

Developing Y chromosome-linked genome editors as
a self-limiting strategy for population control of the
malaria vector *Anopheles gambiae*

Ignacio Tolosana Labarta

Department of Life Sciences

Imperial College London

Thesis submitted for the degree of
Doctor of Philosophy

2023

Acknowledgments

During the last four years, there have been many people that were part of this journey and helped me become a better scientist and hopefully a better person. I will be forever grateful for that.

First, I want to thank my supervisors Professor Andrea Crisanti and Dr Tony Nolan, for giving me the opportunity to work on a project and a field that I found fascinating, and for their support and their guidance through these years.

I would also like to thank my progress review panel, Professor Austin Burt and Professor Karen Polizzi, for their feedback and suggestions during my degree.

Thank you, Federica, because it was not easy, but you managed. And you helped all of us on the way.

The last four years have been so much easier thanks to colleagues that became good friends very soon. Anna, Nicole, Nace, Matteo, Silvia, I can't thank you enough for the long and always interesting scientific and existential discussions, and for the life outside the lab. And thanks to the rest of the Crisanti group for their support.

A mis padres, Jesús y Gemma, que han sido y son orgullo y ejemplo. Por vuestro apoyo en todas mis decisiones y por enseñarme y ayudarme a perseguir lo que quiero. Y a mis hermanos, Juan y Jaime, que han hecho más fácil desconectar en cada videollamada.

A Isabel, la persona que ha sido un pilar diario. No tengo suficientes palabras para expresar mi agradecimiento y mi admiración. Gracias de corazón por el apoyo constante, que empezó mucho antes que este proyecto.

A mis granjeros del Imperial y mis Enter Room, por hacer que Londres se sintiera casa. Gracias, de verdad.

Y gracias a mis amigos, por hacer que en cada vuelta a casa acabase recargado de energía y con otra perspectiva.

Statement of Originality

I hereby declare that the project presented in this thesis is my own work and the result of grouped discussions with my supervisors and my progress review panel.

Mathematical modelling highlighting the relevance of the developed technology in this thesis was performed by Katie Willis at Austin Burt's group.

Bioinformatic analysis of whole genome sequencing and RNA sequencing data was carried out by Crisanti lab member Nace Kranjc.

Additional sources of information that have been used are duly cited.

Copyright Declaration

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

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

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

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

Abstract

Anopheles gambiae mosquitoes are the major vector of malaria in Africa, the most affected region by this disease that causes more than 600,000 deaths annually. In the last decades, important reductions in malaria incidence and mortality have been achieved, primarily driven by vector control programmes based on the use of insecticides and bed nets. Nevertheless, the decrease in malaria cases has recently come to a standstill and there is a consensus that new technologies will be needed to achieve malaria eradication. Genetic control strategies to suppress mosquito populations arise as very promising tools due to their species-specificity and versatility. While effective self-sustaining strategies have been developed in *Anopheles* mosquitoes, the development of an efficient technology that is geographically and/or temporally restricted would present advantages that include a higher prospect for control and regulatory approval. This thesis describes the generation of such technology: a Y chromosome-linked genome editor (YLE), only present in males, that transmits a female-specific sterility-causing dominant mutation to over 90% of the offspring. The YLE is a Cas9-based system that is always inherited by all the males and targets the female isoform of the *doublesex* gene, involved in sex determination. It is the first reported genetic control strategy linked to the Y chromosome in *Anopheles*. Alternative target genes to induce female lethality in the offspring of YLE males were explored by attempting to validate putative haplo-insufficient genes located in the X chromosome. A different YLE strategy approach might be based on depositing female-specific toxins in the embryos. Transcripts present in the early *Anopheles* embryos that had been potentially contributed from the male progenitor were identified. Future studies on the regulatory sequences of these transcripts will determine whether they can be used to induce paternal deposition of a female-specific toxin in the form of mRNA or snRNA.

Table of Contents

Statement of Originality	5
Copyright Declaration.....	7
Abstract.....	9
Table of Contents	11
List of Figures	17
List of Tables	21
List of Acronyms.....	23
1 Introduction.....	25
1.1 An overview on malaria.....	26
1.1.1 The malaria burden.....	26
1.1.2 Cause and symptoms of malaria	27
1.1.3 Malaria vectors.....	27
1.1.3.1 Mosquito species and geographical distribution.....	27
1.1.3.2 <i>Anopheles</i> mosquito biology and life cycle.....	29
1.2 Malaria control interventions	31
1.2.1 Treatment and prevention of the <i>Plasmodium</i> infection.....	31
1.2.1.1 Antimalarial drugs	31
1.2.1.2 Vaccines	32
1.2.2 Vector control.....	33
1.2.2.1 Habitat removal.....	33
1.2.2.2 Insecticide-based interventions: Insecticide-Treated Nets (ITNs) and Indoor Residual Spraying (IRS).....	33
1.2.2.3 Novel approaches: the use of microorganisms (<i>Wolbachia</i> , <i>Microsporidia</i> , <i>Metarhizium</i>).....	34
1.2.2.4 Looking forward.....	37

1.3	Genetic control of mosquito vectors.....	38
1.3.1	The advent of the genomic era in vector control	38
1.3.1.1	The CRISPR/Cas9 system.....	39
1.3.2	Outcome and persistence of genetic control strategies.....	40
1.3.3	Self-limiting genetic control strategies	42
1.3.3.1	Sterile Insect Technique (SIT).....	42
1.3.3.2	Release of Insects carrying a Dominant Lethal gene (RIDL)	44
1.3.3.3	Autosomal sex ratio distorter (X-shredder).....	45
1.3.4	Self-sustaining genetic control strategies.....	47
1.3.4.1	Driving Y chromosome	47
1.3.4.2	Homing-based gene drives	48
1.3.4.3	CRISPR/Cas9-based gene drives.....	51
1.3.4.4	Sex distorter gene drive	52
1.3.4.5	Toxin-antidote-based strategies.....	53
1.3.5	Regulation and social implications.....	54
1.3.6	The need for new powerful self-limiting strategies	56
2	Relevant considerations to developing putative Y chromosome-linked genome editors	57
2.1	Requirements for a Y chromosome-linked genome editor.....	58
2.2	The nature of the Y chromosome.....	62
2.2.1	Evolution of the Y chromosome	62
2.2.2	The <i>Anopheles gambiae</i> Y chromosome.....	64
2.2.3	Transgene integrations on the Y chromosome.....	65
2.2.3.1	The use of chromatin insulators to stabilize transgene expression	66
2.3	Regulatory sequences for transgene expression in the <i>An. gambiae</i> male germline	67
2.3.1	Transgene expression in the <i>An. gambiae</i> male germline from the Y chromosome.	70
2.4	Causing dominant female-specific sterility or lethality	72
2.4.1	Identified dominant negative mutations in insects	73

2.5	Concluding remarks	74
3	Aims.....	75
4	Creating and evaluating a proof-of-concept Y chromosome-linked genome editor in <i>Anopheles gambiae</i>.....	77
4.1	Designing a genetic construct for the creation of a Y chromosome-linked genome editor in <i>Anopheles gambiae</i>	78
4.1.1	The <i>doublesex</i> gene as a target of a YLE strategy.....	78
4.2	Generating a proof-of-concept Y chromosome-linked genome editor targeting the female isoform of <i>doublesex</i>	80
4.3	Evaluating the Cas9 activity levels in the male germline from the Y chromosome-linked transgenes.....	82
4.3.1	Y-linked nuclease activity leads to homing of <i>dsx</i> null alleles	83
4.3.2	Homing of a female-specific dominant negative mutation represents a viable strategy for a Y chromosome-linked genome editor	86
4.4	The YLE ^{dsx} transgenes do not pose a fertility cost on males	91
4.5	Attempting to stabilize the Cas9 activity in the YLE ^{dsx} -A line through the use of <i>gypsy</i> chromatin insulators.....	92
4.6	Mathematical modelling of YLE ^{dsx} -B male releases confirms its potential as an efficient self-limiting genetic control tool.....	93
4.7	Discussion.....	98
4.8	Methods	102
4.8.1	Design and cloning of plasmids.....	102
4.8.1.1	p16510-eCFP with <i>gypsy</i> chromatin insulators (p16510-eCFP[<i>gypsy</i>])	102
4.8.1.2	pY_16510-eCFP[<i>gypsy</i>] for integration by homologous recombination	104
4.8.1.3	Helper plasmids for pY_16510-eCFP[<i>gypsy</i>] integration.....	105
4.8.2	Plasmid purification.....	105
4.8.2.1	Bacterial transformation	105
4.8.2.2	Colony PCR screen.....	106
4.8.2.3	Miniprep	106

4.8.2.4	Maxiprep.....	107
4.8.3	Generation of transgenic lines.....	107
4.8.3.1	Embryo microinjections	107
4.8.3.2	Maintenance of <i>An. gambiae</i> mosquito strains	109
4.8.4	Evaluating Cas9 activity in the YLE ^{dsx} lines	109
4.8.4.1	Homing of a marked allele in <i>dsxF</i>	109
4.8.4.2	Homing of a dominant negative mutation in <i>dsxF</i>	110
4.8.5	Amplicon sequencing of the <i>dsxF</i> target.....	111
4.8.6	Fertility assays.....	112
4.8.7	Mathematical modelling.....	112
5	Identification of female-specific dominant mutations in the <i>doublesex</i> gene of <i>Anopheles gambiae</i>.....	115
5.1	The <i>doublesex</i> gene in <i>Anopheles gambiae</i>	116
5.1.1	Gene structure and splicing mechanisms	116
5.1.2	Protein structure and role of the conserved domains.....	117
5.1.3	Phenotypic effect of mutations in the <i>doublesex</i> gene.....	120
5.2	An 11 base-pair deletion in the coding sequence of <i>Agdsx</i> exon 5 is a female-specific dominant negative mutation	121
5.2.1	The <i>dsxF^{A11}</i> mutation is highly penetrant and causes complete sterility in heterozygosity in chromosomally female individuals.....	123
5.2.2	The <i>dsxF^{A11}</i> allele is transcribed and would produce a truncated DSX ^F protein	126
5.3	A mutagenesis screen reveals the existence of other dominant and recessive mutations in the female-specific exon of <i>doublesex</i>	128
5.4	Discussion.....	133
5.5	Methods	138
5.5.1	Phenotype evaluation.....	138
5.5.1.1	Fertility assay	138
5.5.1.2	Dissections of females showing an intersex phenotype	138

5.5.2	Reverse Transcription PCR (RT-PCR) of the <i>dsx</i> ^{Δ11} allele	138
5.5.3	Mutagenesis screen in the <i>doublesex</i> target site	139
6	Validating candidate haplo-insufficient genes in <i>Anopheles gambiae</i>	141
6.1	Haploinsufficiency.....	142
6.1.1	Haploinsufficient genes in <i>Drosophila melanogaster</i>	143
6.1.2	Haploinsufficient genes in <i>Anopheles gambiae</i>	144
6.1.2.1	Dosage compensation of the X chromosome in <i>Anopheles gambiae</i>	145
6.2	Identification of candidate haploinsufficient genes in <i>Anopheles gambiae</i>	146
6.3	Experimental plan.....	149
6.4	Selection of the target sequences	150
6.5	Generation of the gRNA lines.....	154
6.6	Genetic crosses with Cas9-bearing strains.....	155
6.6.1	Crosses with <i>vas2::Cas9</i> strain.....	155
6.6.2	Crosses with <i>β2::Cas9</i> strain.....	158
6.6.3	Cross with Y-linked <i>vas2::Cas9</i> strain.....	160
6.7	Discussion.....	162
6.8	Methods	166
6.8.1	Selection of the target sequences.....	166
6.8.2	Cloning of plasmids	166
6.8.3	Generation of transgenic lines.....	167
6.8.4	Molecular characterization of the transgenic lines	167
6.8.5	Genetic crosses to evaluate the haploinsufficiency of the candidate genes.....	168
6.8.5.1	Crosses with <i>vas2::Cas9</i> strain	168
6.8.5.2	Crosses with <i>β2::Cas9</i> strain	168
6.8.5.3	Cross with YLE ^{dsx} strain	169
6.8.6	Amplicon sequencing of the target sequences	169
7	Discovery of regulatory sequences to induce paternal deposition of transcripts in <i>Anopheles</i> embryos	171

7.1	The need to identify regulatory sequences for paternal deposition	172
7.1.1	Paternal deposition of transcripts and proteins in other organisms.....	173
7.1.2	Zygotic genome activation in <i>Anopheles</i>	174
7.2	Experimental design to detect paternal deposition in <i>Anopheles</i> embryos	175
7.3	Identifying the origin of deposited transcripts in the early <i>Anopheles</i> embryos	177
7.3.1	Ribosomal RNA depletion with a customised kit is extremely efficient in <i>Anopheles</i> samples.....	177
7.3.2	RNA sequencing with 100 million reads per sample allows to detect extremely rare variants.....	178
7.3.3	No signs of zygotic genome activation in 1-hour-old <i>Anopheles</i> embryos.....	180
7.3.4	Assigning the parental origin to transcripts present in early <i>Anopheles</i> embryos	181
7.3.5	Identified transcripts with signs of paternal deposition.....	183
7.3.5.1	Candidate genes for additional studies to achieve paternal deposition of a transgene	183
7.4	Discussion.....	186
7.5	Methods	188
7.5.1	Sample collection	188
7.5.2	RNA extraction and RNA sequencing.....	189
7.5.3	Creation of a spiked RNA sample	189
7.5.4	Design of the customised probe pool for ribosomal RNA depletion.....	190
7.5.5	Analysis of the WGS and RNA sequencing data.....	190
	Conclusions and future directions.....	191
	Bibliography.....	195
	Appendix	221
	Licences and Permissions	221

List of Figures

Figure 1.1 – Distribution of the most dominant malaria vectors and <i>Plasmodium falciparum</i> incidence rate in Africa.....	28
Figure 1.2 – The life cycle of the mosquito.....	30
Figure 1.3 – <i>Anopheles gambiae</i> male and female pupae.....	30
Figure 1.4 – The CRISPR/Cas9 system.....	40
Figure 1.5 – Classification of genetic control strategies.....	42
Figure 1.6 – Self-limiting strategies effect in the target population.....	46
Figure 1.7 – A driving Y chromosome	48
Figure 1.8 – Homing endonuclease genes' spread mechanism.....	49
Figure 1.9 – Homing-based gene drives for population suppression.....	50
Figure 2.1 – The Y chromosome-linked genome editor strategy	58
Figure 2.2 – Modelling on population control with different self-limiting strategies.....	61
Figure 2.3 – Expression levels of <i>vasa</i> , <i>zpg</i> , <i>nos</i> , <i>exu</i> and $\beta 2$ genes in the male germline	69
Figure 4.1 – Transcript comparison of the doublesex gene in <i>Drosophila melanogaster</i> and <i>Anopheles gambiae</i>	79
Figure 4.2 – Generation of the YLE ^{dsx} -A strain by integrating the p16510-eCFP plasmid in the Y-AttP line.....	81
Figure 4.3 – Generation of the YLE ^{dsx} -B strain by integrating the p16510 plasmid in the 3-2A2B line.....	81
Figure 4.4 – Homing of the <i>dsxF</i> - allele to evaluate Cas9 activity from the YLE ^{dsx} transgenes.....	82
Figure 4.5 – Inheritance rates of the <i>dsxF</i> - allele caused by the YLE ^{dsx}	83
Figure 4.6 – Inheritance rates of a modified <i>dsx</i> allele caused by the YLE ^{dsx}	87
Figure 4.7 – The fertility of YLE ^{dsx} males	91
Figure 4.8 – Modified p16510-eCFP plasmid to incorporate <i>gypsy</i> chromatin insulator sequences	93

Figure 4.9 – Comparison of the effect of releases of an idealised YLE and the developed YLE ^{dsx-B} strain	95
Figure 4.10 – Release rates required to achieve 95% suppression of a target population within 36 generations with the developed YLE ^{dsx} as a function of various parameters	97
Figure 5.1 – The structure of the <i>doublesex</i> gene in <i>Drosophila melanogaster</i> and in <i>Anopheles gambiae</i>	117
Figure 5.2 – Protein domains of the DSX ^M and DSX ^F isoforms in <i>An. gambiae</i> and comparison with <i>D. melanogaster</i>	119
Figure 5.3 – The <i>dsx</i> ^{F^{Δ11}} allele was probably generated by microhomology end-joining	122
Figure 5.4 – Intersex phenotype of chromosomally female individuals at pupal stage	123
Figure 5.5 – Intersex phenotype of chromosomally female individuals at adult stage	124
Figure 5.6 – Internal organs of chromosomally female individuals with intersex phenotype	125
Figure 5.7 – Females showing an intersex phenotype are completely sterile	125
Figure 5.8 – Comparison of the amino acid sequence of the female specific region of DSX ^F and DSX ^{FΔ11}	126
Figure 5.9 – Predicted structure of the homodimers and heterodimers formed by DSX ^F and DSX ^{FΔ11} proteins	127
Figure 5.10 – Identified mutations at the <i>dsx</i> target in the mutagenesis screen	130
Figure 5.11 – Comparison of the amino acid sequences potentially generated by mutations occurring in the coding sequence of exon 5	132
Figure 6.1 – Diagram showing the possibilities to target haploinsufficient genes with a YLE ...	145
Figure 6.2 – Germline expression of the putative haploinsufficient genes	147
Figure 6.3 – Comparison of the amino acid sequences of the candidate haploinsufficient genes and their paralogs in <i>An. gambiae</i>	148
Figure 6.4 – Diagram of the experimental plan to verify the haploinsufficiency of the candidate genes	149
Figure 6.5 – Conservation levels across the candidate haploinsufficient genes	152
Figure 6.6 – Comparison of the coding sequences of the candidate haploinsufficient genes and their paralogs in <i>An. gambiae</i>	153
Figure 6.7 – Sex ratio in the progenies of <i>vasa::Cas9 ; gRNA⁰⁹⁵³</i> males	156

Figure 6.8 – Fertility of $\beta 2::Cas9$; $gRNA^{HI}$ males and sex distortion in their progeny.....	159
Figure 6.9 – Fertility of YLE^{dsx} ; $gRNA^{0739}$ males and sex distortion in their progeny.....	161
Figure 6.10 – Diagram of a potential experiment to corroborate the haploinsufficiency of the candidate genes <i>0739</i> and <i>0952</i>	165
Figure 7.1 – Experimental design to find paternally deposited transcripts	176
Figure 7.2 – Level of ribosomal RNA depletion in the embryonic RNA samples.....	178
Figure 7.3 – Frequencies of <i>An. gambiae</i> -specific variants in the spiked control sample.	179
Figure 7.4 – Detected transcript levels of early zygotic genes and maternally deposited genes.	180
Figure 7.5 – Classification of transcripts based on their marker SNPs.....	182

List of Tables

Table 4.1 – Frequencies of new end joining generated mutations at the <i>dsx</i> target in the progenies of <i>YLE^{dsx} ; dsxF^{-/+}</i> males.....	85
Table 4.2 – Frequencies of new end joining mutations at <i>dsx</i> target in the progenies of <i>YLE^{dsx-A}; dsxF^{Δ11/+}</i> males	89
Table 4.3 – Frequencies of new end joining mutations at <i>dsx</i> target in the progenies of <i>YLE^{dsx-B}; dsxF^{Δ11/+}</i> males	90
Table 4.4 – Release rates required to achieve certain suppression levels in a target population with an idealised YLE, the developed <i>YLE^{dsx}</i> and other idealised self-limiting strategies	96
Table 4.5 – Primers used to amplify by PCR the fragments to clone p16510-eCFP[gypsy].	103
Table 4.6 – Primers used to PCR amplify the fragments to clone pY_16510-eCFP[gypsy]	104
Table 4.7 – The gRNA spacer and target sequences used for homology integration of pY_16510-eCFP[gypsy] in the Y-AttP line.....	105
Table 4.8 – Injection mixes to create YLE lines	108
Table 4.9 – Primers used for amplicon sequencing of the <i>dsxF</i> target.....	111
Table 4.10 – Model inheritance parameters.....	113
Table 5.1 – Primers used for amplification of the <i>dsxF</i> target site to verify the expression of the <i>dsxF^{Δ11}</i> allele.....	139
Table 6.1 – Designed gRNA sequences to target the candidate genes	150
Table 6.2 – Inferred genomic location of the randomly integrated transgenes in the gRNA strains	154
Table 6.3 – Frequency of mutations in the target gene in the progenies of <i>vasa::Cas9 ; gRNA⁰⁹⁵³</i> males	157
Table 6.4 – Frequency of mutations in the target gene in the progenies of <i>β2::Cas9 ; gRNA^{HI}</i> males	160
Table 6.5 – The gRNA spacer and target sequences selected to target the candidate haploinsufficient genes	166

Table 6.6 – Injection mixes to create the lines bearing the gRNA targeting each of the haploinsufficient candidates.....	167
Table 6.7 – Primers used for amplicon sequencing of the candidate haplolethal genes target sequence.....	169
Table 7.1 – Classification of marker SNPs based on the most plausible progenitor origin.....	181
Table 7.2 – Genes with transcripts that showed signs of paternal deposition.....	184

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

cDNA – complementary DNA

CDS – Coding sequence

CI – Cytoplasmic Incompatibility

CRISPR – Clustered Regularly Interspaced Short Palindromic Repeats

Dmel – *Drosophila melanogaster*

DSB – Double-Strand Break

dsx – *doublesex*

eCFP – enhance Cyan Fluorescent Protein

eGFP – enhanced Green Fluorescent Protein

EJ – End Joining

FPKM – Fragments Per Kilobase of transcript per Million mapped reads

FSR – Female-Specific Region

fs-RIDL – Release of Insects carrying a female-specific Dominant Lethal gene

gRNA – guide RNA

HDR – Homology-Directed Repair

HEG – Homing Endonuclease Gene

IRS – Indoor Residual Spraying

ITNs – Insecticide-Treated Nets

LLINs – Long-Lasting Insecticidal Nets

MAGs – Male Accessory Glands

MMEJ – Microhomology-Mediated End Joining

MSR – Male-Specific Region

NHEJ – Non-Homologous End Joining

OD – Oligomerisation Domain

PAM – Protospacer Adjacent Motif

PCR – Polymerase Chain Reaction

qPCR – quantitative Polymerase Chain Reaction

RFP – Red Fluorescent Protein

RIDL – Release of Insects carrying a Dominant Lethal gene

RT-PCR – Reverse Transcription Polymerase Chain Reaction

SIT – Sterile Insect Technique

SNP – Single Nucleotide Polymorphism

TPM – Transcripts Per Million

UTR – Untranslated Region

WT – Wild-type

YLE – Y chromosome-linked genome editor

1 Introduction

1.1 An overview on malaria

1.1.1 The malaria burden

Since the start of this PhD thesis project in May 2019, malaria has claimed over 2.4 million lives, 1.8 million of which were children under the age of 5. At the current incidence rate, by the time the reader reaches the end of the Introduction chapter, there will be over 28,000 new cases of malaria and over 70 people will have died as a result of this disease (World Health Organization, 2022).

Despite being preventable and curable, in 2021 alone there were 247 million cases of malaria, causing 619,000 deaths. Remarkably, malaria is a disease that affects particularly children under 5 years-old, who account for 77% of malaria-related deaths. Pregnant women and HIV patients are among other high-risk groups. (World Health Organization, 2022). This disease poses an enormous burden especially in the African continent, where 95% of the total cases and 96% of global deaths occur. Even within Africa, malaria prevalence is not evenly distributed and four countries account for over half of the global malaria deaths: Nigeria (31.3%), the Democratic Republic of the Congo (12.6%), the United Republic of Tanzania (4.1%) and Niger (3.9%) (World Health Organization, 2022).

A considerable international effort has been made in the last decades to reduce malaria incidence globally, resulting in an estimated 12 million malaria deaths averted since the year 2000. However, the progress achieved in the decline of malaria cases and deaths has stalled since 2015. This has also occurred in the funding allocated for malaria control and elimination programmes. While the total amount invested has increased from 2019 (US\$ 3.0 billion) to 2021 (US\$ 3.5 billion), the funding gap between the total investment and the resources needed to achieve the milestones set by the World Health Organization has widened (from US\$ 2.6 billion in 2019 to US\$ 3.8 billion in 2021). The continuation and widening of this gap jeopardize the accomplishment of these milestones that aim to reduce malaria cases and deaths by 75% before the end of 2025 and by 90% before the end of 2030 (World Health Organization, 2022).

1.1.2 Cause and symptoms of malaria

Malaria is caused by parasites of the *Plasmodium* genus, which are transmitted to humans through the biting of infected female *Anopheles* mosquitoes. There are five species of *Plasmodium* known to infect humans and cause malaria: *P. falciparum*, *P. vivax*, *P. malariae*, *P. ovale* and *P. knowlesi*, of which *P. falciparum* and *P. vivax* are responsible for the vast majority of malaria cases (World Health Organization, 2022).

The first symptoms of malaria usually appear 10-15 days after the infection and are nonspecific and flu-like. These include fever, headache, chills, and muscular pain. If not treated properly, and particularly in non-immune subjects, the parasite load can increase and cause the patient to develop severe malaria. Severe malaria symptoms comprise coma (cerebral malaria), anaemia, hypoglycaemia, renal failure or pulmonary oedema, and it is almost always fatal if untreated (World Health Organization, 2015).

1.1.3 Malaria vectors

1.1.3.1 Mosquito species and geographical distribution

Over 460 species of *Anopheles* mosquitoes have been identified, from which approximately 70 are able to transmit the human *Plasmodium* parasites (Hay et al., 2010; Sinka et al., 2012). However, not all of them pose the same threat to public health. Differences in geographical location, abundance, propensity to feed on humans versus on animals, adult longevity and ability for parasite transmission make some species more predominant vectors of *Plasmodium* parasites (Hay et al., 2010; Sinka et al., 2010; White et al., 2011).

Four species are the most important vectors of malaria: *Anopheles gambiae sensu stricto* (referred to as *Anopheles gambiae* hereafter), *Anopheles coluzzii*, *Anopheles arabiensis* and *Anopheles funestus*. The first three belong to the *Anopheles gambiae* complex, which is a group of closely related sibling species that are morphologically indistinguishable but with some ecological and behavioural differences (Sinka et al., 2012, 2010; White et al., 2011). The four mentioned species are highly efficient vectors of *Plasmodium* parasites

and are widespread across sub-Saharan Africa (Figure 1.1), causing the huge malaria burden in this region (Hay et al., 2010; Sinka et al., 2012, 2010).

Worryingly, the presence of a new vector species has been recently detected in Africa. *Anopheles stephensi*, the major malaria vector in the Asian continent, was found in the Horn of Africa and caused severe malaria outbreaks (Sinka et al., 2020). *An. stephensi* is more adapted to urban locations than the endemic African species, and therefore its spread across the African continent increases considerably the number of people living under the risk of malaria, seriously threatening malaria control programmes in Africa (Mnzava et al., 2022; Sinka et al., 2020; World Health Organization, 2022).

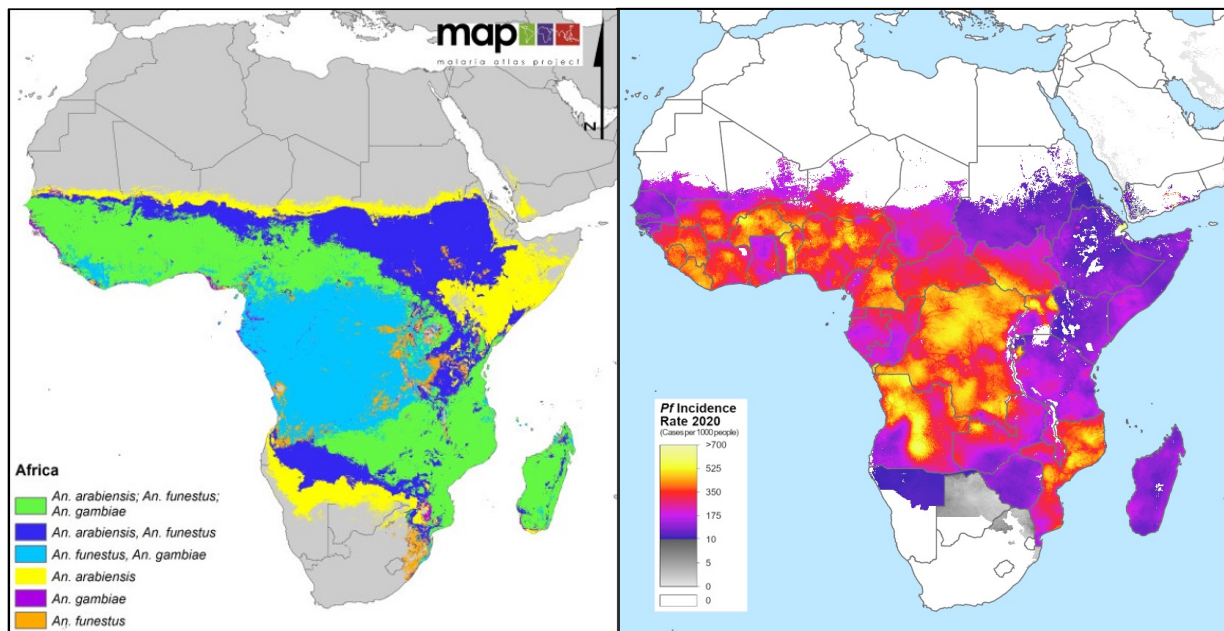


Figure 1.1 - Distribution of the most dominant malaria vectors and *Plasmodium falciparum* incidence rate in Africa. *Left:* Map showing the geographical distribution of the *An. gambiae*, *An. arabiensis* and *An. funestus* species in the African continent (Sinka et al., 2012). Coexistence of the three species is shown in green, presence of *An. arabiensis* and *An. funestus* is represented in dark blue, presence of *An. funestus* and *An. gambiae* is marked in light blue, areas with only *An. arabiensis* are shown in yellow, zones with only *An. gambiae* are represented in purple and regions with only *An. funestus* are shown in orange. *Right:* Map showing the cases of malaria caused by *Plasmodium falciparum* per 1,000 people (incidence rate) in Africa in the year 2020. The incidence rate ranges from 5 (in grey), passing through 175 (in light purple) and 350 (in red), and reaching 525 (in dark yellow) and >700 (in light yellow). Both maps were generated by The Malaria Atlas Project and are reproduced with permission of the rights holders Springer Nature; Sinka et al., 2012, and The Malaria Atlas Project.

1.1.3.2 *Anopheles* mosquito biology and life cycle

Anopheles mosquitoes belong to the Diptera order and to the Culicidae family, which comprises over 3,500 mosquito species divided in the Anophelinae and Culicinae subfamilies.

The life cycle of *Anopheles* mosquitoes encompasses both aquatic and terrestrial stages. The egg, larval and pupal phases are completely aquatic, while the adult phase is the only terrestrial one (Figure 1.2). Adult female mosquitoes typically lay 50-200 eggs on the water surface, where they hatch into the larval stage after 2-3 days. Larvae develop through four instar stages (L1-L4) during approximately 6-8 days, when they become pupae. At pupal stage, the sex of the individual can be identified upon morphological examination, as the male pupae show a longer and more developed genital lobe than females (Figure 1.3). The metamorphosis at the pupal phase lasts around 36 hours, after which adult mosquitoes emerge from the water. Morphological differences between males and females at the adult stage become more apparent. Males show characteristic plumose antennae with long flagellomeres and at the end of the abdomen they have a pair of claw-like structures called claspers which are used during mating, while females present pilose antennae with short flagellomeres and instead of claspers they have two very short cerci at the end of their abdomen (Becker et al., 2020; Clements, 1992).

Mating occurs generally at dusk, when adult males form swarms and adult females enter these swarms to mate. *Anopheles* female mosquitoes typically mate only once in their lives and store enough sperm of the mated male in their spermatheca to go through several gonotrophic cycles in their lifetime (blood meal, egg maturation and oviposition). While both sexes can survive feeding on nectar, females require a blood meal for egg development and maturation. The anthropophilic species of the *Anopheles gambiae* complex have a preference to bite indoors, normally during the night. Around 2-3 days after the blood meal, the females look for bodies of calm water for oviposition, where the larvae can develop (Becker et al., 2020; Clements, 1992; White et al., 2011).

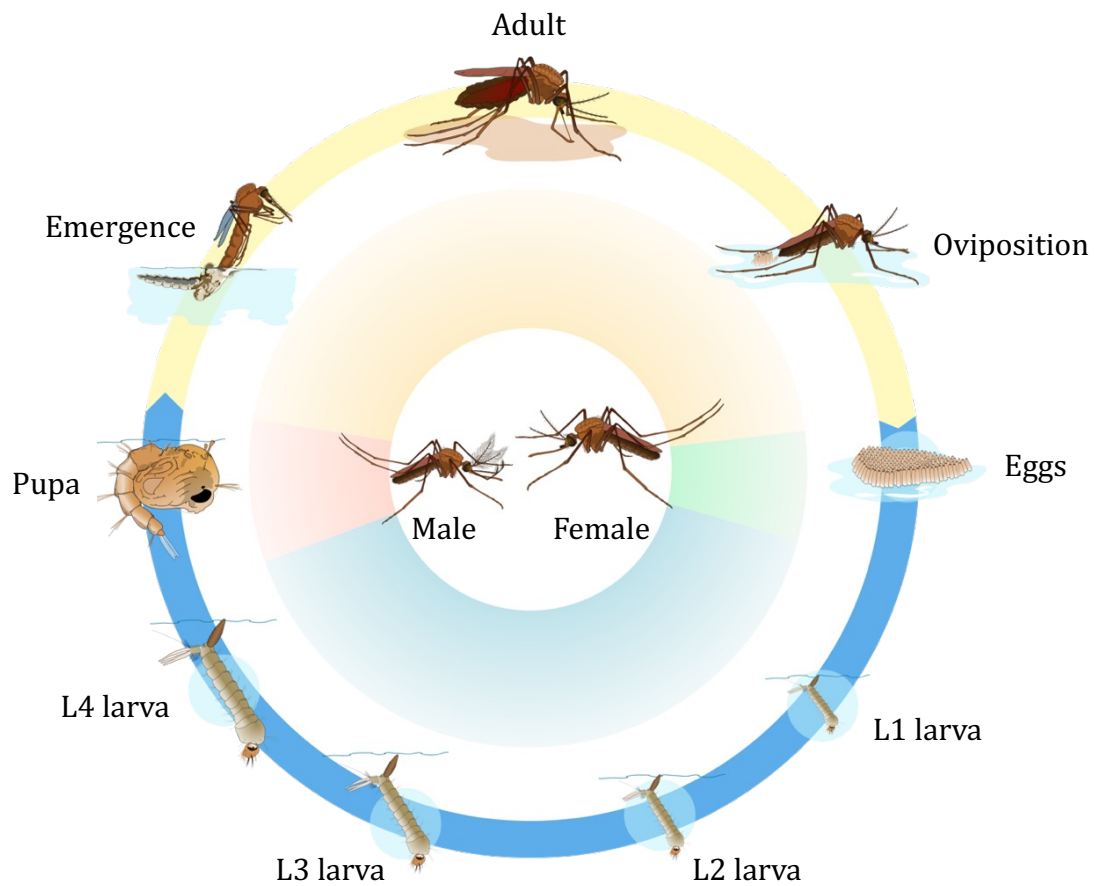


Figure 1.2 – The life cycle of the mosquito. Eggs are laid on the surface of stagnant water bodies. After approximately 2 days, the larvae hatch and they develop through 4 instar stages (L1-L4) during 6-8 days before pupating. Adult mosquitoes emerge after ~36 hours. Adult male and female mosquitoes can be distinguished by the naked eye based on the male-characteristic plumose (bushy) antennae and the pilose (slender) antennae that is characteristic of females. Image created by Mariana Ruiz Villarreal and released into the public domain.

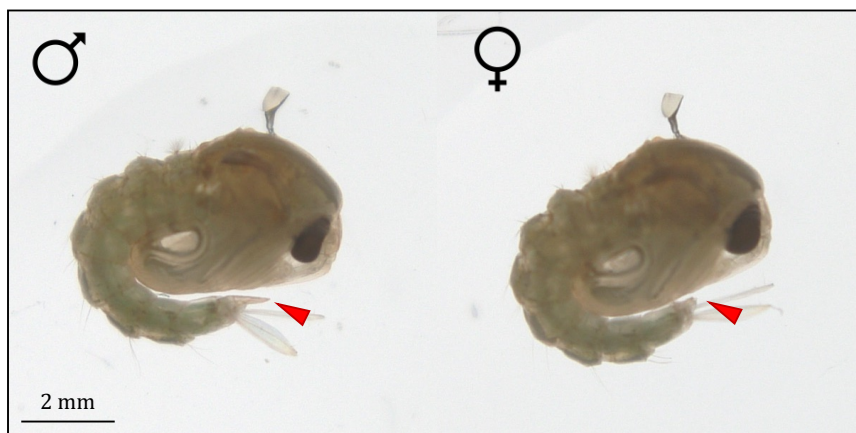


Figure 1.3 – *Anopheles gambiae* male and female pupae. The male pupae (left) have a longer and more developed genital lobe (red arrow) compared to the much shorter one in female pupae (right).

1.2 Malaria control interventions

The large global decrease of malaria cases and deaths achieved in the last decades has been possible thanks to an integrative approach that included some interventions targeting the parasite and others focused on the mosquito vector control. The first ones aim to treat the infection by eliminating the *Plasmodium* parasites from the human body or preventing their development, while the second ones aim to reduce malaria incidence by avoiding mosquito biting or reducing the mosquito populations near human dwellings (Bhatt et al., 2015; World Health Organization, 2022).

1.2.1 Treatment and prevention of the *Plasmodium* infection

1.2.1.1 Antimalarial drugs

Discovered in the 17th century, quinine is one of the oldest known antimalarial drugs, which has been used for centuries and it is still being used in certain circumstances (Achan et al., 2011; World Health Organization, 2015). Much later, in the 1940s, different antimalarial drugs were synthetically produced, among which chloroquine was particularly effective and became the first-line treatment of malaria. However, its widespread use led to the emergence of resistance in the *Plasmodium* parasites, decreasing its efficiency. Several other effective antimalarial drugs were developed across the 20th century, unfortunately suffering the same fate as chloroquine, since parasite resistance appeared soon after they were widely applied (Blasco et al., 2017; Haldar et al., 2018; Wongsrichanalai et al., 2002).

Currently, artemisinin-based combination therapies (ACTs) are the cornerstone of antimalarial treatment. They involve a combination of an artemisinin derivative and a partner drug that helps for both treating the parasite and preventing or retarding the emergence of resistance against artemisinin. ACTs are used to treat diagnosed malaria cases but also as preventive treatment in specific scenarios, such as intermittent treatment during pregnancy and in infants (under a year of age), or for seasonal malaria chemoprevention in areas where malaria transmission is highly seasonal (Haldar et al., 2018; Hanboonkunupakarn and White, 2022; World Health Organization, 2015).

However, resistance to artemisinin was detected over a decade ago in Southeast Asia (Ashley et al., 2014; Noedl et al., 2008). While resistance to ACTs is not currently predominant in Africa, recent studies found signs of resistance in *Plasmodium falciparum* parasites in Rwanda, and the spread or further emergence of resistance across Africa could be devastating for the continent (Uwimana et al., 2020).

New antimalarial drugs are being developed and several of them are in clinical trials, but their approval and affordable production will still take years (Ashley and Phyo, 2018; Hanboonkunupakarn and White, 2022).

1.2.1.2 Vaccines

The development of an effective vaccine against *Plasmodium falciparum* and *Plasmodium vivax* has been hampered, among other factors, by the complex life cycle of the parasite and the great genetic diversity in *Plasmodium* populations (Laurens, 2018). However, significant advances have been made in the recent years, to the point that in 2021 WHO recommended the use of the RTS,S vaccine for children in sub-Saharan areas with moderate to high malaria transmission (World Health Organization, 2022). Clinical trials in different African regions showed that the RTS,S vaccine has an efficacy of around 30%. Therefore, while it can be a very useful tool in the fight against malaria and pilot studies have been implemented in Ghana, Kenya and Malawi, the low efficacy of the vaccine will prevent it from becoming a silver bullet (Adepoju, 2019; White et al., 2015; World Health Organization, 2022).

A promising alternative vaccine, named R21, is currently in phase 3 trial. The new candidate has shown an efficacy of 77%, which is higher than the 75% efficacy goal set by WHO (Dattoo et al., 2021). While it is still to be recommended by WHO, Ghana and Nigeria have provisionally approved the use of the R21 vaccine, although the funding of the vaccination programmes remains unclear (BBC News, 2023; Grover et al., 2023; Reuters, 2023).

1.2.2 Vector control

Equally to other vector-borne diseases, malaria has been historically controlled through vector control (Wilson et al., 2020). These approaches aim to reduce mosquito populations in human-populated areas and to prevent mosquito biting, hence decreasing malaria transmission. Even in recent times, vector control has been the most efficient way to reduce malaria incidence: it has been estimated that between 2000 and 2015, around 80% of the decline in the number of malaria cases in Africa has been due to interventions targeting the mosquito vector (Bhatt et al., 2015).

1.2.2.1 Habitat removal

Before the 1940s, when the use of insecticide became widespread due to the discovery of dichloro-diphenyl-trichloroethane (DDT), most of the vector control interventions focused on environmental management (Wilson et al., 2020). Based on understanding the ecology and behaviour of mosquitoes, this approach aimed to eliminate larval habitats near human dwellings. Environmental management included draining marshes, filling water bodies and clearing vegetation (Wilson et al., 2020).

These interventions were widely applied from the 1900s to 1950s, leading to massive reductions in malaria incidence and virtual elimination of malaria in many countries and regions around the globe. However, habitat removal has a huge impact on non-target species and it is not feasibly applicable at a continental scale (Wilson et al., 2020).

1.2.2.2 Insecticide-based interventions: Insecticide-Treated Nets (ITNs) and Indoor Residual Spraying (IRS)

Upon the discovery of DDT in the early 1940s, the use of insecticides in malaria control campaigns became widespread. Two main insecticide-based interventions have been and still are the pillar of malaria incidence reduction. The first one is the use of insecticide-treated nets (ITNs), which consist in bed nets impregnated with insecticide that avoid mosquito biting during the night while also killing female mosquitoes trying to feed on humans. An improvement of ITNs are long-lasting insecticidal nets (LLINs), which have a lifespan of up to 3 years. According to WHO, in 2021 around 68% of the households in the

sub-Saharan region had access to an ITN or LLIN, resulting in 47% of the population at risk of malaria sleeping under a bed net (Wilson et al., 2020; World Health Organization, 2022, 2012). The second major insecticide-based intervention is indoor residual spraying (IRS), consisting in spraying insecticide on the walls and ceiling of the houses to kill female mosquitoes that would be resting there before or after a blood meal (Wilson et al., 2020; World Health Organization, 2022). Both strategies have proven to be highly cost-effective and enormously efficient. Between the years 2000 and 2015, malaria incidence in Africa decreased by 40%, equivalent to 663 million cases averted, of which 68% can be attributed to ITNs and 13% to IRS (Bhatt et al., 2015; World Health Organization, 2012; Yukich et al., 2008).

However, resistance to at least one of the 4 classes of insecticides currently approved for these interventions (pyrethroids, organochlorines, organophosphates and carbamates) had been reported in 88 countries by 2020. Importantly, 29 of those countries confirmed resistance to all of the 4 classes in at least one of the major vector species of malaria (World Health Organization, 2022). This could have serious implications for malaria control, as it has been estimated that the lack of pyrethroids efficacy could result in the loss of 56% of the vector control benefits and failing to avert 120,000 yearly deaths (World Health Organization, 2012).

While the benefits of sleeping under an insecticide-treated bed net are still significant even with insecticide resistance arising, other issues such as the slow replacement rate of old nets and limitations in the bed net distribution in some areas also threaten the efficacy of these interventions and contribute to the stall in the malaria incidence reduction. This, together with the behavioural plasticity of *Anopheles* vector mosquitoes to bite at earlier times and/or outdoors, or the presence of species that predominantly feed and rest outdoors, evidence that additional approaches for vector control are needed (Kleinschmidt et al., 2018; Ranson et al., 2011; World Health Organization, 2022, 2012).

1.2.2.3 Novel approaches: the use of microorganisms (*Wolbachia*, *Microsporidia*, *Metarhizium*)

The search for new strategies to target *Anopheles* vector species has led to investigate the use of microorganisms to reduce mosquito populations or to halt the *Plasmodium*

parasites transmission. The most advanced lines of research revolve around the use of endosymbionts, such as *Wolbachia* or *Microsporidia*, or the use of the entomopathogenic fungi *Metarhizium*.

Wolbachia is a maternally transmitted intracellular bacterial symbiont that naturally infects over 60% of insect species (Werren et al., 2008), and its use as a vector control tool has been widely studied in *Aedes aegypti* mosquitoes. In this species, *Wolbachia*-infected individuals have a highly reduced vectorial capacity, since *Wolbachia* induces resistance against certain viruses, including Dengue and Zika viruses (Hoffmann et al., 2011; Moreira et al., 2009; Walker et al., 2011). Additionally, *Wolbachia* biases its own inheritance by inducing cytoplasmic incompatibility (CI): if a *Wolbachia*-infected male mates with an uninfected female, no progeny is produced. However, a *Wolbachia*-infected female will produce viable offspring irrespective of their mate status and will transmit the symbiont to its progeny (Werren et al., 2008). These two characteristics allowed the successful establishment of *Wolbachia*-infected *Ae. aegypti* mosquitoes in urban locations to decrease dengue incidence (Nazni et al., 2019; Nguyen et al., 2015; Utarini et al., 2021). Besides this use, the release of only *Wolbachia*-infected *Ae. aegypti* males led to the suppression of wild populations by hampering the reproduction of wild females that mated with these males and did not generate viable offspring due to CI (Crawford et al., 2020).

