of invasion. Nonetheless it represented a useful experimental model to investigate the long term dynamics of the *de novo* generation of target site mutations and their selection at the expense of a gene drive imposing a large reproductive load. We therefore maintained this cage experiment for 25 generations and used the presence of the RFP marker in the gene drive construct as a proxy to estimate the frequency of individuals containing it. The frequency of gene drive progressively

increased in both cages, peaking at around generation 6, and thereafter we observed a gradual and continuing decrease such that by G_{25} the frequency of individuals with the gene drive was less than 20%.

Sequencing the gene drive target site revealed nuclease-induced mutations

To investigate whether the gradual decline in the gene drive frequency observed in the cage experiment was due to the selection of pre-existing variant target sites in the population or the generation and selection of nuclease-resistant indels, we used deep sequencing of a PCR amplicon comprising sequences flanking the target site on pooled samples of mosquitoes from early (G_2) and late (G_{12}) generational time points (Fig 1B). The expected amplified region from the original wild type sequence was 320bp long, with the putative cleavage point within the target site residing after nucleotide 208 (Fig 2A). Ultra-deep sequencing of PCR reactions were performed on pooled DNA under non-saturating conditions so that the number of reads corresponding to a particular allele at the target site is proportional to its representation in the pool. We developed a computational method to analyse the sequences edited by the CRISPR-based gene drive close to the nuclease target site. In the colony of mosquitoes that we used there are a number of pre-existing single nucleotide polymorphisms (SNPs) within the amplicon that do not overlap the nuclease target site and are present at varying frequencies (S1 Fig). Our method identified small insertions and deletions introduced by the repair system and used the presence of surrounding SNPs to characterize the haplotypes on which they arose. Because the PCR only amplifies the non-drive allele, the frequencies reported below refer to their frequency within this class, rather than within the population as a whole.

Mapping amplicon sequences reconstructed from the sequenced against the *Anopheles gambiae* reference genome (PEST strain, AgamP4, Vectorbase) we observed a large repertoire of deletions already in the G_2 generation, with a wide range of sizes and centred around the predicted nuclease cleavage site after nucleotide 208 (Fig 2A—one cage trial shown as a representative example), and a lower proportion of small insertions, consistent with the known mutational activity of the nuclease. By contrast, ten generations later we observed a much reduced diversity of indels. We then considered all alleles that reached a frequency of at least 1% in any sample, classified these as to whether the indel caused a frameshift in the coding sequence of the target gene or was in-frame, and analysed their frequency over time (Fig 2B). The predominant target allele in the G_2 was still the reference (non-mutated) allele at 63% and 48% in cages 1 and 2, respectively (Fig 2B and S1 Table), while the second major class (at least 15% in each replicate) was represented by a wide range of non-reference alleles, each present at low frequency (<1%), consistent with the stochastic generation of a broad range of indels. Thus at a time when the gene drive was still increasing in frequency there was a significant accumulation of mutations at the target site that would likely render it refractory to the homing mechanism of copying. Of note, three separate indels causing in-frame deletions of 3- or 6bp (202-TGAGGA, 203-GAGGAG, 203-GAG; where 203 refers to the starting site of the indel in the reference amplicon and "-"means deletion) were present among a large number of indels at low but appreciable frequencies in the G_2. Such short in-frame deletions may result in only minimal disruption to the final encoded protein while at the same time proving resistant to the gene drive. Indeed these three deletions, plus a 6bp in frame insertion (207+AAAGTC), had increased significantly in frequency to make up the 4 most abundant non-drive alleles in the G_{12}, almost to the exclusion of the reference allele (present at 6% and 0.4%; S1 Table). At the same time, a wide range of frameshift indels that were present in G_2 had fallen in frequency in G_{12} to either below the 1% threshold or were not detected at all (Fig 2B and S1 Table). The