However, while there are reported evidences of *Wolbachia* infections in natural *Anopheles* populations (Ayala et al., 2019; Baldini et al., 2018; Hughes et al., 2011; Jeffries et al., 2021, 2018), the benefits of this symbiont for malaria control are not as apparent as for dengue or zika control. An artificially *Wolbachia*-infected *Anopheles stephensi* colony was successfully established and promising results were reported regarding the induced CI and the reduction in *Plasmodium* development (Bian et al., 2013). Nonetheless, it seems that both artificial and natural *Wolbachia* infections can have a big negative effect in the fitness of infected *Anopheles* mosquitoes, highlighting that the *Wolbachia* effect on *Plasmodium* development remains unclear and seems to vary significantly depending on the symbiont strain, host species and even environmental conditions (Hughes et al., 2014, 2012; Mouillaud et al., 2023; Murdock et al., 2015; Shaw et al., 2016; Zélé et al., 2014).

Therefore, much research is still needed to find and evaluate *Wolbachia* strains that might be used as a tool for malaria control.

Microsporidia is an obligate intracellular symbiont related to fungi known to be present in some mosquito species (Han and Weiss, 2017). Recently, a microsporidian species (*Microsporidia MB*) was found in natural populations of *Anopheles arabiensis* and *Anopheles gambiae* (Herren et al., 2020; Nattoh et al., 2023). *Microsporidia MB* arises as a very promising vector control tool, as it has been shown to block *Plasmodium falciparum* transmission in *An. arabiensis* mosquitoes. When *Microsporidia MB* bearing mosquitoes were actively infected with *P. falciparum*, it was not possible to detect the parasite in the head and thorax of those individuals, suggesting that *Microsporidia MB* prevents colonisation of the salivary glands by *Plasmodium* parasites, hence averting transmission (Herren et al., 2020). Additionally, this symbiont is maternally transmitted to the offspring in *An. arabiensis* and *An. gambiae*, although the efficiency of this transmission is highly variable (it is transmitted to 45-100% and 0-100% of the progeny, respectively) and seems to depend on the density of the symbiont in the progenitor female. Sexual horizontal transmission has also been observed even though at lower rates (Herren et al., 2020; Nattoh et al., 2023, 2021).

These characteristics make *Microsporidia MB* a very interesting candidate for vector control. However, further research is needed to thoroughly examine the symbiont and to fully understand its behaviour in different malaria vector species, the stability of the infections and level of protection.

Metarhizium, on the other hand, is an entomopathogenic fungus. Some strains of *Metarhizium spp.* are pathogenic to adult mosquitoes and their effects have been studied in *Anopheles* species, which renders them as candidates for vector control strategies (Scholte et al., 2005). Spores of *Metarhizium* can be applied in an insecticide-like manner, by impregnating cloth or cotton sheets and placing them in the indoor walls or ceiling of houses to infect females resting there after a blood meal. Scholte et al. (2005) showed that *Anopheles* mosquitoes infected with *Metarhizium* in this fashion had a significantly shorter lifespan. However, wild-type strains of *Metarhizium spp.* have a low persistence

on surfaces and a low virulence, reflected in the slow killing of the mosquitoes, which severely reduced its potential for population control (Bilgo et al., 2017).

In order to develop an efficient tool, a strain of *Metarhizium pingshaense* was engineered to produce an arthropod toxin, with a strong lethal effect (Bilgo et al., 2017). This modified strain required a lower dose of spores to inflict lethality in the infected mosquitoes and it further reduced their lifespan to 5.19 days after infection compared to 8.73 days if the mosquitoes were infected with the wild-type strain (Lovett et al., 2019). Additionally, a negative impact on blood feeding in infected females was observed too, indicating an already pre-lethal effect (Bilgo et al., 2017; Lovett et al., 2019). Remarkably, these results were observed in both caged and semi-field conditions, and the transgenic *Metarhizium* strain was effective against both insecticide-resistant and insecticide-susceptible *An. gambiae* and *An. coluzzii* mosquitoes. Hence, the use of *Metarhizium* arises as a very promising tool in the field of vector control and researchers aim to further develop the mentioned strain and to apply it in open field interventions to reduce malaria incidence (Lovett et al., 2019).

1.2.2.4 Looking forward

The experience on malaria control interventions shows that the future work towards malaria eradication must combine approaches targeting the parasite and the vector mosquitoes in order to increase the possibilities of being successful. Furthermore, the continuous emergence of antimalarial and insecticide resistances proves that control strategies cannot rely on a single tool, and that it is not responsible or practical to wait until resistance arises against novel technologies to develop alternatives.

Malaria is strongly linked with poverty and rural areas, with poor dwelling conditions and deprived access to diagnosis and treatment (Laxminarayan, 2004; Sachs and Malaney, 2002; Teklehaimanot and Paola Mejia, Awash, 2008). It is thus important to consider this when developing new tools to reduce malaria incidence. While a new vaccine showed highly promising results and might lead to a turning point in the fight against malaria, vector control has been the most efficient approach up to now, reaching a high frequency of the population at risk (Bhatt et al., 2015; World Health Organization, 2022; Yukich et al., 2008).

In this framework, bacterial symbionts and pathogenic fungi are very promising tools that might be used in the near future. Equally, genetic vector control strategies have surged in the last years and show remarkably interesting characteristics that are discussed below.

1.3 Genetic control of mosquito vectors

Genetic control strategies are based on the propagation of desired heritable elements (or genetic modifications) through mating and vertical transmission. In this way, modified mosquitoes introduce genetic elements in the target population that reduce their vectorial capacity by either reducing the number of biting mosquitoes or their ability to transmit the *Plasmodium* parasite (Alphey, 2014; Burt, 2014).

An important feature shared by all the genetic control methods is that, being based on mating, they are highly species-specific and environmentally friendly, which minimises harmful effects on other insect species that are affected by insecticide-based interventions. Additionally, released modified mosquitoes will self-disperse and seek wild populations of their species, which makes genetic approaches effective across the whole area of application, irrespective of people's wealth or education. This same characteristic facilitates interventions in large areas and in geographically inaccessible regions that might be challenging to treat with current tools, as well as targeting at the same time outdoors-/indoors-/daytime-/night-time-biting mosquitoes (Alphey, 2014; Burt, 2014). These advantages render genetic control as a powerful and promising tool for future malaria control interventions.

1.3.1 The advent of the genomic era in vector control

Different strategies for genetic control of pest and vector species were first proposed, and some of them applied, in the 1940s-1960s (Curtis, 1968; Hamilton, 1967; Knippling, 1959). However, the progress in genome engineering and genome sequencing experienced in the

last decades has generated increasing interest in the field of genetic control and has enormously catalysed the development of new technologies.

Regarding malaria control, the generation of the first transgenic *Anopheles* mosquito in the year 2000 (Catteruccia et al., 2000) started a revolution in this field that has been in constant expansion. Another major breakthrough was the first complete assembly of the *Anopheles gambiae* genome (Holt et al., 2002), which has been complemented over the years with the whole genome sequencing of wild mosquitoes from different regions of sub-Saharan Africa (The *Anopheles gambiae* 1000 Genomes Consortium, 2017). These data are publicly available at the website of the *Anopheles gambiae* 1000 Genome project, providing useful information about the genetic variation in natural populations, insecticide resistance and enormously facilitating gene function studies.

Among all the discoveries and developments in genetic engineering, there is one that undoubtedly stands out for the revolution that it entailed: the CRISPR/Cas9 system, whose mechanism and relevance for vector control are explained below.

1.3.1.1 The CRISPR/Cas9 system

Originally discovered as an adaptive immune system in Bacteria and Archaea against viruses (Barrangou et al., 2007; Bhaya et al., 2011) and then fully characterised and harnessed for biotechnology (Jinek et al., 2012), the Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) and CRISPR associated nucleases (Cas) system has been widely adopted as the predominant genetic engineering tool due to its simplicity, flexibility and specificity.

The system is formed by Cas nucleases and small RNA molecules, termed guide RNAs (gRNAs), that direct the nucleases to a specific genomic location to induce double strand breaks (DSBs). This particular location is complementary to a short 'spacer' sequence (17-20 base pairs long) present in the gRNAs. Therefore, in order to target a different sequence, the only required modification is the spacer sequence of the gRNA. Further specificity of the CRISPR/Cas9 system is provided by the obligatory presence of a protospacer adjacent motif (PAM) at the 3' end of the target sequence (and absent in the spacer one), consisting in 3 nucleotides (NGG) (Figure 1.4) (Bhaya et al., 2011; Jinek et al., 2012; Wiedenheft et al., 2012).

The most commonly used system for genomic modification is a Cas9 nuclease derived from the bacterium *Streptococcus pyogenes* (Jinek et al., 2012). This tool has immensely facilitated gene function studies and the generation of transgenic organisms. Furthermore, it has been incorporated in vector control strategies (which are described below), exploiting in different ways the possible outcomes of the DSBs generated.

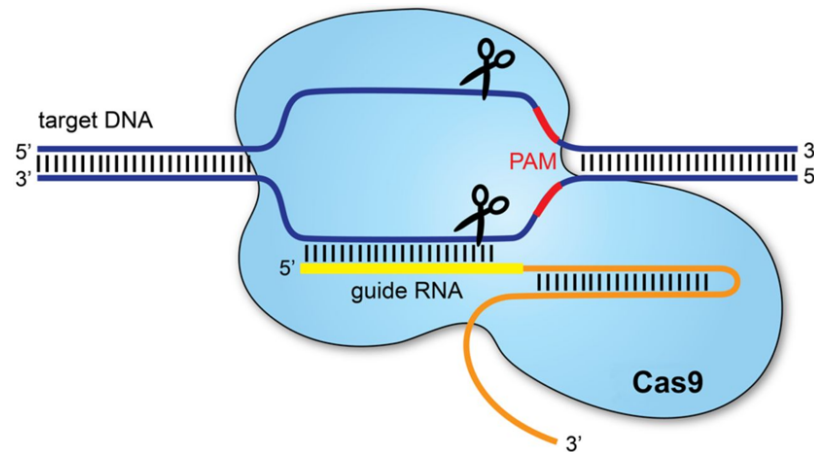


Figure 1.4 – The CRISPR/Cas9 system. The system is formed by a Cas9 nuclease and a guide RNA (gRNA) molecule. A synthesised gRNA recognises an 18-20 bp long sequence and directs the Cas9 nuclease to its target. The specific protospacer adjacent motif (PAM) sequence at the 3' end of the target locus is required for the cleavage. Image reproduced with permission of the rights holder, BMJ Publishing Group Ltd., Redman et al., 2016.

1.3.2 Outcome and persistence of genetic control strategies

Depending on the intended outcome, genetic control strategies can be classified in two groups: those that aim to replace or enrich target populations with mosquitoes that have a reduced ability to transmit parasites (**population replacement**) and those whose purpose is to reduce the overall number of mosquitoes in the target population (**population suppression**) (Figure 1.5) (Alphey, 2014; Burt, 2014). Because the female mosquitoes represent the reproductive potential of a population and are the actual vectors of the parasites, population suppression strategies generally aim to reduce the number of females in the target population, or to bias the population towards males (Alphey, 2014; Burt, 2014).

Another fundamental distinction is the expected dynamics of the modification introduced in the target population. In **self-limiting strategies**, the modification tends to disappear from the population, while in **self-sustaining strategies** the modification is designed to increase its frequency in the target population and persist over generations at high frequencies (Figure 1.5) (Alphey, 2014; Burt, 2014; Grilli et al., 2021). Both types of strategies present advantages and disadvantages, and the preference to use one over the other will depend on multiple factors, such as the target organism, its behaviour and ecology, or the geographic area where the intervention is applied, besides the social and regulatory limitations.

Self-limiting strategies generally require large (inundative) and continuous releases to have a maintained effect in the target population and an epidemiological impact, which entails high costs associated to mass production of insects and releases. Because the introduced modification disappears on its own, they are readily controlled and reversed, and this usually makes them temporarily and/or geographically restricted, which highly decreases the chance of invading neighbouring populations (Alphey, 2014; Burt, 2014). Self-sustaining strategies, on the other hand, need only one or few relatively small (inoculative) releases to spread and have an impact in the target population. As such, the associated costs to insect rearing and releases are much lower, but these technologies are generally not geographically nor temporarily restricted, and thus have a higher probability of invading neighbouring populations (Alphey, 2014; Burt, 2014). Within each group, different strategies have different threshold frequencies that determine their levels of persistence or invasiveness.

In this section, several genetic control strategies will be summarised, focusing on the population suppression strategies that are most relevant for the project of this thesis.

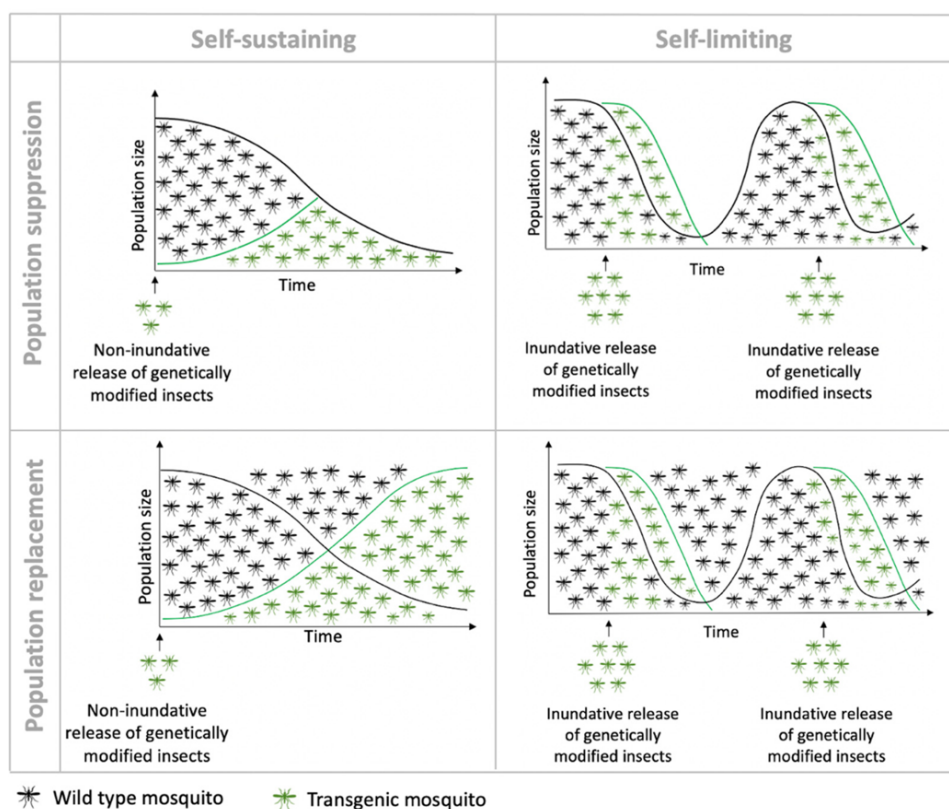


Figure 1.5 – Classification of genetic control strategies. Self-sustaining strategies (left) require very low initial frequencies (inoculative releases) to spread the modification in the target population to induce population suppression (top-left) or population replacement (bottom-left). Conversely, self-limiting strategies (right) require repetitive mass releases (inundative releases) to have a maintained effect in the population, either for suppression (top-right) or replacement (bottom-right) and the modification tends to disappear. Wild type mosquitoes are shown in black and transgenic mosquitoes bearing the desired modification are shown in green. Plots show the population size across time. Image reproduced with permission of the rights holder, MPDI, Grilli et al., 2021.

1.3.3 Self-limiting genetic control strategies

1.3.3.1 Sterile Insect Technique (SIT)

The sterile insect technique (SIT) consists in releasing sterile males in the target population, which compete with wild males and upon mating with females do not produce offspring, hence contributing to reduce the population. SIT is the most self-

limiting strategy because the modification and effect in the population disappears within a generation unless additional releases are performed (Alphey, 2014; Knippling, 1959).

Traditionally, SIT approaches used ionising radiation and chemosterilisation to induce DNA-damaging modifications that resulted in male sterilisation. This method has been successfully applied to suppress different pest populations across the globe. Examples of successful SIT programs are the eradication of the New World screwworm fly, *Cochliomyia hominivorax*, in North and Central America, and Libya (Lindquist et al., 1992; Wyss, 2000), or the elimination of the tsetse fly (*Glossina austeni*), vector of the sleeping sickness, from the island of Zanzibar (Vreysen et al., 2000).

One of the biggest limitations of the SIT interventions is the need for continuous mass sexing of male insects prior to releases. However, advances in insect transgenesis have provided improvements such as high-throughput sorting based on sex-specific expression of a fluorescent marker (Marois et al., 2012; Papathanos et al., 2009). Furthermore, a recently developed version of SIT (precision-guided SIT – pgSIT) removes the need of separating males and females. pgSIT has been developed in *Ae. aegypti* (vector of human viruses such as Dengue or Zika) and it is based on the CRISPR/Cas9 system for the disruption of genes essential for male fertility and female flight, hence generating sterile males and flightless females (Kandul et al., 2019; Li et al., 2021). In this way, unsorted eggs can be deployed in the field, given that females will be unable to flight and thus unable to bite or reproduce. Alternatively, flying sterile male mosquitoes can be collected at adult stage upon emerging to be released.

Despite the SIT successful interventions, its application for malaria control is challenging, since SIT requires the release of enormous amounts of sterile males to have a significant impact in the target population. Mathematical modelling performed by Li et al. (2021) suggests that population elimination of *Ae. aegypti* would require sustained releases of at least 100-200 pgSIT eggs per wild adult, and Maïga et al. (2014) estimated that gamma irradiated sterile *An. coluzzii* males would need to be released in ratios of at least 5:1 (sterile to wild males) to have an impact in the target population. The number of *Anopheles* males that would be required to produce and release across a vast area like the sub-Saharan Africa renders current SIT approaches as unsuitable for malaria control.

1.3.3.2 Release of Insects carrying a Dominant Lethal gene (RIDL)

A modification of SIT methods is the release of insects carrying a dominant lethal gene (RIDL). It involves the use of a repressible toxic effector or lethal gene that is inherited by the offspring of released males and expressed at certain developmental stages. This provides flexibility regarding the time at which lethality is induced to improve the efficiency of the strategy. For example, when targeting mosquitoes, inducing lethality in late larval stages or pupal stage would be beneficial, as modified larvae would compete with wild type counterparts contributing to further reduce the target population (Alphey, 2014; Thomas et al., 2000). Similar to SIT, the effect of RIDL in the population disappears after a generation, and mass sexing is required prior to releases.

A tetracycline-repressible RIDL system was developed in *Ae. aegypti* (Phuc et al., 2007). Mosquitoes bearing the RIDL construct were reared in the presence of tetracycline, and the gene is activated in the offspring of released males in the wild in the absence of the antibiotic. This RIDL system has been extensively tested and led to the successful suppression of local *Ae. aegypti* populations in Brazil (Carvalho et al., 2015) and in the Grand Cayman islands (Harris et al., 2012).

A variant of RIDL is female-specific RIDL (fs-RIDL), in which the toxic gene is only expressed in females. Males are viable and inherit the genetic construct, which will be passed to the following generation in a Mendelian fashion, hence fs-RIDL has a more prolonged effect in the population. A modification was introduced in the above-mentioned tetracycline-repressible RIDL system in *Ae. aegypti*, whereby lethality was only induced in females, and for which a field trial has been successfully performed in Brazil to suppress local populations (Spinner et al., 2022). A similar fs-RIDL system was developed in *Ae. albopictus*, by targeting a gene required for female flight (Labbé et al., 2012), and later adapted to the malaria vector *An. stephensi* (Marinotti et al., 2013). While fs-RIDL releases in the latter organism have not been performed yet, the still highly self-limiting feature of this strategy might make it suitable to target *An. stephensi* in urban, localised settings, but it is not optimal to treat big rural regions in Africa.

1.3.3.3 Autosomal sex ratio distorter (X-shredder)

A different approach for population suppression is inducing sex ratio distortion in the offspring of released males. In organisms with heteromorphic sex chromosomes, this can be achieved by favouring the inheritance of one of the sex chromosomes over the other.

In *An. gambiae*, homogametic individuals are females (XX chromosomes), and heterogametic individuals are males (XY chromosomes). Sex ratio distortion systems have been developed in this organism by generating transgenic individuals that target the X chromosome in the spermatogenesis process, leading to non-viable X-bearing sperm and thus males producing virtually only males in the offspring (Galizi et al., 2016, 2014; Windbichler et al., 2008). This strategy has been achieved by targeting ribosomal DNA (rDNA) repeats that in *An. gambiae* are exclusively located in the X chromosome (X-shredding). Two nucleases have been successfully used to create this system: I-PpoI, which was found to selectively cleave the rDNA repeats (Galizi et al., 2014; Windbichler et al., 2008), and Cas9, for which a gRNA targeting these repeats was designed (Galizi et al., 2016). In both cases, transgenic males produced progenies in which ~95% of the individuals were males, and because the selection of Y-bearing sperm occurs before fertilisation, the number of offspring produced is not reduced (Galizi et al., 2016, 2014).

The developed X-shredder strategies in *An. gambiae* are based in autosomal constructs, and consequently half of the progeny will inherit the transgene. Hence, in common with an fs-RIDL strategy, an autosomal sex ratio distorter generates only male viable progeny, and the genetic construct is inherited by half of the offspring. However, there is a key distinction. Because the X-shredder acts pre-zygotically, 50% of the progeny (eggs and larvae produced) will have an effect in the following generation; while for the fs-RIDL, causing post-zygotic lethality, only 25% of the progeny (eggs and larvae produced) will have an effect in the following generation. Therefore, while both strategies are self-limiting and the genetic constructs tend to disappear, it has been modelled that the described autosomal sex ratio distorter would be more efficient for population suppression than a fs-RIDL approach (Galizi et al., 2014).

While field trials of the autosomal X-shredder strategies for *An. gambiae* have not been performed yet, important studies have determined its stability and genomic location, as well as its functionality in other *Anopheles* backgrounds, including *An. arabiensis* and

colonies of *An. coluzzii* isolated in Mali and Burkina Faso (Bernardini et al., 2019; Pollegioni et al., 2023; Vitale et al., 2023).

Among the described self-limiting strategies (Figure 1.6), the autosomal X-shredder is the most powerful one. However, these technologies still require continuous and large releases of males to have a significant effect in the target population. Thus, they are generally more suited for local interventions, while their use in wide regions remains challenging. For this reason, researchers have sought to develop self-sustaining strategies that can be used for population suppression of *An. gambiae* mosquitoes in big areas of Africa with relatively small and few releases. These strategies are described below.

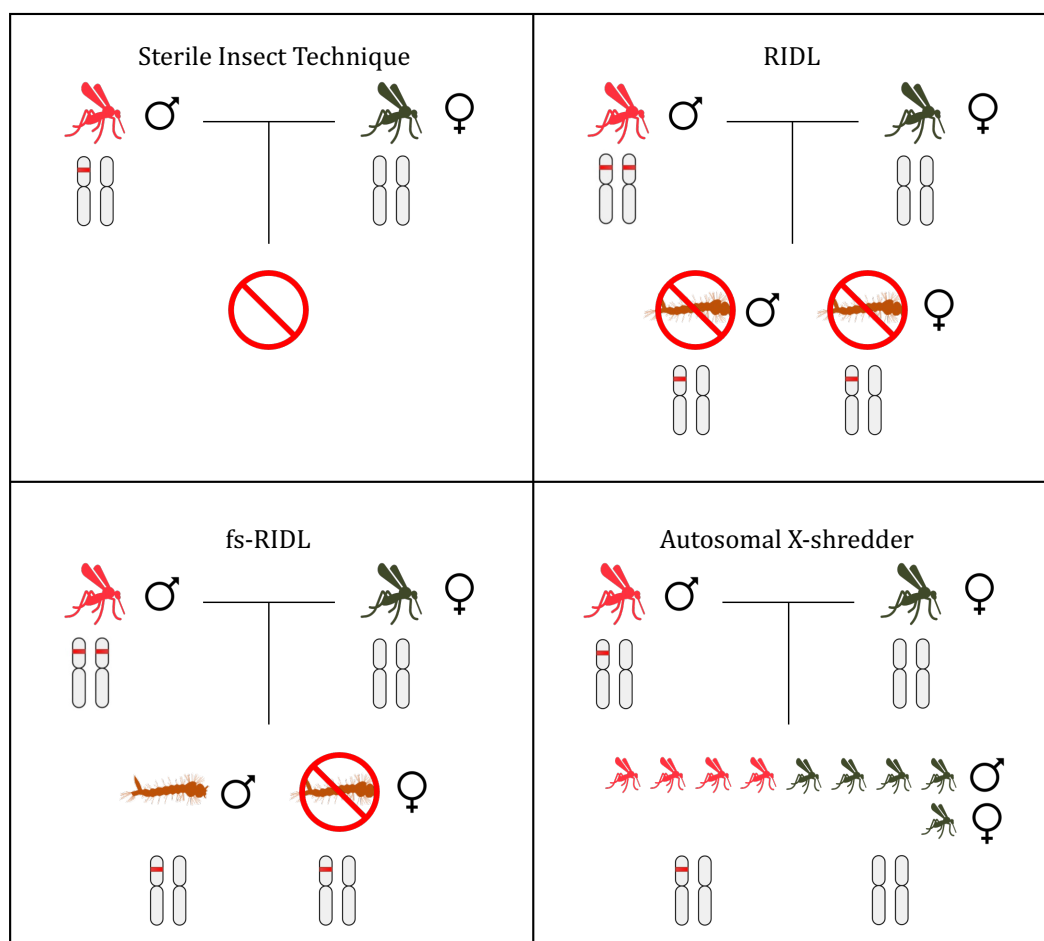


Figure 1.6 – Self-limiting strategies effect in the target population. Mode of action of the sterile insect technique (top-left), RIDL (top-right), fs-RIDL (bottom-left) and autosomal X-shredder (bottom-right). The genetic construct or genetic modification is shown in red. The strength of the autosomal X-shredder over both RIDL strategies can be observed by the fact that the release of heterozygous males would have the same effect in the number of females as the release of homozygous RIDL males. Consequently, the effect in the following generations will also be stronger.

1.3.4 Self-sustaining genetic control strategies

Contrary to self-limiting strategies, self-sustaining strategies are based on genetic constructs that spread in the target population and, in theory, could persist indefinitely. When an element biases its own inheritance and increases its frequency in a population over generations, it is referred to as a 'driving' element or 'gene drive'. This may occur if the genetic element is inherited at a frequency higher than the common Mendelian 50% inheritance, or if it decreases the fitness of alternative genotypes without directly affecting Mendelian inheritance (Alphey et al., 2020).

Many of the self-sustaining technologies developed are based on natural selfish genetic elements and have been synthetically improved in a way that they can be used in vector control strategies.

1.3.4.1 Driving Y chromosome

The idea of a Y chromosome that biases its own inheritance to constantly generate a male bias in the population had already been proposed as a method for population control in the 1960s (Hamilton, 1967). Following on the last described strategy, if an X-shredder construct is placed on the Y chromosome, it would automatically turn it into a driving Y chromosome, whereby all the progeny of these individuals would be males, and more importantly, all these males would inherit the Y chromosome with the X-targeting element (Figure 1.7). The modelled dynamics of such strategy show that it would be a highly efficient approach for population suppression (Deredec et al., 2008). In fact, mathematical modelling predicted that yearly releases of just 10 transgenic males bearing a driving Y chromosome, in ~420 settlements across Burkina Faso, would result in a 93.6% population suppression in the country after 4 years, and this suppression would remain stable after the halt of releases, even slightly increasing to 94.1% after 8 years. As a comparison, yearly releases of 50,000 SIT males at every settlement would cause only 0.4% population suppression in 4 years (North et al., 2019).

However, despite the outstanding potential of a driving Y chromosome for population control, attempts to generate this technology in *An. gambiae* have been unfruitful. The nature of the Y chromosome and the transcriptional silencing that it undergoes during

meiosis in the germline have hampered the engineering of Y-linked strategies. These issues will be addressed in more detail in Chapter 2, since they are particularly relevant for this thesis.

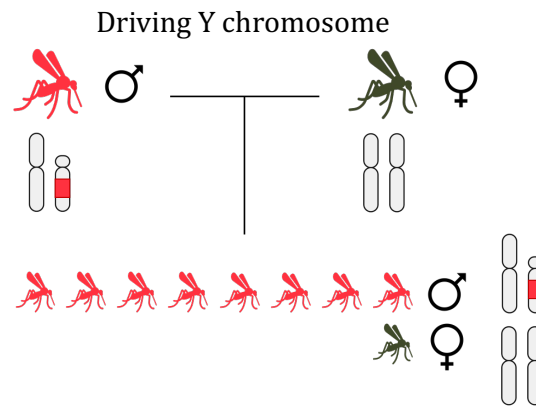


Figure 1.7 – A driving Y chromosome. A male bearing a driving Y chromosome would always generate a male bias in the progeny. Since all the males would inherit the Y chromosome, this would be a powerful self-sustaining strategy to rapidly suppress a target population. The genetic construct linked to the Y chromosome and causing the male bias is represented by the red square.

1.3.4.2 Homing-based gene drives

Homing endonuclease genes (HEGs) are naturally occurring selfish genetic elements that bias their own inheritance and increase their frequency among populations (Chevalier and Stoddard, 2001; Goddard et al., 2001). Based on this feature, Burt (2003) proposed the engineering of HEGs for the purpose of pest and vector population control.

HEGs encode nucleases that recognize and cleave specific DNA sequences (20-30 bp long). Importantly, HEGs are located within this recognition sequence, preventing the HEG-bearing chromosome from being cut. A double-strand cleavage is induced in the wild-type chromosome, which is generally repaired by the cell using the homologous chromosome as a template (homology directed repair – HDR). As a result, a copy of the HEG is incorporated and both chromosomes contain the HEG (Figure 1.8). This process is termed ‘homing’ and, in diploid organisms, converts heterozygous individuals into homozygous ones. If homing occurs in the germline of said organism, the HEG will be inherited by fundamentally 100% of the offspring instead of the typical 50% (Burt, 2003).

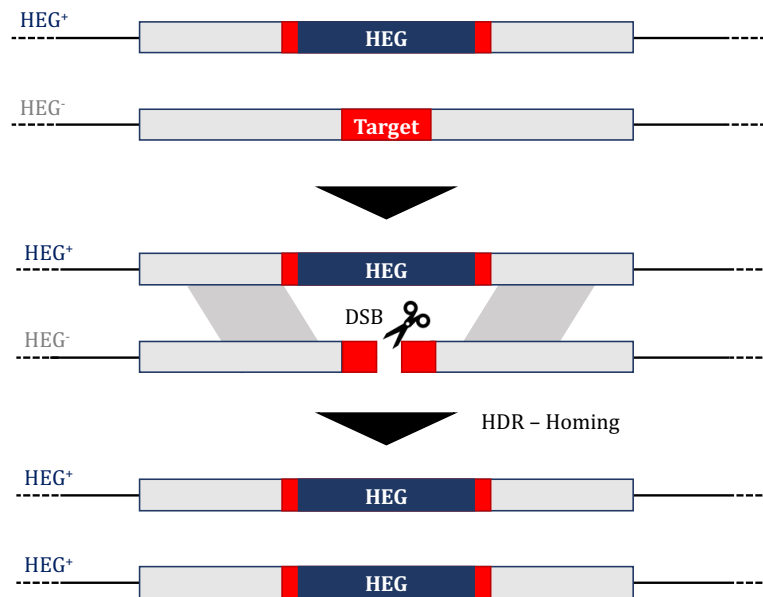


Figure 1.8 – Homing endonuclease genes’ spread mechanism. The homing endonuclease gene (HEG, in blue) is placed within its own target sequence (in red). When the HEG is active, it induces a double strand break (DSB) in the wild-type (HEG⁻) chromosome. The cell repairs this DSB by homology-directed repair (HDR) using the uncut homologous chromosome that carries the HEG. During this homing process, the originally heterozygous cell gets converted into a homozygous cell for the HEG. When this process occurs in the germline, the HEG biases its own inheritance to a higher frequency than the expected 50%.

Burt (2003) suggested that a HEG could be adapted for population modification or population suppression. For the latter one, a HEG could be engineered to recognize and cleave a gene essential for female fertility or female viability, and inserting itself into the middle of the sequence, thereby disrupting the gene. In this case, the target gene should be somatically expressed (i.e., not needed in the germline), and its disruption in heterozygosity should have no or little phenotypic effect, while being highly deleterious in homozygosity. Additionally, the HEG should be under the control of a germline-specific promoter to ensure that the homing process occurs only in the germline, leaving heterozygous individuals unaffected in the rest of the somatic tissues where the target gene is required, while the genetic construct is transmitted to most of their gametes. The spread of the HEG would eventually lead to most of the individuals being homozygous, generating sterile or non-viable females and ultimately collapsing the target population (Figure 1.9). For population modification, the same mechanism could be applied by disrupting a gene essential for parasite transmission or by carrying an antiparasitic effector gene.

A major breakthrough for the development of HEG-based control strategies was achieved by Windbichler et al. (2011), who created a synthetic gene drive system based on a HEG that recognised an artificial target site deliberately placed in the *An. gambiae* genome. Windbichler et al. (2011) demonstrated that a genetic element containing regulatory regions from the *An. gambiae* genome and a HEG increased its transmission to the offspring in these mosquitoes and invaded cage populations rapidly, in agreement with what had been modelled before (Burt, 2003). This showed that homology-directed repair occurs at high frequencies when a double strand break is induced in the *An. gambiae* germline, opening a gateway to use modified HEGs as vector control tools. Since the spreading mechanism is not an inherent property of the HEG, but it relies on the cell's repairing machinery, in principle it could be possible for any site-specific nuclease to be copied in such a fashion, allowing to use easier-to-engineer nucleases than HEGs or the traditionally used transcription activator-like effector nucleases (TALENs) or zinc finger nucleases (ZFNs).

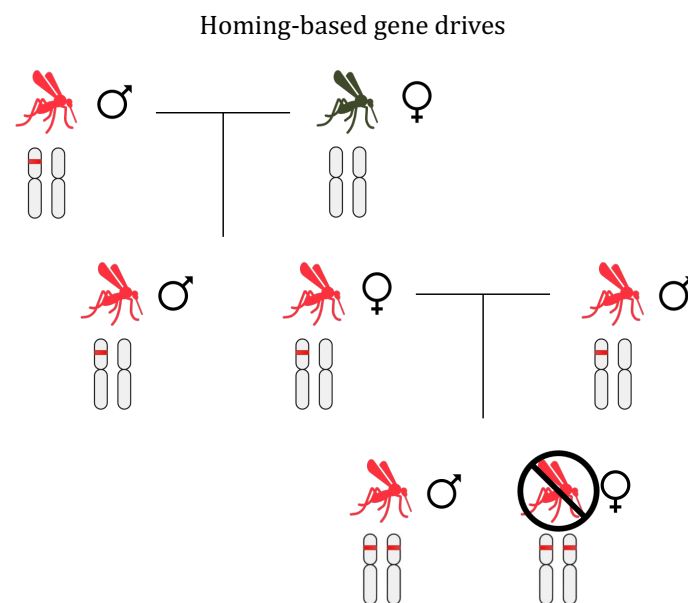


Figure 1.9 – Homing-based gene drives for population suppression. A gene drive designed to spread via homing and inserted within a gene involved in female fertility, disrupting it, would rapidly invade the target population. Heterozygous males and females would transmit the gene drive construct (in red) to virtually all their offspring. The increase in frequency of the gene drive construct would lead to the generation of homozygous individuals. Homozygous males would be unaffected and would transmit the construct to the following generation, but homozygous females would be sterile, which would result in the suppression of the target population.

1.3.4.3 CRISPR/Cas9-based gene drives

Soon after the discovery and development of CRISPR/Cas9 as a genetic engineering tool, a highly efficient proof-of-principle Cas9-based gene drive was developed in *An. stephensi* (Gantz et al., 2015). The adaptability of this system led to the generation of numerous Cas9-based gene drives for both population replacement and population suppression in *An. stephensi* and *An. gambiae* (reviewed in Bier, 2022). Focusing on population suppression technologies developed in *An. gambiae*, Hammond et al. (2016) developed Cas9-based gene drive systems targeting genes involved in female fertility. These were designed to mimic HEGs: the constructs were placed within the target genes, disrupting them and causing female sterility only in homozygosity. The genetic constructs contained a gene coding for Cas9 under an early germline promoter, and a gRNA ubiquitously expressed targeting the corresponding gene. The inheritance rates of the driving constructs were on average 84.6 – 99.2%, confirming the potential of this technology and that homology-directed repair was the most common repair mechanism in the early stages of the *Anopheles* germline. However, the invasion and suppression of caged mosquito populations was prevented by the emergence of resistant alleles in the target sequence (Hammond et al., 2017).

Resistant alleles can appear when instead of homology-directed repair, the double-strand break caused by the nuclease is repaired by non-homologous end joining (NHEJ) or microhomology-mediated end joining (MMEJ), leading to insertion/deletion mutations that make the sequence no longer recognizable by the gRNA and prevent from further cleavage. If these mutations do not restore the gene's function, they will not present a selective advantage and will disappear from the population. But if they rescue the gene's function, they will be strongly positively selected (Hammond et al., 2017).

Kyrou et al. (2018) developed an improved Cas9-based gene drive for population suppression by targeting a gene involved in sex determination in a highly conserved region (hence any mutation in the sequence was less likely to restore the gene's function) and by using a more germline-restricted promoter to express Cas9 (the importance of the germline promoters' characteristics to drive nuclease expression will be addressed in Chapter 2). A single small release (12.5% of the population) of males bearing this gene drive construct led to the total collapse of a caged population within 8 to 12 generations (Kyrou et al., 2018). Importantly, these results were replicated in large cages, with

overlapping generations and a more natural-resembling environment than small cages, confirming the robustness of the strategy (Hammond et al., 2021b). Kyrou et al. (2018) demonstrated that targeting a conserved sequence reduced the likelihood of resistance emergence. However, a comprehensive mutagenesis screen on this gene drive's target site (Ioanna Morianou, manuscript in preparation) revealed the possibility for resistant alleles to arise. Because of the big selective pressure that population suppression strategies impose, it seems inevitable that these technologies will have to be multiplexed (i.e., targeting multiple genes simultaneously or targeting a gene in multiple sequences) to reduce or delay the appearance of resistant alleles (Burt, 2003).

The potential of the Cas9-based gene drive developed by Kyrou et al. (2018) is reflected in the modelling performed by North et al. (2020), who predicted that releases of 5,000 males every year in ~430 sites across Burkina Faso would result in 95% suppression of the female population across the country in 4 years.

Cas9-based gene drives have shown to be one of the most powerful self-sustaining strategies developed to date to suppress *An. gambiae* populations, providing a very promising tool for malaria control interventions across the sub-Saharan Africa.

1.3.4.4 Sex distorter gene drive

Another powerful technology for population suppression was developed in *An. gambiae* by combining a Cas9-based gene drive element located within a female-specific region of a gene involved in sex determination with the previously described X-shredder construct (Simoni et al., 2020). This sex distorter gene drive (SDGD) has very similar dynamics to a driving Y chromosome, as it biases its own inheritance at the same time that it causes male-biased sex distortion.

Cage trial experiments showed that males bearing the SDGD construct generated a progeny with a male bias of 93.1%, and the SDGD element was inherited by 96 – 99% of the offspring (Simoni et al., 2020). However, later experiments revealed that the male-biased sex distortion ratio dropped to 60% at temperatures of 28°C, compared to the previously analysed at lower temperatures (Alekos Simoni, personal communication). While still a promising and powerful tool, the highly temperature sensitive sex distortion diminishes the SDGD potential in its current conformation.

1.3.4.5 Toxin-antidote-based strategies

An alternative way of biasing the inheritance of a genetic construct is to decrease the fitness of alleles not bearing it. In order to achieve this, different population control strategies based on toxin-antidote systems have been proposed and developed.

An example of such strategies is the maternal effect dominant embryonic arrest (*Medea*), which consists of a genetic construct with two genes: a gene that expresses a toxin in the female germline and deposits it in the embryo, and another gene that expresses the antidote (Chen et al., 2007). In this way, only the individuals that inherit the *Medea* element survive, and thus the genetic construct spreads in the population. *CleaveR* is another selfish genetic element based on a similar system based on cleaving an essential gene and rescuing the construct with a recoded copy of the target gene (Oberhofer et al., 2019). However, these strategies are more suitable for population replacement strategies than for population suppression (Alphey et al., 2020).

All the self-sustaining strategies described until now were low invasion threshold systems (i.e, they can spread and invade a population from a very low initial frequency). High invasion threshold strategies require to be present at a higher frequency in the population to spread, and if they are found below this threshold they tend to disappear, which makes them suitable for local control with a low risk of invading neighbouring populations (Alphey, 2014). Synthetic **underdominance** and a **split drive killer-rescue** systems have been proposed as high threshold-dependent drives for both population modification and population suppression. Both strategies are based on toxin-antidote systems in which the toxin gene is not linked to the antidote one, hence only the individuals in the offspring that inherit both genetic constructs will be viable (Akbari et al., 2013; Champer et al., 2016; Edgington et al., 2020; Edgington and Alphey, 2018). Nonetheless, even though proof-of-principle systems have been created for these strategies, they have not been implemented in *Anopheles* mosquitoes, partly due to the lack of suitable known toxins and antidotes, as well as the limited knowledge on how to effectively apply RNAi-mediated silencing in the mosquito germline (Champer et al., 2016).

1.3.5 Regulation and social implications

The characteristics of certain genetic control strategies make them promising tools that can be implemented for malaria control. While these technologies should not be seen as a silver bullet, the combination of newly developed genetic control strategies and current interventions could have a great impact to reduce malaria incidence globally (Wilson et al., 2020; World Health Organization, 2022). The vast areas of sub-Saharan Africa heavily affected by this disease render self-sustaining strategies, such as the developed Cas9-based gene drives in *An. gambiae*, more appealing than self-limiting alternatives. This is partly because of the big reduction in costs and efforts associated to the relatively low initial frequencies required by gene drives to suppress mosquito populations, in comparison to the costly requirements of the self-limiting counterparts.

Nonetheless, gene drives present some disadvantages too. Precisely because of its own efficacy and the low initial frequency needed to spread in a population, these genetic elements could easily invade neighbouring populations. While anti-drive technologies have been developed to potentially stop the spread of a gene drive (Taxiarchi et al., 2021), they are not temporarily or geographically restricted, and hence limiting or recalling them after release becomes challenging. Moreover, windborne long-distance migration of *Anopheles* mosquitoes has been reported in sub-Saharan regions (Huestis et al., 2019). This windborne migration can displace mosquitoes up to 300 km in one night, which denotes that gene drives could spread not only within the country of release, but also to neighbouring countries.

Because of the potential impact of gene drives in the natural ecosystem, even before the development of efficient gene drive elements, scientists agreed on the need to establish pathways to ensure the controlled research on gene drives and the conditions under which they should be developed and tested. The discussed pathways included the evaluation of gene drives' efficacy and stability in caged trials, working with transparency and engaging with local communities that would benefit from the use of gene drives (Annas et al., 2021; James et al., 2018; Long et al., 2020; Oye et al., 2014). Furthermore, some authors have published a set of recommendations for environmental risk assessment of gene drive applications (Connolly et al., 2022, 2021).

Due to the novelty of gene drive technologies and the lack of similar precedents, current regulation is not up to date with these advances and presents some gaps and inconsistencies (James et al., 2023; Marshall, 2010; Ostera, 2011; Oye et al., 2014). Most of the countries where gene drive strategies would be applied signed the Cartagena Protocol, which does not allow the release of genetically modified organisms that can cross international boundaries. However, scientists pointed out some ambiguities in the protocol that need to be addressed (James et al., 2023; Marshall, 2010). An additional obstacle might be found in the differences in regulatory processes, accountability laws, and evidence requirements demanded by each country for approval purposes (James et al., 2023). In any case, it seems clear that international agreements among the countries that would be potentially affected will be required in a process that should involve governments, regulatory bodies and local communities, and that will take years (James et al., 2023; Marshall, 2010; Ostera, 2011; Oye et al., 2014).

On the other hand, since self-limiting strategies are temporarily and/or geographically restricted, their potential impact in neighbouring populations is heavily reduced and their dynamics are more predictable. There are numerous precedents in the use of self-limiting genetic control strategies, as several SIT interventions have been approved in the previous decades (Lindquist et al., 1992; Vreysen et al., 2000; Wyss, 2000). Additionally, the already mentioned field trials of RIDL and fs-RIDL releases have been providing further information on the dynamics of released mosquitoes (Carvalho et al., 2015; Garziera et al., 2017; Harris et al., 2012). Altogether, new self-limiting strategies are more likely to meet current international regulations and face fewer obstacles towards approval. Moreover, the release of self-limiting genetic control elements would allow to gain experience in the field, gather useful data to better understand the behaviour of released males and their offspring, and evaluate the efficiency of the strategy as well as corroborating that no unexpected outcomes are produced. All the knowledge collected along this process would also be highly valuable for a potential release of a self-sustaining element in the future, while giving opportunities for engagement with local communities and contributing to gaining public acceptance (Adelman et al., 2017; Carter and Friedman, 2016; Neuhaus and Caplan, 2017).

1.3.6 The need for new powerful self-limiting strategies

The unresolved regulatory and social concerns of self-sustaining gene drives, and the impracticality of currently available self-limiting tools for the control of *Anopheles* populations, highlight the need to develop more powerful self-limiting strategies that can achieve local suppression in an efficient way while preserving the low risk of invading neighbouring populations. In recent years, different CRISPR/Cas9-based technologies that would fit in this new category have been proposed.

An example of such strategy would be an fs-RIDL element that causes female-specific sterility or lethality that has been modified to drive in males via homing. This genetic construct would be inherited by virtually all the offspring of modified males, triggering female sterility in the progeny and allowing the male progeny to replicate this effect in the following generations (Burt and Deredec, 2018; Thomas et al., 2000). The use of a gene drive system with their elements (Cas9 nuclease and gRNA) separated in two different constructs has also been suggested for efficient local suppression and proof-of-concept systems have been developed in *Drosophila* (Terradas et al., 2021).

Another powerful self-limiting strategy that has been proposed and modelled is a Y chromosome-linked genome editor (YLE) (Burt and Deredec, 2018). The development of this strategy for population suppression in *An. gambiae* is the focus of this thesis, and it will be comprehensively described in the following chapter, where the requirements and obstacles for the generation of such technology will be properly addressed, together with its advantages over other genetic control strategies.

2 Relevant considerations to developing putative Y chromosome-linked genome editors

The focus of this project is the Y chromosome-linked genome editor strategy (YLE). All the essential features to develop such technology are addressed in this chapter, as well as the difficulties and barriers that need to be overcome in the process to generate this genetic control tool in *An. gambiae*.