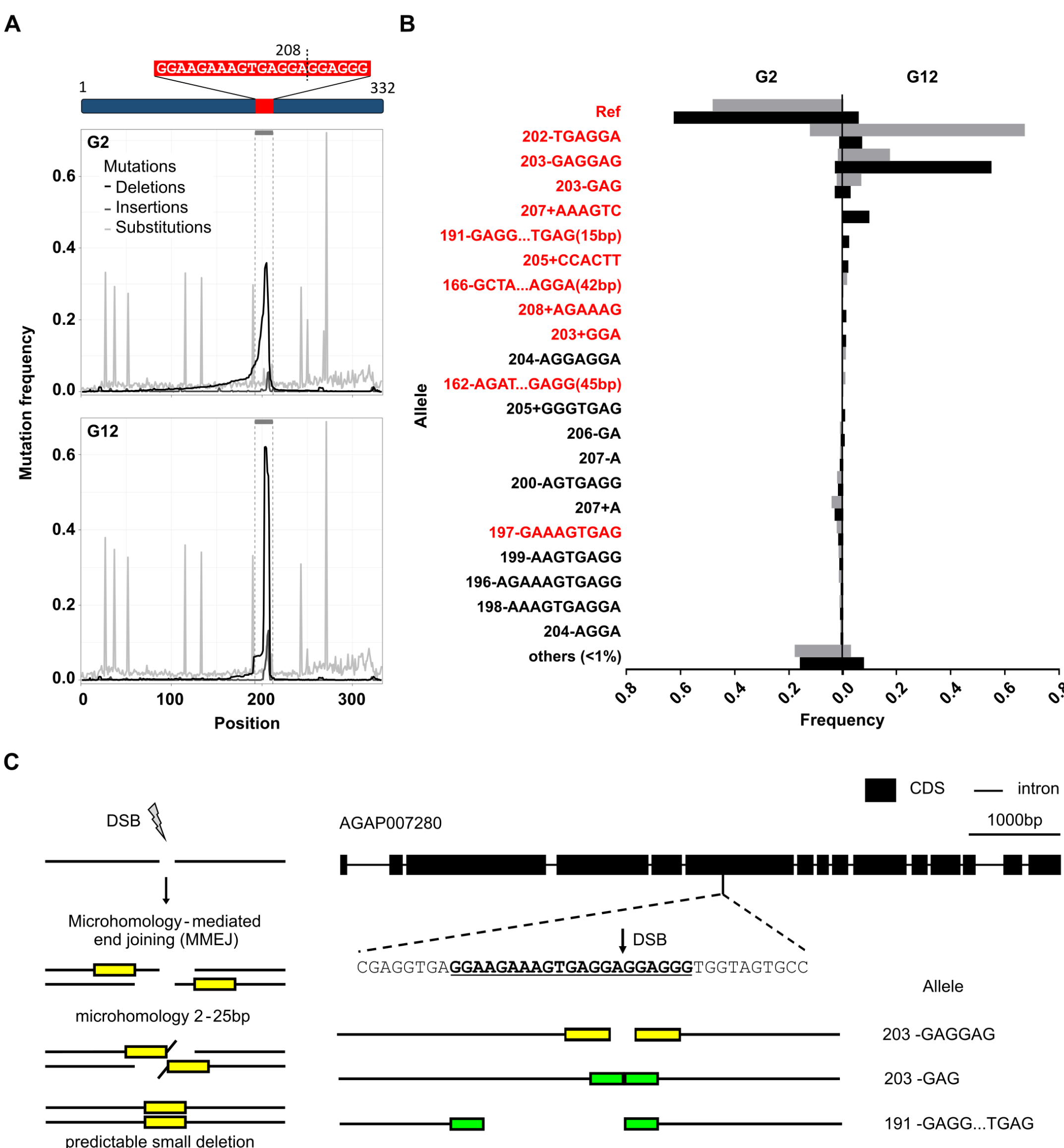


Fig 2. Nuclease-generated target site mutations show selection over time. (A) Location of the target site of the CRISPR-based gene drive relative to the region amplified for pooled sequencing (highlighted in red), with the double stranded break occurring after nucleotide 208. Cumulative frequencies of deletions (black), insertions (dark grey) and substitutions (light grey) at each nucleotide position along the amplicon were plotted. Numerous substitutions outside of the target site represent SNPs circulating in the lab strain of mosquito in which the gene drive was introduced and were used for analysis of haplotypes of target site mutations (S1 Fig). One cage trial (CT1) shown as a representative example. **(B)** Frequency of individual target site alleles in both G_2 and G_{12} for cage 1 (black) and cage 2 (grey) with in-frame indels highlighted in red. All alleles never exceeding 1% frequency in at least one condition were grouped together as one class (<1%). **(C)** Microhomologies flanking the double stranded break at the target could explain some of the most frequent deletions observed. Microhomologies of GAGG (yellow) or GGA (green) and the resultant deletion alleles arising in the event of MMEJ are shown.

https://doi.org/10.1371/journal.pgen.1007039.g002

most parsimonious explanation for these results is that a wide range of frameshift and in-frame indels was created by the gene drive, yet only short in-frame indels were selected for because they restore functionality to the target gene while protecting the sequence from gene drive activity. These ‘restorative’ mutations are likely to be most strongly selected when the

frequency of the gene drive is high in the population—when the majority of individuals are homozygous for the driver, the relative gain in viable offspring from an individual with a gene drive balanced by a resistant restorative mutation is that much higher.

Small in-frame deletions can arise by either classical non-homologous end-joining (NHEJ), or an alternative form of end-joining (microhomology-mediated end-joining, MMEJ) that relies on alignment of small regions of microhomology, as little as 2 base pairs, on either side of the DSB, resulting in loss of the intervening sequence [16]. Consistent with this latter possibility, at least three of the most frequent alleles in the G_{12} generation can be explained by MMEJ via 3bp repeats (Fig 2C).

To investigate whether the most common indels at G_{12} had single or multiple origins, we used the naturally occurring SNPs in the sequences flanking the recognition sequence. In cage 1 the deletion 203-GAGGAG was present on 10 separate haplotypes in G_{12}, with each haplotype being present at ratios broadly similar to their ratios in the starting population (S1 Fig), suggesting that the same deletion was generated at least 10 times independently and that there was no detectable selective advantage to any particular haplotype surrounding the deletion. In cage 2 the predominant allele was 202-TGAGGA (68% of all non-reference alleles), an end-joining deletion that shows no apparent features of a MMEJ event, and was found on 5 separate haplotypes.