2.1 Requirements for a Y chromosome-linked genome editor

The idea of a self-limiting YLE was first proposed by Burt and Deredec (2018), based on a previously suggested way to increase the efficacy of a driving Y chromosome (Deredec et al., 2011). The YLE strategy is based on a construct located on the Y chromosome that creates germline mutations in genes that induce sterility or lethality when inherited by the female offspring but have no effect in the originator male or the male offspring (Figure 2.1).

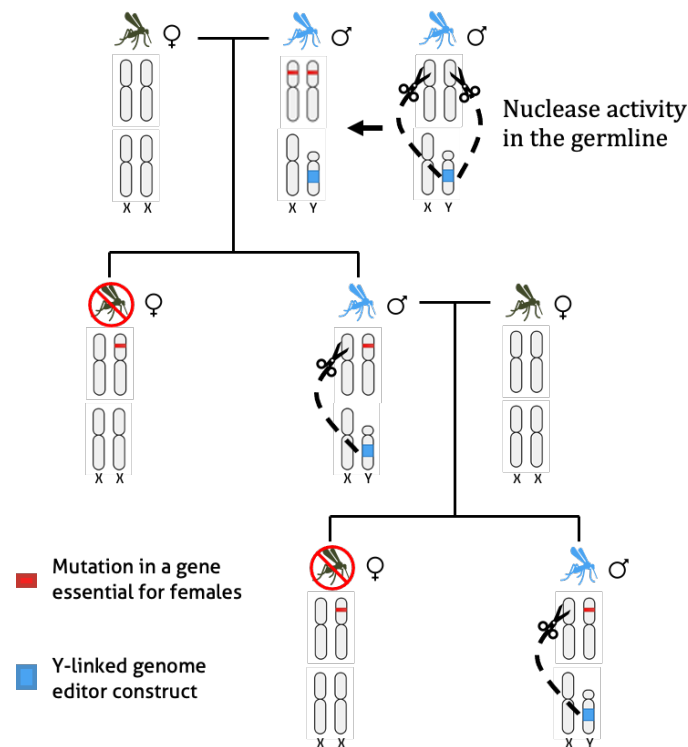


Figure 2.1 – The Y chromosome-linked genome editor strategy. The Y chromosome-linked genome editor construct (in blue) generates mutations in genes required for female survival/fertility (in red) in the male germline through nuclease activity. These mutations are detrimental when inherited by females (causing lethality or sterility) but do not affect males. Because the genetic construct is located on the Y chromosome, which is always inherited by males, all the males in the offspring inherit it, allowing this process to be maintained through generations.

Contrary to a driving approach, the modified Y chromosome would not increase its frequency in the population, but it is expected to remain at the released frequency because it is inherited in Mendelian fashion. In theory, since the YLE construct only causes harm in the female offspring, where the Y chromosome is not present, and there is no alternative allele or chromosome in the YLE-bearing males to compete with, the selection against the YLE is absent. This lack of selection against the YLE has a pivotal consequence: if releases of males carrying the transgene are halted, the population would not immediately recover (as occurs with many other self-limiting strategies such as SIT and RIDL) but it would remain suppressed indefinitely (Figure 2.2 – Top-middle). In practice, there may be other fitness costs that reduce the competence of males bearing the YLE construct. If the YLE poses a small fitness cost in males and is gradually selected against, the population will recover after stopping releases but at a much slower pace than with other self-limiting strategies (Figure 2.2 – Top-right). This long-lasting effect in the population once releases are halted would in theory allow a phased suppression, adding an extra element of control. On this premise, if YLE males are released at a 10% frequency of the initial male population, half of their offspring (i.e., the female portion) will be sterile or die each generation, resulting in a persistent 5% reduction in the population. If the same number of initial YLE males is released in a future generation, it will represent a higher overall frequency and, in a ratchet-like manner, cause a further decrease in the total population due to a higher proportion of females being eliminated.

Such a YLE would have geographically restricted impacts, since the initial frequency of immigrants escaping to a new population is expected to be low and the YLE should not increase in frequency (Burt and Deredec, 2018). Furthermore, a Y chromosome-linked technology would have a higher level of containment than other alternatives; in accordance with Haldane's rule, heterogametic F1 hybrid males from different species of the *An. gambiae* complex are sterile (Haldane, 1922). Thus, a YLE would be much more restricted to the species in which it is released than strategies based on autosomal constructs that have the potential to invade other species of the complex due to the gene flow that exists through hybrid females.

Burt and Deredec's modelling (2018) showed that, in an idealized case, a YLE tool is considerably more efficient than other existing self-limiting approaches such as the previously described bisexual-RIDL (bi-RIDL, equivalent to SIT), females-specific RIDL

(fs-RIDL) or an autosomal X-shredder, and equally good as a hypothetical fs-RIDL that drives in males (fs-RIDL-drive; the female-specific lethal gene affects only females, and males transmit it to 100% of the offspring). A YLE would be the only of these tools (together with fs-RIDL-drive) that is powerful enough to totally suppress a population by performing continuous releases of males that are equivalent to the 10% of the initial male population, while the rest of these technologies could not reduce the population size in half under those conditions (Figure 2.2 – Bottom). The differences among these technologies are also reflected in the release rates of males required to achieve the same level of suppression in the population, which are generally 10 to 50 times higher for RIDL, fs-RIDL or an autosomal X-shredder than for a YLE (Burt and Deredec, 2018). Hence, an efficient YLE would enormously reduce the costs associated to mass rearing and would facilitate the logistics for releases.

To build such YLE system, there are several requirements:

- A genetic construct must be integrated in a region of the Y chromosome that is transcriptionally active in the germline.
- It must modify the target gene in the male germline in order that the modification gets inherited by the offspring.
- The modification must be detrimental only for females, and it must cause dominant lethality or dominant sterility. This is because the offspring will inherit a mutagenized allele from the YLE male and an unmodified allele from the female.
- In the case that the modification itself is not dominant, deposition of the nuclease in the embryos would be required to induce mutagenesis of the unmodified allele.
- Potentially, the Y-linked editor could generate an X chromosome-linked allele that is even dominant in both sexes, since only the female offspring of such a male would receive the dominant allele. However, there are several conditions and caveats on building such a strain (see Chapter 6).

Alternatives to the use of nucleases might be used if paternal deposition can be achieved, such as a female-specific toxin to induce female embryo lethality. This possibility would show the same dynamics as the above-described YLE and is explored in more detail in Chapter 7.

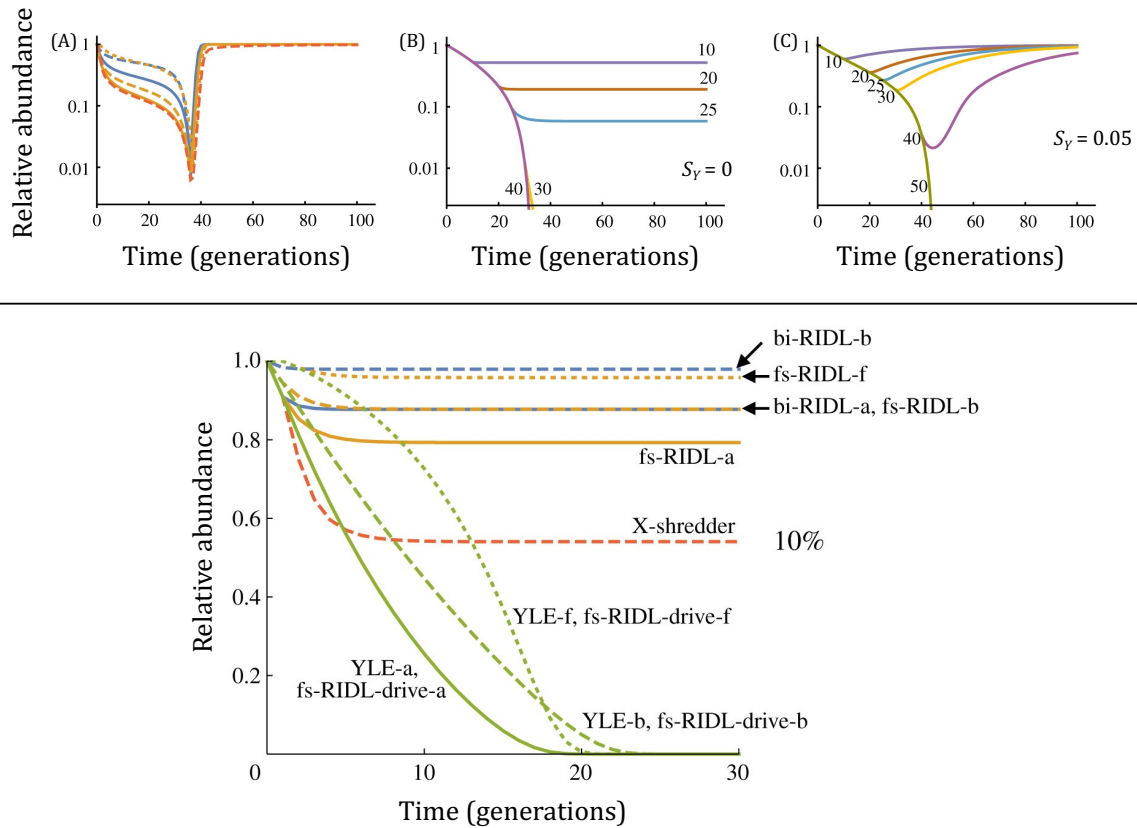


Figure 2.2 – Modelling on population control with different self-limiting strategies.

Top: The effect on the target population of halting releases of males with different self-limiting strategies. (A) Relative abundance of females after 36 generations of releases of males bearing one of the following constructs: bi-RIDL (blue), fs-RIDL (orange) or autosomal X-shredder (red). Alternative mode of action for these strategies include death after density-dependent mortality – i.e., at late larval/pupal stage in mosquito populations (suffix -a, solid lines), death before density-dependence – i.e., at embryonic stage in mosquito populations (-b, dashed lines) or female sterility (-f, dotted lines). For each construct, the release rates were such to cause 99% suppression after 36 generations. In all cases populations recover rapidly to their original levels after releases stop. (B, C) Relative abundance of females after releases of YLE-a constructs with releases stopping after the indicated number of generations. S_Y indicates the fitness cost associated to the YLE construct. The release rates are 5% of the initial male population.

Bottom: Population suppression with different self-limiting strategies. Relative abundance of females in the target population across the number of generations of releases. In each generation, a number of males equivalent to the 10% of the initial male population is released. The colour and line schemes are the already described ones for the plots on the top, with the addition of the fs-RIDL-drive and YLE (indistinguishable, green). Only these last two strategies would achieve complete suppression of the population with the indicated release rate.

The intrinsic rate of increase of the population (R_m) considered in these scenarios is 6. R_m is defined as the expected number of daughters produced by a female in the absence of density-dependent mortality. Figure adapted from Burt and Deredec (2018). Reproduced with permission of the Royal Society under the terms of the Creative Commons Attribution License.

2.2 The nature of the Y chromosome

Until now, genetic control strategies linked to the Y chromosome had not been developed in *Anopheles*, and only some of the necessary elements for these technologies have been recently integrated in the *Drosophila melanogaster* Y chromosome (Gamez et al., 2021). In this section, the structure of the Y chromosome is described, as well as the difficulties that it poses for its engineering and manipulation, with special focus on the *An. gambiae* Y chromosome.

2.2.1 Evolution of the Y chromosome

In many organisms, sex determination is controlled by chromosomes that are morphologically and genetically different. When the male is the heterogametic sex, the chromosomes are identified as X and Y chromosome (XX females, XY males), while when the heterogametic sex is the female, the chromosomes are referred to as Z and W (ZZ males, ZW females). Sex chromosomes evolved independently on many occasions in plants and animals (Charlesworth, 1991; Rice, 1996) and, as comprehensively reviewed by Bachtrog (2013), they derived from a pair of identical autosomes, where the Y (or W) chromosome underwent a process known as Y chromosome degeneration. Presumably, the first step in the evolution of sex chromosomes is the acquisition of a sex-determining locus (e.g., a male determining gene) by one of the originally identical autosomes, hence becoming a proto-sex chromosome. Next to this sex-determining region, sexually antagonistic mutations (e.g., male-advantageous and female-deleterious) arise and accumulate, restricting chromosomal recombination within this area to prevent the generation of sterile individuals. As a mean to impede recombination, chromosomal rearrangements (such as chromosomal inversions) might occur on the proto-sex chromosomes. The insertion of transposable elements and the emergence of novel mutations with sexually antagonistic fitness effects can expand the non-recombining region. The consequences of the lack of recombination include the multiplication of repetitive sequences but also the accumulation of loss-of-function mutations in Y-linked genes (while these processes are eluded in the X chromosome, for which recombination

still occurs in the females). As a result, large fragments of the Y chromosome comprising non-functional DNA can be lost (even the whole chromosome has been lost in some species), reducing its size and introducing additional morphological differences between both sex chromosomes. This evolutionary process leads to a situation where the X chromosome remains similar to the autosome from which it originated (with the potential increase of female-advantageous genes), whereas the Y chromosome has lost most of its genes and accumulated repetitive sequences. For these reasons, it is extremely difficult to sequence the Y chromosome and many species lack an assembly of this chromosome.

Despite the crucial role of the Y chromosome in gene regulation, spermatogenesis and male fertility (Cortez et al., 2014; Lawlor et al., 2021; Lemos et al., 2010; Sackton et al., 2011), its structural characteristics have hampered in-depth studies of this chromosome and most of the research has been performed in mammals and *Drosophila* (Bachtrog, 2013). One of the main differences between the mammalian and the *Drosophila* Y chromosome is the male determining role. In mammals, the presence or absence of the Y chromosome determines the sex of the individual, whereas *Drosophila* sex determination is controlled by the number of X chromosomes (Erickson and Quintero, 2007). Another difference between these two groups is the Y chromosome structure. In mammals, this chromosome has partial homology with the X chromosome and a big proportion of the male-specific region is euchromatic (30% in humans and 99.9% in mice) (Skaletsky et al., 2003; Soh et al., 2014), while in *Drosophila* the Y chromosome is entirely heterochromatic and does not bear homology to the X (Bachtrog, 2013; Carvalho et al., 2003). This is also reflected in the number of protein-coding genes, since the mouse male-specific region of the Y chromosome contains ~700 of these genes, whilst humans have 78 protein-coding genes on the Y chromosome (compared to ~800 on the X chromosome) and the *Drosophila* Y chromosome contains only 13 protein-coding genes (compared to ~2,200 on the X chromosome) (Bachtrog, 2013; Skaletsky et al., 2003; Soh et al., 2014).

2.2.2 The *Anopheles gambiae* Y chromosome

The Y chromosome in the *Anopheles* genus shares some characteristics with the mammalian one and others with the *Drosophila* one. In *Anopheles*, like in mammals, the sex of the individual is determined by the presence or the absence of the Y chromosome (Baker and Sakai, 1979), and a male-determining gene has been identified in *An. stephensi* (*Guy1*) and in the *An. gambiae* complex (*Yob*) (Criscione et al., 2016, 2013; Hall et al., 2016; Krzywinska et al., 2016). However, in common with *Drosophila*, cytogenetic studies concluded that the *Anopheles* Y chromosome is heterochromatic in its entirety, and posterior genomic analysis revealed its highly repetitive nature, which has prevented a complete assembly of this chromosome (Bonaccorsi et al., 1980; Hall et al., 2016; Holt et al., 2002; Krzywinski et al., 2004; Marchi and Mezzanotte, 1990).

Nonetheless, given the significance of *An. gambiae* mosquitoes for the human health as malaria vectors, together with the advantages presented by certain genetic control strategies linked to the Y chromosome, Hall et al. (2016) carried out a comprehensive analysis of the Y chromosome content in the different species of the *An. gambiae* complex, estimated to have diverged ~0.5 million years ago (Thawornwattana et al., 2018). This study revealed that ~49% of the *An. gambiae* Y chromosome is satellite DNA (highly repetitive tandems of DNA) and that transposable elements account for ~43.5% of the content on the Y. Furthermore, another ~5.5% are repetitive sequences that do not belong to the previous groups, and only 5 genes have been confirmed to be located on the Y chromosome, most of which seem to be multicopy (Hall et al., 2016, 2013; Krzywinska et al., 2016).

Strikingly, despite the relatively young species radiation of the *An. gambiae* complex, Hall et al. (2016) found large differences in the chromosome length, gene content and copy number variation not only among the different species of the complex but also between strains of the same species. An example of this is that among the 5 genes confirmed to be Y-linked in *An. gambiae*, only the male-determining one is consistently found on the Y chromosomes of all the species of the complex. However, in spite of the substantial differences on the chromosomal structure of the Y among sibling species of the *An. gambiae* complex, its function seems to be conserved and they can be interchanged at least between some of the species. Inasmuch as F1 hybrid males are sterile, the introgression of a Y chromosome in a different species of the complex did not seem likely.

Nevertheless, Hall et al. (2016) found evidences of this introgression and Bernardini et al. (2017) showed that a Y chromosome can be introgressed from an *An. gambiae* background to an *An. arabiensis* one without obvious effects on male fecundity, male competitiveness or large scale alterations in gene expression.

Contrary to *Drosophila*, the X and Y chromosome in *Anopheles* have partial homology and crossing-over between the sex chromosomes had been historically reported in certain anopheline species (Hall et al., 2016; Mitchell and Seawright, 1989; Sakai et al., 1979). Evidence of this process has also been found in *An. gambiae*, although the stability of X and Y-linked transgenes and translocations seems to indicate that this recombination is rare in this species (Bernardini et al., 2014; Hall et al., 2016). As it occurs in most organisms with heteromorphic sex chromosomes, the *An. gambiae* X and Y chromosomes undergo a process of epigenetic silencing in the male germline known as meiotic sex chromosome inactivation (MSCI) (Turner, 2007). This occurs during the meiotic phase of spermatogenesis to transcriptionally silence chromosomes that did not pair successfully or chromosomes with unsynapsed regions. Taxiarchi et al. (2019) provided evidence of MSCI in the X chromosome in the male germline of *An. gambiae* mosquitoes, adding to previous studies that showed silencing during meiosis of X and Y-linked transgenes, which will be addressed more in detail later (Alcalay et al., 2021; Galizi et al., 2014; Magnusson et al., 2012).

2.2.3 Transgene integrations on the Y chromosome

Exploring and manipulating the Y chromosome of *An. gambiae* has an extraordinary value for the development of vector control technologies. These include marking the Y chromosome with a transgene that allows to easily separate male and female larvae in high numbers for different studies or releases (such as the expression of a fluorescent protein), and the generation of genetic control tools linked to the Y chromosome that bias its inheritance and reduces the number of females in a population (such as the previously described Y-drive) (Burt and Deredec, 2018; Deredec et al., 2011; Geci et al., 2022).

However, the heterochromatic and highly repetitive structure of the chromosome has historically hampered its modification in a precise manner and the insertion of

transgenes has largely relied on random occurring integrations (Alcalay et al., 2021; Bernardini et al., 2014). The advent of CRISPR/Cas9 systems has provided a flexible and efficient way to generate transgenics in a very directed fashion, which allowed the site-specific directed insertion of a transgene in two different regions of the Y chromosome in *D. melanogaster* (Buchman and Akbari, 2019). Nonetheless, transgene expression from the Y chromosome in *An. gambiae* has only been reported twice for randomly integrated constructs (Alcalay et al., 2021; Bernardini et al., 2014). Upon confirmation that one of those loci was transcriptionally active, secondary integrations were performed within that locus (Bernardini et al., 2014).

Both in *Drosophila* and *Anopheles*, Y-linked transgene expression is highly affected by the structure of the chromosome. The highly repetitive heterochromatic areas cause that construct integrated in the Y chromosome undergo position effect variegation (PEV), by which transgenes are not consistently expressed but they are silenced in a stochastic pattern (Elgin and Reuter, 2013; Lloyd et al., 1999; Wallrath and Elgin, 1995). In *An. gambiae*, this has been observed in the variable expression of the fluorescent marker among males bearing the same Y chromosome (Bernardini et al., 2014).

2.2.3.1 The use of chromatin insulators to stabilize transgene expression

A series of studies in *Drosophila* showed that chromatin insulators can be used to prevent PEV. As reviewed by Bushey et al. (2008), there are multiple insulators that by different mechanisms protect gene expression from position effects. One of the best characterized chromatin insulators in *Drosophila* is known as *gypsy*, which is derived from the *gypsy* retrotransposon (Valenzuela and Kamakaka, 2006). The *gypsy* insulator sequences recruit the Suppressor of Hairy Wing proteins and, when two *gypsy* sequences are located in opposite orientations flanking a gene, the recruited proteins form a loop that isolates the elements within the loop from the chromatin effects of the surrounding regions (Bushey et al., 2008; Chen and Corces, 2001; Roseman et al., 1993; Valenzuela and Kamakaka, 2006; West et al., 2002). *gypsy* insulator sequences isolated from the *D. melanogaster gypsy* retrotransposon have been shown to improve transgene expression in *Drosophila melanogaster* but also in such a different organism as *Arabidopsis thaliana* (Markstein et al., 2008; She et al., 2010). More relevantly for the development of vector

control tools in *Anopheles*, Carballar-Lejarazu et al. (2013) provided evidence that these same sequences increased and stabilized transgene expression in *Anopheles stephensi*. However, all these experiments have been performed in autosomal loci, and the efficiency of *gypsy* insulators to increase expression of sex chromosome-linked transgenes remains to be conclusively validated. Two examples of the use of chromatin insulators in the Y chromosome of *D. melanogaster* and *An. gambiae* are described at the end of the following section.

2.3 Regulatory sequences for transgene expression in the *An. gambiae* male germline

One of the essential aspects of a YLE is to drive expression of the transgene in the male germline. Regardless of the specific design of the YLE, expression in the germline is key to have an effect in the embryos either by delivery of a toxin or by inheritance of a modified allele in a nuclease-based system. The recent burst in the creation of novel genetic control strategies has brought the exploration of regulatory sequences for expression in the germline. This segment contains a brief description of the promoters and terminators that have been tested in vector control tools developed in *An. gambiae* and for which transgene expression in the male germline has been reported.

Regulatory sequences for germline expression in these genetic control strategies are usually designed by selecting the upstream and downstream sequences of genes expressed in the germline. The length and the parts of the flanking sequences that are used in a transgene affect and modify the pattern and levels of expression. A clear example of this are the regulatory sequences of the *vasa* gene (AGAP008578), from which two different promoters were designed: *vas1*, with a longer upstream sequence but lacking two exons that are part of the 5'UTR; and *vas2*, with a shorter upstream sequence but including the whole 5'UTR of the gene (Papathanos et al., 2009). These two promoters, paired with the same *vasa* terminator sequence, were used to express GFP in two different constructs. *vas1* drove expression in the male germline but not in the female germline,

and even though it caused expression of GFP in the early spermatogenesis, it did not drive expression in the male germline stem cells. On the other hand, *vas2* directed expression in both male and female germline, caused an overall higher expression of GFP than *vas1* and drove expression in the male germline stem cells (Papathanos et al., 2009).

There are three major aspects that determine the suitability of a germline promoter for a certain genetic strategy: the stages at which it drives expression, the tissue specificity, and whether it induces deposition of transcripts or proteins in the embryos. Considering the transcriptional silencing that the Y chromosome undergoes during meiosis, a candidate promoter for a YLE strategy should be expressed in the early stages of spermatogenesis. If the target is female-specific, somatic expression would not be problematic, and paternal deposition might be advantageous in some circumstances.

Most of the technologies developed in *An. gambiae* are based on nuclease expression under the control of the regulatory sequences of one of the following genes: *vasa* (AGAP008578) (using the *vas2* promoter), *exu* (AGAP007365), *nos* (AGAP006098), *zpg* (AGAP006241), or $\beta 2$ (AGAP008622) (Carballar-Lejarazú et al., 2020; Galizi et al., 2016, 2014; Hammond et al., 2021a, 2016; Kyrou et al., 2018; Simoni et al., 2020; Windbichler et al., 2008, 2007). Figure 2.3 shows the expression levels of these genes in the male germline across four cell populations at different stages of spermatogenesis, based on RNA sequencing data obtained by Taxiarchi et al. (2019). From these genes, *vasa* is the one with the highest expression in premeiotic cells. The promoters *vas2*, *zpg*, *nos* and *exu* have been used in homing-based gene drives located within haplosufficient genes that have a key role in female fertility, providing practical information regarding the deposition, somatic expression, and homing rates motivated by each promoter (Carballar-Lejarazú et al., 2020; Hammond et al., 2021a, 2016; Kyrou et al., 2018; Simoni et al., 2020).

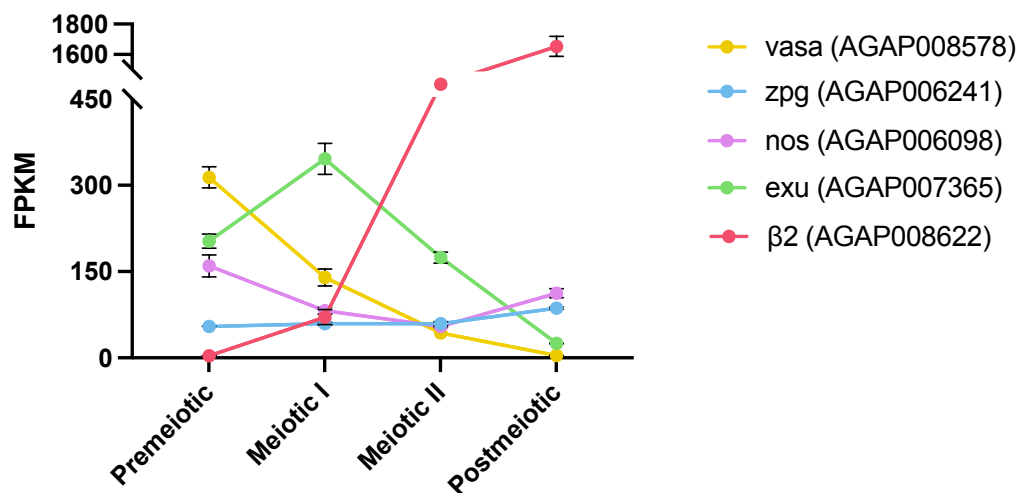


Figure 2.3 – Expression levels of *vasa*, *zpg*, *nos*, *exu* and $\beta 2$ genes in the male germline.

Expression levels of the genes whose regulatory sequences have been used in *Anopheles* for germline expression of transgenes, across four cell populations of the male testes at different stages of spermatogenesis: a premeiotic population, two meiotic populations and a postmeiotic population. For $\beta 2$, the data shown represents the transgene expression levels generated with the commonly used modified $\beta 2$ promoter instead of the endogenous $\beta 2$ gene because the differences in the early stages of spermatogenesis are exceptionally big. Expression levels were normalised and represented by number of Fragments Per Kilobase of transcript per Million mapped reads (FPKM). Data obtained from Taxiarchi et al. (2019).

Using the *vas2* promoter, remarkably high homing rates have been achieved from males and females in different loci. Depending on the locus, individuals bearing homing constructs based on Cas9 expression under *vas2* transmit the transgene to 87.3 – 99.6% of their progeny, and the rate of creating new mutations in the target is very low (Hammond et al., 2021a, 2016). It is noteworthy that *vas2* has been shown to cause maternal (but not paternal) deposition of transcripts and proteins in the embryos (Hammond et al., 2021a; Papathanos et al., 2009). Furthermore, the strong fertility cost observed in females bearing only one copy of the *vas2::Cas9* gene drive construct suggests “leakiness” of the *vas2* promoter, that drives expression in some somatic tissues where Cas9 activity cleaves the unmodified allele, heavily impairing female fecundity.

Gene drives based on *zpg* promoter to drive Cas9 have achieved homing rates similar to those obtained with *vas2*, and fertility costs in heterozygous females seem to indicate some degree of somatic leakiness, although much lower than *vas2* promoter (Hammond

et al., 2021a; Kyrou et al., 2018). Signs of paternal deposition have been reported in some cases, but this has not been confirmed and it seems to be locus-dependent (Hammond et al., 2021a; Kyrou et al., 2018).

Very similarly to *vas2*, homing rates obtained with the *nos* promoter are exceptionally high and it causes maternal deposition (Carballar-Lejarazú et al., 2020; Hammond et al., 2021a). On the other hand, the use *exu* promoter led to very low homing rates in males (~60% of inheritance of the homing construct) and no significant homing in females (Hammond et al., 2021a).

In the cases described above, the homing constructs targeted genes that are only required for females, hence the fertility of the males was not affected despite the somatic leakiness of some of the promoters.

Characteristically different from the rest, $\beta 2$ expression surges during meiosis and is particularly high in the later stages of spermatogenesis, which made it an attractive option to bias the inheritance of the Y chromosome by targeting the X chromosome in the late spermatogenesis without affecting the Y-bearing sperm (Galizi et al., 2016, 2014; Windbichler et al., 2008).

2.3.1 Transgene expression in the *An. gambiae* male germline from the Y chromosome

A major barrier for germline expression of transgenes inserted on the Y chromosome is the already described meiotic sex chromosome inactivation (MSCI). The silencing of this chromosome during meiosis reduces the spectrum of possibilities when choosing regulatory regions to drive expression in the germline from the Y. From the described promoters above, only *vas2* and $\beta 2$ have been reportedly tested in transgenes located on the Y chromosome of *An. gambiae*, with opposite results.

MSCI has prevented the development of a Y-drive strategy. A construct containing the same elements used in the autosomal X-shredder (a *I-PpoI* nuclease under the control of the $\beta 2$ promoter targeting repeated regions of the X chromosome) was used to independently generate two transgenic strains bearing this transgene on the Y

chromosome (Alcalay et al., 2021; Galizi et al., 2014). Males from both lines showed somatic expression of the fluorescent marker in the same way as in autosomal insertions. Nonetheless it was not possible to detect the nuclease transcript in the testes of these males, and no bias of the Y chromosome was observed in their progeny (Alcalay et al., 2021). This is in consonance with the transgene expression pattern under the $\beta 2$ promoter (Figure 2.3), that takes place during meiosis, when the Y chromosome is transcriptionally silenced.

On the other hand, *vas2* promoter has been successfully used to drive expression of GFP in the early stages of spermatogenesis from a transgene inserted in the Y chromosome, by site-specific integration in the Y-AttP line (Bernardini et al., 2014). In *D. melanogaster*, a *vasa* promoter was effectively used in a Y-linked construct to drive expression of Cas9 (Gamez et al., 2021). In this study, the levels of mutagenesis induced with the Y-linked Cas9 were lower than those of an autosomal *vasa::Cas9* construct, suggesting lower expression of the nuclease from the Y chromosome possibly due to position effect variegation. Nevertheless, Gamez et al. (2021) proved that sufficient Cas9 expression was achieved in the germline by homing an autosomal construct in a split-drive system. Interestingly, even though the construct designed in this study was flanked by a *gypsy* and a CTCF chromatin insulators, these did not prevent inconsistent expression of the fluorescent marker in different males and an overall lower Cas9 expression when compared to equivalent autosomal constructs.

The use of chromatin insulators on the *An. gambiae* Y chromosome has not yet been reported. However, unpublished studies used *gypsy* sequences in an effort to escape MSCI and achieve transgene expression from the Y chromosome during meiosis (Dario Meacci, personal communication). The results of this study are unclear on whether these chromatin insulators provide a significant increase on the expression of a Y-linked transgene.

Therefore, despite of being used in *D. melanogaster* and *An. gambiae*, it remains to be decisively demonstrated whether the use of chromatin insulators significantly improves the expression of transgenes in the Y chromosome.

2.4 Causing dominant female-specific sterility or lethality

The fact that there are no known regulatory sequences that cause significant paternal deposition in *An. gambiae* rules out the aim of depositing a nuclease or a female-specific toxin in the embryos as a first approach to develop a YLE. Additionally, available information on transgene expression from the Y chromosome has been regularly reported to be lower and less consistent than their equivalent constructs located in autosomes, which suggests that a YLE system targeting an essential non-sex-specific haploinsufficient target and relying on a rescue copy inserted on the Y chromosome might not be a sensible first approach either. Thus, the remaining alternative is finding a gene that can be targeted to create dominant mutations that cause female sterility or female lethality, while leaving males unaffected. There are two main ways of causing such dominant phenotypes: knocking out of an haploinsufficient gene or creating dominant negative mutations.

A gene is defined as haploinsufficient when the knockout of one allele, or the generation of a null allele, leads to an altered phenotype (Johnson et al., 2019; Veitia et al., 2018). These genes are also termed “dosage-sensitive”, as the amount of gene product generated by one allele is not enough to preserve the normal phenotype. In Chapter 6, haploinsufficiency is more deeply addressed, together with examples of haploinsufficiency that lead to lethality (haplolethality).

Conversely, dominant negative mutations generate a truncated gene product that interferes with the normal one produced by the wild-type allele, and consecutively trigger an alteration in the phenotype (Herskowitz, 1987; Veitia, 2002). The most common example are oligomeric proteins, where the truncated proteins interact with the unmodified ones rendering the oligomers non-functional. Thus, dominant negative mutations usually produce proteins with an intact functional domain but with another protein domain that is missing or strongly modified (Herskowitz, 1987). If such dominant negative mutations could be induced in genes with an essential role only in females, these could potentially be integrated as a target of a YLE strategy.

2.4.1 Identified dominant negative mutations in insects

Dominant negative mutations in female-specific genes have been described in different species of Diptera. An example of this is found in the gene coding for indirect flight muscle actin in *Aedes aegypti* and *Culex quinquefasciatus* (*AeAct4* and *CxAct4*, respectively) (Navarro-Payá et al., 2020; O’Leary and Adelman, 2020). These genes, with female-specific expression, code for actin filaments that polymerize to form indirect flight muscles required for flying. Navarro-Payá et al. (2020) found that certain in-frame mutations within the coding sequence of these genes triggered a female-specific dominant loss of flight phenotype. However, complete knock-out of *AeAct4* resulted in a recessive phenotype, showing that the flightless phenotype was not caused by the reduced dosage of wild-type protein, but rather to the interaction of normal actin filaments with the modified ones produced by the mutated allele, preventing the correct assembly of the indirect flight muscles.

Other examples of female-specific dominant negative mutations are found in the *doublesex* gene (*dsx*) in *D. melanogaster*. This gene has two sex-specific isoforms and codes for a transcription factor that dimerizes and plays an essential role in sex determination (Burtis and Baker, 1989; Hildreth, 1965). In this case, the reported dominant negative mutations do not generate a truncated protein, but they prevent the splicing of the female isoform and induce the expression of the male one, causing an intersex phenotype in chromosomally female individuals (Nagoshi and Baker, 1990; Nöthiger et al., 1987). The *dsx* gene, the phenotypes caused by mutant alleles and the mechanism that triggers such phenotypes are described and analysed in the subsequent chapters, given their paramount relevance for this thesis.

Nevertheless, despite these examples in relatively closely related species, no female-specific dominant negative mutations that cause lethality, sterility or a phenotype beneficial for a genetic control tool (such as loss of flight), have been reported in *Anopheles*. Equally, no female-specific haploinsufficient genes have been described in this genus.

2.5 Concluding remarks

The highly repetitive and heterochromatic nature of the *An. gambiae* Y chromosome has prevented the creation of Y-linked genetic control technologies up to now, and it heavily restricts the course of action to develop a proof-of-concept YLE. The scarce information on transgene expression from the Y chromosome in *An. gambiae* and the difficulty to find non-repetitive regions encourage the use of strains with transcriptionally active transgenes for secondary integrations within those loci over integrations in new areas of the chromosome. Furthermore, the lack of known regulatory sequences that cause paternal deposition in the embryos, together with the position effect variegation observed in Y chromosome-linked transgenes, discard the use of a toxin-antidote system to induce lethality in female embryos as a practical approach at this stage. Consequently, the use of nucleases to generate mutations in female-specific targets appears as the most sensible proposal for a proof-of-concept YLE. In this framework, the CRISPR/Cas9 system arises as the best option thanks to its flexibility and its vastly demonstrated efficacy in genetic control strategies.

Another obstacle is found in the transcriptional silencing of the Y chromosome during meiosis, which reduces the potential regulatory regions that can be used to induce germline expression of Cas9 from a Y-linked transgene. The *vas2* promoter and *vasa* terminator are the only regulatory sequences that have been successfully used to drive transgene expression from the Y chromosome in *An. gambiae*. This, together with the germline expression pattern of alternative promoters and terminators implemented in other genetic control tools, suggest that *vas2* would be a rational first choice to express Cas9 in a YLE construct. While there is evidence that shows that *vas2* causes certain expression in somatic tissues, this should not be a concern for the fitness of the males as long as the target of the YLE is only essential for females.

Nevertheless, in order to develop an efficient YLE, it will be necessary to find a female-specific target gene with dominant negative mutations or an haploinsufficient gene that has a vital role only in females.

3 Aims

The main aim of this project was to explore the necessary aspects to develop a self-limiting Y chromosome-linked genome editor, with the ultimate goal of integrating them in a genetic tool that can be used for population suppression in *Anopheles gambiae*.

These essential features were investigated as described below and addressed in the subsequent chapters:

- **Generating germline nuclease activity from a transgene integrated on the Y chromosome.** This was performed by creating a proof-of-concept YLE with a Cas9 gene under the *vas2* promoter and a gRNA targeting the female isoform of *dsx*, which allowed to explore in parallel the potential use of *dsx* as a target of this strategy.
- **Validating candidate haplo-insufficient target genes located in the X chromosome,** with the goal of finding alternative targets that cause dominant lethality in the female offspring.
- **Discovering regulatory sequences that could be used to achieve paternal deposition of transcripts in the embryos.** These could eventually be integrated in a YLE that deposits female-specific toxins in the progeny.

4 Creating and evaluating a proof-of-concept Y chromosome-linked genome editor in *Anopheles gambiae*

4.1 Designing a genetic construct for the creation of a Y chromosome-linked genome editor in *Anopheles gambiae*

Following on what has been described in Chapter 2, a YLE genetic construct was designed based on Cas9 expression under the *vas2* promoter and *vasa* terminator. Regarding the target, a gRNA recognising the female isoform of the *doublesex* gene was incorporated in the construct.

4.1.1 The *doublesex* gene as a target of a YLE strategy

First identified in *Drosophila melanogaster* in 1965, *doublesex* (*dsx*) is a gene with an essential role in sex determination (Hildreth, 1965). Since then, orthologs to *dsx* have been found in many insect species, where it is highly conserved (Price et al., 2015). In multiple species, including *D. melanogaster* and *An. gambiae*, *dsx* has two sex-specific isoforms that code for transcription factors with conserved DNA-binding domains and dimerization domains (An et al., 1996; Burtis and Baker, 1989; Price et al., 2015; Scali et al., 2005). In *D. melanogaster*, the two sex-specific spliced isoforms share the exons 1, 2 and 3, while exon 4 is only present in the female transcript (*Dmdsx^F*), and exons 5 and 6 are exclusive of the male isoform (*Dmdsx^M*) (Figure 4.1 - Top). Similarly, in *An. gambiae* (*Agdsx*) the two sex-specific transcripts of *dsx* share exons 1, 2, 3 and 4, whereas exon 5 is exclusive of the female transcript and exons 6 and 7 are present in both isoforms (Figure 4.1 - Middle). Thus, *Agdsx* has a female-specific exon (exon 5), but not a male-specific one. The coding sequence of both transcripts starts in exon 2 and, while the male isoform ends at exon 6, the female one ends at the beginning of exon 5. A big portion of exon 5 and the whole of exons 6 and 7 constitute the 3' UTR of the female isoform.

This gene, with such a crucial role and a female-specific exon, arises as a very interesting target for genetic control strategies. However, the disruption of the exon 5 coding sequence is recessive and only causes an intersex phenotype in homozygosity in chromosomally female individuals (Kyrou et al., 2018), which makes it an appropriate target for a gene drive strategy but, a priori, unsuitable for the YLE described here. Nonetheless, the availability of a transgenic line with a modified allele containing a gene

coding for GFP under a *3xP3* promoter at the beginning of the coding sequence of exon 5, makes it a convenient target to evaluate Cas9 activity levels by measuring the homing of this allele when in presence of the YLE. In addition, the effects of different insertion-deletion mutations within this sequence in *An. gambiae* when paired with a wild-type allele remain unknown, hence targeting this region might provide useful insight about the female isoform of the *dsx* gene.

For these reasons, the genetic construct designed to generate a proof-of-concept YLE contained a gRNA targeting the coding sequence of the female-specific exon 5 of the *dsx* gene (Figure 4.1 - Bottom).

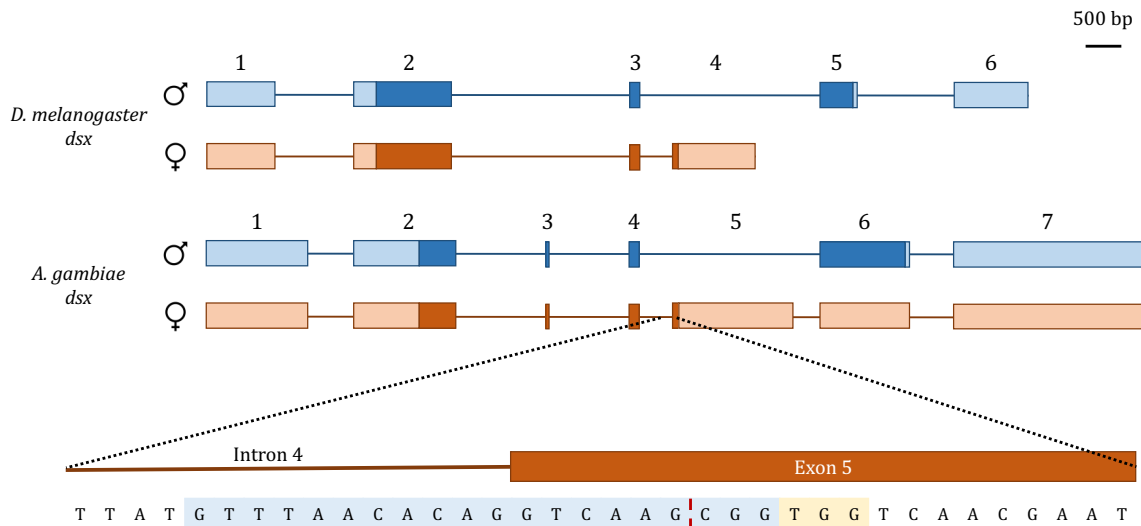


Figure 4.1 – Transcript comparison of the *doublesex* gene in *Drosophila melanogaster* and *Anopheles gambiae*. In *D. melanogaster* (top), the *dsx* gene consists of 6 exons (coloured boxes) and in *An. gambiae* (middle) it is comprised by 7 exons. In both species there are two sex-specific isoforms of the *dsx* gene, with the female isoform (orange) containing an exon that is absent in the male transcript (blue). Transcription of both isoforms starts at a common start codon in exon 2. Light colours represent untranslated regions. Introns are not drawn to scale. The target sequence of the YLE, located at the intron 4 – exon 5 boundary, is shown at the bottom. The gRNA sequence is highlighted in light blue and the PAM of the target site is highlighted in light yellow. The cut site is represented by a red dashed line.

4.2 Generating a proof-of-concept Y chromosome-linked genome editor targeting the female isoform of *doublesex*

Two existing plasmids contained all the necessary elements to create such a YLE (referred to as YLE^{dsx} hereafter). The p16510 plasmid (Kyrou et al., 2018) contains a human-codon-optimized sequence of the *Streptococcus pyogenes* Cas9 gene (*hCas9*) under the control of the *vas2* promoter and *vasa* terminator; the gRNA shown in Figure 4.1 targeting the beginning of the *dsx* female-specific exon driven by the ubiquitously expressed *U6* promoter; and a gene coding for RFP under the *3xP3* promoter. This construct is flanked by two Φ C31 *attB* recombination sequences for site-specific integrations, and by two *piggyBac* inverted repeats for transposon-mediated integration. A modified version of p16510 contains an eCFP marker gene (p16510-eCFP) (Figure 4.2 and Figure 4.3).

In order to insert the YLE^{dsx} construct in the Y chromosome, I used two independently originated strains harbouring Y-linked sequences for site-specific recombination. Both of these lines, with an *An. gambiae* G3 strain background, had shown transgene expression, and a secondary integration in one of them (Y-AttP) successfully expressed GFP in the germline using the *vas2* promoter (Bernardini et al., 2014). The highly repetitive nature of the Y chromosome and the lack of a complete assembly hinder the understanding of the respective locations of the transgenes in the two strains. However, the purpose of using two different integrations was to evaluate whether one of them would be less prone to position effect variegation in transgene expression.

The Y-AttP line created by Bernardini et al. (2014) was used to generate the first of the YLE^{dsx} strains (YLE^{dsx}-A). This line has a Φ C31 *attP* recombination sequence located between the *3xP3* promoter and the gene coding for RFP, in a way that any integration in this site inactivates the expression of that marker (Figure 4.2) (Methods 4.8.3.1). Since the Y-AttP line has only one *attP* recombination sequence and the p16510-eCFP has two *attB* sequences, the transgene could have been integrated in two different orientations, which may influence its expression levels. PCR reactions performed amplifying the *attB* sequences of the plasmid in genomic DNA samples from G1 transgenic males confirmed that all the transgenic males had the construct integrated in the same orientation. Upon this confirmation, the YLE^{dsx}-A was established from one single individual, ensuring that all the males from each strain had the same Y chromosome.

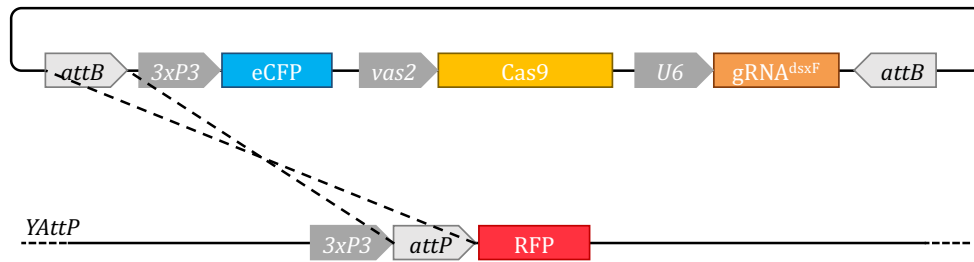


Figure 4.2 – Generation of the YLE^{dsx}-A strain by integrating the p16510-eCFP plasmid in the Y-AttP line. The p16510-eCFP plasmid used to create the YLE^{dsx}-A strain contained a gene coding for Cas9 under the *vas2* promoter, a gRNA targeting the female-specific exon 5 of the *dsx* gene, and an eCFP marker gene. The construct was flanked by two *attB* sequences, one of which was used for ΦC31 integrase mediated recombination with the *attP* sequence located in the Y-linked transgene of the Y-AttP line.