Positively selected target site mutations restore function to the target gene and are resistant to gene drive activity

The progressive increase in frequency of specific mutations at the target site, concomitant with a decrease of gene drive activity, strongly suggested that they conferred resistance to cleavage while still ensuring a normal functional activity of *nudel*. To confirm this hypothesis we crossed individual RFP+ females from G_{20} with wild type males and assessed both their fertility and the transmission rate of the drive. We also sequenced the target site of each parental female to characterize allelic variants at the target locus. This analysis failed to detect wild type sequence at the target site (Fig 3A) among 70 individuals tested; instead every individual showed an indel, indicating that each female tested was heterozygous for the gene drive and balanced by a mutated target site. In cage 1 the 203-GAGGAG 6bp deletion was the predominant allele (23/31 individuals) while in cage 2 another 6bp deletion (202-TGAGGA) was predominant (37/39 individuals). The relative frequency of each allele was consistent with the results obtained using pooled amplicon sequencing performed on the G_{12} individuals.

Of those heterozygous females that could be confirmed as having mated, the vast majority (56/58) generated viable progeny (average clutch size 119 +/- 35.9 eggs, average hatching rate 78.5% +/- 19.9%; Fig 3B and S2 Table) at rates significantly higher than those previously observed in females heterozygous for the gene drive and a wild type allele (90.7% overall reduction in fecundity) [11], suggesting that the mutations detected at the target site substantially restored the functionality of *nudel*. There was no obvious difference in fertility associated with the 5 indels sampled in this assay (Fig 3B). The two major indels across the two cages result in relatively conservative changes to the overall amino acid sequence of the final gene product—two glutamate residues missing (203-GAGGAG) or two glutamates missing and a conversion of a serine to arginine (202-TGAGGA). Finally, the offspring of these individuals showed a non-biased inheritance of the gene drive (50.82% mean +/-5.96% standard error; total RFP + offspring 50.85%), consistent with normal Mendelian segregation (Fig 3B). Thus, as well as restoring *nudel* function, the mutated sequences were also resistant to the gene drive.

A

Allele	Sequence	Frequency Total	Cage Trial 1	Cage Trial 2
WT	GGTGAGGAAGAAAGTGAGGAGGAGGGTGGTAGT	0% (0/70)	0% (0/31)	0% (0/39)
202 -TGAGGA	GGTGAGGAAGAAAG------GGAGGGTGGTAGT	52.9% (37/70)	0% (0/31)	94.9% (37/39)
203 -GAGGAG	GGCGAGGAAGAAAGT------GAGGGTGGTAGT	34.3% (24/70)	74.2% (23/31)	2.5% (1/39)
203 -GAG	GGTGAGGAAGAAAGTGAG---GACGGTGGTAGT	0.1% (1/70)	0% (0/31)	2.5% (1/39)
191 -GAGG...TGAG	GGTCGATGTCGT------GAGGAGGGTGGTAGT	0.1% (1/70)	3.2% (1/31)	0% (0/39)
207 +AAAGTC	GGTGAGGAAGAAAGTGAGGAAAGTCAGGAGGGTGGTAGT	10.0% (7/70)	22.6% (7/31)	0% (0/39)

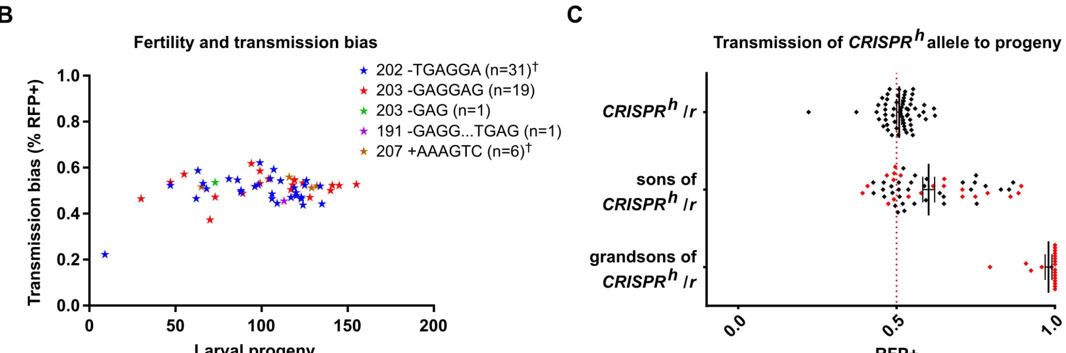


Fig 3. Target site mutations under positive selection are resistant to gene drive activity and restore function to the target female fertility gene. (A) Individual females containing at least one copy of the gene drive (RFP+) were selected from the G_{20} generation and the nature of the target allele was determined by PCR and sequencing. Each class of allele is shown with gene drive target sequence highlighted in red and PAM sequence underlined. **(B)** The fecundity of these females and transmission rates of the gene drive were measured and grouped according to allele class at the target site. **(C)** Each *CRISPR*^{h}/*r* female was used to form a separate lineage and transmission of the gene drive was assessed in sons receiving a maternal copy of the gene drive. A smaller fraction of grandsons receiving a paternal copy of the gene drive were similarly assessed for gene drive transmission. Individual lineages assessed in all three generations are marked in red. † Of 58 mated females one (with deletion 207+AAAGTC) failed to produce eggs while another (202–TGAGGA) produced eggs that failed to hatch.

https://doi.org/10.1371/journal.pgen.1007039.g003

End-joining repair in the embryo by maternally deposited Cas9 nuclease dramatically increases the generation rate of drive resistant alleles

Conceivably a breakdown in the nuclease component (e.g. mutation in the Cas9 coding sequence, or the gRNA sequence) could be an additional explanation for the Mendelian transmission of the gene drive element and restored fertility in heterozygous females that we observed. To assess this possibility we took the male offspring ('sons') of the above crosses that inherited the construct and crossed them in turn to wild type females. We assumed that if the gene drive construct was still functional it should show a biased inheritance when the resistant target site allele had been replaced with a wild type one. Indeed, in these sons we saw a significant increase in the transmission of the gene drive to their progeny, but the observed rate (mean 60.13% +/- 13.9% S.E.; total RFP+ progeny 59.6%) was much lower than that previously observed (~99% inheritance) [11]. A similar phenomenon of reduced homing has been observed in the offspring of another mosquito species [8], and more recently in *Drosophila* [10], when the drive construct was inherited from the mother and when the same *vasa* germline promoter was used to transcribe the Cas9 nuclease. The reduced gene drive activity in the immediate offspring of heterozygous mothers was attributed to the persistence of maternally-deposited Cas9 in fertilized embryos, leading to double stranded DNA breaks being repaired preferentially by end-joining mechanisms in the early zygote possibly before paternal and maternal homologous chromosomes are aligned. Consistent with this explanation, males in

the subsequent generation ('grandsons') that had received only a paternal copy of the gene drive had exceptionally high homing rates, with 97.5% of progeny inheriting the gene drive (Fig 3C). The drop in homing seen in sons receiving a maternal dose of Cas9 (59.6% transmission vs. 97.5% in grandsons) allows us to estimate an 'embryonic end-joining' rate of 79.6% of wild type alleles being converted to cleavage-resistant alleles. This rate of embryonic end-joining is much higher than that observed in the germline at or just prior to meiosis (~1% [11]) and is predicted to reduce the rate of spread of the gene drive, due to a reduced frequency of cleavable alleles [17], and increase the rate at which restorative resistant alleles can arise and be selected.

Cleavage due to maternally deposited Cas9 could potentially be followed by HDR instead of end-joining, effectively leading to 'embryonic homing', where the cleaved allele is converted to the non-cleaved allele. In the case of a resistant allele this could lead to an individual heterozygous for the allele being converted into a homozygote in the early zygote, thereby further accelerating the spread of the resistant allele in the population. One signature of embryonic homing of a resistant allele would be novel hybrid haplotypes due to partial conversion of the haplotype surrounding the wild type allele where the DSB was generated to the haplotype surrounding the resistant allele. Looking in detail at the most abundant resistant allele in each cage, we failed to observe such a signature, and all resistant haplotypes were already pre-existing in the population (S1 Fig), suggesting that if this phenomenon is occurring then the resection following cleavage and resultant conversion encompasses a section longer than the ~300bp covered in our sequenced amplicon.

Incorporating experimental rates of resistance generation into a population model

The key qualitative results from the cage experiments—that a gene drive can increase in frequency in a susceptible population even if it reduces individual fitness, and that the spread of a gene drive can in turn lead to the spread of mutants that are resistant to cleavage and restore individual fitness—are fully consistent with expectations from population genetic models [13, 18, 19]. To investigate how well such models can account for the quantitative details of the cage experiments, we extended the model of Deredec *et al.* [18] to incorporate our observations of embryonic cleavage by maternally derived nuclease (which is independent of inheritance of the gene drive), and have two classes of resistant allele (in-frame functional and frame-shift non-functional; see S1 Text and S1 File). Due to the sex-specific fitness effects of our construct, the model also has a separate treatment of females and males. Using this model and the baseline parameter values from the single-generation crosses, we generated the expected allele frequency dynamics over the 25 generations of the experiment (Fig 4). Again, the qualitative fit to the observed dynamics is good, but there are quantitative differences. For example, the model predicts that at G_{12} the original wild type allele will be 9.3% of all non-drive alleles, while our observed rates were 6% and 0.4% in cages 1 and 2, respectively (Fig 2B, S1 Table). The model also recapitulates the observation that while non-functional resistant alleles initially outnumber the functional ones, because they are produced more frequently, by the end of the experiment it is the functional ones that predominate. Importantly, for the gene drive itself, the model captures the essential aspects of the observed dynamics, showing an initial increase in frequency followed by an eventual loss, though in earlier generations our observed frequencies exceeded the predicted frequencies (Fig 4B). To investigate what might explain this discrepancy, we examined the effect of varying each of the different parameter values individually in the model, and found that small variations in the fertility of females heterozygous for the gene drive had the largest effect in increasing the match between observations