The second YLE^{dsx} strain (YLE^{dsx}-B) was site-directed integrated on a transgenic line provided by Jaroslaw Krzywinski (personal communication). This strain contains a transgene with a GFP marker gene flanked by two ΦC31 *attP* recombination sequences. In this case, mosquito embryos were microinjected with the p16510 plasmid (Figure 4.3) (Methods 4.8.3.1). While it was not verified whether integration in the two possible orientations had occurred, after molecular confirmation of the transgene insertion (by PCR and subsequent sequencing) the YLE^{dsx}-B was established from one single individual to guarantee the presence of the same Y chromosome in all the males of the strain.

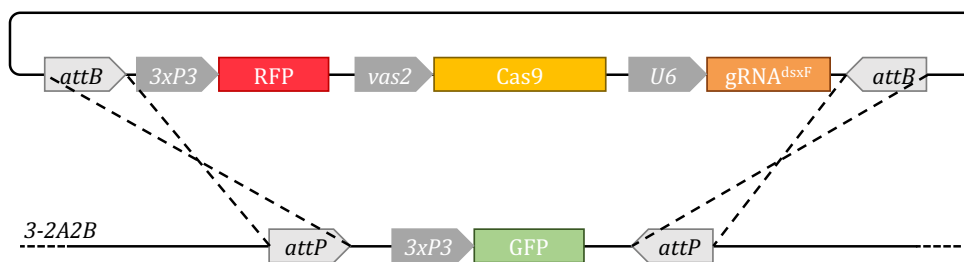


Figure 4.3 – Generation of the YLE^{dsx}-B strain by integrating the p16510 plasmid in the 3-2A2B line. The p16510 plasmid used to create the YLE^{dsx}-B strain contained a gene coding for Cas9 under the *vas2* promoter, a gRNA targeting the female-specific exon 5 of the *dsx* gene, and an RFP marker gene. The construct was flanked by two *attB* sequences, which were used for ΦC31 integrase mediated cassette exchange with the *attP* sequences located in the Y-linked transgene of the 3-2A2B line.

4.3 Evaluating the Cas9 activity levels in the male germline from the Y chromosome-linked transgenes

An essential aspect of this YLE is driving sufficient germline expression of nucleases from the Y chromosome. Since previous research showed that nuclease activity expressed under the *vas2* promoter leads mainly to homology-directed repair (Hammond et al., 2021a, 2016), an assay was designed to measure Cas9 activity based on the inheritance bias of a marked allele. Kyrou et al. (2018) generated a female-specific null mutation in *dsx* by knocking in a gene coding for GFP under a *3xP3* promoter within the coding sequence of the female-specific exon of *dsx*. This allele (referred to as *dsxF⁻* hereafter) lacks the sequence recognized by the gRNA present in the YLE^{*dsx*} constructs, rendering it resistant to cleavage. Hence, this can be leveraged to easily evaluate the Cas9 activity from the YLE^{*dsx*} strains by measuring the homing of *dsxF⁻* when in presence of the YLE transgene (Figure 4.4).

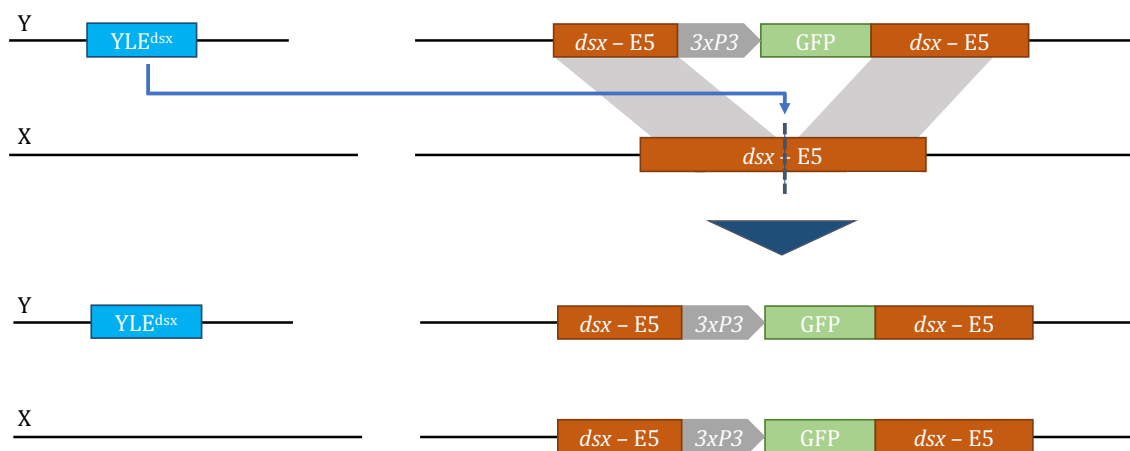


Figure 4.4 – Homing of the *dsxF⁻* allele to evaluate Cas9 activity from the YLE^{*dsx*} transgenes.

The Cas9 activity levels from the YLE^{*dsx*} strains were evaluated by using the *dsxF⁻* allele, that contained a gene coding for GFP within the target sequence in the *dsx* exon 5. Because Cas9 was expressed under the *vas2* promoter, and in the early germline of *Anopheles* the preferred repairing mechanism is homology-directed repair, observing the homing and biasing in the inheritance of the *dsxF⁻* allele was a straightforward approach to measure Cas9 activity levels.

The same procedure was followed for both YLE^{*dsx*}-A and YLE^{*dsx*}-B strains. First, YLE^{*dsx*} males heterozygous for the *dsxF⁻* allele (YLE^{*dsx*} ; *dsxF⁻* / +) were obtained through the

crosses reported in Methods 4.8.4.1. Given that an important aspect of this experiment was to assess the variability in Cas9 activity among different males from the same strain, these YLE^{dsx} ; *dsxF*^{-/+} were then mated with wild-type females as described in Methods 4.8.4.1 to ensure that each batch of progeny analysed for homing has a unique male progenitor. Larvae in the progenies from at least 30 females were screened for the presence of the GFP marker associated to the *dsxF*⁻ allele.

4.3.1 Y-linked nuclease activity leads to homing of *dsx* null alleles

High Y chromosome-linked nuclease activity was observed in both YLE^{dsx} lines. Cas9 cleavage led to 81.5% and 92.8% of the progeny inheriting the *dsxF*⁻ allele in YLE^{dsx}-A and YLE^{dsx}-B strains respectively (Figure 4.5). Whereas the results obtained in the two lines are not significantly different (P=0.0694; Mann-Whitney test), the variability in YLE^{dsx}-A is higher whilst the homing rates in the progenies of different YLE^{dsx}-B males are more consistent. This suggests that the YLE^{dsx}-A locus might be more exposed to position effect variegation and thus to a higher variability in transgene expression.

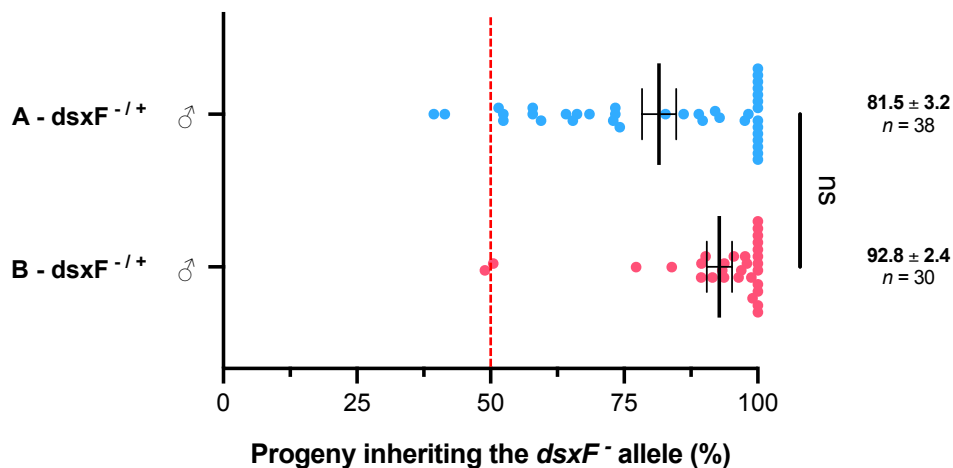


Figure 4.5 – Inheritance rates of the *dsxF*⁻ allele caused by the YLE^{dsx}. Percentage of the progeny from YLE^{dsx}-A ; *dsxF*^{-/+} males (top) and YLE^{dsx}-B ; *dsxF*^{-/+} males (bottom) that inherited the *dsxF*⁻ allele. In both cases, this percentage is considerably higher than the typical Mendelian inheritance of 50% (red line), thus indicating high Cas9 activity in the germline of those males. Each dot corresponds to the progeny of a different couple. A higher variance is observed in the progeny of YLE^{dsx}-A, suggesting higher exposure to position effect variegation. Homing rates induced by both lines are not significantly different (P=0.0694; Mann-Whitney test). The numbers on the right show the mean and the standard error of the mean for each YLE^{dsx} line, and the size of each sample (n).

In agreement with prior observations that homology-directed repair is far preferred over non-homologous end joining (NHEJ) or microhomology-mediated end joining (MMEJ) as the dominant repair pathway in the early stages of the germline of *Anopheles* mosquitoes (Hammond et al., 2021a, 2016), new mutations in the target were found at very low frequencies (Table 4.1). Amplicon sequencing of the target site was performed in progenies from YLE^{dsx} ; *dsxF*^{-/+} males (for both YLE^{dsx}-A and YLE^{dsx}-B strains) with low homing rates to check whether new end joining mutations had been generated and prevented the bias of the *dsxF*⁻ allele, hence explaining the low homing rates.

It is complicated to extrapolate the number of heterozygous individuals from the allelic frequencies obtained, since the *dsxF*⁻ allele cannot be amplified for the amplicon sequencing assay and therefore the results do not show the actual allele frequency in the offspring. However, all the new mutations combined in the progenies of YLE^{dsx}-A and YLE^{dsx}-B males account for less than 1% of the reads (Table 4.1). These results suggest that the low homing rates observed in some progenies were due to low Cas9 activity in the progenitor males of those progenies rather than to the generation of new mutations that prevented the homing of the *dsxF*⁻ allele.

Looking at the data from the YLE^{dsx}-B males' offspring, it seems that progenies with higher homing rates also observe higher presence of other mutations in the target. This could be explained by the fact that higher Cas9 activity levels lead to higher homing rates but also to the generation of other mutations more frequently, or because higher (but still very low) doses of Cas9/gRNA are deposited from those males with higher Cas9 activity in the germline. However, the same pattern is not observed in the YLE^{dsx}-A progenies and therefore this should be explored further for a well-founded explanation.

Table 4.1 – Frequencies of new end joining generated mutations at the *dsx* target in the progenies of YLE^{dsx} ; *dsxF*^{-/+} males. Amplicon sequencing was performed in pooled individuals of the offspring of YLE^{dsx} ; *dsxF*^{-/+} males to corroborate whether new mutations were being generated and inherited by those individuals that did not inherit the *dsxF*⁻ allele. Each row corresponds to the progeny of a different couple (A-1 to A-8 are batches of progenies from YLE^{dsx}-A ; *dsxF*^{-/+} males, and B-1 to B-6 are batches of progenies from YLE^{dsx}-B ; *dsxF*^{-/+} males). The second column indicates the inheritance bias of the *dsxF*⁻ allele (i.e, percentage of the offspring that inherited this allele) as a reference of the homing rates observed in each of these progenies. The third column shows the frequency of all newly generated end joining (EJ) mutations in the progeny that were absent in the progenitor male. The fourth column indicates the frequency of the most common mutation found in each progeny. The frequencies of new mutations are remarkably low, suggesting that the observed lack of *dsxF*⁻ allele homing in some progenies was due to lack of (or reduced) Cas9 activity rather than to the generation of new mutations that prevented homing of said allele.

Progeny	Inheritance of <i>dsxF</i> ⁻ allele (%)	Frequency of new EJ mutations (%)	Frequency of most common mutation (%)
A-1	73.33	0.09	0.03
A-2	68.50	0.16	0.04
A-3	64.10	0.38	0.30
A-4	59.46	1.10	1.04
A-5	52.38	0.15	0.06
A-6	52.38	0.27	0.11
A-7	51.51	0.09	0.07
A-8	41.46	0.03	0.01
Average A	-	0.29	0.21
B-1	89.47	2.89	1.97
B-2	89.39	0.53	0.06
B-3	83.93	0.58	0.08
B-4	77.22	0.20	0.02
B-5	50.51	0.04	0.01
B-6	48.94	0.01	0.01
Average B	-	0.71	0.36

4.3.2 Homing of a female-specific dominant negative mutation represents a viable strategy for a Y chromosome-linked genome editor

Upon the generation of the YLE^{dsx} transgenic males, a female-specific dominant negative mutation in *dsx* was found in both YLE^{dsx}-A and YLE^{dsx}-B strains. This mutation (referred to as *dsxF^{A11}* hereafter) is an 11 base-pair deletion within the coding sequence of the female-specific exon of *dsx*, making it unrecognisable by the gRNA of the YLE^{dsx} construct. It causes an intersex phenotype in chromosomally female individuals that are heterozygous for the mutation, rendering them completely sterile, while it does not have any effect in males. The *dsxF^{A11}* mutation and its discovery and characterisation is described in detail in Chapter 5, while here its potential role in an efficient YLE strategy is analysed.

Given that Cas9 activity in both YLE^{dsx} strains led mainly to homing of a modified allele in the target rather than to the creation of new mutations, I investigated whether it was possible to bias the inheritance of a female-specific dominant mutation driven by the same Y-linked Cas9 sources, as this could represent a viable strategy to cause female sterility in the progeny of YLE^{dsx} males.

The same procedure was again followed for both YLE^{dsx}-A and YLE^{dsx}-B strains. YLE^{dsx} males that were heterozygous for the dominant *dsxF^{A11}* allele (YLE^{dsx} ; *dsxF^{A11}/+*) were isolated following the procedure described in Methods 4.8.4.2. These males were mated with wild-type females and their progenies were screened at pupal stage to estimate the frequency of females with an intersex phenotype. All the females from each batch of progenies were then pooled to perform genomic DNA extraction and amplicon sequencing on the target site to confirm that the *dsxF^{A11}* deletion was present at the expected frequencies. It was found that, similarly to what occurred with *dsxF⁻*, the *dsxF^{A11}* allele was inherited by an average of 72.7% and 94.5% of the offspring of YLE^{dsx}-A and YLE^{dsx}-B males respectively (Figure 4.6). This shows that an efficient YLE strategy could be planned around the homing of a dominant mutation that is only detrimental for females.

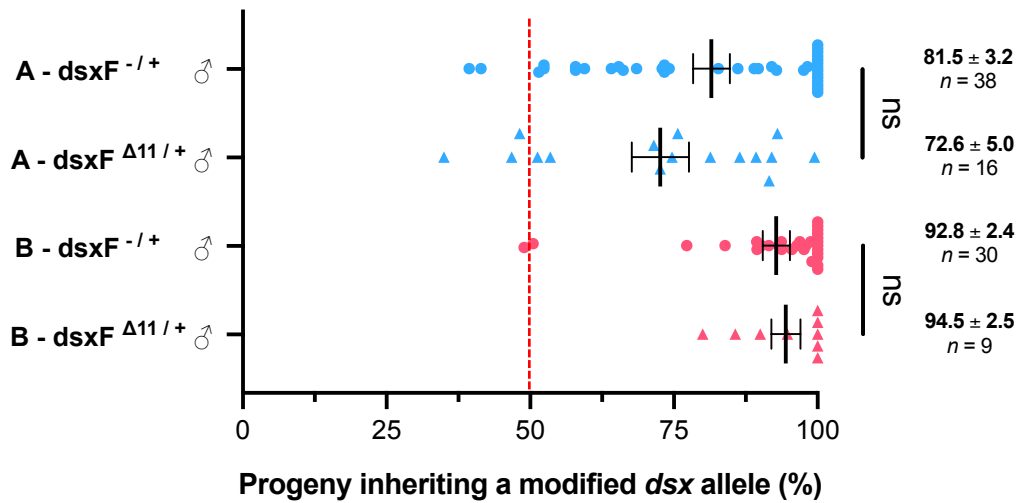


Figure 4.6 – Inheritance rates of a modified *dsx* allele caused by the YLE^{dsx} . Percentage of progeny from YLE^{dsx} -A males (blue) and YLE^{dsx} -B males (pink) that inherited the *dsxF*⁻ (circles) or the *dsxF*^{Δ11} allele (triangles). In all the cases, the percentage is considerably higher than the typical Mendelian inheritance of 50% (represented by the red line). Each dot corresponds to the progeny of a different couple. Homing rates of the *dsxF*⁻ and the *dsxF*^{Δ11} alleles are not significantly different in YLE^{dsx} -A males ($P=0.0641$; Mann-Whitney test) nor in YLE^{dsx} -B males ($P=0.7714$; Mann-Whitney test). The numbers on the right show the mean and the standard error of the mean, and the size of each sample (n). The higher differences observed in the mean homing rates of the different alleles in YLE^{dsx} -A males is in consonance with a higher exposure of this locus to position effect variegation than the YLE^{dsx} -B locus.

In order to be a viable strategy, the YLE^{dsx} males must faithfully home the dominant *dsxF*^{Δ11} allele continuously rather than generating new end joining mutations. Amplicon sequencing was performed to determine the rate of generating novel alleles. Consistently with the previous results reported, new mutations at the target site were found at very low frequencies. Evaluating the progenies from YLE^{dsx} ; *dsxF*^{Δ11/+} males, I found that the aggregate of all other mutations was present in ~1% of individuals (Table 4.2 and Table 4.3). Since progenies from a single female rarely contain more than 200 individuals, if a mutation is present in a frequency lower than 0.25%, it seems to be more likely due to either contamination or a low load of paternal deposition of the complex Cas9/gRNA that created mutations in the zygotes rather than to the presence of individuals that inherited

this mutation from the YLE^{dsx} progenitor. This is because 50% of the alleles in the progeny will be wild-type contributed by the female progenitor, hence a frequency of 0.25% of a certain mutation would correspond to 1 individual heterozygous for said mutation in a progeny of 200 individuals. In 13 out of 16 progenies from YLE^{dsx}-A males, all the newly created mutations were found in a frequency $\leq 0.07\%$, which could not account for the presence of an individual heterozygous for said mutation and therefore it is not likely to have been generated in the male progenitor's germline. Only in 3 progenies a mutation was found in a frequency higher than 0.25% (0.59, 1.20 and 2.14%), which would be equivalent to 1.18, 2.40 and 4.28% of the individuals in these progenies carrying those mutations in heterozygosity (Table 4.2). Interestingly, these 3 progenies also showed high homing rates of *dsxF^{A11}* (>91% of individuals inherited *dsxF^{A11}*).

In the case of YLE^{dsx}-B, the amplicon sequencing was performed separating normal females and females with an intersex phenotype. Comparable to what was observed for YLE^{dsx}-A, only one progeny of YLE^{dsx}-B ; *dsxF^{A11}/+* males had a different mutation in a frequency higher than 0.25%. One mutation was found in 28.36% of the intersex individuals in the progeny B-5, suggesting that it was another dominant mutation (Table 4.3). The two females identified as wild-type from progeny B-6 were actually heterozygous for the *dsxF^{A11}* allele, which seems to indicate that it was homed in the male progenitor rather than generated again. It is unclear whether the *dsxF^{A11}* allele did not trigger the intersex phenotype in these two females or if they were wrongly classified as wild-type. In any case, the average of individuals with new mutations among the intersex pool was ~3%, but this percentage would be slightly reduced by including the data from the pool of females with a normal phenotype. Overall, the results in the two YLE^{dsx} strains are highly similar and show that the frequency of new mutations generated in the male germline is very low, which is highly beneficial for the YLE^{dsx} strategy based on homing a female-specific dominant mutation. The analysis of additional dominant mutations in the target site, as well as recessive ones, is shown in Chapter 5.

Table 4.2 – Frequencies of new end joining mutations at *dsx* target in the progenies of YLE^{dsx-A}; *dsxF*^{Δ11/+} males. Amplicon sequencing was performed in pooled individuals of the progenies of YLE^{dsx-A}; *dsxF*^{Δ11/+} males to corroborate if new mutations were being generated in the male germline and inherited by the offspring. Each row corresponds to a different set of offspring. The second column indicates the frequency of the *dsxF*^{Δ11} allele in the progeny. The third column shows the frequency of the combination of all the end joining (EJ) mutations found in each progeny. The fourth column indicates the frequency of the most common mutation (besides *dsxF*^{Δ11}) in each progeny. Because a wild-type allele will always be inherited from the female, the percentage of individuals bearing each mutation corresponds to the double of the frequency of said mutation (i.e., if the frequency of the *dsxF*^{Δ11} allele is 50%, this reflects that 100% of the individuals bear the *dsxF*^{Δ11} allele). The sum of all mutations in progeny A-1 is >50%; this might be explained by either contamination or very low loads of nuclease deposition. Only in 3 progenies (highlighted in light yellow – A-1, A-2, A-4) there was a new mutation found in a frequency high enough to represent heterozygous individuals (0.59 – 2.14%).

Progeny	Frequency of <i>dsxF</i> ^{Δ11} allele (%)	Frequency of new EJ mutations (%)	Frequency of most common mutation (%)
A-1	49.73	1.58	1.20
A-2	46.51	0.72	0.59
A-3	46.01	0.27	0.04
A-4	45.79	2.81	2.14
A-5	44.64	0.34	0.04
A-6	43.23	0.44	0.07
A-7	40.68	0.39	0.05
A-8	37.83	0.17	0.02
A-9	37.33	0.29	0.03
A-10	36.31	0.20	0.03
A-11	35.74	0.22	0.03
A-12	26.75	0.11	0.03
A-13	25.64	0.17	0.02
A-14	24.10	0.16	0.02
A-15	23.39	0.13	0.02
A-16	17.49	0.09	0.01
Average	-	0.51	0.27

Table 4.3 – Frequencies of new end joining mutations at *dsx* target in the progenies of YLE^{dsx-B}; *dsxF^{Δ11/+}* males. Amplicon sequencing was performed in pooled females of the progenies of YLE^{dsx-B}; *dsxF^{Δ11/+}* males, separating the ones showing an intersex phenotype from those that had a normal phenotype. Each row corresponds to a different set of offspring. The second column shows the frequency of the *dsxF^{Δ11}* allele in the intersex pool, and the third column indicates the frequency of all newly created end joining (EJ) mutations in the male germline and inherited by females showing an intersex phenotype. The fourth column shows the frequency of the wild-type allele in the pool of females with a normal phenotype (expected to be 100% unless they inherited a newly generated recessive mutation), and the fifth column indicates the frequency of the aggregate of mutations in this pool. Because a wild-type allele will always be inherited from the female, the percentage of individuals bearing each mutation corresponds to the double of the frequency of said mutation (i.e., if the frequency of the *dsxF^{Δ11}* allele is 50%, this reflects that 100% of the individuals bear the *dsxF^{Δ11}* allele). Only one progeny (B-5, in yellow) had a different mutation to *dsxF^{Δ11}* in a high enough frequency to represent heterozygous individuals, which was found in the intersex pool, suggesting that it was a dominant negative mutation. Only one progeny (B-6, in orange) had a mutation in high frequency in the pool of normal-looking females; this corresponds to 2 individuals that were heterozygous for the *dsxF^{Δ11}* allele, suggesting that they had inherited the allele from their progenitor and were either misclassified as wild-type or showed no intersex phenotype at pupal stage. The average results seem to indicate that the generation of new mutation in the germline of YLE^{dsx-B} males is very low. The average of mutations found in the normal pool does not include the progeny where the *dsxF^{Δ11}* allele was found, since it seems more likely it was homed than created by MMEJ (*).

Progeny	Freq of <i>dsxF^{Δ11}</i> allele (%) – intersex pool	Freq of new EJ mutations (%) – intersex pool	Freq of wild-type allele (%) – normal pool	Freq of new EJ mutations (%) – normal pool
B-1	50.05	0.05	-	-
B-2	50.87	0.07	-	-
B-3	51.49	0.07	-	-
B-4	50.23	0.12	-	-
B-5	37.51	14.26	-	-
B-6	50.88	0.06	49.07	50.93
B-7	46.51	0.07	99.99	0.01
B-8	51.33	0.05	99.99	0.01
B-9	51.39	0.06	99.98	0.02
Average	-	1.65	-	0.01*

4.4 The YLE^{dsx} transgenes do not pose a fertility cost on males

Upon the release of YLE males, the transgene persistence in the wild and its suppression effect in the population over generations will be affected by potential fitness costs associated to the YLE construct. If the YLE does not pose fertility costs to the males, selection against the transgene will be absent and the target population would in theory remain suppressed indefinitely; but if the YLE has a small fertility cost in males, it would be gradually selected against, and the population would recover to its initial size. A high fitness cost would severely limit this strategy, potentially rendering it impractical.

The fertility of the YLE^{dsx} males from both lines was evaluated by crossing them to wild-type females and recording the number of eggs produced and hatched larvae (Methods 4.8.6). In standard insectary conditions, the fecundity of YLE^{dsx}-A and YLE^{dsx}-B males is not significantly different from the fecundity of wild-type males from the same colony (Figure 4.7).

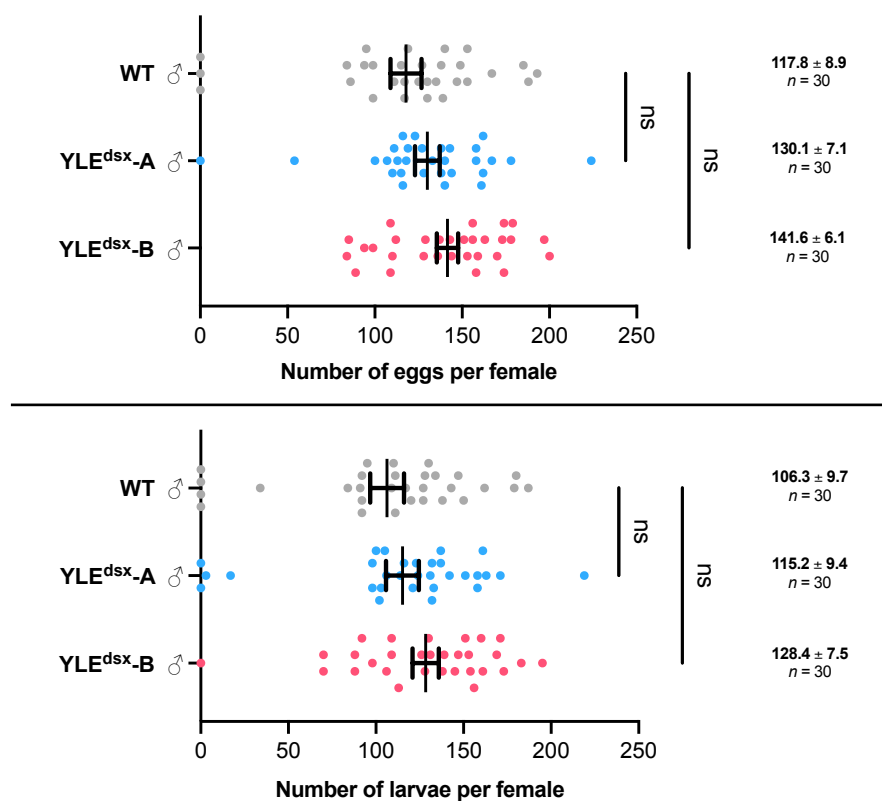


Figure 4.7 – The fertility of YLE^{dsx} males. Females mated to YLE^{dsx} males produced an equivalent number of eggs (top) and larvae (bottom) as those mated to wild-type males. A Kruskal-Wallis test revealed no significant differences with the wild-type control in the production of eggs ($P = 0.1548$) or in the number of larvae ($P = 0.2322$), indicating that the YLE^{dsx} transgene does not pose a fecundity cost on males. Each dot represents the progeny of a different female. The numbers on the right show the mean and the standard error of the mean, and the size of each sample (n).

4.5 Attempting to stabilize the Cas9 activity in the YLE^{dsx}-A line through the use of *gypsy* chromatin insulators

The results of the Cas9 activity levels in different males of the YLE^{dsx}-A strain showed considerable variability, presumably due to variegation on the transgene expression. Therefore, I attempted to increase and stabilize the expression of a transgene from this locus by using *gypsy* chromatin insulators.

In order to do that, I used the same exogenous *gypsy* sequences from *D. melanogaster* that Carballar-Lejarazu et al. (2013) successfully applied to increase and stabilize transgene expression from different autosomal loci in *An. stephensi*. The p16510-eCFP plasmid was modified to introduce two *gypsy* chromatin insulator sequences in opposite orientations flanking the cassette containing *3xP3::eCFP*, *vas2::Cas9* and *U6::gRNA^{dsxF}* (p16510-eCFP[*gypsy*]) (Figure 4.8) (Methods 4.8.1.1). The aim was to integrate the new plasmid in the same locus, in the same orientation and with exactly the same content as in the YLE^{dsx}-A strain, with the only difference of the *gypsy* insulator sequences to properly compare expression both in the soma (with the fluorescent marker) and in the germline (with the Cas9 protein). A total of ~3,100 eggs from the Y-AttP line were injected with the modified p16510-eCFP[*gypsy*] and a helper plasmid with a Φ C31 integrase gene. However, no transgenic males were recovered.

After failing to generate transgenics with the described approach, I created another plasmid to integrate the same elements in the same locus by using Cas9 to cleave the transgene of the Y-AttP line and trigger homology-directed integration with homology arms incorporated to the plasmid (pY_16510-eCFP[*gypsy*]) (Methods 4.8.1.2). Approximately 1,900 eggs from the Y-AttP line were injected with the pY_16510-eCFP[*gypsy*]. Nonetheless, this approach did not result in the generation of transgenic males either.



Figure 4.8 – Modified p16510-eCFP plasmid to incorporate *gypsy* chromatin insulator sequences. The p16510-eCFP[*gypsy*] plasmid is a modification of p16510-eCFP, with a central part constituted by a Cas9 gene under the *vas2* promoter, a gRNA targeting the female-specific exon of the *dsx* gene and an eCFP marker gene, all flanked by two *gypsy* sequences in opposite orientations, which are in turn flanked by two *attB* sequences for Φ C31 integrase mediated recombination with the *attP* sequence located in the Y-linked transgene of the Y-AttP line. Unfortunately, no transgenics could be generated with this genetic construct. Another plasmid containing these same elements and including homology sequences with the Y-linked transgene of the Y-AttP line was cloned with the same purpose of stabilizing Cas9 activity from this locus. However, no transgenic individuals were found.

4.6 Mathematical modelling of YLE^{dsx}-B male releases confirms its potential as an efficient self-limiting genetic control tool

To better understand the potential of this YLE being used as a tool to suppress a population, Kate Willis performed mathematical modelling with the parameters obtained for the YLE^{dsx}-B line (referred to as YLE^{dsx} in this subsection) and compared it to the idealized scenario. The homing rates are based on the values of inheritance bias of the *dsxF^{A11}* allele, which were slightly higher but still very similar to those of the *dsxF⁻* one. This work has been included in this thesis since it highlights the relevance of the developed YLE as a potentially useful self-limiting tool for genetic control. In this simulation, YLE^{dsx} males are released bearing a female-specific dominant negative mutation in *dsx* with total penetrance that causes sterility. This would be an almost identical scenario to the release of YLE^{dsx} males heterozygous for the *dsxF^{A11}* allele.

An experiment described in Chapter 5 revealed the existence of other female-specific dominant mutations in *dsx* (the penetrance of which remains to be properly assessed), as well as recessive ones that can be created by targeting the same sequence as the YLE^{dsx} does. Additionally, the rate of generating new mutations in a YLE^{dsx} male that has two unmodified *dsx* alleles has not been experimentally determined. For these reasons, a conservative approach was taken with the assumption that every male that did not inherit the released dominant mutation would generate a recessive one that would be transmitted to their progeny in the same described fashion.

As shown in Figure 4.9 – Top, a single release of YLE^{dsx} males equivalent to the initial number of males in the target population would reduce the population to less than 75% of its original size. Assuming that YLE males do not have a fitness cost, as it has been observed in insectary conditions, this suppression would be maintained through generations and the YLE genetic construct would remain at a constant 50% frequency in the male population. However, the initial effect (that suppressed the population to 50% of its original size) would be slightly diminished by the emergence of recessive mutations that would replace the dominant one in the population, as females heterozygous for these alleles would be fertile and would contribute to their transmission.

If instead of a single big release, YLE^{dsx} males are released every generation at a rate of 20% of the initial number of males, the target population would be virtually suppressed after ~30 generations, in comparison to suppression before the 20th generation in the idealised case (Figure 4.9 – Middle). The Y chromosome bearing the YLE construct would get to nearly fixation prior to collapsing the population. In an alternative scenario where these releases occur for only 6 generations, an optimal YLE would reduce the population to ~20% of its original size and the suppression would be indefinitely maintained at this level; while for the currently existing YLE^{dsx}, the population would experience a strong initial decline and a later recovery, but it would remain partially suppressed at 50-60% of its original size (Figure 4.9 – Bottom). In this case, the Y chromosome with the YLE transgene would increase to almost 75% frequency in the male population and would remain stable at that rate.

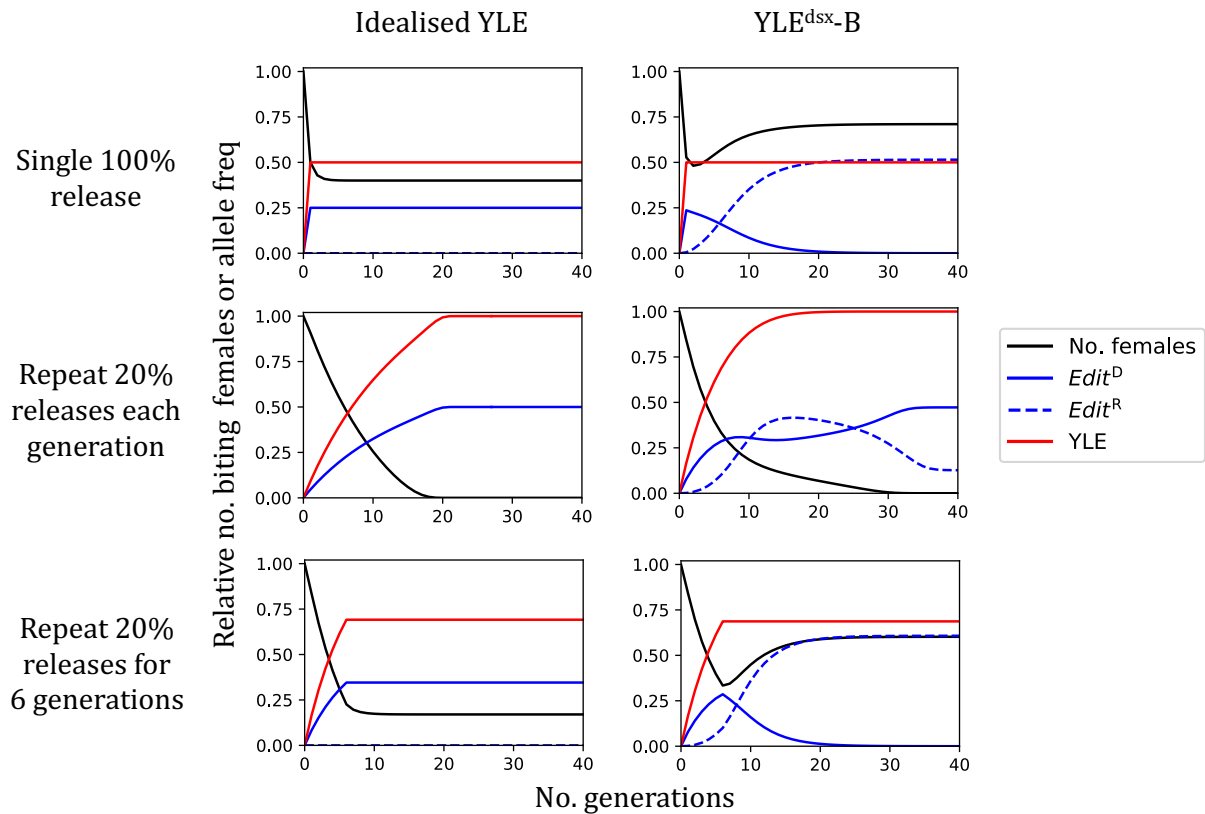


Figure 4.9 – Comparison of the effect of releases of an idealised YLE (left) and the developed YLE^{dsx-B} strain. In both scenarios, released males contain a dominant negative mutation with total penetrance that causes female-specific sterility. The relative number of biting females across generations is shown by the black line; the frequency of males bearing the YLE construct is represented in red; the frequency of the released dominant mutation is shown by the continuous blue line (*edit^D*); and the frequency of emerging recessive mutations is represented by the dashed blue line (*edit^R*). Recessive mutations never appear in the idealised scenario because the optimal YLE males transmit the dominant mutation to 100% of the offspring. The top row shows the effects of a single release of YLE males equivalent to the initial number of males in the target population (100% release). The middle row illustrates the effects of generational releases of YLE males at a rate equivalent to 20% of the initial number of males in the target population, where population collapse is achieved in both scenarios. The bottom row illustrates the effects of performing releases of YLE males at a rate of 20% of the initial number of males during 6 generations and then the releases are halted.

Looking at the required release rates to achieve certain levels of suppression within 36 generations (Table 4.4), they are 3 to 5 times larger for the developed YLE^{dsx} than for the optimal case. Nonetheless, these rates are still 2 to 20 times lower than the required ones for the optimal version of other self-limiting strategies.

Analysing the different parameters that affect the developed YLE^{dsx} efficacy, the rise of recessive mutations is what causes the biggest reduction in efficiency (Figure 4.10). This figure shows that to decrease the release rate required (i.e., to improve this YLE^{dsx}), the main focus should be to promote the generation of dominant mutations and to prevent the emergence of recessive ones. The rest of the parameters of the YLE^{dsx} are very close to the optimum values.

Table 4.4 – Release rates required to achieve certain suppression levels in a target population with an idealised YLE, the developed YLE^{dsx} and other idealised self-limiting strategies. Release rates required to achieve a suppression level of 67%, 95% and 99% in a target population assuming releases for 36 generations. Release rates were calculated as a proportion of the initial number of males in the target population. Release rates were estimated for an optimal YLE (idealised scenario) causing female-specific sterility, an optimal SIT, an idealised RIDL (causing unspecific lethality after density-dependence), an optimal fs-RIDL (causing female-specific lethality after density-dependence), an idealised autosomal X-shredder, and the developed YLE^{dsx} line (specifically the YLE^{dsx}-B strain) with the observed parameters. Numbers in square brackets indicate how many times larger the release rate would have to be compared to the release of optimal YLE to achieve the same level of suppression.

Suppression level (%)	Optimal YLE	YLE ^{dsx} -B	Optimal SIT	Optimal RIDL	Optimal fs-RIDL	Optimal X-shredder
67	0.022	0.065 [3.0]	1.286 [58.4]	0.419 [19.1]	0.473 [21.5]	0.240 [10.9]
95	0.037	0.171 [4.6]	1.297 [35.1]	0.438 [11.9]	0.668 [18.1]	0.396 [10.7]
99	0.041	0.183 [4.5]	1.3 [31.7]	0.440 [10.7]	0.678 [16.5]	0.403 [9.8]

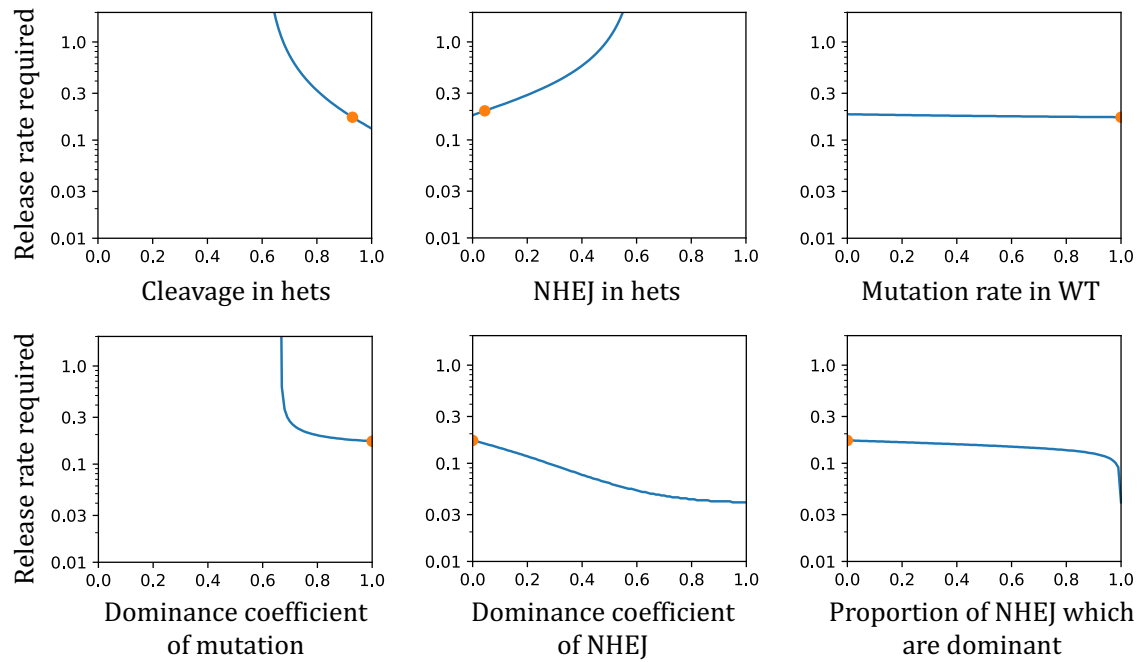


Figure 4.10 – Release rates required to achieve 95% suppression of a target population within 36 generations with the developed YLE^{dsx} as a function of various parameters. Different aspects that affect the efficiency of the YLE were studied to understand the strong and weak points of the current YLE^{dsx} and address those features where a bigger improvement can be achieved. The blue line illustrates the release rates required (proportion of the initial number of males in the target population) as a function of each of the parameters. The yellow dot indicates the current value of the developed YLE^{dsx} for each parameter. The analysed parameters shown in the top row (left to right) are the cleavage rate in YLE males heterozygous for a mutation (i.e., probability of cleaving the unmodified target in YLE males bearing one mutated copy), the probability of generating a new mutation in the target by non-homologous end-joining (NHEJ) given cleavage in YLE males heterozygous for a mutation, and the probability of creating a mutation in the target site in YLE males that are wild-type homozygous. The first two aspects have a big impact in the efficiency of the YLE, but the values of the current YLE^{dsx} are very close to the optimal ones. The rate of generating new mutations in wild-type homozygous was not measured in insectary conditions, and therefore it was assumed to be 1; however, this parameter scarcely affects the release rate required. In the bottom row (left to right), the analysed aspects are the dominance coefficient of the released mutation (i.e., if the mutation carried by released YLE^{dsx} males is fully penetrant and always causes sterility in females, the dominance coefficient will be 1), the dominance coefficient of new dominant mutations generated by NHEJ, and the proportion of new mutations generated by NHEJ that are dominant. The dominance coefficient of the released female-specific dominant mutation has a vital impact in the efficiency of the YLE, but once again the *dsxF^{Δ11}* allele that would be released in YLE^{dsx} males is virtually fully penetrant, being indistinguishable from the optimal values. Since the dominance coefficient of other mutations in the *dsx* target was not deeply studied, it was assumed that any newly generated mutation by NHEJ would be recessive. Therefore, this is the weakest parameter of the YLE^{dsx} and in order to improve this strategy, the focus should be on the reduction of recessive mutations emergence. The intrinsic rate of increase of the population (R_m) considered in these scenarios is 6; R_m is defined as the expected number of daughters produced by a female in the absence of density-dependent mortality.

4.7 Discussion

The most essential requirement for the development of a YLE is the successful expression of transgenes in the germline from the Y chromosome. Regardless of whether it is based on generating mutations with nucleases, or on deposition of toxins or mRNAs in the embryos, achieving sufficient and relatively stable germline expression from the Y chromosome is imperative. To investigate whether this was possible, a YLE was designed based on the CRISPR/Cas9 system. The selected genetic constructs contained a gene coding for Cas9 under the control of the *vas2* promoter and a gRNA targeting the female isoform of *dsx*. These were integrated in two independently created strains with sequences for site-specific recombination on the Y chromosome.

The results revealed remarkably high homing rates of the *dsxF*⁻ allele in both YLE^{dsx} lines, indicating high Cas9 activity in the germline of the progenitor males. It is of special relevance that these results were replicated for the homing of a female-specific dominant negative mutation (*dsxF*^{Δ11}), since this enormously improves the efficiency of the strategy. It would be desirable to evaluate the homing of the *dsxF*^{Δ11} in the YLE^{dsx}-B strain in a higher number of progenies, given the potential relevance of this technology and the promising results in this YLE line. Nonetheless, the pattern observed is consistent with the results obtained when homing the *dsxF*⁻ allele. Overall, these experiments confirmed that sufficient nuclease activity can be achieved in the germline of *An. gambiae* from a construct located on the Y chromosome when using the *vas2* promoter. Furthermore, the homing rates observed in YLE^{dsx}-B males are comparable to those of homing constructs located in autosomes that use the same *vasa* regulatory sequences to express Cas9 (Hammond et al., 2021a, 2016). Consistently with what had been observed in previous experiments with these autosomal homing constructs based on Cas9 expression under *vasa*, chromosomal cleavage in the early stages of spermatogenesis is mainly restored by homology-directed repair, and new mutations by NHEJ or MMEJ are created at a very low frequency.