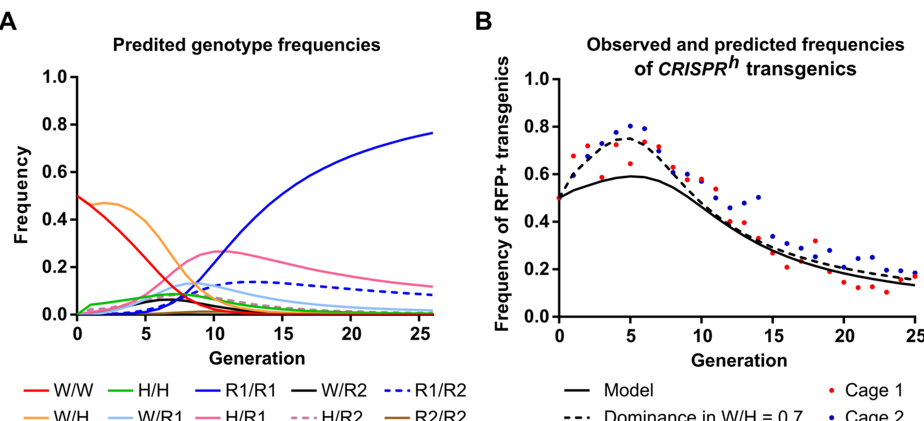


Fig 4. Comparison of observed data with model predicting frequencies of gene drive and resistance alleles. (A) Expected genotype frequencies according to the model described in the text and considering the four following target site alleles: wild type (w), *CRISPR*^{h} gene drive (h), resistant and in-frame (r1), resistant and out of frame (r2). We used our best experimental estimates of the considered parameters: homing rate (*e*) as 0.984, the dominance of the fertility effect due to leaky somatic expression in females heterozygous (*w/h*) for the gene drive as 0.907, meiotic end-joining rate (γ_m) as 0.01, embryonic end-joining rate (γ_e) as 0.796. **(B)** Our observed gene drive frequencies were compared against model predictions using our best experimental estimates (solid black line) and using the best-fit value (0.70 cf 0.907) for dominance of the heterozygous fertility effect in females (dashed black line).

https://doi.org/10.1371/journal.pgen.1007039.g004

and expectations. Keeping the experimental estimates of all other parameters unchanged, the least squares best fit occurred at a dominance coefficient of 0.70 (Fig 4B and S1 File), compared to our previous direct estimate of 0.9, with lower confidence limit of 0.86 [11]. We also used our model to investigate the potential impact of HDR after embryonic cleavage caused by maternal deposition of Cas9, and found this parameter has little effect on the expected rate of resistance emergence when the rates of meiotic homing are as high as we observe.

Discussion

We have analyzed the dynamics of a gene drive deliberately designed to impose a fitness load on a population, and characterized the resistant or compensatory mutations which it generated and selected for. As with any control approach aimed at suppressing an organism, 'push back' from the target organism is to be expected. One of the advantages of the modular gene drives investigated here is that contingency in planning for and overcoming resistance can be foreseen and built into the system in a number of ways. First, the use of multiple gene drives targeting separate sequences has long been considered an essential pre-requisite for any gene drive intended as a functional vector control tool [19] and the ease with which the guide RNA expression constructs can be multiplexed lends CRISPR-based gene drives this flexibility [12, 20]. Second, it will be useful to target sites at which sequence changes are likely to destroy function. The nuclease target site in the gene AGAP007280 described in this report was not chosen according to any prioritisation based on high levels of sequence conservation that would imply functional constraint—a feature expected to mean that resistant mutations are less likely to restore function of the gene. Clearly the choice of the target should be guided by the extensive genomic data that is now available on sequence conservation both across different *Anopheles* species [21] and within *An. gambiae* [22]. This data revealed *a posteriori* that for the target site in AGAP007280 used here there is pre-existing variation in this species at at least 8 of the 20 nucleotides covered by the gRNA. Going forward low tolerance of sequence variation at the

target site should be a key criterion for designing a gene drive. Third, our results show that one of the key drivers in the generation of resistant alleles is the nuclease activity itself, followed by end-joining, and a significant proportion of these alleles are created as a result of maternally deposited nuclease in the early zygote where end-joining repair predominates over homology-directed repair. We suggest this maternal effect may be suppressed either through the use of more tightly regulated promoters to restrict nuclease expression to the early germline or through the addition of destabilising modifications to the nuclease, either of which are expected to reduce perdurance in the embryo. Fourth, an additional consideration in the choice of target site may take into account the propensity for a particular double strand break to be repaired more readily into a resistant, restorative (R2) allele, for example due to microhomology either side of the cleavage site that more readily re-creates an in-frame allele than a frameshift allele. Where MMEJ is the predominant end-joining repair pathway, this feature could be incorporated into target site choice to ensure the most likely end-joining repair event is an out of frame allele and therefore not likely to be selected.

Our approach of pooled sequencing of a targeted region allowed us to reliably detect even low frequency signatures of gene drive activity and reveal the complex dynamics of different genotypes emerging over time. Certainly for the future improvement of gene drives it will be important to have a faster method to triage for the most robust gene drives least prone to resistance without a multi-generational cage experiment, a laborious and time consuming process that should be reserved for more extensive evaluation of the best candidates. A simple way to do this would be to apply the method of amplicon sequencing described here in a screen where all generated mutant alleles are balanced against a known null allele to see if they restore function to the target gene.

The potential for rapid emergence and spread of resistance highlights not only one of the technical challenges associated with developing a gene drive, but also how intentionally releasing a resistant allele in to a population could be a simple and effective means of reversing the effects of a gene drive, if it fully restores function [23].