The differences observed between the two YLE^{dsx} lines, in both overall Cas9 activity and variability in this activity among males with the same Y chromosome, seem to indicate that some parts of this chromosome are less affected by position effect variegation. This is particularly relevant for the development of future genetic control strategies linked to

the Y chromosome in *An. gambiae*, and an effort to locate the relative positions of the two integrations could provide useful information for the search of Y chromosome regions that are more transcriptionally active and less prone to silencing. Additionally, a deeper study of expression patterns from other loci would provide highly valuable information for the same purpose.

Even though the direct association between expression levels of Cas9 and Cas9 activity levels has not been addressed in this study, there are some indications that point towards a correlation between the two as the most plausible explanation for the variability detected. Firstly, males from each strain share the same Y chromosome, hence all the elements of the genetic construct are identical. It has also been observed high Cas9 activity in the male offspring of males that showed no Cas9 activity at all, which dismisses the possibility of some males having a mutation that makes the Cas9 or the gRNA inactive. Secondly, variegation of the fluorescent marker expression was observed too among sibling males, in concordance with what has been reported for Y-linked transgenes both in *An. gambiae* and in *D. melanogaster* (Bernardini et al., 2014; Gamez et al., 2021). Correlation between the expression level of the fluorescent marker in somatic tissues and the Cas9 in the germline has not been explored either.

Failed attempts to recreate the YLE^{dsx}-A strain incorporating *gypsy* chromatin insulators prevented the analysis of their capability to improve and stabilize transgene expression from the Y chromosome. Previous studies were successful in integrating genetic constructs with *gypsy* insulator sequences flanking a fluorescent marker both by ΦC31 integrase-mediated recombination and by *piggyBac* random integration in *An. stephensi* (Carballar-Lejarazu et al., 2013). Moreover, a transgene containing these same sequences was effectively integrated in the *An. gambiae* Y-AttP line (Dario Meacci, unpublished, personal communication). With these precedents, it seems unlikely that the presence of *gypsy* insulators would have prevented the insertion of the construct, or that it would have negatively affected the expression of the fluorescent marker rendering the transformed larvae impossible to detect. Therefore, while transgenesis on the Y chromosome is proven to be difficult, considering the high numbers of eggs injected, it is challenging to find a plausible explanation for the lack of success other than the plasmids not being completely purified and the carrying over of endotoxins that inhibited a successful transformation.

A major breakthrough for this YLE^{dsx} was the discovery of a female-specific dominant negative mutation in *dsx*. What initially was designed to be just a proof-of-concept to investigate if enough Cas9 expression could be achieved from the Y chromosome, became an efficient self-limiting tool by linking the YLE^{dsx} with the dominant negative *dsx*^{F^{Δ11}} allele.

The modelling performed with the observed parameters in the developed YLE^{dsx}-B strain showed that it is a powerful self-limiting strategy that can lead to significant reductions in the size of the target population even with only one release. The versatility of this tool allows to maintain a certain suppression level over generations, or to achieve further suppression with more releases, until potentially reaching the virtual collapse of the population. While the release rates required for YLE^{dsx}-B are 3 to 5 times larger than for the optimal YLE, these are still much lower than the required ones for the optimal version of other self-limiting strategies, which evidences the efficacy of the YLE^{dsx}-B. This is particularly relevant given the cost and logistic problems associated with mass rearing of mosquitoes for extra releases.

However, the emergence of recessive mutations in the target will limit the efficiency of the YLE. In order to overcome this, it would be necessary to find a target where no recessive mutations can be created. It remains to be explored whether this could be achieved by targeting the female exon of *dsx* downstream from the current target site (a more detailed explanation for this suggestion is described in Chapter 5). Alternative targets could also be explored by looking for genes which proteins form oligomers and have an essential role only in females, such as the *Act4* gene in *Aedes aegypti* and *Culex quinquefasciatus* mosquitoes, for which different in-frame deletions in the CDS cause a dominant flightless phenotype in females (Navarro-Payá et al., 2020; O’Leary and Adelman, 2020).

Another limitation that could affect the YLE^{dsx} is the generation of alleles resistant to cleavage that restore the function of the gene. Ioanna Morianou (manuscript in preparation, personal communication) found that such an allele could naturally appear in the target sequence of the YLE^{dsx}. Even if extremely rare, the fact that a resistant allele can eventually appear poses a threat for the efficacy of the YLE, as it would be rapidly selected for in the population. A conceivable approach to reduce the possibility of resistance would be multiplexing. If a female-specific dominant negative mutation is

found in another gene, it could be possible to drive both mutations at the same time by including the corresponding gRNA in the YLE^{dsx} construct.

Looking forward, additional studies should be performed on the stability of the transgene and Cas9 activity to ensure that the lines are not silenced due to the variegation existing in the Y chromosome. Focusing on the YLE^{dsx}-B strain, the homing of *dsx*^{F411} should be evaluated after a certain number of generations to verify that it remains stable. In addition, the capability of the YLE^{dsx}-B males to suppress a caged population in insectary conditions should be evaluated to corroborate the mathematical modelling and as a previous stage to further checks that would pave the way for potential field tests. Regarding the stability of the construct and its potential translocation to another chromosome, Alcalay et al. (2021) concluded that this was highly unlikely, and this has not been observed in over 35 generations for the YLE^{dsx}-A strain or over 15 generations for the YLE^{dsx}-B line. Previous integrations in the loci where the YLE^{dsx} constructs were inserted have been equally stable since the moment they were generated.

In essence, the YLE^{dsx}-B strain presented here is a self-limiting genetic control tool for *An. gambiae* that is considerably more effective than other self-limiting technologies: it has a stronger effect, and this effect is maintained in the population for a longer period of time. At the same time, it is less invasive than self-sustaining approaches and is expected to be more geographically confined, with a much lower risk of being transferred to a sibling species of the *An. gambiae* complex thanks to the sterility of the male hybrids. This expands the toolbox of genetic control strategies and offers an option halfway between previously existing self-limiting technologies and gene drives.

4.8 Methods

4.8.1 Design and cloning of plasmids

4.8.1.1 p16510-eCFP with gypsy chromatin insulators (p16510-eCFP[gypsy])

A two-step approach was followed to modify the p16510-eCFP plasmid by inserting *gypsy* chromatin insulators surrounding the construct (*vas2::Cas9*; *U6::gRNA^{dsxF}*; *3xP3::eCFP*).

Step 1

In the first step, a *gypsy* chromatin insulator was inserted between the eCFP marker cassette and one of the *attB* recombination sequences. In order to do that, p16510-eCFP was linearized by digesting it with PvuI restriction enzyme (NEB #R0150). Antarctic Phosphatase (NEB #M0289) was added into the digestion reaction to dephosphorylate the 5' ends. The linearized plasmid was purified using the Wizard® SV Gel and PCR Clean-Up System (Promega).

The *gypsy* chromatin insulator sequence was obtained from the plasmid PB[3xP3_GFP]gypsyUas-b2 (Dario Meacci, personal communication), and is the same sequence that was used by Carballar-Lejarazu et al. (2013). A fragment of 458 bp containing a *gypsy* insulator sequence was amplified using the KAPA HiFi HotStart ReadyMix (Roche) with the primers IT62 and IT63 (Table 4.5). These primers contained overhangs that overlapped for 27 bp to each side of the linearized p16510-eCFP. The *gypsy* fragment was purified by gel extraction using the Wizard® SV Gel and PCR Clean-Up System (Promega).

Both fragments were assembled using the Gibson Assembly® Master Mix (NEB) and the reaction was used to transform NEB 5-alpha Competent *E. coli* cells. Successfully transformed colonies were selected, and the plasmid was purified and sequenced prior to the second cloning step (these procedures are described below).

Step2

In the second step, a *gypsy* insulator was inserted between the gRNA cassette and the plasmid backbone. Due to the restriction sites present in the plasmid, unnecessary parts of the backbone were lost (one of the *attB* recombination sequences and one of the *piggyBac* repeats).

The newly cloned plasmid was digested with BspTI and SgsI FastDigest restriction enzymes (ThermoFisher) and the fragment (11,785 bp) containing the plasmid backbone, the marker eCFP cassette and the *vas2::Cas9* sequence was purified by gel extraction using the Wizard® SV Gel and PCR Clean-Up System (Promega).

The gRNA cassette was amplified (359 bp) from the original p16510-eCFP with the primers IT70 and IT71 (Table 4.5). This fragment contained overlapping sequences with the *vasa* terminator (60bp) and with the *gypsy* insulator (46 bp).

The primers IT51 and IT60 (Table 4.5) were used to amplify the *gypsy* sequence. This fragment (445 bp) included overlapping parts with the gRNA cassette (46 bp) and the p16510-eCFP backbone (25 bp).

Both PCR amplifications were performed with the KAPA HiFi HotStart ReadyMix (Roche) and fragments were purified by gel extraction with the Wizard® SV Gel and PCR Clean-Up System (Promega).

The three fragments were assembled with the Gibson Assembly® Master Mix (NEB) to generate the p16510-eCFP[*gypsy*] plasmid.

Table 4.5 – Primers used to amplify by PCR the fragments to clone p16510-eCFP[*gypsy*].

Overhangs (underlined) were designed to generate overlaps and insert the fragments in the correct orientation. No overhangs were needed for the primer IT70 because the sequence that was amplifying was overlapping already with the destination fragment.

Fragment	Primer name	Primer sequence
<i>gypsy</i> insulator 1	IT62	<u>CCGGGCGCGTACTCCACCTCACC</u> CATGCGATGCTAGCTCACGTAATAAGTGTGCG
	IT63	<u>CAAAC</u> TCATCAATGTATCTTTCCGGAGCGATCGCAATTGATCGGCTAAATGGTATGG
gRNA cassette	IT70	GCATGTTAGATGGTTGATACATCG
	IT71	<u>TGCCATACCATTTAGCCGATCAATTCTCGAG</u> GGGACCCACGCGTAAAAAAGCA
<i>gypsy</i> insulator 2	IT51	<u>TTTTACGCGTGGGTCCCTCGAGA</u> ATTGATCGGCTAAATGGTATGG
	IT60	<u>CTTTCTGTGTTGTAAAACGACGGCCAGTCTTAAGCTCGT</u> CACGTAATAAGTGTGCGTTGA

4.8.1.2 pY_16510-eCFP[gy] for integration by homologous recombination

The plasmid pY_16510-eCFP[gy] was cloned to attempt to integrate on the Y chromosome the components present in the p16510-eCFP[gy] plasmid, by using homology-directed integration instead of integrase-mediated recombination.

p16510-eCFP[gy] was digested with AasI and NheI restriction enzymes (FastDigest ThermoFisher). The fragment (10,243 bp) containing *vas2::Cas9*; *U6::gRNA^{dsx}*; *3xP3::eCFP* (all of it flanked by *gy* insulator sequences) was purified by gel extraction with the Wizard® SV Gel and PCR Clean-Up System (Promega).

The homology arms used were fragments of the plasmid originally used to create the Y-AttP docking line (3xP3[AttP]RFP plasmid, Bernardini et al., 2014). The primers IT77 and IT78 (Table 4.6) were used to amplify the whole 3xP3[AttP]RFP plasmid, removing only part of the RFP marker gene and the 3xP3 promoter (KAPA HiFi HotStart ReadyMix, Roche). Gel extraction was performed to purify the 5,662 bp fragment (Wizard® SV Gel and PCR Clean-Up System, Promega), which included 30 bp sequences on each side that overlapped with the fragment amplified from p16510-eCFP[gy].

Both fragments were cloned into a plasmid using the Gibson Assembly® Master Mix (NEB).

Table 4.6 – Primers used to PCR amplify the fragments to clone pY_16510-eCFP[gy]. Overhangs (underlined) were designed to generate overlaps and insert the fragments in the correct orientation.

Fragment	Primer name	Primer sequence
Backbone (3xP3[AttP]RFP)	IT77	<u>GAATAAATTC</u> AACGCACACTTATTACGTGAGCTAGCGATCAAACAAACGCGAGATAC
	IT78	CCGCCCCCTGACGAGCATCACA <u>AAAAATCGACGCTC</u> AGAACGTCATCAAGGAGTTCA

4.8.1.3 Helper plasmids for pY_16510-eCFP[gypsy] integration

Because pY_16510-eCFP[gypsy] already has a gene coding for Cas9 under a germline promoter, the helper plasmids for the insertion through homologous recombination only need to express a gRNA targeting the specific site on the Y chromosome.

A plasmid named p125 contains a gRNA scaffold cassette under a *U6* promoter where insertion of desired spacer sequences can be made by Golden Gate assembly. Two gRNAs were designed to target the Y-AttP docking line at the beginning of the RFP marker gene and at the beginning of the 3xP3 promoter. These gRNA spacers were created using 20 bp oligonucleotides (Table 4.7) with overhangs for directional insertion in the p125 plasmid after digestion with Eco31I for Golden Gate assembly.

Table 4.7 – The gRNA spacer and target sequences used for homology integration of pY_16510-eCFP[gypsy] in the Y-AttP line. The sequences of 20 nucleotides forming the gRNA spacers are homologous to the target sequence present in the Y-AttP line. The target sites include the PAM sequence (underlined) that allows the Cas9 cleavage.

Target	gRNA spacer sequence	Target sequence
3xP3 promoter	CTCGCGTTTGTTCGATCGCA	CTCGCGTTTGTTCGATCGC <u>ACGG</u>
RFP gene	CTCCTTGATGACGTTCTTGG	CTCCTTGATGACGTTCTTGG <u>AGG</u>

4.8.2 Plasmid purification

4.8.2.1 Bacterial transformation

2 µL of the assembly reaction were used to transform chemically competent cells. For the plasmids assembled with the Gibson Assembly® Master Mix (NEB), NEB 5-alpha Competent *E. coli* cells were used, while OneShot® Top10 competent cells were used for the rest of the plasmids.

The transformation was performed as indicated by the manufacturer and transformed bacteria were spread on agar plates containing the pertinent antibiotic (for which the plasmid bore a resistant gene, usually ampicillin or chloramphenicol). Plates were incubated overnight at 37°C.

4.8.2.2 Colony PCR screen

Upon bacterial transformation, PCR reactions were performed on the resultant colonies using primers located in different fragments of the plasmids to verify their correct assembly. Reaction mixes using the QIAGEN Fast Cycling PCR Kit were loaded in standard PCR tubes. Colonies were picked with a 10 µL pipette tip, streaked on a new plate (properly labelling each colony), and added to one of the PCR tubes with the reaction mix. PCR conditions were set according to the kit manufacturer suggestions. 10 µL of the PCR products were loaded into a 1% agarose gel for gel electrophoresis and correct assembly of the plasmids was corroborated by the number and size of the PCR fragments.

The streaked plate with the screened colonies was left in a 37°C incubator for ≥6 hours.

4.8.2.3 Miniprep

Bacterial colonies with the plasmid properly assembled were grown in a mix of 5 mL of lysogeny broth (LB) medium and 5 µL of the corresponding antibiotic, left at 37°C in a shaking incubator overnight. Precipitation of the bacterial cells and plasmid purification was performed following the indications of the QIAprep Spin Miniprep Kit (QIAGEN). The plasmid was eluted in 30 µL of DNase/RNase-free distilled water and its concentration was measured using a NanoDrop Spectrophotometer.

The fragments junctions were Sanger sequenced to verify that no mutations were introduced during the assembly.

4.8.2.4 Maxiprep

Once the Sanger sequencing confirmed that the plasmid did not contain mutations in the junctions, the selected bacterial colonies were grown in a mix of 100 mL of LB medium and 100 μ L of the corresponding antibiotic, left at 37°C in a shaking incubator overnight. Precipitation of the bacterial cells and plasmid purification was performed following the indications of the ZymoPURE II Plasmid Maxiprep Kit (ZYMO RESEARCH). The plasmid was eluted in 200 μ L of DNase/RNase-free distilled water and its concentration was measured using a NanoDrop Spectrophotometer.

The relevant parts of the plasmid were Sanger sequenced prior to microinjections to confirm that there were no mutations that could impair expression or activity of the elements in the construct.

4.8.3 Generation of transgenic lines

4.8.3.1 Embryo microinjections

Microinjections of mosquito embryos were performed as described in detail by Fuchs et al. (2013), and briefly explained here.

Adult mosquitoes were left in a cage for 5 days post-emergence. Females were blood fed 3-4 days prior to microinjection. On the injection day, adults were placed in the dark to stimulate oviposition. Embryos 20 min to 2 hours-old were uniformly aligned against a moistened filter paper to be injected with a mix containing the plasmid to be inserted in the genome and the helper plasmid. Embryos were injected in their posterior part, where the pole cells that will develop into germ cells are located. The components of the injection mixes used for each plasmid are shown in Table 4.8.

Surviving L1 larvae were screened for transient expression of the fluorescent marker in the anal papillae and parts of the ventral nerve chord. Larvae with and without this transient expression were reared separately and crossed to wild-type mosquitoes at adulthood. The progeny from these crosses was screened for expression of the fluorescent marker in the expected pattern depending on the promoter used (in the case of *3xP3*

promoter, the expression is visible in the optic lobe of the head and the dorsal ganglia across the larval body).

To create the **YLE^{dsx-A} strain**, a total of ~2,500 eggs from the Y-AttP line were injected and over 80 males were found to show transient expression of the CFP marker. Successful integration events in the offspring of these males were detected by the exchange of the fluorescent marker in over 30 G1 male individuals that were RFP⁻ and CFP⁺.

For the generation of the **YLE^{dsx-B} strain**, around ~1,800 eggs from the 3-2A2B line were injected, from which 52 males showed transient expression of the RFP marker. Among their progeny, only G1 individuals that were GFP⁻ and RFP⁺ were selected, as this indicated complete transgene exchange and removal of all the elements of the previous construct.

The **p16510-eCFP[gypsy]** and a helper plasmid with a Φ C31 integrase gene were injected in ~3,100 eggs from the Y-AttP line. 142 males showed transient expression of the CFP marker, but the exchange of fluorescent marker from RFP to CFP indicating a successful integration of the transgene was not detected in any individual in the progeny.

Approximately 1,900 eggs from the Y-AttP line were injected with the **pY_16510-eCFP[gypsy]** plasmid and two helper plasmids containing a gRNA each targeting the Y-AttP construct. Among the survivals, 24 males showed transient expression of the CFP marker. No transgenics were found among the offspring of these males.

Table 4.8 – Injection mixes to create YLE lines. The donor and helper plasmids (and their concentrations) included in each injection mix. Ranges of concentrations are shown when different injections rounds were done with different concentrations. Lines in which each mix was injected are also shown. All the mixes included a diluted volume of the stock 10x microinjection buffer (50 mM KCL, 1 mM Na phosphate buffer pH 7.2).

Donor plasmid	Helper plasmid	Line injected
p16510 (200 ng/μL)	vas2::integrase (400 ng/μL)	3-2A2B (GFP+)
p16510-eCFP (50-200 ng/μL)	vas2::integrase (200-400 ng/μL)	Y-AttP (RFP+)
p16510-eCFP[gypsy] (50-200 ng/μL)	vas2::integrase (200-400 ng/μL)	Y-AttP (RFP+)
pY_16510-eCFP[gypsy] (100-200 ng/μL)	p125_3xP3 (200-400 ng/μL)	Y-AttP (RFP+)
	p125_RFP (200-400 ng/μL)	

4.8.3.2 Maintenance of *An. gambiae* mosquito strains

An. gambiae mosquitoes were maintained in a contained facility at Imperial College London, where the temperature and the humidity levels were monitored and set at 27°C and 65% humidity. Larvae were kept in a controlled density in trays (~150 larvae per tray) with 500-600 mL of demineralised water and fed with fish food pellets. At pupal stage, the individuals were collected and left to emerge in a bowl with water within a mesh cage. Adult mosquitoes were fed with glucose solution. Bovine blood was provided with a Hemotek to produce offspring. A bowl with filter paper and water was placed in the cage 3 days after blood-feeding to allow the females to lay eggs. Once the larvae hatched, the progeny was transferred to trays.

YLE^{dsx} lines were maintained by sexing transgenic males at pupal stage and crossing them to wild-type females every generation.

4.8.4 Evaluating Cas9 activity in the YLE^{dsx} lines

4.8.4.1 Homing of a marked allele in *dsxF*

An assay was designed to evaluate Cas9 activity from the YLE lines based on the homing of a marked allele. The *dsxF*⁻ allele contains a gene coding for a GFP within the female-specific exon of *dsx* and lacks the sequence recognised by the gRNA present in the YLE construct. An important aspect in this experiment was to assess the variability in Cas9 activity between different males. For that, it is key to ensure that each progeny analysed for homing is coming from a different male. Typically, individual crosses of mosquitoes are not very efficient due to their mating behaviour, therefore the experimental cross was performed in the same way as described by Rogers et al. (2008), and explained below. The same procedure was used for both YLE^{dsx} lines created.

YLE^{dsx} males were initially mated with females heterozygous for the *dsxF*⁻ allele. Individuals in the progeny with expression of the marker linked to the YLE and GFP (linked to *dsxF*⁻) were selected. To ensure that YLE^{dsx} males used in the experimental cross would not have inherited other mutations in the target site that would prevent homing of *dsxF*⁻, the selected males were mated with wild-type females. Around 600

double-positive individuals were selected in the offspring at pupal stage and placed in a large cage. 100 females from the *An. gambiae* wild-type colony were sexed at pupal stage and placed in a separate cage. Six days after emergence, at the time of the programmed dusk, females were introduced into the male cage in groups of 6-8. Mating pairs falling to the bottom of the cage were isolated by gently placing a 25 mL plastic glass over them until copulation terminated. Once the couple separated, the plastic glass was brought over a thin piece of cardboard previously left on the bottom of the cage and the mated mosquitoes were transferred to another cage. Mated females were kept in this separate cage and males were removed to avoid a new mating with another female. 50-60 mated females were captured. The following day to the mating, females were fed with blood and 3 days later they were allowed to lay eggs individually. Hatched L1 larvae were screened for the presence of GFP. Larvae from progenies with low homing rates of *dsxF*⁻ were pooled to perform genomic DNA extraction (Wizard® Genomic DNA Purification Kit, Promega) and subsequently do amplicon sequencing on the *dsx* target (as described below) to check for the presence of mutations that could have prevented the homing.

While the procedure was the same for both YLE^{dsx} lines, the YLE^{dsx}-B line was screened by eye, but the YLE^{dsx}-A line was screened using the COPAS system to unambiguously detect co-expression of GFP and CFP, which are difficult to distinguish by eye under the fluorescence microscope. Homing of the *dsxF*⁻ allele in the YLE^{dsx}-A line was assessed by eye under the fluorescence microscope and evaluated only in females to avoid confusion with CFP+ only males.

4.8.4.2 Homing of a dominant negative mutation in *dsxF*

Measuring the homing rates of a dominant negative mutation in *dsx* from the YLE lines was done in a different fashion. In this case, the homed allele is markerless, which makes very difficult to achieve high numbers of YLE^{dsx} males guaranteeing that they bear the dominant mutation.

The *dsxF*^{A11} deletion was present in the YLE^{dsx} lines, but to ensure that the males used for the experimental cross had this mutation, the offspring of different females from these lines were reared separately. Individuals were sexed at pupal stage, looking for progenies where all the females showed the intersex phenotype (indicating that all of them had the

dominant mutation, and therefore their sibling males did too). The males of those progenies were selected to perform the experiment: 50 YLE^{dsx} males supposedly bearing the dominant mutation were allowed to mate for 5 days to 60 wild-type females in a cage. Females were then fed blood and allowed to lay eggs individually 3 days later. Progenies were raised separately and sexed at pupal stage. Females from each progeny were then pooled to perform genomic DNA extraction (Wizard® Genomic DNA Purification Kit, Promega) and subsequently do amplicon sequencing on the *dsxF* target (as described below) to check the frequency of the *dsxF*^{A11} allele in the offspring.

4.8.5 Amplicon sequencing of the *dsxF* target

Using as a template 100 ng of the extracted genomic DNA, a PCR reaction was set with KAPA HiFi HotStart ReadyMix (Roche) to amplify the *dsxF* target. A fragment of 291 bp was amplified using primers that contained Illumina partial adapters for sequencing (Table 4.9). Two replicates were made for every reaction. For the first replicate, the PCR reaction was stopped after 20 cycles to avoid altering the frequencies of the alleles present in the pool. For the second replicate, the reaction was performed for another 20 cycles to later confirm by gel electrophoresis that the amplification had occurred as expected. The fragments from the first replicates were then purified using Wizard® SV Gel and PCR Clean-Up System (Promega) and sent for amplicon sequencing (Amplicon-EZ, GENEWIZ).

Table 4.9 – Primers used for amplicon sequencing of the *dsxF* target. The primers amplified a fragment of 291 bp around the target site in *dsxF*. Illumina partial adapter sequences are underlined.

Primer name	Primer sequence
Illumina-AmpEZ-4050-F1	<u>ACACTCTTTCCCTACACGACGCTCTTCCGATCT</u> ACTTATCGGCATCAGTTGCG
Illumina-AmpEZ-4050-R1	<u>GACTGGAGTTCAGACGTGTGCTCTTCCGATCT</u> GTGAATCCGTCAGCCAGCA

Amplicon sequencing data were analysed using CRISPResso2 (version 2.0.29) (Clement et al., 2019). This software performs quality filtering of the reads, adapter trimming, read merging and alignment to the reference sequence provided. It then generates a set of graphs with the frequencies of the variants found in the reads.

4.8.6 Fertility assays

YLE^{dsx} males from both lines were crossed with wild-type females in groups of 50. They were left to mate for 5-7 days before being blood fed. A minimum of 30 females were allowed to lay eggs individually by placing them in separate cups 3 days after blood feeding. After 2 nights, the number of eggs laid by each female was recorded. The number of larvae hatching from each progeny in the following days was recorded as well.

4.8.7 Mathematical modelling

The model created by Katie Willis to predict the population dynamics of the developed YLE^{dsx} has a deterministic structure, mirroring the one developed by Burt and Deredec (2018). It simulates a single, infinite-sized population where mating occurs randomly, with discrete and non-overlapping generations. The model includes two life stages (juveniles and adults), where juvenile survival is density-dependent, following the Beverton-Holt model. The model considers a Y-linked locus (with two possibilities: wild-type or YLE) and the autosomal target locus (with three possibilities: wild-type, cleavage-resistant recessive mutation and dominant mutation). No cleavage-resistant functional alleles were considered.

Released males carry a YLE construct and a dominant female-specific cleavage-resistant allele. The parameters that influence the inheritance of this mutation and the generation of novel alleles are shown in Table 4.10.

Fitness effects were also considered. Since the targeted locus is a female-specific one, the male fitness is not affected (i.e., is standardised to 1). Fitness of homozygous wild-type females is also standardised to 1 (fully fertile). The fitness of heterozygous females for the released dominant mutation was 0 (fully sterile). The fertility of heterozygous females for new recessive mutations was 1 (fully fertile) but homozygous females for recessive mutations had a fitness of 0 (fully sterile).

Table 4.10 – Model inheritance parameters. Parameters that determined the inheritance of the released dominant mutation and the generation of new end-joining mutations (NHEJ). The description of each parameter, the data source and the values for the modelling of the YLE^{dsx} are displayed.

Parameter	Description	Data source	YLE^{dsx} values
d	Proportion of offspring which carry the desired mutation	Experiment	0.945
v	Proportion of offspring which carry an alternative mutation	Experiment	0.020
e	Probability of homing	$2d - 1$	0.890
u	Proportion of non-homed chromosomes which are NHEJ	$\frac{v}{1 - d}$	0.455
c	Probability of cleavage (in heterozygotes)	$e + (1 - e)u$	0.930
j	Probability of NHEJ given joining (in heterozygotes)	$\frac{(1 - e)u}{e + (1 - e)u}$	0.043
μ	Probability of mutation (in wild-type homozygotes)	Unknown	1
p	Proportion of dominant NHEJ (in heterozygotes)	Unknown	0
q	Proportion of dominant NHEJ (in wild-type homozygotes)	Unknown	0

5 Identification of female-specific dominant mutations in the *doublesex* gene of *Anopheles gambiae*

5.1 The *doublesex* gene in *Anopheles gambiae*

5.1.1 Gene structure and splicing mechanisms

Briefly introduced before, *doublesex* (*dsx*) is a gene involved in the sex determination process with sex-specific spliced isoforms (*dsxF* and *dsxM*). Deeply studied in *D. melanogaster* (*Dmdsx*), Scali et al. (2005) identified the orthologue of *Dmdsx* in *An. gambiae* (*Agdsx*). While the gene structure in these two species is very similar (Figure 5.1), the pathway that leads to the production of sex-specific isoforms varies significantly.

In *D. melanogaster*, the sex of the individual is determined by the dose of X chromosomes. In females, the presence of two X chromosomes leads to a female-specific splice pattern of the *Sex-lethal* (*Sxl*) gene, producing an active form of Sxl protein (Bell et al., 1988; Bopp et al., 1991). Subsequently, Sxl splices the *transformer* (*tra*) pre-mRNA to generate a functional Tra protein (Valcárcel et al., 1993) which, together with Tra2, acts as a splicing regulator of *dsx*. The Tra/Tra2 complexes bind to a group of six 13-nt sequences [UC(U/A)(U/A)CAAUCAACA] on the pre-mRNA of *dsx*, located at the 3'UTR in the female-specific exon (Figure 5.1), and activate the preceding weak 3' acceptor site located 295-566 bp from the Tra/Tra2 binding sites (Inoue et al., 1992; Lynch and Maniatis, 1995). This results in the production of the *dsx* female isoform. In males, the absence of Sxl prevents the splicing of *tra*, thus generating a non-functional protein (Valcárcel et al., 1993). Since the Tra/Tra2 complex is not present to bind to the *dsx* pre-mRNA, the weak 3' acceptor site is not recognized by the splicing machinery, which splices out the female-specific exon, and the male isoform is produced (Inoue et al., 1992; Lynch and Maniatis, 1995).

Even though the sex determination pathway in *An. gambiae* is not as well characterized as in *D. melanogaster*, Krzywinska et al. (2021) discovered the *femaleless* (*fle*) gene while searching for *tra2* homologous genes based on its RNA recognition motif (RRM). Sequences with high similarity to the 13-nt elements to which Tra2 binds in *Dmdsx* are also found in the 3' UTR of the female isoform in *Agdsx*. Given that the RRM of Fle and Tra2 share substantial similarity, these were described as putative Fle binding sites (Krzywinska et al., 2021). However, in *Agdsx* these sequences are located near the end of the female-specific exon (Figure 5.1), and there is not a weak 3' acceptor splice site preceding this exon (Scali et al., 2005). This suggests that when Fle binds these

recognition sequences, it activates the 5' donor site of intron 5, allowing the production of the female isoform. In males, it has been proposed that the male-determining factor *Yob* suppresses *Fle* by a yet unknown mechanism, preventing the activation of the 5' donor site of intron 5 and causing the splicing out of the female-specific exon 5, thereby producing the male isoform (Krzywinska et al., 2021). This hypothesis is supported by the fact that knocking down *fle* triggered the expression of the male transcript in female individuals (Krzywinska et al., 2021).

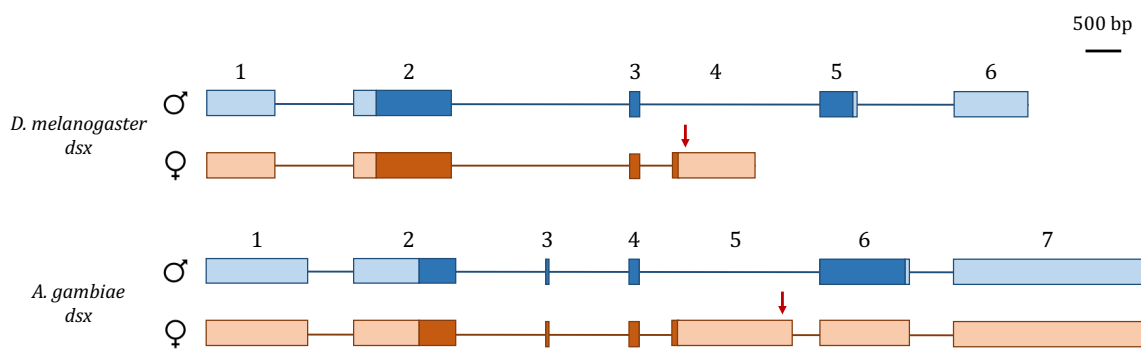


Figure 5.1 – The structure of the *doublesex* gene in *Drosophila melanogaster* and in *Anopheles gambiae*. The *Drosophila melanogaster dsx* gene consists of 6 exons (top) and the *Anopheles gambiae* is comprised by 7 exons (bottom). In both species there are two sex-specific isoforms of the *dsx* gene, with the female transcript (orange) containing an exon that is absent in the male isoform (blue). Translation of both isoforms starts at a common start codon in exon 2. Red arrows indicate the location of the binding sites of Tra2 (top, in the female isoform of the *D. melanogaster dsx* gene) and the putative binding sites of Fle (bottom, in the female isoform of the *An. gambiae dsx* gene). Light colours represent untranslated regions. Introns are not drawn to scale.

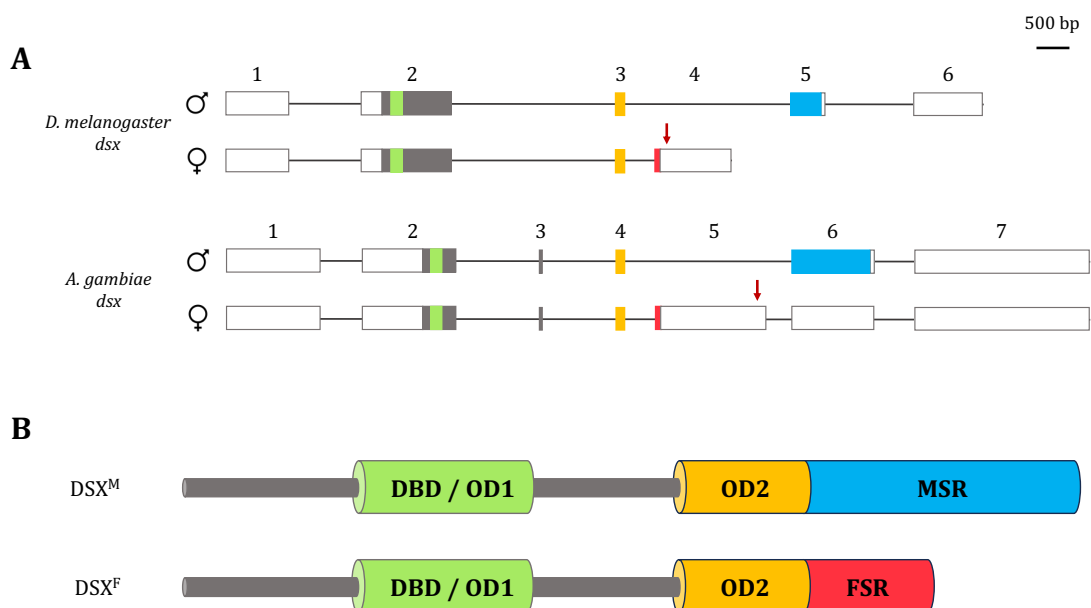
5.1.2 Protein structure and role of the conserved domains

The sex-specific spliced isoforms of *dsx* produce two transcription factors (DSX^M and DSX^F) with two conserved domains. This structure is similar across multiple insect species, including *D. melanogaster* and *An. gambiae* (Price et al., 2015; Verhulst and van de Zande, 2015). The first domain (OD1 or DBD/OD1) is a zinc finger-like domain that comprises the DNA-binding domain and one oligomerization domain. This is common to both isoforms and is located near the N-terminal (Figure 5.2 - B). Another oligomerization

domain (OD2) is situated in the C-terminal section, preceding the sex-specific parts of the protein (Verhulst and van de Zande, 2015). The OD2 contains a shared region by the two isoforms and a sex-specific part. As shown in Figure 5.2 - C, the OD1 domain, the non-sex-specific part of the OD2 domain and the female-specific region of DSX are extremely similar in *D. melanogaster* and in *An. gambiae* (and among other insect species, as shown by the alignment displayed at Scali et al., 2005). Contrarily, the longer male-specific part is poorly conserved across insect species (Scali et al., 2005).

Previous research in *D. melanogaster* showed that the DNA-binding domain is what determines the sequence to which DSX dimers bind (An et al., 1996; Cho and Wensink, 1998, 1997; Erdman and Burtis, 1993; Erdman et al., 1996). Since this domain is shared between the sex-specific isoforms, DSX^M and DSX^F bind to the same targets but have antagonistic regulatory effects by interacting with the transcriptional machinery in a different manner through the sex-specific parts of the protein. This was first observed in *D. melanogaster* for the transcription of the *yolk protein* genes, which are repressed by DSX^M and activated by DSX^F (Coschigano and Wensink, 1993). Additional examples of these dynamics have been since identified in several insect species and reviewed by Verhulst and van de Zande (2015).

Therefore, the OD1 and OD2 seem to have a common function in the dimerization of the protein and a specific one in which OD1 is required for DNA binding and OD2 together with the sex-specific parts of the protein are required for the interaction with the transcriptional machinery (Cho and Wensink, 1998, 1997).



DSX^M[illegible]DSX^F[illegible]

Figure 5.2 – Protein domains of the DSX^M and DSX^F isoforms in *An. gambiae* and comparison with *D. melanogaster*. (A) Representation of the *dsx* gene transcripts in *D. melanogaster* and *An. gambiae*. The region highlighted in green is the sequence that codes for the DNA binding and oligomerization domain 1 (DBD/OD1 or OD1). The region highlighted in yellow codes for the shared oligomerization domain 2 (OD2). In red, it is highlighted the sequence coding for the female-specific region of DSX^F (FSR), and in blue, the part coding for the male-specific region of DSX^M (MSR). Red arrows indicate the location of the binding sites of Tra2 and Fle. (B) Graphical representation of the OD1, OD2, MSR and FSR of the DSX proteins. Colour code is the same as described in A. (C) Amino acid sequence comparison and similarities between the DSX^M (top) and DSX^F (bottom) of *D. melanogaster* (*Dmel*) and *An. gambiae* (*Agam*). The protein domains are highlighted in the same colour code as described above. The shared domains OD1 and OD2, and the FSR are highly conserved between the two species, while the MSR is more diverged. The grey line over the FSR indicates the female-specific part of the OD2 domain. Colon (:) and period (.) indicate strong and weak conservation of residue properties respectively. Sequence alignment was performed with the Needleman-Wunsch algorithm from the EMBOSS Needle tool (EMBL-EBI).

5.1.3 Phenotypic effect of mutations in the *doublesex* gene

Since the developed YLE strategy aims to target the *dsxF* isoform, it is important to understand the probable effects of mutations in the sex determination of females.

Female-specific dominant mutations in *dsx* in *D. melanogaster* were identified and well-characterised in the 1980s-1990s (Nagoshi and Baker, 1990; Nöthiger et al., 1987). Four alleles (*dsx^D*, *dsx^M*, *dsx^T* and *dsx^S*) are known to cause an intersex phenotype (simultaneously showing female and male-specific morphological characteristics) in heterozygosity in chromosomally female individuals while they do not affect males. As described by Nagoshi and Baker (1990), these alleles both prevent the production of the female isoform and trigger the expression of the male one. *dsx^D*, *dsx^M* and *dsx^T* are insertions of transposable elements of 7.7 kb, 3.9 kb and 4.5 kb respectively, all of them occurring at the beginning of the 3' UTR sequence in the female-specific exon. It is thought that these mutations exert their effect by displacing the Tra2 binding sites, which end up extremely far from the weak 3' acceptor splice site that Tra/Tra2 activate to produce the *dsx* female isoform. As a consequence, the default male isoform is produced from this allele, resulting in an intersex phenotype due to the expression of both sex-specific isoforms. The same is predicted to occur for *dsx^S*, which is a deletion of ~500 bp at the beginning of the 3' UTR in the female-specific exon and removes the six 13-nt sequences recognised by Tra2. More recently, Yadav et al. (2023) also showed that the insertion of a transgene disrupting the female exon of *dsx* led to dominant female sterility.

Likewise, female-specific recessive mutations have also been found and studied in *D. melanogaster*. Initially described as *dsx^f* (Nöthiger et al., 1987), this mutation only causes an intersex phenotype in chromosomally female individuals when carried in homozygosity, while males are always unaffected. *dsx^f* entails a single missense mutation that changes glycine 398 to aspartic acid (Erdman et al., 1996). This modification is located within the female-specific part of OD2, and it has been shown to completely prevent dimerization of the protein (Erdman et al., 1996). In heterozygosity, the altered DSX^F does not dimerize and does not interfere with the regular DSX^F protein, thus not causing any phenotype. However, in homozygosity and therefore lack of normal DSX^F, an intersex phenotype is observed, equivalent to the one found in females that have two null *dsx* alleles.

Nevertheless, in *An. gambiae*, no dominant mutations in *dsx* had been described to date. Disruption of the coding sequence of the female-specific exon (*dsxF*⁻) is recessive (Kyrou et al., 2018); *dsxF*^{-/-} XX individuals present abnormal male-like traits, such as plumose antenna and claspers (although they are inverted compared to their position in males), as well as alterations in the proboscis that make them unable to bite and feed on blood. These individuals also lack fully developed ovaries and spermathecae, but develop male accessory glands and some of them form testes-resembling organs (Kyrou et al., 2018). All these characteristics render them completely sterile. On the other hand, males are unaffected by the modification, both in heterozygosity and in homozygosity.

The *dsxF*⁻ allele consists of a ~1.5 kb insertion, containing a transcription unit that codes for GFP under a *3xP3* promoter, at the beginning of the coding sequence of exon 5. It is still unclear whether this insertion interferes with the correct splicing of the female isoform, but contrarily to what occurs in *D. melanogaster*, it does not lead to the production of the male isoform (Nace Kranjc, unpublished, personal communication). Nonetheless, no phenotype effect is observed in females heterozygous for the *dsxF*⁻ allele. Since the developed YLE^{dsx} strains target the same region of the female-specific exon where *dsxF*⁻ contains the modification, the expectation was that the majority of mutations created in this area would also be recessive.

5.2 An 11 base-pair deletion in the coding sequence of *Agdsx* exon 5 is a female-specific dominant negative mutation

Upon the generation of the YLE^{dsx} strains, among the offspring of G1 transgenic males crossed to wild-type females I unexpectedly found a high frequency of females showing an intersex phenotype, similar to the one observed in females homozygous for a *dsxF* null mutation (Kyrou et al., 2018), even though in this case the mutations were expected to be in heterozygosity. Since previous evidence showed that *Agdsx* is an haplosufficient gene, and that the disruption of the female-specific exon is recessive, there were two most likely explanations for this. The first one is bi-allelic conversion of both wild-type alleles to a

null mutation in the embryos by both germline activity in the male progenitors and via paternally deposited nuclease. The second plausible explanation is the generation of a dominant negative mutation that can produce an intersex phenotype even in females heterozygous for the mutation. Consistent with the latter possibility, sequencing of over 400 individuals showing the intersex phenotype revealed the presence in heterozygosity of an 11 base-pair deletion within the coding sequence of the *dsx* female-specific exon (referred to as *dsxF^{A11}* hereafter) (Figure 5.3). Sequencing of the G1 progenitor males that gave rise to the intersex individuals showed that they were heterozygous for the *dsxF^{A11}* allele, suggesting that this mutation had occurred during the embryo microinjection process and G1 transgenic males had inherited it together with the YLE^{dsx} transgene from surviving injected individuals. The fact that this mutation was found in both YLE^{dsx} strains might be explained by the microhomology of 6 bp-long sequences (GGTCAA) flanking the cut site in exon 5 (Figure 5.3), which might have triggered microhomology end-joining and favoured the generation of this mutation over alternative ones.

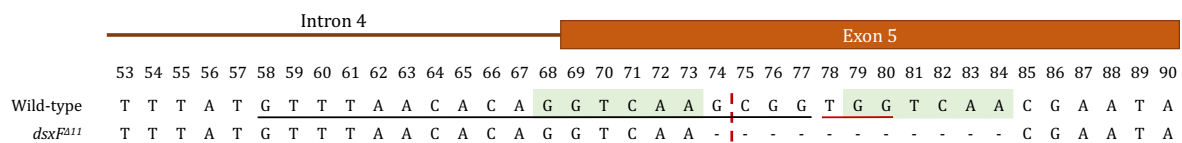


Figure 5.3 – The *dsxF^{A11}* allele was probably generated by microhomology end-joining. Representation of the target sequence environment in the intron 4 – exon 5 boundary of the *dsx* gene. The gRNA (and target sequence) is underlined in black, the PAM at the target site is underlined in red, and the cut site is shown by a dashed red line. The 6 bp-long sequences flanking the cut site that could trigger microhomology are highlighted in green (GGTCAA). Sequencing of over 400 individuals with an intersex phenotype (combining Sanger sequencing of individual samples and amplicon sequencing of pooled samples) confirmed the presence in heterozygosity of the *dsxF^{A11}* mutation.

5.2.1 The *dsxF^{A11}* mutation is highly penetrant and causes complete sterility in heterozygosity in chromosomally female individuals

The strong association between the presence of this mutation and the intersex phenotype suggests very high penetrance of the mutation. Besides it being found in every sequenced female with an intersex phenotype, from over 200 females with a normal phenotype, siblings of the females bearing the *dsxF^{A11}* mutation, only 4 were found to have this mutation (through a combination of Sanger sequencing of individual samples and amplicon sequencing of pooled samples). The low frequency at which this deletion is found in the pool of females with a normal phenotype may represent either a not complete penetrance of the *dsxF^{A11}* mutation or a misidentification of some females' phenotype at pupal stage, when they were classified. At this stage, the intersex females show a diverse and incomplete development of the genital lobe (Figure 5.4), which makes some of them ambiguous to classify.