Methods

Gene drive multi-generation cage experiments

These experiments were essentially as described before in Hammond *et al.* [11] Briefly, in the starting generation (G_0) L1 mosquito larvae heterozygous for the $CRISPR^h$ allele at AGAP007280 were mixed within 12 hours of eclosion with an equal number of age-matched wild-type larvae in rearing trays at a density of 200 per tray (in approx. 1L rearing water). The mixed population was used to seed two starting cages with 600 adult mosquitoes each. For subsequent generations, each cage was fed after 5 days of mating, and an egg bowl placed in the cage 48h post bloodmeal to allow overnight oviposition. After allowing full eclosion a random sample of offspring were scored under fluorescence microscopy for the presence or absence of the RFP-linked $CRISPR^h$ allele, then reared together in the same trays and 600 were used to populate the next generation. After a generation had been allowed the opportunity to oviposit, a minimum of 240 adults were removed and stored frozen for subsequent DNA analysis.

PCR of target site and deep sequencing library preparation

For the sequence analysis, a minimum of 240 adult mosquitoes taken at generations G_2 and G_{12} of the cage trial experiments were pooled and extracted en masse using the Wizard Genomic DNA purification kit (Promega). A 332 bp locus containing the target site was amplified from 40 ng of genomic material from each pooled sample using the KAPA HiFi HotStart Ready Mix PCR kit (Kapa Biosystems), in 50 μl reactions. Specially designed primers that

carried the Illumina Nextera Transposase Adapters (underlined), 7280-Illumina-F (TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGGAGAAGGTAAATGCGCCAC) and 7280-Illumina-R (GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGCGCTTCT ACACTCGCTTCT) were used to tag the amplicon for subsequent library preparation and sequencing. The annealing temperature and time were adjusted to 68˚C for 20 seconds to minimize off-target amplification. In order to maintain the proportion of the reads corresponding to particular alleles at the target site, the PCR reactions were performed under non-saturating conditions and thus they were allowed to run for 20 cycles before 25 μl were removed and stored at -20˚C. The remnant 25 μl were run for an additional 20 cycles and used to verify the amplification on an agarose gel. The non-saturated samples were used to prepare libraries according to the Illumina 16S Metagenomic Sequencing Library Preparation protocol (Part # 15044223 Rev.A). Amplicons were then purified with AMPure XP beads (Beckman Coulter) followed by a second PCR amplification step with dual indices and Illumina sequencing adapters using the Nextera XT Index Kit.

After PCR clean-up via AMPure XP beads and validation performed with Agilent Bioanalyzer 2100, the normalized libraries were pooled and loaded at a concentration of 10 pM on Illumina Nano flowcell v2 and sequenced using the Illumina MiSeq instrument with a 2x250bp paired-end run.

Deep sequencing analysis

Sequencing data of the amplified genomic region were analysed using available tools and developed scripts in R v3.3.1. Raw reads were cleaned up for low quality and trimmed for the presence of adapters using Trimmomatic v.0.36 [24].

Paired-end reads were merged together in order to reconstruct the whole amplicon sequence using PEAR v0.9.10 [25]. Resulting assembled identical fragments were then clustered using fastx_collapser module from the FASTX v0.0.13 suite (http://hannonlab.cshl.edu/fastx_toolkit/) and aligned to the reference amplicon with vsearch tool v2.0.3 [26]which implements a global alignment based on the full dynamic programming Needleman-Wunsch algorithm. We considered for downstream analysis only sequences represented by at least 100 reads in each dataset. The blast6 output files from the alignment phase were parsed by ad hoc written R scripts to identify sequence variants containing insertions and/or deletions in the target site. The quantification of each allelic variant was measured as relative alternative allele frequency by summing up the reads representing that particular variant in the dataset. Finally, for each identified variant, we examined the single nucleotide variants (SNVs) along the full amplicon and selected the ones with a minimum alternative allele frequency of 2.5% for the purposes of haplotype calling.

Haplotype estimation for *A. gambiae* G3 laboratory colony

As part of a sequencing effort one year prior to the start of this experiment 12 males and 12 females from our *A. gambiae* G3 laboratory colony were subjected to individual genome resequencing. Mosquitoes were chosen randomly as pupae of differing ages from separate trays of a large cohort of the colony population (census population size >2000) in order to minimise biased sampling from a reduced number of founders. After emerging as adults whole mosquitoes were individually homogenised and genomic DNA was extracted with Promega Wizard Genomic DNA extraction kit. Paired-end reads (2 x 100bp) obtained from the Illumina HiSeq 2000 sequencing were aligned to *A. gambiae* PEST reference genome assembly (AgamP4, VectorBase) using BWA-MEM (Li and Durbin 2009, v0.7.15). Alignments were sorted using Samtools (v1.5) and raw SNPs and indels were called using HaplotypeCaller tool from Genome

Analysis Toolkit (GATK, v3.7) for each of the 24 samples in the same 320bp region around the nuclease target site in AGAP007280 gene that was used for the pooled amplicon sequencing. Raw SNPs were then merged using GATK GenotypeGVCFs tool. No indels were observed in this step for the selected region. We used SHAPEIT2 (Delaneau et al. 2013, v2) for the final haplotype estimation from previously obtained unphased genotypes of 24 individuals. Raw sequencing reads from the 24 individuals have been submited to NCBI Sequence Read Archive (SRA) under project accession number PRJNA397539.

Modelling

We use a discrete generation deterministic model to explore how the dynamics of gene-frequencies depend on underlying parameters. We suppose there are four possible alleles: Wild-type (W), driver allele (H), and two mutant alleles that are resistant to homing, R1 which is fully functional and R2 which is recessive but non-functional (i.e., H/R2 and R2/R2 type females are sterile). We assume alleles segregate at meiosis according to Mendelian inheritance except in W/H males and females, where segregation may be distorted by cleavage followed by either homing or non-homologous repair. Our model also allows for the possibility that eggs from females with at least one H allele will contain the driver nuclease (regardless of the egg's own genotype), in which case cleavage and repair may occur in the embryo. The mathematical details of the model are given in the S1 Text, and model outputs from user-defined parameters values can be seen using a computable document format (Wolfram CDF Player) available as a file(S1 File). Baseline parameter values for the model are provided in the legend accompanying Fig 4.

Individual assays of mosquito fertility and transmission of gene drive

Individual females containing at least one copy of the RFP-linked *CRISPR*^{h} gene drive were selected as virgins from the G_{20} generation and allowed to mate with 5 wild type male mosquitoes, essentially as in Hammond *et al.* [11]. The fecundity of females and transmission of the gene drive was measured by counting larval offspring positive for the RFP marker. To check mating status of females, spermathecae were dissected and examined for the presence of sperm. Unmated females were censored from the fertility assay.

Sons of each gene drive mother from the G_{20} generation were kept together and allowed to mate in groups of approximately 5 males with an equal number of wild type females and assessed for rates of transmission of the gene drive. The male offspring of these sons (grandsons) inheriting the drive from their fathers were in turn assessed in the same way, keeping lineages separate.

Supporting information

S1 Fig. Resistant target site alleles showing selection are formed independently on a wide range of haplotype backgrounds. The presence of polymorphic SNPs surrounding the target site and circulating at various frequencies in the laboratory wild type colony allowed us the resolution to identify a variety of haplotypes on which target site indels may have been formed. The most prominent target site indel in each cage replicate in the G_{12} generation was analysed and the number of haplotypes containing the respective indel and the frequency of each haplotype was calculated. A measure of the diversity of pre-existing reference haplotypes present in the colony was obtained by examining the nature and frequency of known haplotypes in the original wild type colony based on resequencing of 24 individuals (48 haplotypes) and the haplotypes surrounding the wild type target site allele in the early G_2 generation of the cage experiment. In both replicates there were no unique haplotypes containing the indel that were not

already pre-existing in the starting population. The relative frequency of haplotypes surrounding a given target site allele are also displayed.
(TIF)

S1 Table. Target site allele frequencies at G_2 and G_{12} generations.
(XLSX)

S2 Table. Gene drive transmission in offspring of *CRISPR^h/R* receiving a maternal copy of gene drive, and subsequent paternal copies. Individual females containing at least one copy of the RFP-linked *CRISPR^h* gene drive were selected from the G_{20} generation and allowed to mate with 5 wild type male mosquitoes. Offspring were kept as separate lineages and sons and grandsons were usually mated as groups of at least 5 with wild type females. Transmission of the gene drive was monitored by screening for the presence of the linked RFP gene. The target site allele present in the initial G20 female founder of each lineage is also displayed. Summary data is shown on the left, data for each cage shown on the right.
(XLSX)

S1 Text. Mathematical details of the model used to predict gene drive dynamics.
(PDF)

S1 File. Computable document format (Wolfram) allowing user-entered values of parameters that may affect gene drive performance.
(CDF)

Acknowledgments

We are grateful to M. Astrakhan for technical assistance, to N. Windbichler, P. Papathanos for useful discussion and to A. Burt for helpful feedback on the manuscript.

Author Contributions

Conceptualization: Andrew M. Hammond, Tony Nolan.

Data curation: Marco Bruttini, Nace Kranjc, Francesco M. Carpi, Romina D'Aurizio.

Formal analysis: Kyros Kyrou, Marco Bruttini, Ace North, Romina D'Aurizio, Tony Nolan.

Funding acquisition: Andrea Crisanti.

Investigation: Andrew M. Hammond, Kyros Kyrou, Roberto Galizi, Xenia Karlsson, Tony Nolan.

Methodology: Andrew M. Hammond, Roberto Galizi, Tony Nolan.

Project administration: Tony Nolan.

Software: Ace North.

Supervision: Tony Nolan.

Visualization: Andrew M. Hammond, Tony Nolan.

Writing – original draft: Tony Nolan.

Writing – review & editing: Andrew M. Hammond, Ace North, Andrea Crisanti, Tony Nolan.