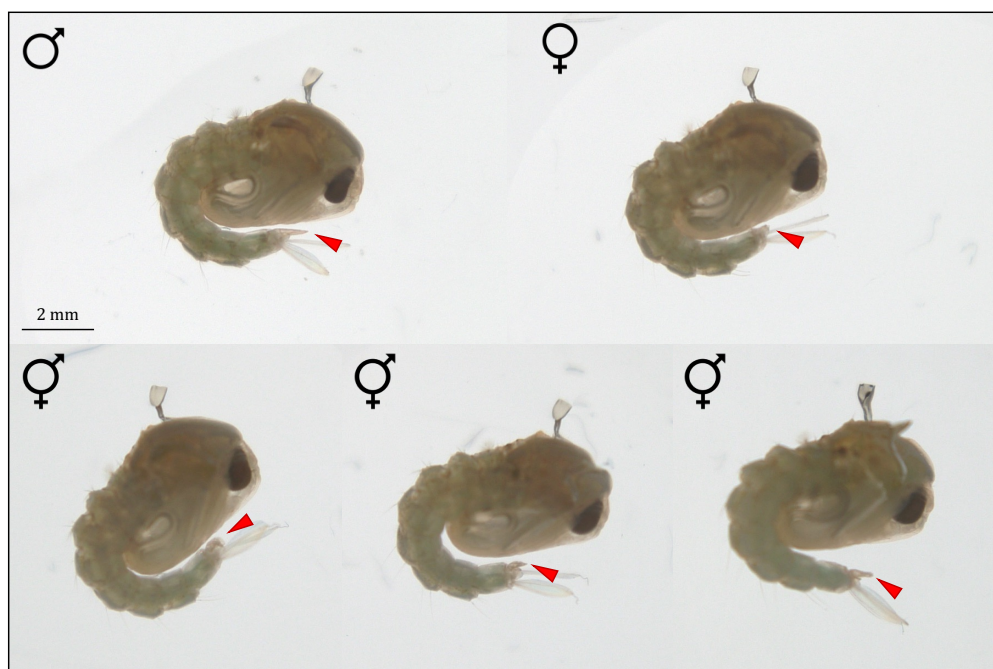


Figure 5.4 – Intersex phenotype of chromosomally female individuals at pupal stage. Females bearing one copy of the *dsxF^{A11}* mutation showed an intersex phenotype at pupal stage (bottom row), which could be observed in the varied development of the genital lobe (indicated with the red arrows), which was much more developed than in a normal female (top-right) but less developed than in wild-type males (top-left). In some individuals, this phenotype was less pronounced (bottom-left), making some pupae difficult to classify.

The dissection of 15 female adults showing this intersex phenotype revealed the presence of other male-specific traits in all of them, such as incompletely formed claspers and organs resembling male accessory glands, while at the same time they had female-like antennae and apparently normal ovaries and spermathecae (Figure 5.5 and Figure 5.6).

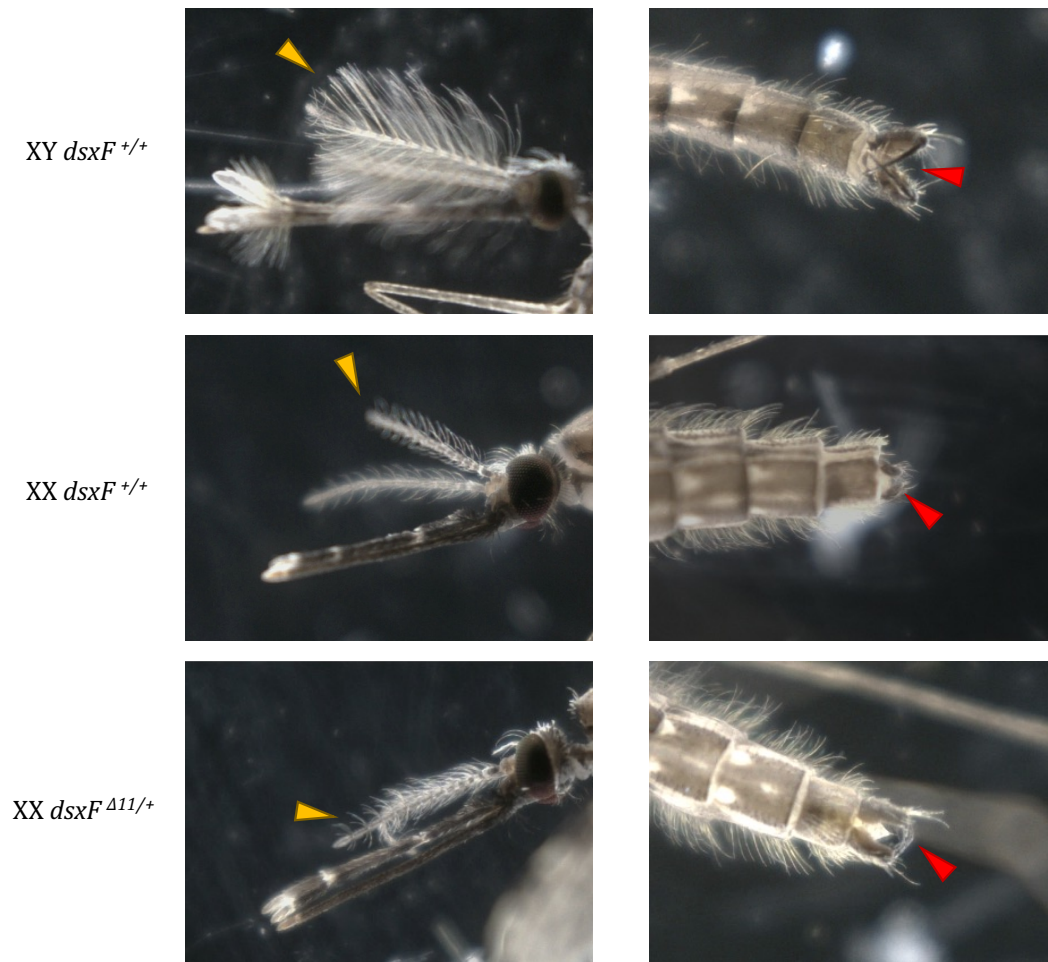


Figure 5.5 – Intersex phenotype of chromosomally female individuals at adult stage. Females bearing one copy of the *dsxF^{Δ11}* mutation (bottom) have pilose antennae (yellow arrows) that resemble those of wild-type females (middle), while males (top) have plumose antennae. On the other hand, while normal females lack claspers (red arrows), the genotypic females showing an intersex phenotype have a pair of claspers that are dorsally rotated in comparison to the claspers carried by males.

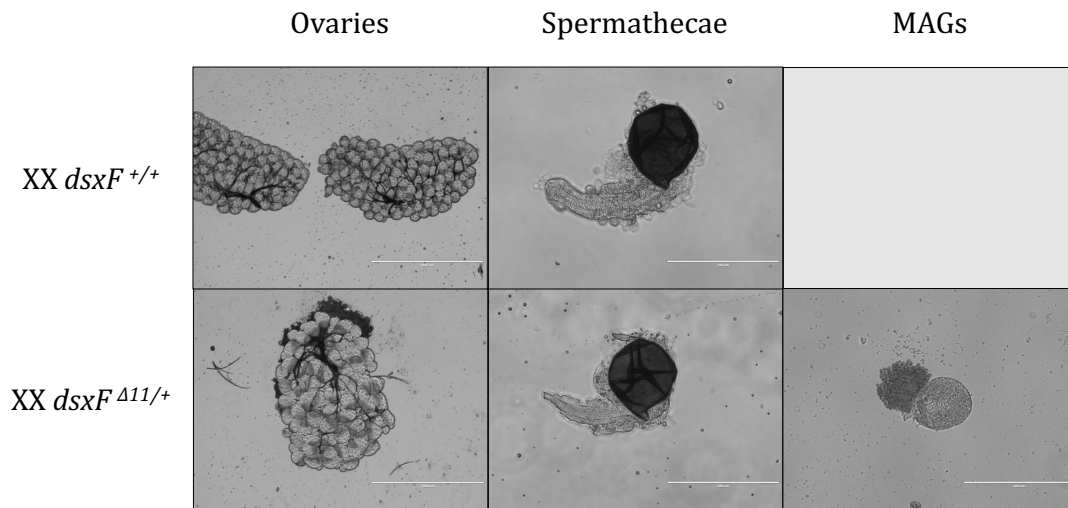


Figure 5.6 – Internal organs of chromosomally female individuals with intersex phenotype. Females bearing one copy of the *dsxF^{Δ11}* mutation (bottom) have ovaries and spermathecae that are visually similar to those of normal females (top). Simultaneously, they have organs resembling male accessory glands (MAGs). The white bar is equivalent to 200 μ m.

A fertility assay was performed by crossing females showing an intersex phenotype with wild-type males and recording the number of eggs produced and hatched larvae. The ability to feed on blood differed among intersex females; all of them tried to feed on blood but only some of them were able to feed. More importantly, the results confirmed that these genetically female individuals with an intersex phenotype are completely sterile, as none of them were able to lay eggs (Figure 5.7).

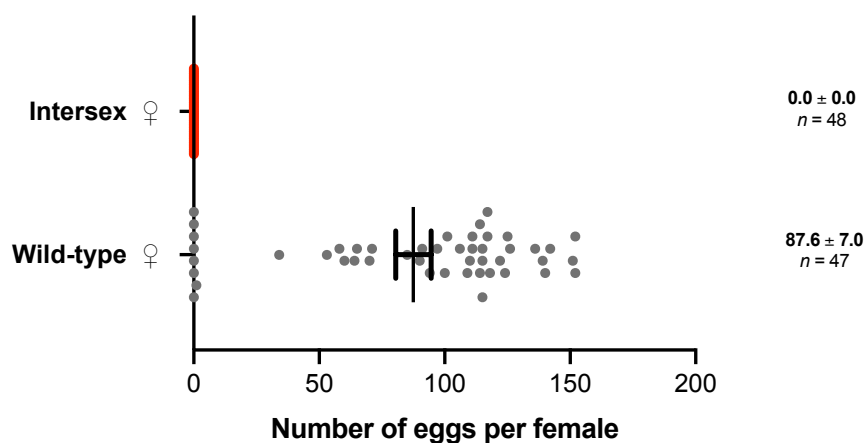


Figure 5.7 – Females showing an intersex phenotype are completely sterile. Females with an intersex phenotype bearing a copy of the *dsxF^{Δ11}* allele were unable to produce any eggs upon allowing them to mate with wild-type males and were given blood to feed. Each dot represents the progeny of a different female. The numbers on the right show the mean and the standard error of the mean, and the size of each sample (*n*).

5.2.2 The *dsxF^{Δ11}* allele is transcribed and would produce a truncated DSX^F protein

To verify that the *dsxF^{Δ11}* allele did not affect the female splicing of the *dsx* gene and that it was being transcribed, RNA was extracted from intersex individuals. Through RT-PCR, band extraction and subsequent Sanger sequencing of the *dsx* target sequence, it was corroborated that *dsxF^{Δ11}* was transcribed in heterozygous chromosomally female individuals. Given that this mutation is located at the beginning of exon 5, which constitutes the highly conserved female-specific part of the second DSX^F dimerization domain (OD2), the potentially translated *dsxF^{Δ11}* allele introduces a frameshift that would produce a truncated protein that would be missing a big part of this region and modifying the remaining amino acids, while conserving the rest of the protein intact (Figure 5.8).

To better understand the potential impact of this mutation in the structure of the DSX^F dimers, I used the ColabFold software with AlphaFold2 (Mirdita et al., 2022) to obtain a predicted model of the homodimers that this truncated protein could potentially form, as well as the heterodimers with a normal DSX^F protein (Figure 5.9). The predicted structure of the dimers DSX^F/DSX^{FΔ11} and DSX^{FΔ11}/DSX^{FΔ11} shows that the OD1 remains generally unaltered while the conformation of the OD2 and in particular the female-specific part is significantly distorted. The OD2 of DSX^{FΔ11} is considerably shorter than the OD2 of DSX^F and is missing a large portion of the C-terminal region of the protein, which seems to interact with other parts of the protein and the protein that it is dimerizing with.

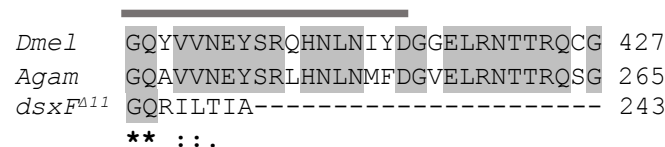


Figure 5.8 – Comparison of the amino acid sequence of the female specific region of DSX^F and DSX^{FΔ11}. The female specific region (FSR) of the DSX protein is highly conserved in *D. melanogaster* (*Dmel*) and *An. gambiae* (*Agam*). In DSX^{FΔ11}, the DSX^F version putatively produced by the *dsxF^{Δ11}* allele, the FSR is only 8 aa-long instead of the wild-type 30 aa-long, and only conserves the first two amino acids. This shows the big differences introduced by this mutation in the region that interacts with other proteins in the sex-determining pathway. The grey line indicates the female-specific part of OD2. Conserved amino acids in *D. melanogaster* and *An. gambiae* are highlighted in grey. Colon (:) and period (.) indicate strong and weak conservation of residue properties respectively. Sequence alignment was performed with the Needleman-Wunsch algorithm from the EMBOSS Needle tool (EMBL-EBI).

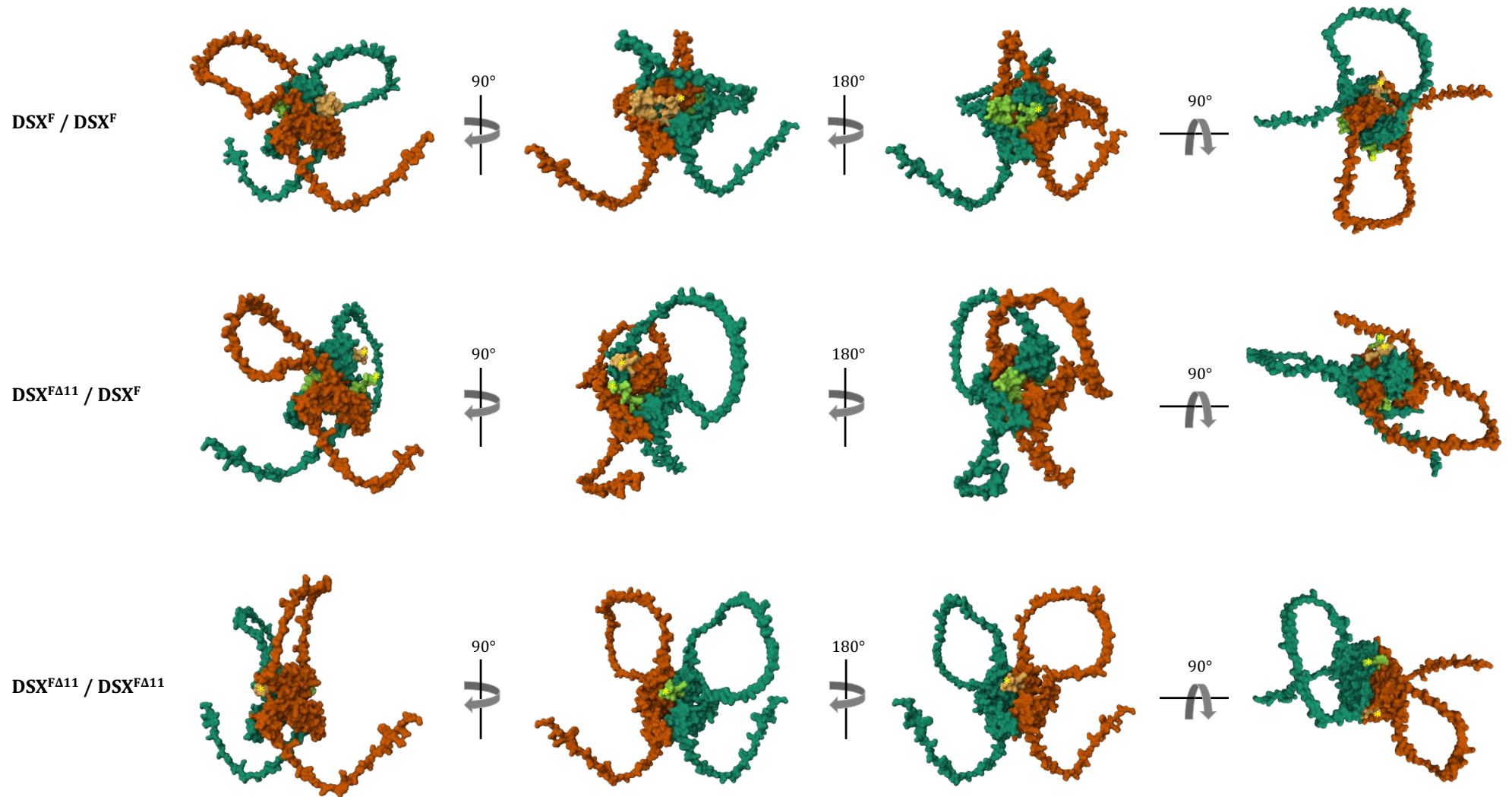


Figure 5.9 – Predicted structure of the homodimers and heterodimers formed by DSX^F and DSX^{FA11} proteins. The top row shows the predicted structure of the wild-type DSX^F homodimer (DSX^F/DSX^F), the middle one illustrates the predicted structure of the heterodimer (DSX^F/DSX^{FA11}) and the bottom row shows the predicted structure of the homodimer formed by two DSX^{FA11} proteins (DSX^{FA11}/DSX^{FA11}). Monomers are represented in different colours, one in green and one in orange. Lighter colours in the structure illustrate the female specific region of DSX proteins, where the modifications by the *dsx*^{FA11} allele are introduced. The C-terminal ends are indicated with a yellow star when shown from each perspective. In the heterodimer, the DSX^F protein is represented in green, and the DSX^{FA11} one is shown in orange. While the OD1 domain (half bottom part in the first three columns) is generally conserved among the three dimers, considerable structural changes can be observed in the OD2 and especially in the female specific region of the proteins.

5.3 A mutagenesis screen reveals the existence of other dominant and recessive mutations in the female-specific exon of *doublesex*

The existence of female-specific dominant negative mutations at the *dsx* locus in *An. gambiae* is of paramount relevance for the YLE strategy because it opens up the possibility to use the female isoform of *dsx* as a target to cause dominant sterility in the female offspring. In the previous chapter it has been shown that an efficient genetic control tool can be created based on the homing of the *dsxF^{Δ11}* allele with the Cas9 linked to the YLE^{dsx} construct. However, the mathematical modelling performed showed that the prospect of generating other dominant or recessive mutations in the target will influence the efficiency of the YLE^{dsx}.

To elucidate whether *dsxF^{Δ11}* is the only female-specific dominant mutation within the targeted region and to obtain information about the potential mechanisms that determine whether mutations in this area are dominant or recessive, I performed a mutagenesis screen on the target site of the YLE^{dsx} strains. In order to generate multiple mutations in the target site at the female-specific exon of *dsx* and study their effects in heterozygosity, I used a mosquito strain with a construct that expresses Cas9 under the male-specific $\beta 2$ late germline promoter and uses the exact same gRNA as the YLE^{dsx} lines. The reason for using this line is that the $\beta 2$ promoter is active at a later stage where Cas9-induced DNA breaks are more likely to be repaired by end-joining repair, compared to breaks that occur in the early germline (which are mainly repaired by homology-directed repair), and therefore more likely to generate increased variety of alleles.

Transgenic males were crossed (4 replicates) to wild-type females. Individuals in the offspring were sexed, and females were divided according to their phenotype (wild-type or intersex). From a total of ~1200 pupae sexed, only 13 females with an intersex phenotype were found. Amplicon sequencing was performed in a pool of 10 of these females and in 5 separate pools of 100 normal looking females, as well as in 3 single individuals from each group to verify that the phenotype was not caused by deposition of the nuclease and modification of both alleles in the embryos. The results confirmed that the females had half of their alleles modified, in coherence with lack of paternal Cas9

deposition in the embryo, and therefore revealed the existence of more female-specific dominant mutations (Figure 5.10).

Even though some alleles might be stochastically under or overrepresented in the amplicon sequencing results from pooled samples, I extrapolated the number of individuals bearing each of the mutations found. While it cannot be assumed that the numbers shown in Figure 5.10 are the exact number of individuals with these mutations in the sample, they provide an estimation to help understanding the effect and frequencies of these alleles. Taking into account both groups, only 87 females out of the total 519 present in the sequenced samples had a mutated allele, indicating that the mutation rate in the male progenitors' germline was approximately 17%. Despite these low numbers, there are interesting patterns that can be detected, especially looking at the location of the mutations in the target site.

Among the group of females with an intersex phenotype, I only found mutations that occurred within the coding sequence of exon 5. These mutations that in heterozygosity caused an intersex phenotype were classified as dominant. Except for one of the identified dominant alleles, the same mutations were also found in pools of normal looking females, reflecting a reduced penetrance. Whether these occurrences in the absence of any visible phenotype produced fertile females or females with internal intersex morphologies was not determined. Remarkably, two deletions within the coding sequence of exon 5 were found in the pool of wild-type females but not in the intersex one. Further research would be needed to understand whether this is due to the low numbers of individuals with a mutated allele in this study or to the fact that these are recessive or functional resistant mutations.

On the other hand, mutations that disrupted the intron 4 – exon 5 boundary, and thus the splicing acceptor site (NYAG), were only found in females with a normal phenotype, suggesting that these are recessive mutations. The effect of these mutations in homozygosity was never evaluated, but the fact that they remove the splice site preceding the female-specific exon and they do not infer any phenotype points towards the creation of null alleles.

		Intron 4																Exon 5																						Normal females	Intersex females
		53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90		
	Wild-type	T	T	T	A	T	G	T	T	T	A	A	C	A	C	A	G	G	T	C	A	A	G	C	G	G	T	G	G	T	C	A	A	C	G	A	A	T	A	-	-
	<i>dsxF^{Δ11}</i>	T	T	T	A	T	G	T	T	T	A	A	C	A	C	A	G	G	T	C	A	A	-	-	-	-	-	-	-	-	-	-	-	C	G	A	A	T	A	-	-
	* 75-CGG-78	T	T	T	A	T	G	T	T	T	A	A	C	A	C	A	G	G	T	C	A	A	G	-	-	-	T	G	G	T	C	A	A	C	G	A	A	T	A	1	0
	D 74-G-75	T	T	T	A	T	G	T	T	T	A	A	C	A	C	A	G	G	T	C	A	A	-	C	G	G	T	G	G	T	C	A	A	C	G	A	A	T	A	3	4
	D 73-AG-75	T	T	T	A	T	G	T	T	T	A	A	C	A	C	A	G	G	T	C	A	-	-	C	G	G	T	G	G	T	C	A	A	C	G	A	A	T	A	5	1
	D 72-AAGC-76	T	T	T	A	T	G	T	T	T	A	A	C	A	C	A	G	G	T	C	-	-	-	-	G	G	T	G	G	T	C	A	A	C	G	A	A	T	A	11	2
	* D 72-AAG-75	T	T	T	A	T	G	T	T	T	A	A	C	A	C	A	G	G	T	C	-	-	-	C	G	G	T	G	G	T	C	A	A	C	G	A	A	T	A	11	1
	D 71-CAAGCGGT-79	T	T	T	A	T	G	T	T	T	A	A	C	A	C	A	G	G	T	-	-	-	-	-	-	-	G	G	T	C	A	A	C	G	A	A	T	A	1	1	
	70-TCAAG-75	T	T	T	A	T	G	T	T	T	A	A	C	A	C	A	G	G	-	-	-	-	-	C	G	G	T	G	G	T	C	A	A	C	G	A	A	T	A	6	0
	* D 69-GTCAAG-75	T	T	T	A	T	G	T	T	T	A	A	C	A	C	A	G	-	-	-	-	-	-	C	G	G	T	G	G	T	C	A	A	C	G	A	A	T	A	0	1
	68-GGTCAAG-75	T	T	T	A	T	G	T	T	T	A	A	C	A	C	A	-	-	-	-	-	-	-	C	G	G	T	G	G	T	C	A	A	C	G	A	A	T	A	2	0
	67-AGGTCAAGCG-77	T	T	T	A	T	G	T	T	T	A	A	C	A	C	-	-	-	-	-	-	-	-	-	-	G	T	G	G	T	C	A	A	C	G	A	A	T	A	1	0
	67-AGGTCAAGC-76	T	T	T	A	T	G	T	T	T	A	A	C	A	C	-	-	-	-	-	-	-	-	-	G	G	T	G	G	T	C	A	A	C	G	A	A	T	A	8	0
	67-AGGTCAAG-75	T	T	T	A	T	G	T	T	T	A	A	C	A	C	-	-	-	-	-	-	-	-	C	G	G	T	G	G	T	C	A	A	C	G	A	A	T	A	7	0
	65-ACAGGTCAAGC-76	T	T	T	A	T	G	T	T	T	A	A	C	-	-	-	-	-	-	-	-	-	-	-	G	G	T	G	G	T	C	A	A	C	G	A	A	T	A	10	0
	65-ACAGGTCAAG-75	T	T	T	A	T	G	T	T	T	A	A	C	-	-	-	-	-	-	-	-	-	-	C	G	G	T	G	G	T	C	A	A	C	G	A	A	T	A	1	0
	63-ACACAGGTCAAG-75	T	T	T	A	T	G	T	T	T	A	-	-	-	-	-	-	-	-	-	-	-	-	C	G	G	T	G	G	T	C	A	A	C	G	A	A	T	A	1	0
	62-AACACAGGTCAAGCGGT-79	T	T	T	A	T	G	T	T	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	G	T	C	A	A	C	G	A	A	T	A	1	0	
	61-TAACACAGGTCAAG-75	T	T	T	A	T	G	T	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	C	G	G	T	G	G	T	C	A	A	C	G	A	A	T	A	4	0
	56-ATGTTTAACACAGGTCAAG-75	T	T	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	C	G	G	T	G	G	T	C	A	A	C	G	A	A	T	A	1	0

Figure 5.10 – Identified mutations at the *dsx* target in the mutagenesis screen. Representation of the target sequence environment in the intron 4 – exon 5 boundary of the *dsx* gene. The gRNA (and target sequence) is highlighted in blue, the PAM at the target site is highlighted in yellow, and the cut site is shown by a dashed red line. All the mutations found were deletions. The estimated number of individuals bearing each of the mutations in heterozygosity is shown in the right, divided in two groups: females with a normal phenotype and females showing an intersex phenotype. This number was extrapolated based on the frequencies of each mutation found in each pool of individuals and in the number of females present in each of those pools. Mutations that were found to cause an intersex phenotype in heterozygous females are marked with a red 'D' (left). Mutations that generated in-frame deletions in the coding sequence of exon 5 are marked with a red asterisk (left). None of the mutations disrupting the intron – exon boundary (and therefore the splicing acceptor site – NYAG) was found in females with an intersex phenotype, suggesting that these are always recessive.

The amplicon sequencing results of one of the individually sequenced females with intersex phenotype showed no mutations in the target site. This could be explained by a misclassification of the phenotype at pupal stage, or it could be due to the creation of a big deletion that lacked the sequence to which one of the primers was binding (55 bp downstream of the CDS of exon 5), thus preventing amplification of this allele.

Similar to the *dsxF^{A11}* allele, the majority of the other found dominant mutations would also generate proteins that either miss the female-specific part of the OD2 domain or completely alter the amino acid sequence of this region (Figure 5.11). Interestingly, two of these dominant deletions would only introduce small changes in the amino acid sequence:

- The 72-AAG-75 deletion would change the glutamine (Q) and alanine (A) at positions 237 and 238 respectively for a proline (P). Since the proline has a high conformational rigidity, this small change might affect the conformation of the OD2 domain, rendering it non-functional but still able to dimerize.
- The 69-GTCAAG-75 deletion would result in the elimination of the glycine (G) and glutamine (Q) at positions 236 and 237. These are two of the highly conserved amino acids in *Dmdsx* and *Agdsx*, thus their elimination might modify the OD2 conformation and functionality.

The two mutations that occurred within the coding sequence of exon 5 but were only found in the pool of wild-type females would have a very different effect in the amino acid sequence:

- The 75-CGG-78 deletion would only eliminate the alanine (A) at position 238, leaving the rest of the sequence intact.
- The 70-TCAAG-75 deletion would introduce much bigger changes in the amino acid sequence and would introduce an early stop codon.

		Exon 4				Exon 5																										Normal females	Intersex females				
		232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	262			262	263	264	265
	Wild-type	R	I	D	E	G	Q	A	V	V	N	E	Y	S	R	L	H	N	L	N	M	F	D	G	V	E	L	R	N	T	T	R	Q	S	G	-	-
	<i>dsxF^{Δ11}</i>	R	I	D	E	G	Q	R	I	L	T	I	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	* 75-CGG-78	R	I	D	E	G	Q	-	V	V	N	E	Y	S	R	L	H	N	L	N	M	F	D	G	V	E	L	R	N	T	T	R	Q	S	G	1	0
	D 74-G-75	R	I	D	E	G	Q	R	W	S	T	N	T	H	D	C	I	I	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	3	4
	D 73-AG-75	R	I	D	E	G	H	G	G	Q	R	I	L	T	I	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	5	1
	D 72-AAGC-76	R	I	D	E	G	R	W	S	T	N	T	H	D	C	I	I	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	11	2
	* D 72-AAG-75	R	I	D	E	G	P	-	V	V	N	E	Y	S	R	L	H	N	L	N	M	F	D	G	V	E	L	R	N	T	T	R	Q	S	G	11	1
D 71-CAAGCGGT-79		R	I	D	E	G	G	Q	R	I	L	T	I	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1
	70-TCAAG-75	R	I	D	E	G	G	G	Q	R	I	L	T	I	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	6	0
	* D 69-GTCAAG-75	R	I	D	E	-	-	A	V	V	N	E	Y	S	R	L	H	N	L	N	M	F	D	G	V	E	L	R	N	T	T	R	Q	S	G	0	1

Figure 5.11 – Comparison of the amino acid sequences potentially generated by mutations occurring in the coding sequence of exon 5.

5.4 Discussion

While female-specific dominant negative mutations in *dsx* had been reported in *D. melanogaster*, this had not been the case for *An. gambiae* prior to this study. Since the disruption of the coding sequence of the female-specific exon in *An. gambiae* led to a recessive mutation, it was not anticipated to find a dominant mutation within the same target sequence. The 11 bp deletion reported here had been previously found by Kyrou et al. (2018) when assessing the appearance of resistant alleles arising in response to a gene drive targeting the same sequence as used here. However, indel mutations found in that study were always paired with a gene drive construct disrupting the other allele, thus even though they could determine that the *dsx*^{Δ11} allele could not restore the normal protein function, its dominant effect could not be predicted. This deletion was the most commonly found one, which is in consonance with the fact that the *dsx*^{Δ11} allele was generated independently in both YLE^{dsx} strains.

This female-specific dominant mutation found in *dsx* in *An. gambiae* (*dsx*^{Δ11}) causes an intersex phenotype in females when present in heterozygosity, but contrary to the known dominant mutations in *D. melanogaster*, *dsx*^{Δ11} does not prevent the correct splicing of the female isoform nor does it lead to expression of the male isoform. This allele carries an 11 bp deletion within the coding sequence of the female-specific exon, leaving all the splicing sites and factor binding sequences intact. It has been shown to be transcribed, and its translation would result in a truncated protein (243 amino acids, instead of the 265 amino acids of the regular DSX^F). The resulting truncated protein would be missing virtually all the highly conserved female-specific part of the OD2, while conserving the rest of the protein intact. As explained before, DSX^F and DSX^M proteins bind to the same target sequences in the genome, which suggests that the female and male-specific parts of OD2 play a key role to interact with different transcription factors and other proteins to trigger different sex determination mechanisms in females and males. It has been demonstrated in *D. melanogaster* that this female-specific part of DSX is essential for the binding of the Intersex (IX) protein, which is required for the transcriptional regulatory function of DSX^F, and for which an *An. gambiae* ortholog exists (Amos et al., 2022; Garrett-Engle et al., 2002; Verhulst and van de Zande, 2015; Yang et al., 2008). At the same time,

being part of this dimerization domain, and as the predicted structure suggests, it seems that this female-specific region might contribute to the stabilization of the dimers. Therefore, the following hypotheses describe two non-mutually exclusive ways in which the truncated protein potentially produced by the *dsx^{Δ11}* allele that is missing this part of the domain could induce an intersex phenotype in heterozygosity:

- Homodimers of DSX^{FΔ11} form and bind to the target sequences, preventing DSX^F dimers from doing it. Because they lack the female-specific part of OD2, these homodimers would not be able to interact with the rest of regulatory elements to correctly continue the female differentiation process.
- DSX^{FΔ11} proteins form heterodimers with DSX^F, sequestering them from forming DSX^F homodimers. These heterodimers are likely to be less stable, as one of the dimers is missing an important part of OD2. While the OD1 domain is responsible for the DNA binding of DSX, it has been shown that dimerization and dimer stability substantially affect the DNA binding affinity and specificity through thermodynamically linked effects (Cho and Wensink, 1998), which could potentially interfere with the activation of the female differentiation process.

These mechanisms would also explain the observed differences in the external phenotype in different females, since the level of interference would be a stochastic process and the phenotype would vary among females.

The translation of the DSX^{FΔ11} protein, as well as the formation of homodimers and heterodimers, could potentially be experimentally determined. While the differences in molecular weight and isoelectric point between DSX^{FΔ11} and DSX^F are too small to separate them through gel electrophoresis (~2 kDa of difference and pI of 8.32 and 8.04 respectively), it might be possible to distinguish them by mass spectrometry. Alternatively, the proteins could potentially be tagged and expressed in insect cell lines for subsequent purification and analysis.

If the phenotype triggered by *dsx⁴¹¹* is indeed caused by the generation of a truncated protein that interferes with the normal DSX^F, three potential explanations to why the exon 5-disrupting *dsxF⁻* allele is recessive are proposed here. The first one is that the ~1.5kb insertion in the coding sequence of exon 5 prevents the correct splicing of the female isoform. The second one is that it produces a protein that cannot interact with the normal DSX^F protein, similar to the *dsxf* allele in *D. melanogaster* described earlier. In this case, the intersex phenotype would only appear in the absence of any functional DSX^F protein (i.e., when the *dsxF⁻* allele is present in homozygosity). A third plausible explanation is that the *dsxF⁻* transcript is not translated and therefore no protein is generated. There are different mechanisms to ensure the quality of the transcripts and nonsense-mediated RNA decay (NMD) is a conserved one among eukaryotic organisms (Nickless et al., 2017). Commonly, NMD eliminates transcripts with premature stop codons to prevent the production of aberrant proteins, but it also plays a role in the expression of certain genes (Metzstein and Krasnow, 2006; Nickless et al., 2017). While the *dsxF⁻* allele does not bear a premature stop codon per se, the mechanism by which these are detected might trigger the recognition of the *dsxF⁻* stop codon as premature. One of the models reviewed by He and Jacobson (2015) to explain this mechanism of identification is that an abnormally long 3' UTR downstream of a premature stop codon might lead to the accumulation of the UPF1 ATPase/helicase and trigger the transcript decay. The 3' UTR of *dsxF* is already considerably long, but in the case of *dsxF⁻*, the distance between the stop codon and the end of the mRNA would increase in ~1.4 kb. Therefore, this could induce its identification as a premature stop codon, as it is located relatively further upstream compared to the non-mutated transcript. In contrast, the *dsx⁴¹¹* mutation introduces a premature stop codon in the gene, but the difference in the distance between the stop codon and the end of the transcript in the normal allele and in *dsx⁴¹¹* is only 66 bp (in an mRNA that is 8,667 bp long). This could potentially be the reason why this transcript is not eliminated by NMD. Even though there are known canonical scenarios for the recognition of premature stop codons and the identification of transcripts as targets of NMD, there is not enough knowledge to effectively predict, based on a certain transcript sequence, whether a NMD will act on it or will leave it intact (Hug et al., 2016; Karousis et al., 2016).

Regarding the rest of mutations that were found to be dominant, since they all occur within the coding sequence of exon 5 and in the same region as *dsx*⁴¹¹, it seems likely that they would produce the phenotype following the same mechanism (i.e., generating a truncated protein that interferes with the wild-type DSX^F protein). The differences in the penetrance level of each mutation could be affected by the extent to which the generated proteins can dimerize in a stable manner, although this is speculation and would require further experiments to be confirmed. The fact that mutations disrupting the intron 4 – exon 5 boundary are only found in females with a normal phenotype reinforces the idea that dominant mutations generate a truncated protein that interferes with the normal function of DSX^F. While the effect on *dsx* expression of these recessive alleles was never evaluated, their lack of phenotype suggests that the removal of the intron 4 splice acceptor sites does not lead to the expression of the male isoform. A hypothesis that could explain this is that these mutations retain the putative binding sites for the Fle splicing factor at the end of exon 5, which might prevent the production of the male isoform and the absence of an intersex phenotype.

Two deletions within exon 5 were only found in the pool of normal females. This could be explained by the low number of individuals with a mutation obtained in the mutagenesis screen study, which did not allow to detect them in females with intersex phenotype, or because the changes produced prevent the dimerization of the protein, which would make these mutations recessive. A third potential explanation would be that the in-frame 75-CGG-78 deletion, which results in the elimination of only one amino acid, restores the function of the protein. A deeper analysis of this mutation would be needed, as it could entail a resistant allele that would be strongly selected for in a genetic control strategy that targets this region of the *dsx* gene. Given the changes introduced by the 70-TCAAG-75 deletion, and how similar the truncated protein would be to the one produced by *dsx*⁴¹¹, it seems more likely that this mutation is recessive rather than a resistant allele that restores the wild-type function.

Based on the nature of the female-specific dominant negative mutations identified in *D. melanogaster*, and with the available knowledge on *dsx* splicing mechanism in *An. gambiae*, there might exist another way to create such mutations in the latter species by preventing the transcription of the female isoform and forcing the expression of the male

one. In *D. melanogaster*, this can be achieved by displacing or eliminating the Tra2 binding sites that activate the weak 3' acceptor splice site preceding the female-specific exon, a necessary process to produce the female isoform. In *An. gambiae*, if Fle binding sites are removed (possibly together with the exon 5 – intron 5 boundary), this would inhibit the activation of the 5' acceptor site in intron 5, the female isoform would not be produced, and this might trigger the transcription of the male one. The expression of both DSX^F and DSX^M is expected to cause an intersex phenotype, as occurs in *D. melanogaster* (Nagoshi and Baker, 1990) and with the dominant negative mutations found in this project. However, if this was demonstrated and such mutation causes the expected phenotype, there are important considerations if this was to be used as part of a genetic control strategy. In theory, this allele could be homed in males with a source of Cas9 in the same form as the *dsx*⁴¹¹ allele was homed, causing sterility in the female offspring. In practice, mutations created by NHEJ or MMEJ in chromosomes where homing did not occur are expected to be more tolerated at the 3' UTR of exon 5 or at the beginning of intron 5 than in the coding sequence of exon 5. Therefore, it seems that resistance is more likely to arise if this region is targeted, and it would be essential to design the tool in a way that prevents or limits the emergence of functional resistant alleles. Multiplexing could be a good approach to solve this issue but given that several mutations in this part of the gene could still give rise to functional alleles, it is not guaranteed to prevent resistance occurrence.

To conclude, even though the mutagenesis screen provided useful and interesting information about the effect of several deletions within the *dsxF* target site, this was quite exploratory and more detailed research should be performed to better understand whether normal females with a dominant mutation are fertile, since the most relevant aspect for a genetic control strategy like the developed YLE^{dsx} is to induce female sterility, independently of other morphological alterations.

Nonetheless, the discovery of female-specific dominant mutations in *dsx* in *An. gambiae* is of significant relevance for this project in particular and for the development of genetic control strategies in general, as it provides a way of inducing female sterility in heterozygosity and proves that the use of dominant mutations can be extremely useful if they are incorporated in these technologies.

5.5 Methods

5.5.1 Phenotype evaluation

5.5.1.1 Fertility assay

Females with an intersex phenotype were crossed with wild-type males in groups of 60. They were left to mate for 5-7 days before being blood fed. A minimum of 50 females were allowed to lay eggs individually by placing them in separate cups 3 days after blood feeding. After 2 nights, the number of eggs laid by each female was recorded. The same procedure was followed with wild-type females as a control, for which the number of larvae hatching from each progeny was recorded as well.

5.5.1.2 Dissections of females showing an intersex phenotype

Adult females with an intersex phenotype were anaesthetised on ice for 5-10 minutes. Individuals were then placed on a glass slide with PBS solution and were dissected under a stereomicroscope using two needles. One needle was used to hold the thorax of the females while the other one was placed on the last segment of the abdomen to pull. Gonads and spermathecae were isolated from the rest of the body for visualization and recording under an EVOS XL Core Cell imaging system. Dissections of wild-type males and females were also performed for comparison.

5.5.2 Reverse Transcription PCR (RT-PCR) of the *dsx*⁴¹¹ allele

Three individual females with an intersex phenotype were collected and separately homogenised in a total of 500 µL of TRIzol. RNA was extracted and purified following the manufacturer's indications (Invitrogen, ThermoFisher). cDNA was then synthesised with the High-Capacity cDNA Reverse Transcription kit (Thermofisher). Using the cDNA of these samples as a template and the primers shown in Table 5.1, PCR amplification of the *dsxF* target site was performed (KAPA HiFi HotStart ReadyMix, Roche). The amplified

fragments were purified by gel extraction with the Wizard® SV Gel and PCR Clean-Up System (Promega), and subsequently sent for Sanger sequencing.

Table 5.1 – Primers used for amplification of the *dsxF* target site to verify the expression of the *dsx⁴¹¹* allele. The primers amplified a fragment of 263 bp using cDNA as a template.

Primer name	Primer sequence	cDNA amplicon size	gDNA amplicon size
IT57	CTCGTGCGTTTCGCCAGA	263 bp	33,172 bp
IT58	AATACCACCCGTCAGAGTGGAT		

5.5.3 Mutagenesis screen in the *doublesex* target site

A transgenic line bearing a construct with a gene coding for Cas9 under the $\beta 2$ promoter and a gRNA cassette targeting *dsxF* had been used previously to generate mutations in the target. Reportedly, this line created mutations in 40% of the alleles (Ioanna Morianou, personal communication).

Males from this transgenic line were crossed to WT females to assess the effect of different mutations in *dsxF* in their progeny. 4 replicates of this cross were made to ensure screening the progeny of different males and observe more independent mutagenesis events. In each of the replicates, 25 transgenic males were crossed to 25 WT females. 300 individuals from each cross were raised and sexed at pupal stage. Females with a normal phenotype at that stage were separated from females with an intersex phenotype. Genomic DNA extraction (Wizard® Genomic DNA Purification Kit, Promega) was performed individually for 3 females from each group, for a pool of 10 females showing an intersex phenotype and for 5 pools of 100 females showing a normal phenotype. Amplicon sequencing was performed on the *dsx* target to find mutations present in each group, following the procedure described in Methods 4.8.5.

6 Validating candidate haplo-insufficient genes in *Anopheles gambiae*

In the previous chapters, it has been described the generation of a YLE based on nuclease activity in the male germline that biases the inheritance of a female-specific dominant negative mutation to cause sterility in the female offspring. However, it has been observed that the prospect of creating recessive mutations or resistant alleles in the target sequence jeopardizes the efficacy of the developed YLE. Finding new target genes emerges as a necessary task to either produce a YLE with multiple targets or a substitute YLE less prone to create recessive or resistant alleles. An alternative to the use of dominant negative mutations could be found in targeting haploinsufficient genes that have an effect in the fertility or viability of the individual. In this chapter, I address the identification of putative haploinsufficient genes in *An. gambiae* and the attempts to validate the haploinsufficiency of three candidates to understand whether they could potentially be used as an alternative target of a YLE strategy.

6.1 Haploinsufficiency

Haploinsufficiency is a dominant condition by which the deletion of a gene or the generation of a null allele produce an altered phenotype in heterozygosity (Johnson et al., 2019; Veitia, 2002). Contrary to the previously explored dominant negative mutations, where the phenotype is caused by the production of a truncated protein that interferes with the wild-type one, in this case the phenotype is triggered by the insufficient amount of functional protein generated from only one allele. When this results in the death or sterility of the individual, the genes can be labelled as haplolethal or haplosterile (Johnson et al., 2019; Veitia, 2002).

6.1.1 Haploinsufficient genes in *Drosophila melanogaster*

Given the nature of these genes, their study is not as straightforward as those of haplosufficient ones, for which knockouts can be generated to analyse the phenotype effect in homozygosity. Nonetheless, in the model organism *D. melanogaster*, large deletions are relatively well tolerated in heterozygosity, and the extensive work on chromosome manipulation historically performed in this organism has allowed the investigation of haploinsufficient regions of the *D. melanogaster* genome (Cook et al., 2012; Parks et al., 2004; Ryder et al., 2007, 2004). In fact, Cook et al. (2012) performed an exhaustive deletion screen of the whole *D. melanogaster* genome and reported the existence of 49 genes or short genomic regions that are haplolethal or haplosterile. The vast majority of these genes code for cytoplasmic ribosomal proteins (CRP).

Ribosomal proteins assemble and, together with ribosomal RNA (rRNA), form the ribosomes (Brodersen and Nissen, 2005). Ribosomes are macromolecular complexes that translate all the mRNAs present in a cell and synthesize the proteins, thus the ribosomal protein genes and ribosomal RNA are generally ubiquitously expressed at extremely high levels (Brodersen and Nissen, 2005; Marygold et al., 2007). Many CRP genes have been shown to be haploinsufficient in diverse organisms, including yeast, *D. melanogaster* or humans (Cook et al., 2012; Deutschbauer et al., 2005; Draptchinskaia et al., 1999; Marygold et al., 2007; Ruggero and Pandolfi, 2003). This is in consonance with their function and the high numbers in which their products are needed, leading to a decrease of functional ribosomes in the cell if one of the two alleles is not generating the protein. In *D. melanogaster*, the deletion of certain CRP genes leads to haploinsufficient phenotypes with affected growth, development or fertility, but the most extreme phenotypes result in lethality (Cook et al., 2012; Marygold et al., 2007).

6.1.2 Haploinsufficient genes in *Anopheles gambiae*

In *An. gambiae*, Krzywinska et al. (2021) suggested that the *femaleless* (*fle*) gene might be haploinsufficient with a lethal effect only in female mosquitoes. *fle* seems to control the splicing of the female isoform of *dsx*, at the same time that it has a role in the regulation of the X chromosome transcription (Krzywinska et al., 2021). The knockdown of this gene in *An. gambiae* caused female lethality in strains with highly reduced levels of *fle* and strong masculinization of female individuals in strains with a milder reduction of *fle* transcription levels, whereas males were always unaffected. The strong phenotype observed only in females even with a relatively low knockdown of *fle* pointed towards a potential female-specific haploinsufficient gene, making it an ideal target for a YLE strategy. However, research performed in our group showed that the knockout of *fle* had no effect in females in heterozygosity, showing that one wild-type copy of the gene is enough to preserve a normal phenotype (William Garrood, personal communication).