References

1. Burt A. Site-specific selfish genes as tools for the control and genetic engineering of natural populations. Proc Biol Sci. 2003; 270(1518):921–8. https://doi.org/10.1098/rspb.2002.2319 PMID: 12803906
2. Curtis CF. Possible use of translocations to fix desirable genes in insect pest populations. Nature. 1968; 218(5139):368–9. PMID: 5649682
3. Ribeiro JM, Kidwell MG. Transposable elements as population drive mechanisms: specification of critical parameter values. J Med Entomol. 1994; 31(1):10–6. PMID: 8158612
4. Sandler L, Novitski E. Meiotic Drive as an Evolutionary Force. The American Naturalist. 1957; 91 (857):105–10.
5. Bhatt S, Weiss DJ, Cameron E, Bisanzio D, Mappin B, Dalrymple U, et al. The effect of malaria control on Plasmodium falciparum in Africa between 2000 and 2015. Nature. 2015; 526(7572):207–11. https://doi.org/10.1038/nature15535 PMID: 26375008
6. WHO. WHO World Malaria Report. 2016.
7. Gantz VM, Bier E. Genome editing. The mutagenic chain reaction: a method for converting heterozygous to homozygous mutations. Science. 2015; 348(6233):442–4. https://doi.org/10.1126/science.aaa5945 PMID: 25908821
8. Gantz VM, Jasinskiene N, Tatarenkova O, Fazekas A, Macias VM, Bier E, et al. Highly efficient Cas9-mediated gene drive for population modification of the malaria vector mosquito Anopheles stephensi. Proc Natl Acad Sci U S A. 2015; 112(49):E6736–43. https://doi.org/10.1073/pnas.1521077112 PMID: 26598698
9. Windbichler N, Menichelli M, Papathanos PA, Thyme SB, Li H, Ulge UY, et al. A synthetic homing endonuclease-based gene drive system in the human malaria mosquito. Nature. 2011; 473(7346):212–5. https://doi.org/10.1038/nature09937 PMID: 21508956
10. Champer J, Reeves R, Oh SY, Liu C, Liu J, Clark AG, et al. Novel CRISPR/Cas9 gene drive constructs reveal insights into mechanisms of resistance allele formation and drive efficiency in genetically diverse populations. PLoS Genet. 2017; 13(7):e1006796. https://doi.org/10.1371/journal.pgen.1006796 PMID: 28727785
11. Hammond A, Galizi R, Kyrou K, Simoni A, Siniscalchi C, Katsanos D, et al. A CRISPR-Cas9 gene drive system targeting female reproduction in the malaria mosquito vector Anopheles gambiae. Nat Biotechnol. 2016; 34(1):78–83. https://doi.org/10.1038/nbt.3439 PMID: 26641531
12. Esvelt KM, Smidler AL, Catteruccia F, Church GM. Concerning RNA-guided gene drives for the alteration of wild populations. Elife. 2014;3. https://doi.org/10.7554/eLife.03401 PMID: 25035423
13. Unckless RL, Clark AG, Messer PW. Evolution of Resistance Against CRISPR/Cas9 Gene Drive. Genetics. 2017; 205(2):827–41. https://doi.org/10.1534/genetics.116.197285 PMID: 27941126
14. Noble C, Olejarz J, Esvelt KM, Church GM, Nowak MA. Evolutionary dynamics of CRISPR gene drives. Sci Adv. 2017; 3(4):e1601964. https://doi.org/10.1126/sciadv.1601964 PMID: 28435878
15. Hong CC, Hashimoto C. The maternal nudel protein of Drosophila has two distinct roles important for embryogenesis. Genetics. 1996; 143(4):1653–61. PMID: 8844153
16. Truong LN, Li Y, Shi LZ, Hwang PY, He J, Wang H, et al. Microhomology-mediated End Joining and Homologous Recombination share the initial end resection step to repair DNA double-strand breaks in mammalian cells. Proc Natl Acad Sci U S A. 2013; 110(19):7720–5. https://doi.org/10.1073/pnas.1213431110 PMID: 23610439
17. Papathanos PA, Windbichler N, Menichelli M, Burt A, Crisanti A. 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 Mol Biol. 2009; 10:65. https://doi.org/10.1186/1471-2199-10-65 PMID: 19573226
18. Deredec A, Burt A, Godfray HC. The population genetics of using homing endonuclease genes in vector and pest management. Genetics. 2008; 179(4):2013–26. https://doi.org/10.1534/genetics.108.089037 PMID: 18660532
19. Deredec A, Godfray HC, Burt A. Requirements for effective malaria control with homing endonuclease genes. Proc Natl Acad Sci U S A. 2011; 108(43):E874–80. https://doi.org/10.1073/pnas.1110717108 PMID: 21976487
20. Marshall JM. Overcoming evolved resistance to population-suppressing homing-based gene drives. biorXiv. 2016;088427.
21. Neafsey DE, Waterhouse RM, Abai MR, Aganezov SS, Alekseyev MA, Allen JE, et al. Mosquito genomics. Highly evolvable malaria vectors: the genomes of 16 Anopheles mosquitoes. Science. 2015; 347 (6217):1258522. https://doi.org/10.1126/science.1258522 PMID: 25554792
22. Miles A. Natural diversity of the malaria vector Anopheles gambiae. bioRxiv 2016;096289

23. Vella M, Gunning C, Lloyd A, Gould F. Evaluating Strategies For Reversing CRISPR-Cas9 Gene Drives. bioRxiv 2017;144097. https://doi.org/10.1101/144097.

24. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014; 30(15):2114–20. https://doi.org/10.1093/bioinformatics/btu170 PMID: 24695404

25. Zhang J, Kobert K, Flouri T, Stamatakis A. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics. 2014; 30(5):614–20. https://doi.org/10.1093/bioinformatics/btt593 PMID: 24142950

26. Rognes T, Flouri T, Nichols B, Quince C, Mahe F. VSEARCH: a versatile open source tool for metagenomics. PeerJ. 2016; 4:e2584. https://doi.org/10.7717/peerj.2584 PMID: 27781170