This implies that no haploinsufficient genes have been reported for *An. gambiae*. In order to identify haplolethal or haplosterile genes in this organism, it seems reasonable to select the orthologs of the *D. melanogaster* CRP genes as candidates since they have been reported to be haploinsufficient in such a diverse range of organisms. Nevertheless, these are not sex-specific genes, and therefore there are two principal options to potentially incorporate them as targets of a YLE. The first one, if the target gene is located in an autosome, will require to insert a rescue copy within the YLE construct (Figure 6.1 – Left). The sequence of this copy should be recoded in order to not be recognized by the gRNA targeting the gene. The second option is to target a CRP gene located in the X chromosome, as the altered copy will always be inherited by the female offspring and the male offspring will inherit an unmodified X chromosome from the wild-type females (Figure 6.1 – Right). However, if the gene is needed for the spermatogenesis process, it will also be required to include a recoded rescue copy in the YLE transgene to prevent male sterility.

Since transgenes integrated in the Y chromosome have been constantly reported to be expressed at lower levels than from autosomal loci, and they are generally inconsistently expressed among males and among cells of the same individual (Bernardini et al., 2014; Gamez et al., 2021), it might be desirable to avoid relying on a rescue copy in the Y chromosome and find potential targets on the X chromosome.

2012). In *D. melanogaster*, the X chromosome of males is overtranscribed, leading to X-linked genes being expressed equally in males and females in similar levels to genes located in the autosomes (Lucchesi, 1973). Similarly, in the *Anopheles* genus, dosage compensation through the overexpression of the X chromosome in males has been observed both in *An. stephensi* and *An. gambiae* (Baker and Russell, 2011; Jiang et al., 2015; Rose et al., 2016). These studies showed that the average in expression levels in autosomal and X-linked genes were within the same range, although a slight female-bias in expression of X-linked genes was reported, probably due to the small amount of male-biased genes present in the *Anopheles* X chromosome (Magnusson et al., 2012; Rose et al., 2016). Rose et al. (2016) demonstrated that the *D. melanogaster* orthologs of haplolethal and haplosterile CRP genes are equally expressed in *An. gambiae* males and females. This holds true for both autosomal and X-linked CRP genes.

While the described dosage compensation mechanism in *Anopheles* occurs in somatic tissues and ovaries, it has been shown that dosage compensation is absent in the *An. gambiae* male germline (Baker and Russell, 2011; Rose et al., 2016; Taxiarchi et al., 2019). Additionally, as it has been described already, the X chromosome undergoes transcriptional silencing in the male germline due to meiotic sex chromosome inactivation (MSCI) (Taxiarchi et al., 2019).

Altogether, these studies suggest that it might be possible to find dosage sensitive genes on the X chromosome of *An. gambiae*, which can be explored further for their prospect as YLE targets.

6.2 Identification of candidate haploinsufficient genes in *Anopheles gambiae*

Based on the scarce knowledge on haploinsufficient genes in *An. gambiae* and the common haploinsufficiency associated to CRP genes in different organisms, I sought for haploinsufficient candidates in the orthologs of *D. melanogaster* haplolethal and haplosterile CRP genes. In *An. gambiae*, 3 of these genes are located in the X chromosome:

AGAP000739, AGAP000952 and AGAP000953 (also referred to as *0739*, *0952* and *0953* in this chapter). The orthologs of these 3 genes in *D. melanogaster* are haplolethal (Cook et al., 2012; Dietzl et al., 2007), hence the terms “haplolethal” and “haploinsufficient” to refer to these genes are used interchangeably hereafter.

To understand the magnitude of their potential role in the spermatogenesis process, I checked the expression levels of these 3 genes in the male germline. For this, I used the data obtained by Taxiarchi et al. (2019), who sequenced the RNA of 4 cell populations in the male testes: a premeiotic one containing germline stem cells and spermatogonia, two meiotic populations comprising primary and secondary spermatocytes, and a postmeiotic one formed by spermatids. The expression of the genes *0739*, *0952* and *0953* across these cell populations (Figure 6.2) suggests that they undergo transcriptional silencing due to MSCI, since they are initially highly expressed and these expression levels are strongly reduced in the meiotic populations. The *0739* gene is still transcribed in the late stages of spermatogenesis, even though at a much lower level. On the other hand, the genes *0952* and *0953* are almost completely silenced in the postmeiotic stage. The meiotic silencing (or partial silencing) of these genes indicates that their role in spermatogenesis may not be essential and it might be possible to modify them in the male germline without affecting male fecundity.

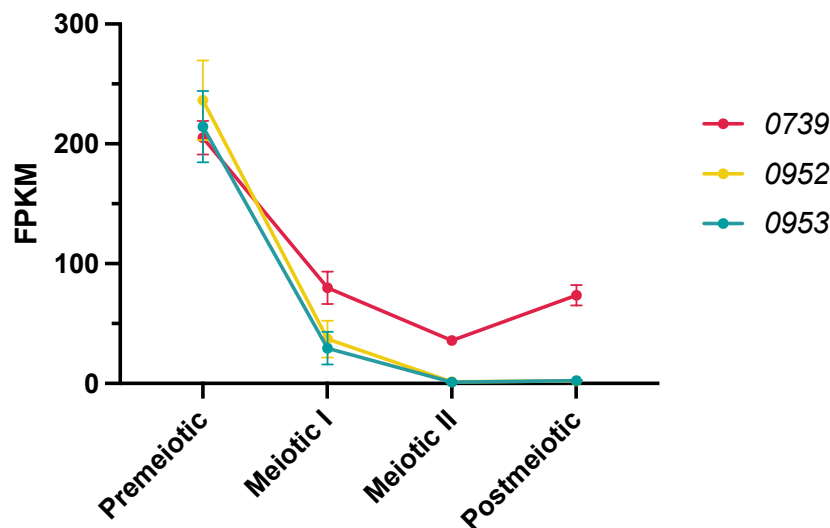


Figure 6.2 – Germline expression of the putative haploinsufficient genes. Expression levels (FPKM) of the candidate genes *0739* (AGAP000739), *0952* (AGAP000952) and *0953* (AGAP000953) in 4 cell populations across the spermatogenesis process. These include a premeiotic stage, two meiotic stages and a postmeiotic one. The considerable reduction of expression levels during and after meiosis seem to indicate that these genes undergo meiotic sex chromosome inactivation (MSCI), similar to most genes on the X chromosome.

Another essential consideration when targeting these genes is the existence of paralogs, since a duplicated gene that conserved the same function might rescue the phenotype that a disrupted copy of the candidate gene would cause otherwise (Koonin, 2005). Using the VectorBase database (Amos et al., 2022), it was possible to determine that *0739* does not have paralogs in *An. gambiae*; *0952* has an autosomal paralog, AGAP011501, whose coding sequence has a 61.7% of identity and similarity with the coding sequence of *0952*, while the amino acid sequence has a 60.4% of identity and 71.4% of similarity (Figure 6.3 – Top); and *0953* has an autosomal paralog, AGAP002395, with a 91.7% of sequence similarity and identity with *0953*, and a 100% identity of the amino acid sequence (Figure 6.3 – Bottom). Even though it seems highly likely that AGAP002395 and *0953* share the same function, it does not necessarily rule out the possibility that they are haploinsufficient genes. Therefore *0953* was included as a candidate in the study.

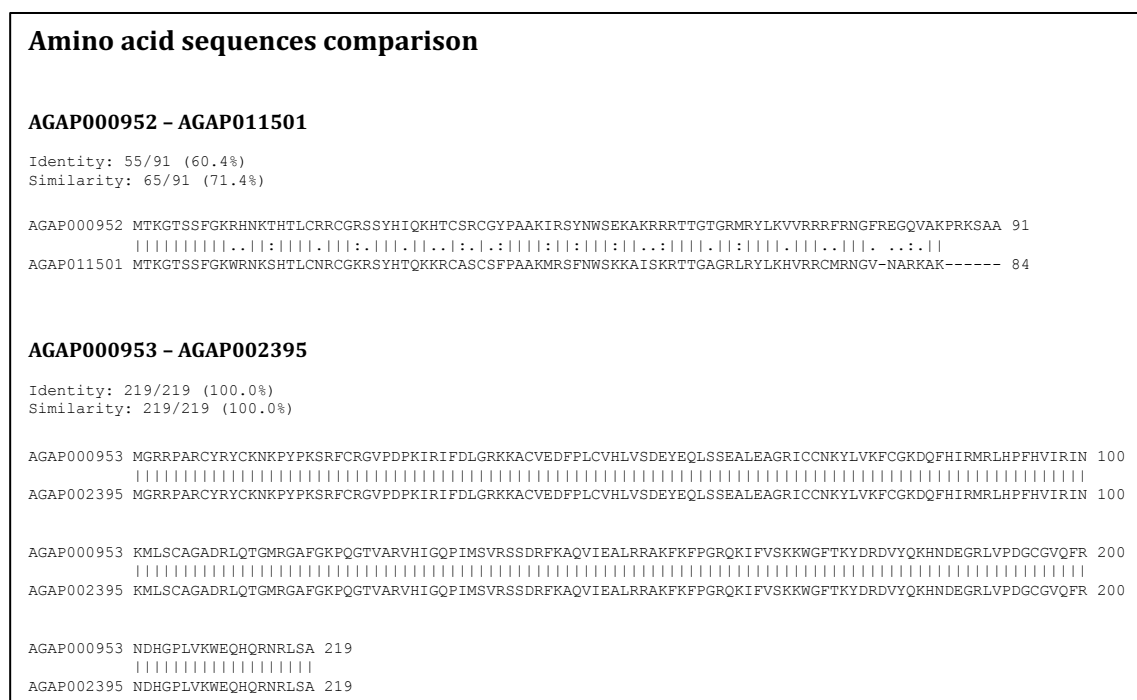


Figure 6.3 – Comparison of the amino acid sequences of the candidate haploinsufficient genes and their paralogs in *An. gambiae*. The candidate X-linked gene *0952* (top) has an autosomal paralog (AGAP011501) with an amino acid sequence with 60.4% identity and 71.4% similarity. The X-linked *0953* gene (bottom) has an autosomal paralog (AGAP002395) with 100% identity and similarity in their amino acid sequence, strongly suggesting that this paralog might rescue the function of *0953*. Colon (:) and period (.) indicate strong and weak conservation of residue properties respectively. Sequence alignment was performed with the Needleman-Wunsch algorithm from the EMBOSS Needle tool (EMBL-EBI).

6.3 Experimental plan

Prior to integrate the 3 candidate genes as YLE targets, the aim was to corroborate their haploinsufficiency. This validation cannot be obtained through the generation of a knockout line, as transgenic individuals would never be viable if the genes are haplolethal. Therefore, the experiment plan was the following: firstly, I would design a gRNA to target each gene in a highly conserved sequence. Secondly, I would separately generate transgenic lines bearing each of these gRNAs. Individuals from these lines would then be crossed with other mosquito strains carrying an autosomal construct with a Cas9 gene under a germline promoter. Males bearing both transgenes (Cas9 and gRNA) would undergo nuclease activity in the germline, and its effect would in theory be observed in the offspring of these males. If these males are sterile, this would be an indication of the candidate genes' essential role in the spermatogenesis process. However, if they are fertile, only the females would inherit the mutated allele, which would potentially cause lethality in these individuals if the target genes were truly haplolethal (Figure 6.4).

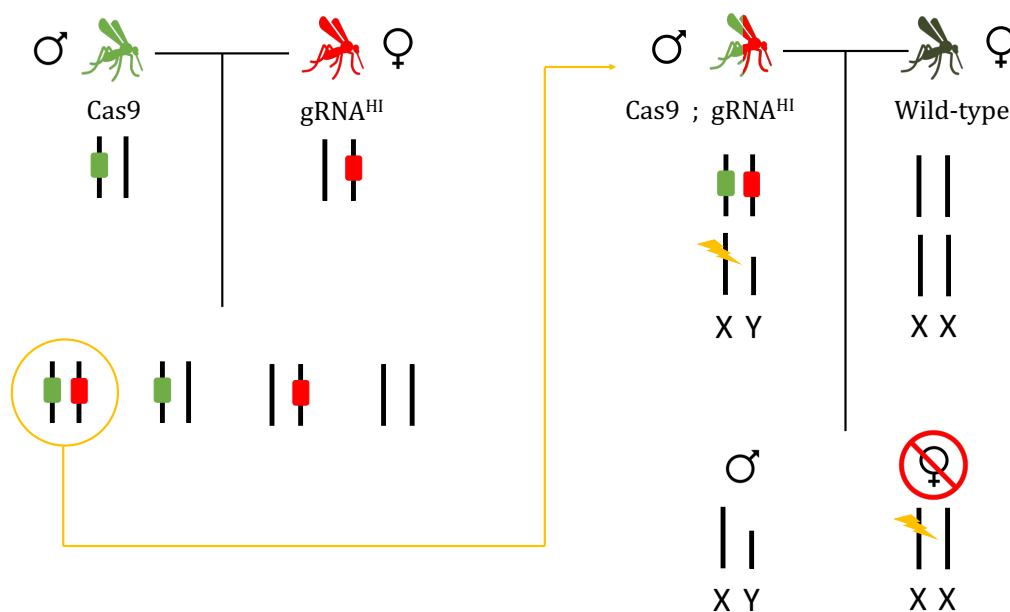


Figure 6.4 - Diagram of the experimental plan to verify the haploinsufficiency of the candidate genes. A two-cross experiment was designed to evaluate the haploinsufficiency of the candidate genes. First, an individual with a transgene coding for Cas9 under a germline promoter (green) would be crossed with an individual carrying a genetic construct with the gRNA targeting one of the candidate genes (red). Males in the progeny bearing both alleles would modify the target X-linked gene in the germline (yellow lightning bolt), and these males would be mated to wild-type females. The modified X chromosome would always be inherited by the females, who would also inherit an unmodified copy from the female progenitor. If the target genes are truly haploinsufficient, these females would not be viable.

6.4 Selection of the target sequences

A hypothetical YLE targeting one of these 3 candidate genes would have to generate new mutations every generation, because the YLE males would always inherit an unmodified X chromosome from a wild-type female. Therefore, it is especially relevant that the gRNA is designed to target them in a highly conserved region, where mutations disrupt the gene function, and possibly with a surrounding environment that would favour the creation of mutations through microhomology-mediated end joining (MMEJ) leading to out-of-frame and/or large deletions.

The software Guido (Kranjc and Thomas, 2023) was used to choose the gRNAs targeting each gene. This tool incorporates a software (Kranjc et al., 2021) that provides information about the conservation level of the sequence within the *An. gambiae* complex, off-target sequences in the genome of the *An. gambiae* G3 colony as well as any SNP that might be present in the gRNA sequence in the genome of this colony. Additionally, it predicts the most likely MMEJ repair outcomes based on small homologous sequences flanking the cut site. To complement this information, the web tool CRISPOR (Concordet and Haeussler, 2018) was used to obtain an additional prediction of the likelihood of generating frameshift mutations within the target. Table 6.1 shows the gRNAs finally selected to target the candidate haploinsufficient genes in *An. gambiae*.

Table 6.1 – Designed gRNA sequences to target the candidate genes. The sequences to target the putative haploinsufficient genes were selected based on conservation level across *Anopheles*, the lack of off-targets in the *An. gambiae* genome and the absence of SNPs at the target sequence in *An. gambiae*. The fifth column indicates the probability of a frameshift caused by any type of insertion or deletion in the target site (provided by the CRISPOR software).

Target gene	gRNA sequence	Off-targets with ≤ 3 mismatches	SNPs at the target sequence	Probability of frameshift mutations
0739	GTGAAGGAGCAG TTCGCC	0	0	81%
0952	GTACTTCCAGCT TTGGAAAG	0	0	87%
0953	GCGAATCACGTG GAACGGA	0	0	72%

The 0739 gene codes for the small subunit ribosomal protein S10 and contains a plectin/S10 domain likely to be involved in RNA binding (Mistry et al., 2021; Paysan-Lafosse et al., 2023). This domain is entirely located in the anterior half of exon 2 and is the most conserved region of the gene, hence the selected gRNA to target this gene is located within this domain (Figure 6.5 – Top).

The gene 0952 encodes a large subunit ribosomal protein L37, with a conserved region among L37 ribosomal proteins located at the beginning of exon 2. Both this area and the end of the coding sequence are the highest conserved ones. Because modifications in the very end of the gene might introduce smaller changes in the protein and may be better tolerated, the chosen gRNA to target 0952 is located at the start of exon 2 (Figure 6.5 – Middle). To avoid undesirable modification of this gene's paralog, the gRNA was designed in a region that would have 5 mismatches in the AGAP011501 gene (Figure 6.6 – Top).

Lastly, the 0953 gene codes for a large subunit ribosomal protein L10 and has a conserved L10/L16 ribosomal protein domain at the end of exon 3 and beginning of exon 4. Despite being located relatively far downstream from the start of the coding sequence, the selected gRNA targets the beginning of the mentioned domain, which at the same time is the most conserved part of the gene (Figure 6.5 – Bottom). To prevent Cas9 activity against the paralog of this gene, I designed a gRNA to target 0953 in a sequence with 3 mismatches in AGAP002395, in addition to another mismatch in the PAM “NGG” sequence, removing this motif in AGAP002395 (Figure 6.6 – Bottom).

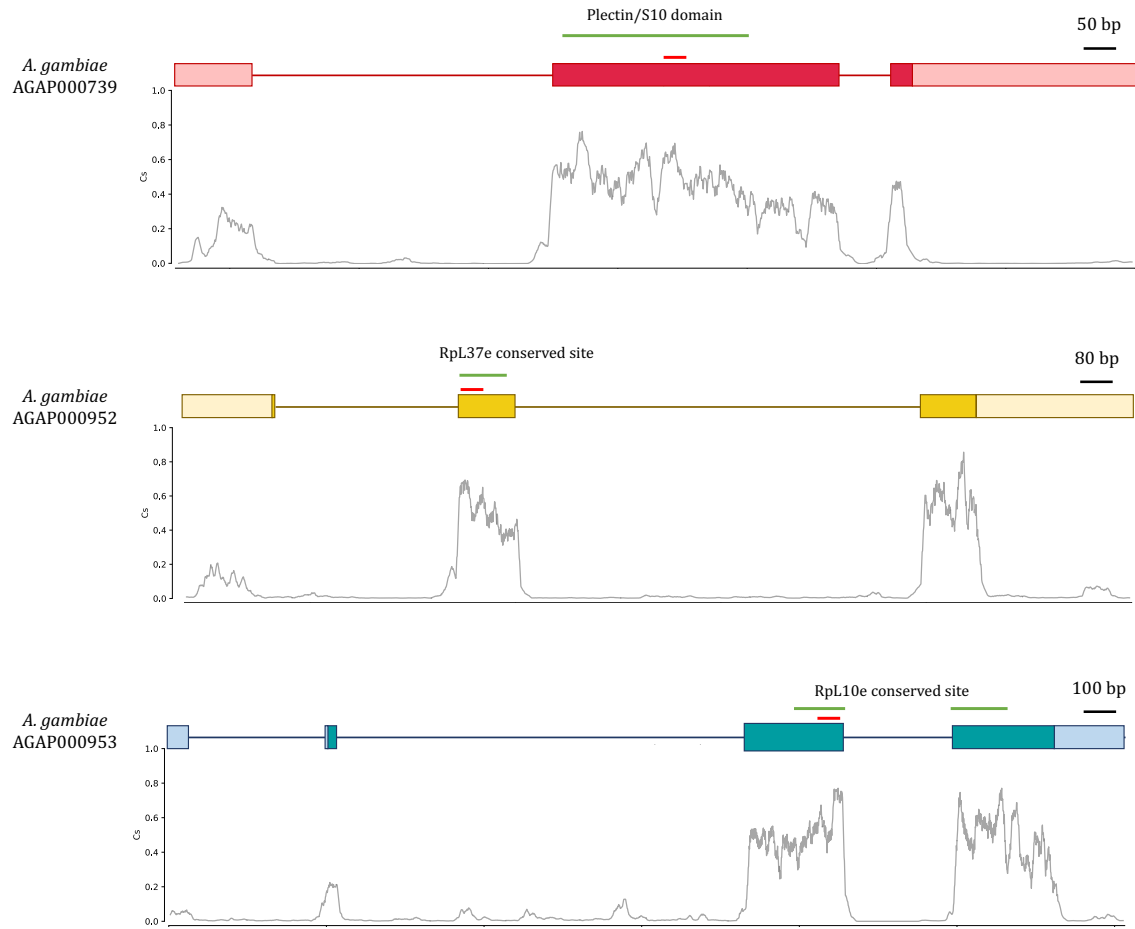


Figure 6.5 – Conservation levels across the candidate haploinsufficient genes. Representation of the conservation levels of the candidate genes, ranging from 0 (least conserved) to 1 (totally conserved). Conservation score is calculated taking into account the genomes of the *Anopheles* genus, as well as other mosquito species (*Ae. darlingi*, *Ae. albimanus*, *Ae. aegypti*, *Culex quinquefasciatus*) and *D. melanogaster* (Kranjc et al., 2021). Exons are represented by boxes, with dark colours indicating the coding sequences and light colours illustrating untranslated regions. Introns are not drawn to scale. The green line shows where conserved domains are located in each gene, and the red line indicates the location of the selected gRNA in each case.

AGAP000952 – AGAP011501

[illegible][illegible]

153

6.5 Generation of the gRNA lines

Upon selection of the gRNA sequences to target each of the 3 candidate genes, they were individually cloned into a plasmid that contained a gRNA cassette under a *U6* promoter. The 3 cloned plasmids were separately microinjected into embryos from the wild-type *An. gambiae* G3 colony for random integration.

Randomly integrated transgenes can land in transcriptionally silenced (or partially silenced) regions or within a gene locus, which could pose a fitness effect due to the disruption of its sequence. For these reasons, for each of the gRNAs, 3 lines were established from 3 independent integration events, to ensure that at least one would show high and stable transgene expression. Besides this, to determine the genomic location where the transgene had been integrated, I performed inverse PCR in the 9 transgenic strains. The results revealed that the strains for which it was possible to determine the integration locus were located in intergenic regions or within big introns (Table 6.2), which reduces the possibility of the transgene entailing a fitness cost.

Table 6.2 – Inferred genomic location of the randomly integrated transgenes in the gRNA strains. The most plausible genomic location of the transgene for each generated line is shown here (when it was possible to infer), together with the information about the genomic environment and how far from coding genes the transgenes were in each case.

Strain name	Most likely location	Genomic environment	Comments on environment
0739-1	2R:58504010	Intergenic region	1 kb from DNA polymerase subunit genes
0739-2	3L:1215994	Intergenic region	20 kb from closest gene
0739-3	3L:35162761	Intergenic region	400 bp from closest gene
0952-1	2L:28244484	Intergenic region	1kb from closest gene
0952-2	Undetermined	-	-
0952-3	3R:47814453	Intronic region (AGAP029234)	Intron length >15kb
0953-1	2R:27910681	Intronic region (AGAP002814)	Intron length >20kb
0953-2	Undetermined	-	-
0953-3	2L:1450977	Intergenic region	9kb from closest gene

6.6 Genetic crosses with Cas9-bearing strains

Following the experiment described earlier, in order to assess the effect of targeting each of the 3 candidate genes, the strains with the designed gRNAs (gRNA^{HI} strains) were crossed with lines carrying autosomal transgenes with a Cas9 gene under different promoters. The purpose of this was to target the candidate genes at early stages of spermatogenesis to evaluate their essentiality in the male germline, and at later stages to generate mutations without affecting the spermatogenesis process and assess the phenotype of the females inheriting a mutated allele.

6.6.1 Crosses with *vas2::Cas9* strain

Since the *vas2* promoter drives expression in the early germline of *An. gambiae* (Papathanos et al., 2009), targeting the candidate genes with a *vas2::Cas9* line would indicate whether they have an essential role in the spermatogenesis process. Additionally, *vas2* is the only germline promoter that has been successfully used for transgene expression from the Y chromosome in *An. gambiae* (Alcalay et al., 2021; Bernardini et al., 2014, Chapter 4 of this thesis), therefore this cross would inform about the feasibility of targeting these genes with a *vas2::Cas9* Y-linked construct.

However, it was not possible to recover males carrying both the *vas2::Cas9* and the gRNA transgene targeting either *0739* or *0952*. A likely explanation for this might be the somatic expression driven by the *vas2* promoter observed in prior experiments in *An. gambiae* (Hammond et al., 2021a, 2016), which would lead to embryonic death due to the nuclease activity in certain somatic tissues. This revealed the essential role of *0739* and *0952* for the viability of the individuals, but it did not allow to understand whether they are required in the male germline or if they are actually haploinsufficient, since the lethality observed might be caused by mutagenesis in both alleles.

Contrarily, it was possible to find double positive males for the lines targeting *0953* (*vas2::Cas9/+ ; gRNA⁰⁹⁵³/+*). Two out of the three *gRNA⁰⁹⁵³* lines produced viable offspring, where the sex ratio was evaluated and revealed the absence of sex bias (Figure 6.7). The two most likely explanations for this seemed to be that either the target gene had not been

modified and the females had inherited a wild-type allele from the progenitor male, or that mutations within the target sequence did not cause lethality in heterozygosity. To answer this, the females were collected and divided in two groups: females carrying both constructs (which would potentially experience additional Cas9 activity in ovaries and somatic tissues) and the rest of females. Then, amplicon sequencing of the target sequence was performed in these two samples of pooled females. The results revealed that virtually all the females that were either wild-type or had inherited only one construct were heterozygous for a frameshift mutation in 0953 (Table 6.3). Females carrying both Cas9 and gRNA⁰⁹⁵³ had more than half of their alleles modified, consistent with additional nuclease activity (Table 6.3). While the fertility of the *vas2::Cas9/+ ; gRNA⁰⁹⁵³/+* males was not accurately analysed, finding viable females with out-of-frame mutations in the 0953 gene in heterozygosity, and moreover viable females with more than half of their alleles modified, suggests that 0953 is not haplolethal and hence it is not a suitable target for a YLE strategy.

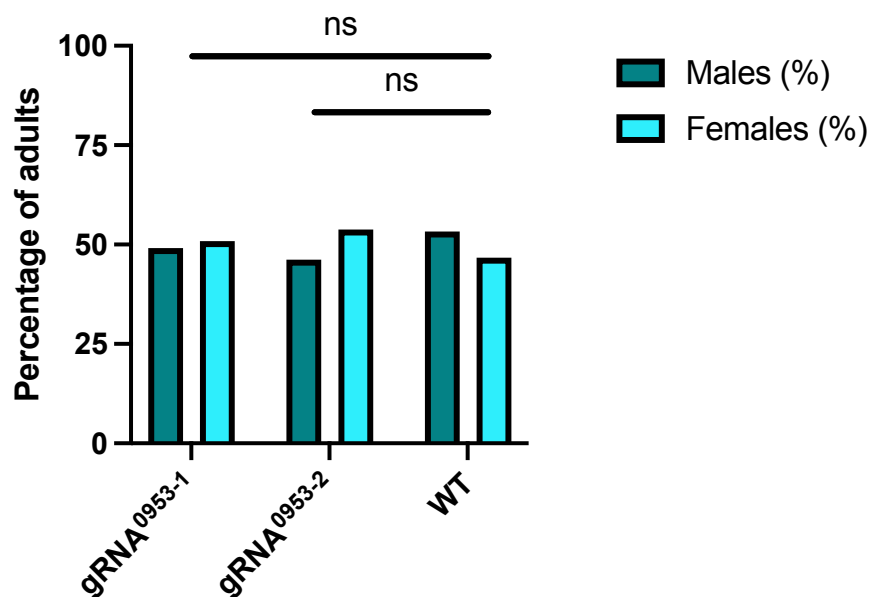


Figure 6.7 – Sex ratio in the progenies of *vasa::Cas9 ; gRNA⁰⁹⁵³* males. Percentage of adult males (dark blue) and adult females (light blue) in the progenies of *vasa::Cas9 ; gRNA⁰⁹⁵³⁻¹* males, *vasa::Cas9 ; gRNA⁰⁹⁵³⁻²* males and wild-type males. No significant differences were observed between the sex ratio in the offspring of those males and the wild-type control ($P = 0.1521$, $P = 0.2210$, respectively, Fisher's test). The significance was calculated over the total number of individuals sexed, but the plot shows the percentages for easier comparison.

Table 6.3 – Frequency of mutations in the target gene in the progenies of *vasa::Cas9* ; gRNA⁰⁹⁵³ males. Females in the offspring of *vasa::Cas9* ; gRNA⁰⁹⁵³ males were pooled according to their genotype for amplicon sequencing of the target site. Because a wild-type allele was always inherited from the female progenitor, the percentage of individuals bearing a mutation corresponds to the double of the frequency of said mutation (i.e., if the frequency of mutations in the progeny is 50%, this reflects that 100% of the individuals carry a mutation). When the sum of all mutations is >50%; this might be explained by either contamination or very low loads of nuclease deposition. In both gRNA⁰⁹⁵³ strains, 80-100% of the females that did not bear both alleles (Cas9 and gRNA⁰⁹⁵³) at the same time carried a mutation in the target gene, and this was always a frameshift mutation. Similarly, in both gRNA⁰⁹⁵³ strains, females that carried both alleles (Cas9 and gRNA⁰⁹⁵³) had more than 50% of their alleles mutated, and most of them were out-of-frame mutations. The fact that these females were viable indicates that the *0953* gene is haplosufficient.

Group	Frequency of mutations in the progeny (%)	Frequency of frameshift mutations in the progeny (%)
0953-1		
Non-transheterozygous females	40.73	40.73
0953-1		
Transheterozygous females	64.86	62.18
0953-2		
Non-transheterozygous females	51.06	51.06
0953-2		
Transheterozygous females	72.10	69.86

Only the lines gRNA⁰⁷³⁹⁻¹, gRNA⁰⁹⁵²⁻¹, gRNA⁰⁹⁵³⁻³ were kept after this experiment (referred to as just gRNA⁰⁷³⁹, gRNA⁰⁹⁵², gRNA⁰⁹⁵³ hereafter).

6.6.2 Crosses with $\beta 2::Cas9$ strain

To evaluate the haploinsufficiency of the *0739* and *0952* genes, targeting them with a $\beta 2::Cas9$ construct presents certain advantages. First, the $\beta 2$ promoter drives expression in the late stages of spermatogenesis, when the candidate genes show a lower expression level, thus modifying them at this stage seems less likely to entail a fertility cost in males. Secondly, the sex ratio distortion systems developed in *An. gambiae* are based on nuclease expression under the $\beta 2$ promoter (Galizi et al., 2016, 2014; Windbichler et al., 2008). These technologies target repetitive sequences on the X chromosome, leading to non-viable X-bearing sperm and unaffected Y-bearing sperm. In my case, the expectation is not to generate non-viable X-bearing sperm, but to create mutations in the X-linked target genes without interfering in the spermatogenesis process and analysing the effect of these mutations in heterozygosity in the female progeny. The results of the cited studies suggest that this might be achievable using a $\beta 2::Cas9$ strain. While it has been shown that the *0953* gene is not haploinsufficient, the gRNA line targeting it was included in the experiment as a control, by targeting an X-linked gene that is not haplolethal.

Males that inherited both transgenes ($\beta 2::Cas9/+$; $gRNA^{HI}/+$) were found and crossed to wild-type females. This time, females were allowed to lay eggs individually and the hatching rates were recorded to spot putative embryonic lethality. The progenies were then separately maintained, and the sex ratio was evaluated at adulthood. As shown in Figure 6.8, there was no significant decrease in the hatching ratio nor sex bias at adulthood in any of the 3 crosses (data provided by Silvia Grilli was added to show that the used $\beta 2::Cas9$ strain is active and generates sex distortion when provided a gRNA targeting the X-linked ribosomal RNA repeats – rRNA).

Again, females in the progeny were collected to perform amplicon sequencing of the targets and verify whether the absence of sex distortion was due to the fact that the target gene had not been modified or that the generated mutations did not cause lethality in heterozygosity. Since the $\beta 2$ promoter is male-specific, in this case it was not necessary to separate the females in two groups because they would not undergo additional nuclease activity. The sequencing results confirmed that the females did not contain any mutation within the target sequence of the 3 candidate genes, including the not haplolethal *0953*

(Table 6.4). The results observed in these crosses suggested absence of Cas9 activity against these targets in the germline of the male progenitors.

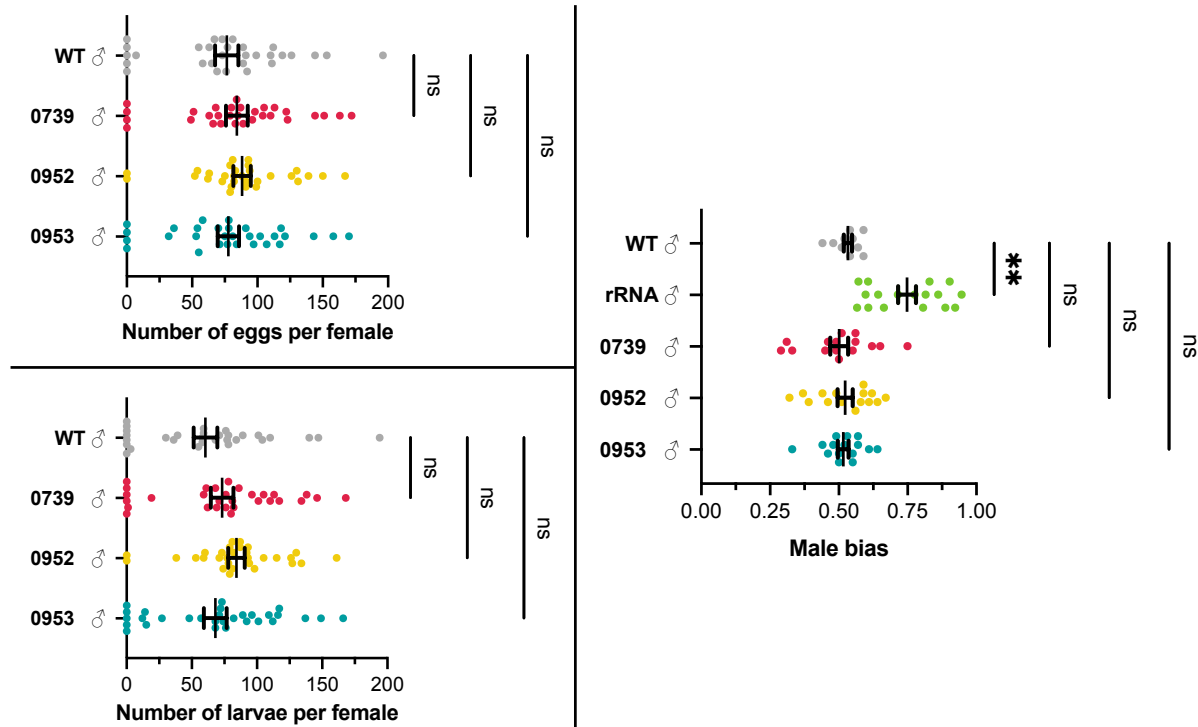


Figure 6.8 – Fertility of $\beta 2::Cas9$; $gRNA^{HI}$ males and sex distortion in their progeny. Males bearing a $\beta 2::Cas9$ genetic construct and a $gRNA^{HI}$ transgene targeting one of the candidate genes (0739, 0952, 0953) were crossed to wild-type females to evaluate the effect of inheriting a mutated allele in the female offspring. The number of eggs and larvae produced by the females mated to these males was not significantly different to those that mated with wild-type males ($P = 0.7029$, $P = 0.2303$, one-way ANOVA). In the progenies of $\beta 2::Cas9$; $gRNA^{HI}$ males there was no male bias; the sex ratio was not significantly different from the sex ratio in the progeny of wild-type males ($P > 0.9999$, Kruskal-Wallis test). Males bearing the same $\beta 2::Cas9$ genetic construct and a gRNA targeting the ribosomal repeats on the X chromosome (rRNA) caused significant male bias in their progenies compared to wild-type males ($P = 0.0015$, Kruskal-Wallis test). Bars and error bars indicate the mean and the standard error of the mean respectively.

Table 6.4 – Frequency of mutations in the target gene in the progenies of $\beta 2::Cas9$; $gRNA^{HI}$ males. Females in the progenies of these males were pooled for amplicon sequencing of the target site. Three progenies from each cross (i.e., three progenies per target gene) were used for this analysis, which revealed the almost total absence of mutations in the target. Just 15 and 1 reads contained a mutation, which could be explained by an artifact in the sequencing process or a really low dose of paternal deposition that generated mutations in very few cells. In any case, none of the frequencies correspond to heterozygous individuals for a mutation.

Target gene	Frequency of mutations in the progeny (%)	Number of amplicon sequencing reads with mutations
0739	0	0/235,496
0952	0.0068	15/217,865
0953	0.0005	1/187,484

6.6.3 Cross with Y-linked $vas2::Cas9$ strain

The crosses with the $\beta 2::Cas9$ strain failed to corroborate the haploinsufficiency of the 0739 and 0952 genes, and the crosses with the $vas2::Cas9$ line only allowed to discard the haploinsufficiency of 0953. I then hypothesised that the $vas2$ promoter might drive less somatic expression from the Y chromosome than from an autosome, and therefore hopefully males with both the $Cas9$ construct and the $gRNA^{HI}$ one would be viable and endure nuclease activity in the germline. For this reason, I performed an exploratory cross of the previously described YLE^{dsx-B} strain (carrying a Y chromosome-linked genetic construct with a $vas2::Cas9$ cassette) with the $gRNA^{0739}$ line.

In this case, it was possible to recover males that inherited both transgenes, consistent with the hypothesis that $vas2$ might cause less somatic leakiness from the Y chromosome. When the males carrying both constructs were mated with wild-type females, I observed that their fecundity was severely affected, since only 7 out of 25 females mated to these males produced viable larvae (Figure 6.9 – Left). These progenies were separately

maintained and the sex ratio at adulthood showed no significant sex bias (Figure 6.9 – Right). These results suggest that the 0739 gene is needed at least in the early spermatogenesis, and that targeting it in the early germline triggers male sterility.

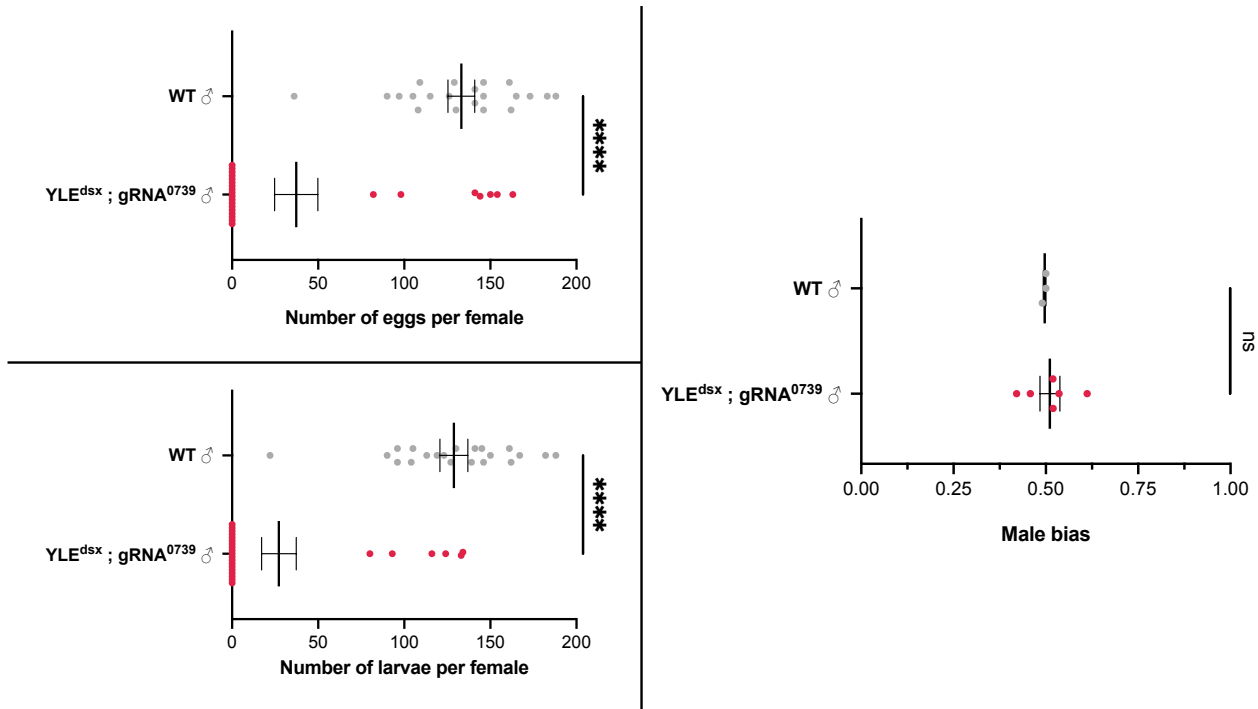


Figure 6.9 – Fertility of *YLE^{dsx}; gRNA⁰⁷³⁹* males and sex distortion in their progeny. Males bearing a Y chromosome-linked construct with a *vas2::Cas9* genetic construct and an autosomal transgene with a gRNA targeting *0739* had a strongly reduced fertility, revealing a potential relevant role of *0739* in the male germline. Females mated to these males (n=25) produced an average of 37.3 eggs (± 12.6 s.e.m) and 27.2 larvae (± 10.1 s.e.m) compared to the females that mated with wild-type males (n=21) that produced on average 133.2 eggs (± 7.8 s.e.m) and 128.9 larvae (± 8.1 s.e.m). **** indicate a P value <0.0001 in a Mann-Whitney test. The progenies generated by *YLE^{dsx}; gRNA⁰⁷³⁹* males were analysed for sex distortion, but the male bias was not significantly different from the wild-type control (P = 0.5119, Mann-Whitney test). Bars and error bars indicate the mean and the standard error of the mean respectively.

6.7 Discussion

The discovery of haploinsufficient genes that cause sterility or lethality can be very advantageous for the development of YLE strategies, as an alternative form of causing a dominant phenotype in the offspring. However, an efficient YLE requires female-specific lethality or sterility, and the haplolethal or haplosterile genes found by previous research in *D. melanogaster* are not sex-specific (Cook et al., 2012; Marygold et al., 2007). For this reason, three X chromosome-linked genes (*0739*, *0952* and *0953*) were selected as candidate haplolethal targets in *An. gambiae*, with the aim of modifying them in the male germline without affecting male fecundity and causing lethality in the female offspring that would inherit the modified X chromosome, leaving the male progeny unaffected. The haploinsufficiency of these candidate genes was assessed by targeting their most conserved regions with a Cas9 nuclease.

Very recently, a study was made public in which they followed the same experimental plan as the stated here, to target a gene located in a haplolethal region in the X chromosome of *D. melanogaster* (Lawler et al., 2023). In this case, the target gene was *wupA*, which encodes several isoforms of Troponin I, involved in muscle contraction and cell proliferation (Casas-Tintó et al., 2016; Casas-Tintó and Ferrús, 2021). Analysing the progeny of males bearing Cas9 under the germline *nanos* promoter and three gRNAs targeting *wupA*, Lawler et al. (2023) observed high male bias and significant decrease of the total number of offspring, consistent with female embryonic lethality induced by mutations in the haplolethal target. The cited study validates the approach used in this project and provides an additional candidate, since the *wupA* ortholog of *An. gambiae* is also found in the X chromosome, with the advantage that it is not needed for the spermatogenesis process. The absence of recessive mutations in the surviving females seems to indicate that *wupA* is completely haplolethal and that no mutations are tolerated, which would make it a perfect target for a YLE strategy.

Targeting the candidate genes with *vas2::Cas9* allowed to discard the *0953* gene as haplolethal. Since it was targeted in the most conserved region and the vast majority of the mutations found in sequenced females were out-of-frame, it seems unlikely that those modifications would be generating a functional protein. To corroborate this, an additional

study could be performed by targeting the gene in a sequence upstream of the one used in this experiment. Given that *0953* has a paralog that produces a protein with 100% of identity in the amino acid sequence, it seems plausible that the paralog might be rescuing the *0953* function.

On the other hand, the haploinsufficiency of *0739* and *0952* could not be evaluated with this cross, since the copresence of the *vas2::Cas9* construct and the gRNA targeting one of these two genes resulted in lethality, likely due to the somatic expression driven by the *vas2* promoter. Lethality triggered by somatic nuclease activity induced by the *vas2* promoter was also observed in *D. melanogaster* by Lawler et al. (2023) when targeting the *wupA* gene. Still, this indicated that *0739* and *0952* are essential for the viability of the individuals, contrary to *0953*.

Interestingly, when the candidate genes were targeted with a Cas9 under the $\beta 2$ promoter, no sign of nuclease activity was found. It was especially surprising to see absence of Cas9 cleavage in the non-haplolethal *0953* gene. While an explanation for these observations was not found, a hypothesis that would elucidate them is that since these genes undergo MSCI (Turner, 2007), the epigenetic silencing experienced in the meiosis and early postmeiosis might restrict the access of Cas9 to their sequences, preventing from mutating them. As a result, no effect in the males' fecundity can be observed and no mutations in the target can be found among the individuals in the offspring. Following this theory, the observed sex distortion when targeting the ribosomal RNA repeats could be explained because there are hundreds of copies and the rRNA is probably needed to some extent at all times, making the target more accessible for Cas9 (Galizi et al., 2016, 2014; Windbichler et al., 2008).

Finally, targeting the *0739* gene with a *vas2::Cas9* construct linked to the Y chromosome suggested that this gene is required in the male germline, as it led to a huge reduction in male fertility. Finding that 6 out of 25 females that mated with males carrying a Y-linked *vas2::Cas9* and the gRNA⁰⁷³⁹ produced viable offspring is in consonance with the variegation observed in Cas9 activity from the Y-linked construct in Chapter 4. It would be interesting to perform this same experiment targeting *0952*, which is almost completely silenced during and after meiosis (while *0739* is still expressed even if at lower levels than in premeiotic populations), to see if the outcome would be different. Lawler et al. (2023) also tried to target *wupA* with a Y-linked *vas2::Cas9* construct. However, the

leakiness of the promoter led to lethality in individuals bearing the Cas9 and the gRNA targeting *wupA*, equally to what occurred here with the autosomal *vas2::Cas9* targeting 0739 and 0952. These results suggest that the leakiness caused by the *vas2* promoter from the Y chromosome is stronger in *D. melanogaster* than in *An. gambiae*, which encourages the idea of using the ortholog of *wupA* as a target of a YLE in *An. gambiae*.

A potential experiment to determine whether the 0739 and 0952 genes are haplolethal would be the following. First, inserting a recoded copy of the gene (or part of the gene) within the endogenous locus, so that it cannot be cleaved by the designed gRNA. This has been done previously in *D. melanogaster*: Champer et al. (2020) developed a homing drive within a haplolethal gene to prevent the emergence of resistant alleles. Then, males bearing the recoded copy and homozygous for the gRNA construct ($gRNA^{HI} / gRNA^{HI} ; X^{recoded} / Y$) should be crossed to females carrying the *vas2::Cas9* cassette ($vas2::Cas9 / vas2::Cas9 ; X / X$). All the females in the offspring would then inherit a wild-type copy of the target gene and the recoded one. Since the *vas2* promoter causes embryonic deposition from females (Papathanos et al., 2009), Cas9 would be deposited in all the embryos, which would also carry the $gRNA^{HI}$ transgene, triggering embryonic nuclease activity only against the wild-type copy of the target. If the candidate gene is indeed haplolethal, this would result in the death of all females of the offspring. On the contrary, if the gene is not haplolethal, all the females should be viable because they would have one functional allele (Figure 6.10).

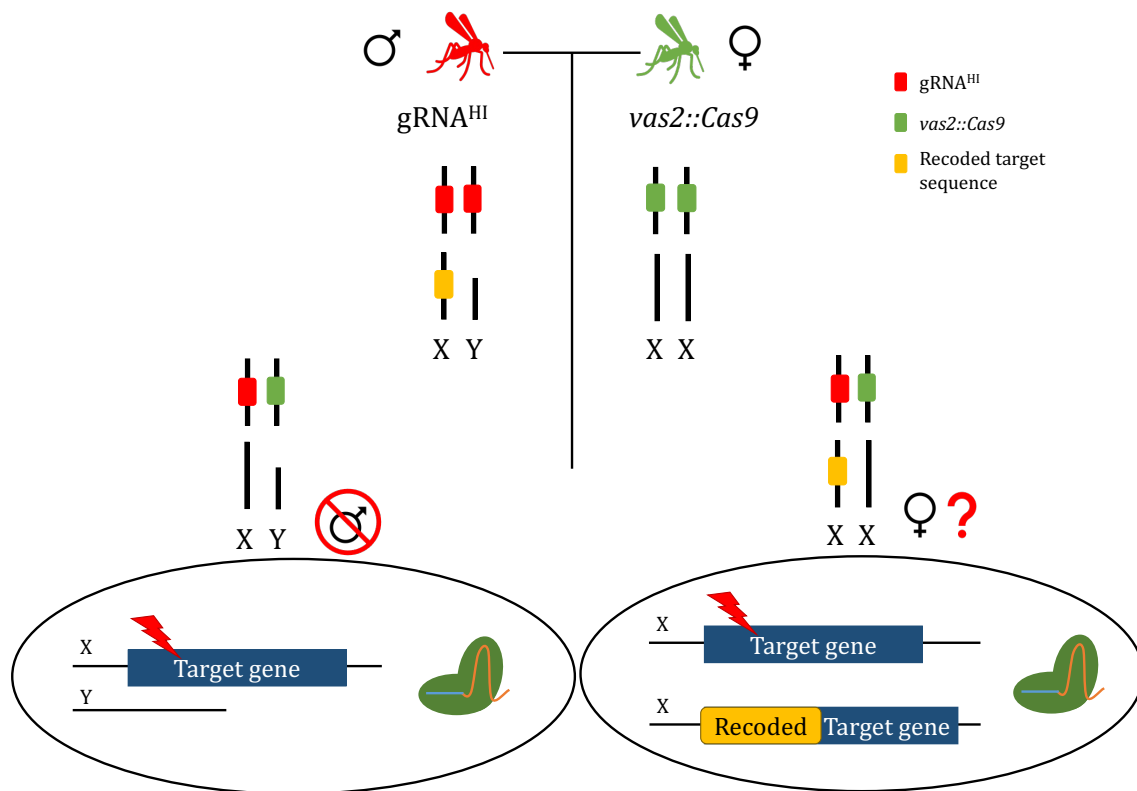


Figure 6.10 - Diagram of a potential experiment to corroborate the haploinsufficiency of the candidate genes 0739 and 0952. A way to finally corroborate the haploinsufficiency of the candidate genes would involve generating a modified copy (in yellow) of the target gene (in blue) in the endogenous locus by recoding part of the sequence so that it is not recognised by the gRNA. Crossing males bearing the recoded version of the gene and the gRNA against the target gene (in red) with females carrying a *vas2::Cas9* construct (in green) would induce maternal deposition of the Cas9 and nuclease activity in the embryos. Because the target gene is on the X chromosome, the recoded copy will always be inherited by the females, who will then have only one copy of the gene that can be recognised by the gRNA. If these females are viable, it would mean that one healthy copy of the target gene is enough, and hence the candidate would not be haploinsufficient. However, if the females are not viable, this would confirm the haploinsufficiency of the candidate gene.

In conclusion, the 0953 gene was discarded as a potential target of a YLE strategy, while further research is needed to understand whether 0739 and 0952 could be integrated as targets in such strategy. Nonetheless, the results seem to suggest that targeting these genes in the male germline without including a recoded rescue copy will be challenging. On the other hand, the results of a very recent study in *D. melanogaster* render the ortholog of the *wupA* gene as a really promising haplolethal target for a YLE strategy in *An. gambiae*.

6.8 Methods

6.8.1 Selection of the target sequences

A target sequence was designed for each candidate gene using Guido (Kranjc and Thomas, 2023), a software that finds gRNA target sites present in a specific region. This tool provides with information regarding the conservation of those sequences within the *An. gambiae* species and within the *Anopheles* genus. A gRNA was selected for each gene based on the absence of off-target sequences in the *An. gambiae* genome, on the absence of SNPs within the target site in the *An. gambiae* G3 colony and verifying that they were located in highly conserved regions of the gene.

6.8.2 Cloning of plasmids

The designed gRNA spacer sequences were inserted into the plasmid p125, containing a gRNA cassette under a *U6* promoter. The gRNA spacers were created using 18-20 bp oligonucleotides (Table 6.5) with overhangs for directional insertion in p125 after digestion with Eco31I for Golden Gate assembly. The plasmid p125 contains *piggyBac* inverted repeats for random integration in the genome.

Table 6.5 - The gRNA spacer and target sequences selected to target the candidate haploinsufficient genes. The sequences of 18-20 nucleotides forming the gRNA spacers are homologous to the target sequence in the candidate genes. The target sites include the PAM sequences (underlined) that allows the Cas9 nuclease to cleave.

Target	gRNA spacer sequence	Target sequence
AGAP000739	GTGAAGGAGCAGTTCGCC	GTGAAGGAGCAGTTCGCC <u>TGG</u>
AGAP000952	GTACTTCCAGCTTTGGAAG	GTACTTCCAGCTTTGGAAG <u>GCGG</u>
AGAP000953	GCGAATCACGTGGAACGGA	GCGAATCACGTGGAACGGAT <u>TGG</u>

Purification of the 3 plasmids was performed as described in Methods 4.8.2.

6.8.3 Generation of transgenic lines

Microinjections of *An. gambiae* embryos were performed as described in Methods 4.8.3.1. The injection mixes used to create these lines are shown in Table 6.6. Surviving larvae were screened for transient expression of the fluorescent marker, reared and their progeny was screened for expression of the fluorescent marker in the expected pattern. For each target gene, 3 transgenic lines were established from single integration events (i.e., from a single transgenic individual).

Table 6.6 – Injection mixes to create the lines bearing the gRNA targeting each of the haploinsufficient candidates. These are the donor and helper plasmids (and their concentrations) included in each injection mix. They were all injected in embryos coming from the wild-type colony. All the mixes included a diluted volume of the stock 10x microinjection buffer (50 mM KCL, 1 mM Na phosphate buffer pH 7.2).

Donor plasmid	Helper plasmid	Line injected
p125_0739 (200 ng/ μ L)	vasa2::transposase (200 ng/ μ L)	Wild-type
p125_0952 (200 ng/ μ L)	vasa2::transposase (200 ng/ μ L)	Wild-type
p125_0953 (200 ng/ μ L)	vasa2::transposase (200 ng/ μ L)	Wild-type

6.8.4 Molecular characterization of the transgenic lines

Inverse PCR was performed in all the transgenic lines to determine the genomic location where the transgene had been integrated. Genomic DNA was extracted from each line (Wizard® Genomic DNA Purification Kit, Promega) and two digestion reactions per sample were set with Sau3AI and HinP1I restriction enzymes. Digested genomic DNA was ligated into circular DNA vectors (T4 DNA Ligase, NEB) and the regions flanking the 5' and 3' *piggyBac* repeats were amplified with two subsequent PCR reactions (Phusion® High-Fidelity PCR Master Mix with HF Buffer, NEB). The amplified fragments were purified (Wizard® SV Gel and PCR Clean-Up System, Promega) and sent for Sanger sequencing (Eurofins Genomics). To confirm the absence of mutations in the gRNA spacer sequences, they were amplified and Sanger sequenced using the same kits and company mentioned above.

6.8.5 Genetic crosses to evaluate the haploinsufficiency of the candidate genes

In order to test the haploinsufficiency of these candidate genes, the transgenic lines bearing the gRNAs were crossed with lines containing a gene coding for Cas9 under different germline promoters. The purpose of these crosses was to obtain males with both transgenes. In the germline of these males, there would be Cas9 activity against one of the target genes located in the X chromosome and the effects would be observed in their offspring. The same procedure was followed for each of the 9 gRNA^{HI} strains (3 strains per target gene).

6.8.5.1 Crosses with *vas2::Cas9* strain

Males from the *vas2::Cas9* line were mated with females from the different gRNA^{HI} lines. The offspring of this cross was screened at pupal stage to select males that inherited both transgenes (*Cas9/+* ; *gRNA^{HI}/+*). It was only possible to recover males carrying both transgenes for the lines targeting *0953*. These males (*vas2::Cas9/+* ; *gRNA⁰⁹⁵³/+*) were then mated with wild-type females, in numbers of 25 males and females per cross. From the crosses that produced viable progeny, ~600 larvae were kept and sexed at adult stage to evaluate the sex ratio.

6.8.5.2 Crosses with *β2::Cas9* strain

Females carrying the *β2::Cas9* transgene were mated with males from the gRNA^{HI} lines. Males that inherited both transgenes (*β2::Cas9/+* ; *gRNA^{HI}/+*) were recovered for the 3 gRNA^{HI} lines, and they were mated with wild-type females in crosses of 60 males and 50 females (1-2 days after emerging). Adults were allowed to mate for 5 days prior to feeding blood to the females. Females were allowed to lay eggs individually 3 days after blood-feeding and the number of eggs laid and number of larvae that hatched from each progeny were recorded. The progenies were then separately maintained, and the sex ratio was evaluated at adulthood.

6.8.5.3 Cross with YLE^{dsx} strain

Males from the YLE^{dsx}-B line were crossed with females from the gRNA⁰⁷³⁹ line. In the progeny, males that inherited both transgenes were selected and crossed to wild-type females (60 males and 50 females, 1-2 days after emerging). Adults were allowed to mate for 5 days prior to feeding blood to the females. Females were allowed to lay eggs individually 3 days after blood-feeding and the number of eggs laid and number of larvae that hatched from each progeny were recorded. The progenies were then separately maintained, and the sex ratio was evaluated at adulthood.

6.8.6 Amplicon sequencing of the target sequences

Amplicon sequencing of the candidate genes was performed following the same procedure as described in Methods 4.8.5, as well as the analysis of the data.

The primers used to amplify each of the target genes are shown in Table 6.7.

Table 6.7 – Primers used for amplicon sequencing of the candidate haplolethal genes target sequence. The primers amplified a fragment of 320 bp, 300 bp and 331 bp around the target site of 0739, 0952 and 0953 respectively. Illumina partial adapter sequences are underlined.

Primer name	Primer sequence
AmpEZ-0739_F	<u>ACACTCTTTCCCTACACGACGCTCTTCCGATCT</u> ATGTTTCATGCCAAAAGCACA
AmpEZ-0739_R	<u>GACTGGAGTTCAGACGTGTGCTCTTCCGATCT</u> GTGCCTGGGTGCGGGCAGCTG
AmpEZ-0952_F	<u>ACACTCTTTCCCTACACGACGCTCTTCCGATCT</u> ATTTTGCTCTTGTTCAAAA
AmpEZ-0952_R	<u>GACTGGAGTTCAGACGTGTGCTCTTCCGATCT</u> GTAAACATGCATATTCATTACA
AmpEZ-0953_F	<u>ACACTCTTTCCCTACACGACGCTCTTCCGATCT</u> CTACCGCTACTGCAAGAACA
AmpEZ-0953_R	<u>GACTGGAGTTCAGACGTGTGCTCTTCCGATCT</u> GTAAATAGCGCAAAAGGCACA

7 Discovery of regulatory sequences to induce paternal deposition of transcripts in *Anopheles* embryos

The previous chapters have focused on the development of a Cas9-based YLE strategy that biases the inheritance of a female-specific dominant negative mutation, and on the search for putative haploinsufficient targets genes on the X chromosome. An alternative to these approaches may be found in the paternal deposition of a female-specific toxin in the embryos to induce female sterility or lethality in the offspring of males bearing a YLE construct. This toxin could be a male-determining factor, in the form of mRNA or protein, or another element that targets RNAs or proteins that are required for female development (Burt and Deredec, 2018). In this chapter, I report the search for paternally deposited transcripts in *Anopheles* embryos to eventually use their regulatory sequences for the development of an alternative YLE strategy.

7.1 The need to identify regulatory sequences for paternal deposition

Among the regulatory sequences generally used to drive germline expression of transgenes in *Anopheles* mosquitoes, *vas2* has been shown to induce maternal deposition of transcripts and proteins in the embryos (Hammond et al., 2021a; Papathanos et al., 2009), and the *nos* promoter and terminator seem to induce certain maternal deposition too (Hammond et al., 2021a). However, there are not known and tested regulatory sequences that cause consistent paternal deposition. While Kyrou et al. (2018) reported signs of nuclease paternal deposition when expressed under the *zpg* promoter, this remains to be confirmed and it has not been observed when the transgene was inserted in other loci, suggesting that the potential deposition from males induced by *zpg* promoter might have a strong position-dependent effect (Hammond et al., 2021; Anna Strampelli, manuscript in preparation). Therefore, in order to induce paternal deposition of transcripts or proteins in *An. gambiae*, new regulatory sequences must be identified.

7.1.1 Paternal deposition of transcripts and proteins in other organisms

The presence of RNA molecules in mature sperm has been widely reported in worms and mammals (Gòdia et al., 2018; Krawetz, 2005; Ostermeier et al., 2004; Stoeckius et al., 2014). Initially considered an artifact, currently there are numerous studies that identified non-coding-RNAs and mRNAs in the sperm of different mammal species and their role in male fertility and embryogenesis is being examined, but little is known about spermatozoal RNA in insects (Barroso et al., 2009; Gòdia et al., 2018; Hosken and Hodgson, 2014; Miller et al., 2005; Miller and Ostermeier, 2006). Interestingly, Fischer et al. (2012) found functional coherence of the sperm transcripts in humans and in *D. melanogaster*, with an enrichment of transcripts coding for proteins involved in translation. More importantly, Fischer et al. (2012) reported the detection of 3 sperm transcripts in 30% of the fertilised embryos, confirming paternal transference of RNA to the oocyte.

More recently, Degner et al. (2019) investigated the proteome and transcriptome of the seminal fluid and sperm in *Aedes aegypti* mosquitoes. Most of the proteins and transcripts identified in the ejaculate transferred to the females show high levels of expression in the male accessory glands (MAG), suggesting that they are transferred in intact MAG cells or vesicles derived from the MAG that fuse with the sperm (Degner et al., 2019). Transfer of mRNA to the females in the ejaculate of *Ae. aegypti* males had been previously demonstrated by Alfonso-Parra et al. (2016). Nonetheless, neither of these studies analysed whether the detected paternal RNA was entering the oocyte upon fertilization.

The lack of *D. melanogaster* and *Ae. aegypti* transcripts reported to be consistently paternally deposited in all or the majority of the embryos discards the possibility of using the regulatory sequences of *An. gambiae* orthologs to corroborate paternal deposition in this organism. An option to find regulatory sequences that cause paternal deposition is to study the transcripts present in the embryos to detect which ones have a paternal origin and which ones have a maternal one. Even though the vast majority of transcripts and proteins in the early embryos are deposited by the female, some studies suggest that the paternal contribution might have been underestimated (Gòdia et al., 2018; Lalancette et al., 2008). In two independent experiments in the *C. elegans* nematode, Stoeckius et al.

(2014) identified hundreds of paternally deposited mRNAs in the zygote, using divergent strains and performing RNA sequencing in the early embryos. To perform a similar experiment in *Anopheles* mosquitoes, it is fundamental to know when the zygotic genome activation occurs, in order that only deposited transcripts are considered.

7.1.2 Zygotic genome activation in *Anopheles*

In animals, the newly formed zygote is transcriptionally inactive, and the early development is controlled by RNAs and proteins that have been predominantly maternally deposited (Lee et al., 2014; Schier, 2007; Tadros and Lipshitz, 2009). The developing embryo then enters the maternal-zygotic transition, during which the maternal RNAs and proteins begin to be degraded and the first genes start to be zygotically transcribed (in a minor wave that is followed by a later major transcriptional wave) (Lee et al., 2014; Schier, 2007; Tadros and Lipshitz, 2009). The absolute time and cell cycle state at which zygotic genome activation occurs vary significantly across different organisms; for instance, mouse embryos commence the minor wave transcription at one-cell stage 10 hours after fertilization and the major wave during cell cycle 2 at ~24 hours post-fertilization, while *D. melanogaster* embryonic minor wave transcription starts after 8 cell cycles ~1 hour after fertilization and the major wave occurs ~2.5 hours post-fertilization at cell cycle 14 (Lee et al., 2014; Tadros and Lipshitz, 2009).

In the major mosquito vector of malaria in Asia, *Anopheles stephensi*, the embryos are not transcriptionally active before 1 hour after oviposition, and the early expressed genes are detected 2-3 hours after egg deposition (Criscione et al., 2013; Wu et al., 2018). Equally, a recent experiment performing RNA sequencing of early *An. gambiae* embryos did not detect expression of genes that mark the onset of zygotic genome activation, such as the *zelda* ortholog (Liang et al., 2008), at 1.5 hours post-oviposition, but they were highly expressed 3 hours after egg deposition, suggesting that *An. gambiae* zygotic genome activation occurs ~2-3 hours after oviposition (Matteo Vitale, manuscript in preparation).

7.2 Experimental design to detect paternal deposition in *Anopheles* embryos

In order to find paternally deposited transcripts in *Anopheles* embryos, I designed an experiment based on the sequencing of RNA molecules present in the early embryos and the genomes of the progenitors. For this, the progenitors should be sufficiently closely related to produce offspring, but their genomes should be different enough to allow the identification of the transcripts' origin based on their divergent variants. Two sibling species of the *An. gambiae* complex, such as *An. gambiae* and *An. arabiensis*, can produce hybrid offspring and therefore provide an excellent opportunity to use the differences in their genomes to identify the origin of deposited transcripts in the early embryos upon sequencing of the parent-offspring trio samples.

Crosses of *An. gambiae* males and *An. arabiensis* females were performed as described in Methods 7.5.1 to isolate mating couples and obtain different batches of offspring while both progenitors were identified. Then, RNA extraction from the embryos was performed before 1.5 hours post-oviposition, and therefore before zygotic genome activation, ensuring that all the transcripts present in the embryos had been deposited from their progenitors. RNA in the embryo samples was sequenced and whole genome sequencing (WGS) was performed in the progenitor individuals. SNPs present in the embryonic transcripts and in the genome of the progenitors were then identified. Genomic positions that have a variant in homozygosity in the *An. gambiae* male progenitor that is different from a variant also present in homozygosity in the *An. arabiensis* female progenitor can be used as “markers” to assign the origin of the transcripts found in early embryos (Figure 7.1):

- If a transcript has been deposited only from the *An. gambiae* male progenitor (paternal deposition), the expectation would be that in the embryos RNA we would only find the variants present in the *An. gambiae* progenitor.
- If a transcript has been deposited only from the *An. arabiensis* female progenitor (maternal deposition), in the RNA from hybrid embryos we would only find the SNPs present in the genome of the *An. arabiensis* progenitor.

- If a transcript has been deposited from both progenitors (hybrid deposition), in the embryos RNA we would find the variants from both *An. gambiae* and *An. arabiensis* progenitors.

Besides these signs of paternal and hybrid deposition in certain transcripts, their deposition levels (i.e., levels at which the transcripts are found in the early embryos) and their expression levels in testes can be taken into consideration to obtain a list of candidate genes that could be studied further prior to testing their regulatory regions to induce transcript deposition from a transgene in the male progenitor.

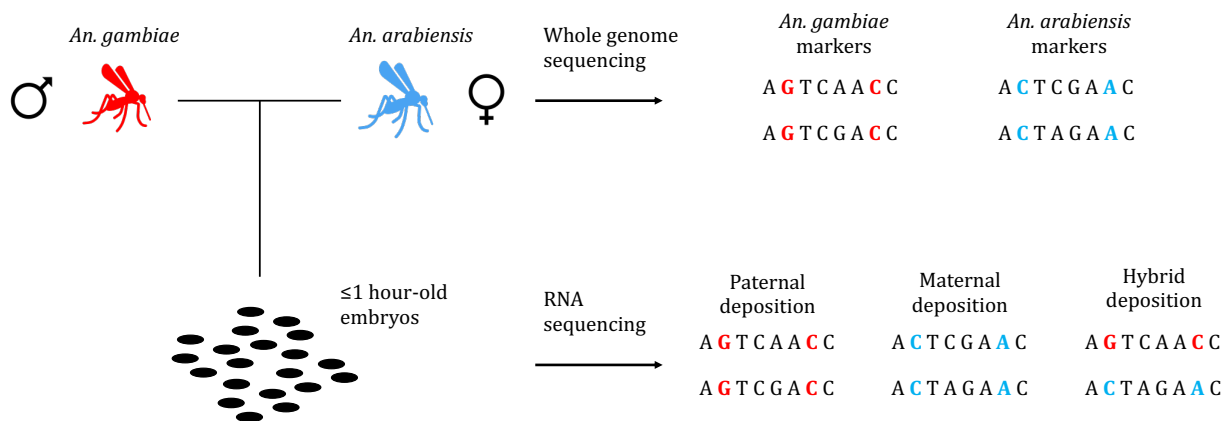


Figure 7.1 – Experimental design to find paternally deposited transcripts. *An. gambiae* males were mated with *An. arabiensis* females, and the produced embryos were collected prior to 1-hour postoviposition (before zygotic genome activation occurs). Whole genome sequencing of the progenitors allowed to identify marker SNPs (variants for which the *An. gambiae* male was homozygous and were different from the variants also present in homozygosity in the *An. arabiensis* female). Marker SNPs are highlighted in red for the *An. gambiae* variant and in blue for the *An. arabiensis* variant. RNA sequencing was performed in the embryo offspring and the origin of each found transcript was assigned based on the identified marker SNPs. The presence of only *An. gambiae* variants in the transcript suggests paternal deposition, the presence of just *An. arabiensis* variants indicates maternal deposition, and the co-presence of *An. gambiae* and *An. arabiensis* variants is a sign of hybrid deposition.

7.3 Identifying the origin of deposited transcripts in the early *Anopheles* embryos

Upon the collection and sequencing of the parent-offspring trio samples (Methods 7.5.1, 7.5.2), the bioinformatic analysis was performed by Nace Kranjc and the applied filters to find paternally deposited candidates were discussed in conjunction.

7.3.1 Ribosomal RNA depletion with a customised kit is extremely efficient in *Anopheles* samples

Previous research showed that spermatozoal RNA of different organisms contains both coding and non-coding RNAs (Gòdia et al., 2018). Thus, the aim of this experiment was to recover protein coding RNAs and long and short non-coding RNAs that might be present in the early *Anopheles* embryos. For this reason, the selected method to remove the highly abundant ribosomal RNA (rRNA) was rRNA depletion instead of the alternative poly-A selection that would not retain most of the non-coding RNAs.

In a first attempt, 5 samples (paternal-offspring trios) were collected. Removal of rRNA was performed with a standard rRNA depletion kit (optimized for human, mouse and rat samples – 7.5.2). However, analysis of the RNA sequencing data revealed that over 95% of the reads corresponded to rRNA (Figure 7.2, Samples E10 – E28), evidencing the lack of rRNA depletion. Given that most of the transcripts found in the early embryos are expected to have a maternal origin, the low number of reads allocated for RNA of interest severely hampered the identification of paternally deposited transcripts that are likely to be found in very low frequencies compared to the maternal counterparts. Consequently, we designed a probe pool for customised rRNA depletion (Methods 7.5.4).

A new embryonic RNA sample was collected to test the levels of rRNA depletion achieved with the customised probe pool, which resulted in the complete removal of rRNA (Figure 7.2, Sample HYB_00). Less than 0.5% of the sequencing reads corresponded to rRNA, showing the extreme efficiency of this method. Following this, 5 new samples (paternal-offspring trios) were obtained. Depletion of the rRNA with the customised probe pool

removed over 95% of rRNA in all the embryonic RNA samples (Figure 7.2, Samples HYB_02 – HYB_09) and hence provided a higher power of detection. We then proceeded with the WGS of the progenitors and the analysis of the data for these 5 samples.

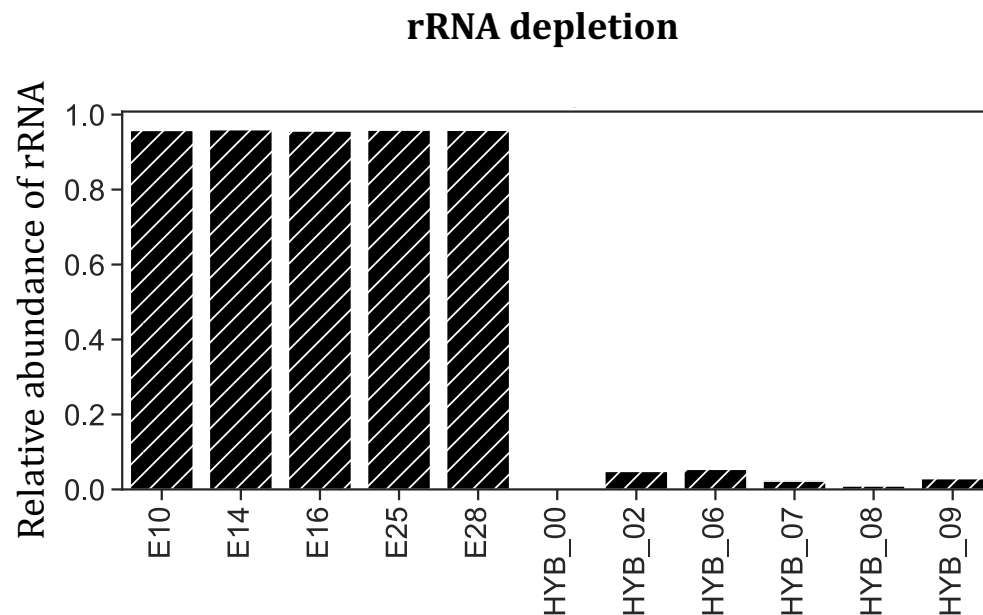


Figure 7.2 – Level of ribosomal RNA depletion in the embryonic RNA samples. Relative abundance of rRNA in each RNA sample obtained. Samples E10 – E28 were treated with a standard rRNA depletion kit (optimised for human, mice and rat samples) and over 95% of the reads corresponded to rRNA sequences. Sample HYB_00 was the test sample used to assess the removal of rRNA with the customised probe pool, resulting in the depletion of over 99.5% of rRNA. Samples HYB_02 – HYB_09 were treated with the same customised pool of probes and are the ones for which the analysis of deposited transcripts was performed.

7.3.2 RNA sequencing with 100 million reads per sample allows to detect extremely rare variants

Following on the expectation that paternally deposited transcripts might be found in relatively low frequency, the embryonic RNA samples were sequenced with a depth of ~100 million reads per sample to maximise the possibilities of detecting them. For comparison, ~20-30 million reads per sample would suffice for an RNA sequencing experiment studying gene expression in *An. gambiae*. In order to have a reference of the detection power provided by this approach, we used a “spiked” sample that contained

10% of RNA from *An. gambiae* embryos and 90% of RNA coming from *An. arabiensis* embryos.

This spiked sample was processed in parallel to the initial ones, using the standard kits for rRNA depletion. However, despite the rRNA not being removed from the sample, the analysis of the RNA sequencing data showed that it was possible to detect *An. gambiae*-specific variants at the expected 10% frequency (Figure 7.3). This was estimated using marker SNPs (i.e., genomic positions that have a variant in homozygosity in the *An. arabiensis* progenitor that is different from a variant also present in homozygosity in the *An. gambiae* progenitor).

The fact that in this sample less than 5% of the reads corresponded to mRNAs and non-coding RNAs and we were still able to detect the *An. gambiae* variants in the expected rate, shows that this approach can detect alleles in extremely low frequencies. Moreover, the power of detection will be considerably higher in the samples where nearly complete rRNA depletion was achieved with the customised probe pool.

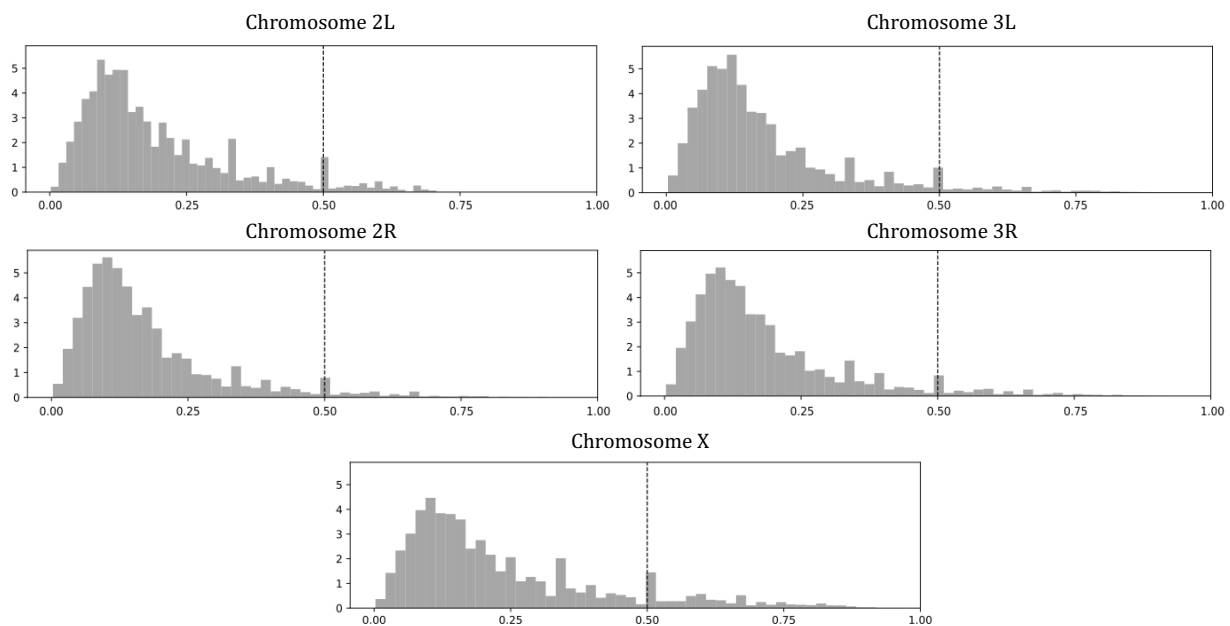


Figure 7.3 – Frequencies of *An. gambiae*-specific variants in the spiked control sample.

Histogram showing the allelic imbalance (i.e., number of *An. gambiae* reads divided by the total coverage in that position) of marker SNPs (per chromosome) in the spiked control sample with 10% *An. gambiae* RNA and 90% *An. arabiensis* RNA. In all the chromosomes, a peak around the 0.10 frequency can be observed, showing that most of the marker SNPs were found at the expected rates. Allelic imbalance of 0 (left) would indicate absence of *An. gambiae* markers, and an allelic imbalance of 1 (right) would imply that all the variants were *An. gambiae*-specific.

7.3.3 No signs of zygotic genome activation in 1-hour-old *Anopheles* embryos

To avoid the misidentification of zygotically expressed genes as paternally deposited transcripts, it is imperative that the zygotic genome activation has not started at the time of embryo collection. Since the *Anopheles* zygotic genome is transcriptionally inactive before 1.5 hours post-oviposition (Criscione et al., 2013; Wu et al., 2018; Matteo Vitale, manuscript in preparation), the embryos were collected and homogenised in TRIzol for posterior RNA extraction prior to 1 hour after oviposition (30 – 55 minutes).

In agreement with previous research, 1-hour-old *Anopheles* embryos showed no expression at all of genes that mark the onset of the zygotic genome activation in *Drosophila*, such as the orthologs of *zelda*, *hunchback* or *even skipped* (Liang et al., 2008; Pritchard and Schubiger, 1996; Tautz et al., 1987) (Figure 7.4). Additionally, genes that have been shown to be maternally deposited and degraded upon the maternal-to-zygotic transition (*vasa*, *orb*) were found in large quantities (Lantz et al., 1992; Papathanos et al., 2009) (Figure 7.4). This evidences that the recovered RNA from embryos had been deposited from the progenitors and not expressed from the zygotic genome.

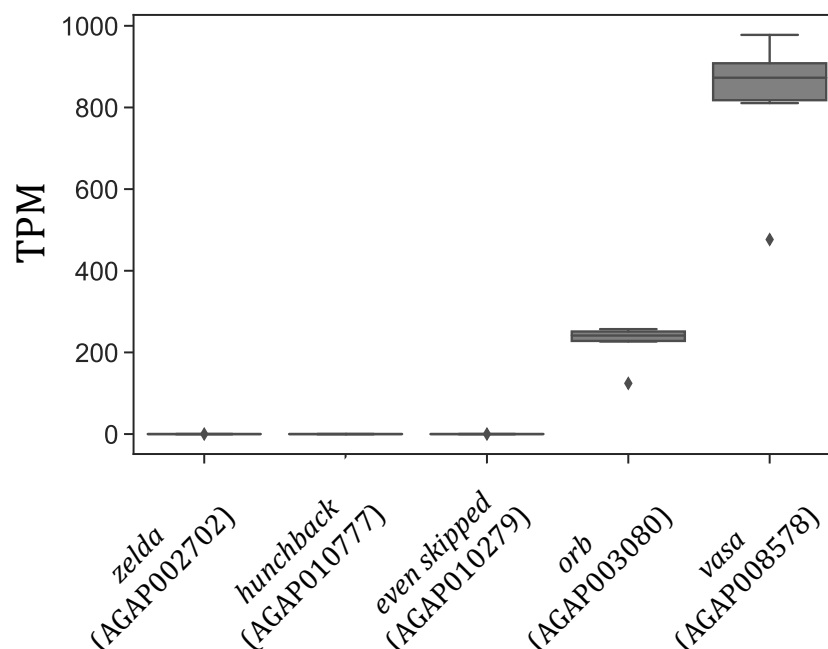


Figure 7.4 – Detected transcript levels of early zygotic genes and maternally deposited genes.

The detected transcript levels (transcripts per million – TPM) of genes expressed in the onset of zygotic genome activation (*zelda*, *hunchback*, *even skipped*) confirmed the lack of expression of these genes, while those of known maternally deposited genes (*orb*, *vasa*) were considerably high.

7.3.4 Assigning the parental origin to transcripts present in early *Anopheles* embryos

Given that the trio samples consisted in WGS of each progenitor and RNA sequencing of the embryo progeny, the first step was to assign the most plausible parental origin to the transcripts in each trio sample. For that, we used marker SNPs, which have been described already as genomic positions that have a variant in homozygosity in the *An. gambiae* progenitor that is different from a variant also present in homozygosity in the *An. arabiensis* progenitor.

In each trio sample, marker SNPs were assigned the most likely origin based on the variants found in the embryos. If the embryonic transcript was homozygous for the *An. arabiensis* allele, it was classified as maternally deposited (M); if it was homozygous for the *An. gambiae* allele, it was classified as paternally deposited (P); and if it was heterozygous, it was classified as deposited from both progenitors (hybrid deposition – H) (Table 7.1).

Table 7.1 – Classification of marker SNPs based on the most plausible progenitor origin. Sequencing data from the progenitors' genomes and the embryonic transcripts were aligned to the *An. gambiae* reference genome. If a variant was homozygous for the reference allele, it was called as 0/0; if it was heterozygous, 0/1; and if it was homozygous for an alternative allele, it was called 1/1. Selected marker SNPs were 0/0 in the male progenitor and 1/1 in the female progenitor, or vice versa. The variants found in the embryonic samples were assigned the most likely parental origin depending on the alleles present in the transcripts.

Sample	Genotype variants in marker SNPs					
Male	0/0	1/1	0/0	1/1	0/0	1/1
Female	1/1	0/0	1/1	0/0	1/1	0/0
Hybrid progeny	1/1	0/0	0/0	1/1	0/1	0/1
Assigned origin	Maternally deposited (M)		Paternally deposited (P)		Deposited from both progenitors (H)	

Subsequently, each transcript was classified based on the number of markers with the same assigned origin. If 90% or more of the markers present in a certain transcript had a maternal origin, the transcript was classified as maternally deposited (M). The equivalent occurred if that was the case for paternal (P) or hybrid (H1) origin (Figure 7.5). An additional less stringent filter was applied to detect transcripts with potential hybrid deposition: if over 40% of the marker SNPs had a hybrid origin and over 40% of the markers had a maternal origin, transcripts were classified as H2. The reasoning behind this filtering was that these transcripts still showed a strong signal of hybrid deposition, and therefore deposition from the male progenitor, but the signal could be diluted by a stronger maternal contribution of the same transcript.

Finally, if a certain transcript was classified in the same group (M, P, H1, H2) in 3 or more of the total 5 samples, the corresponding parental origin was assigned to the transcript.

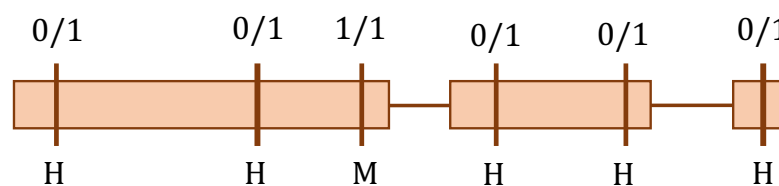


Figure 7.5 – Classification of transcripts based on their marker SNPs. After classifying each marker SNP according to the most plausible origin (maternal, M; paternal, P; hybrid, H), each transcript was classified based on its marker SNPs. If $\geq 90\%$ of the marker SNPs present in the transcript had the same assigned origin, the transcript was classified as such (maternally deposited, M; paternally deposited, P; hybrid deposition, H1). Another filter was applied: if $> 40\%$ of the markers in a transcript had a maternal origin and $> 40\%$ of the markers had a hybrid origin, the transcript was classified as H2, a second category of transcripts with signs of maternal deposition. Boxes represent the exons of an annotated gene and vertical lines indicate the location of marker SNPs and their classification.

7.3.5 Identified transcripts with signs of paternal deposition

Following the described procedure, we identified 29 transcripts with signs of deposition from the *An. gambiae* male progenitor (Table 7.2). Among these transcripts, just one was classified as being paternally deposited, while 3 were classified as H1 and 25 were classified as H2, indicating potential deposition from both progenitors. All these transcripts belong to protein coding genes except for one, which is a small nuclear RNA (snRNA).

Table 7.2 shows the 29 transcripts that were potentially deposited from the male progenitor, together with the deposition level (i.e., the volume at which they were found in the embryos) and the expression level of these genes at different stages of spermatogenesis (data obtained from Taxiarchi et al., 2019), in order to make a more educated decision on the best candidates for further research. In theory, transcripts that show higher deposition levels, higher expression in testes and a higher number of marker SNPs are more likely to be truly paternally deposited and hence represent better candidates to perform additional molecular checks to confirm their deposition and eventually test their regulatory sequences to induce paternal deposition of a transgene. Based on these criteria, 5 protein coding genes and the snRNA, which are highlighted in Table 7.2 and described below, were selected as the best candidates.

7.3.5.1 Candidate genes for additional studies to achieve paternal deposition of a transgene

The basis for the selection of these genes, as well as their function and the function of their orthologs in *D. melanogaster*, are presented here.

AGAP003586 is the only candidate that showed signs of being deposited only from the male progenitor. It was identified as such in 3 out of the 5 samples, with a single marker SNP. This gene has only one reported transcript (Amos et al., 2022) and it shows extremely high expression in testes, especially in the later stages of spermatogenesis. It codes for a phosphate carrier and, based on the available information on the *D. melanogaster* ortholog, the protein is transported to the mitochondria inner membrane, where it has a role in phosphate transmembrane transportation (Gramates et al., 2022).

Table 7.2 – Genes with transcripts that showed signs of paternal deposition. List of genes whose transcripts showed signs of deposition from the male progenitor. H1, H2, M and P columns indicate the number of samples classified in each category. When 3 or more samples were classified in the same category, they were assigned that origin. TPM indicates the deposition levels. Premeiotic, meiotic and postmeiotic expression of each gene is subsequently shown. S2 – S9 markers indicate the number of marker SNPs used in each sample (2-9) to determine the origin of the transcript. Chromosomal location and whether it is a protein coding or non-coding gene are shown in the last columns. The best candidates to use their regulatory regions to induce paternal deposition are highlighted in light yellow.

Gene	H1	H2	M	P	Origin	TPM	Premeiotic (FPKM)	Meiotic I (FPKM)	Meiotic II (FPKM)	Postmeiotic (FPKM)	S2 markers	S6 markers	S7 markers	S8 markers	S9 markers	Chr	Feature
AGAP003586	0	0	0	3	P	3.58	190.24	309.35	445.41	879.19	-	-	1	1	1	2R	Protein coding
AGAP013178	4	0	0	0	H1	56.33	20.60	37.79	76.03	98.14	-	7	9	4	9	2R	Protein coding
AGAP008072	5	0	0	0	H1	0.04	18.76	42.98	62.88	35.25	2	2	2	2	2	3R	Protein coding
AGAP013674	5	0	0	0	H1	621.89	-	-	-	-	2	2	2	2	2	3R	snRNA
AGAP029374	0	4	1	0	H2	12.25	-	-	-	-	13	8	16	3	18	2L	Protein coding
AGAP029373	0	4	1	0	H2	5.80	-	-	-	-	7	26	24	13	24	2L	Protein coding
AGAP002036	0	3	2	0	H2	18.07	47.03	55.21	96.99	135.57	13	7	5	2	6	2R	Protein coding
AGAP002038	0	4	1	0	H2	16.07	9.63	2.44	0.38	0.06	12	4	3	3	4	2R	Protein coding
AGAP002383	0	3	2	0	H2	10.65	6.48	6.55	4.62	2.31	9	4	4	4	5	2R	Protein coding
AGAP013538	0	3	2	0	H2	80.31	18.60	6.78	1.37	0.41	1	1	2	2	9	2R	Protein coding
AGAP013226	0	3	2	0	H2	18.67	12.58	12.25	19.26	30.03	5	9	7	9	5	2R	Protein coding
AGAP002711	0	3	2	0	H2	68.50	30.40	39.70	38.26	37.29	5	5	5	7	7	2R	Protein coding
AGAP003041	0	3	2	0	H2	40.74	17.39	9.75	5.96	3.02	4	4	3	5	7	2R	Protein coding
AGAP010517	0	3	2	0	H2	223.78	72.76	68.56	70.40	89.69	6	8	8	7	4	3L	Protein coding
AGAP011060	0	5	0	0	H2	21.05	104.61	86.10	29.52	8.03	4	7	6	7	6	3L	Protein coding
AGAP011072	0	3	2	0	H2	21.39	5.55	1.87	0.32	0.03	2	2	3	3	2	3L	Protein coding
AGAP028510	0	3	2	0	H2	7.88	48.18	73.28	113.02	285.98	4	13	7	7	7	3L	Protein coding
AGAP029734	0	4	0	0	H2	1.41	-	-	-	-	-	7	3	6	14	3L	Protein coding
AGAP029735	0	4	1	0	H2	27.58	-	-	-	-	8	8	13	8	15	3L	Protein coding
AGAP029742	0	5	0	0	H2	15.74	-	-	-	-	8	7	9	10	12	3L	Protein coding
AGAP028112	0	3	1	0	H2	8.49	14.37	12.18	2.89	1.49	-	5	2	2	2	3R	Protein coding
AGAP008439	0	4	1	0	H2	21.70	2.85	4.13	4.39	4.14	9	3	3	19	7	3R	Protein coding
AGAP029453	0	3	2	0	H2	69.25	-	-	-	-	4	10	6	23	13	3R	Protein coding
AGAP009341	0	3	2	0	H2	28.75	58.63	116.32	252.83	444.19	16	20	18	35	19	3R	Protein coding
AGAP009446	0	3	2	0	H2	18.58	145.18	274.56	425.59	847.16	21	13	9	26	17	3R	Protein coding
AGAP009572	0	4	1	0	H2	322.40	238.17	182.85	167.01	349.26	3	3	2	4	3	3R	Protein coding
AGAP000170	0	3	2	0	H2	43.22	18.72	8.91	5.03	11.91	6	7	7	7	5	X	Protein coding
AGAP000541	0	5	0	0	H2	720.16	14.40	2.33	0.25	0.33	2	6	5	5	9	X	Protein coding
AGAP000759	0	4	1	0	H2	45.41	31.14	6.69	0.78	1.13	12	12	11	12	13	X	Protein coding

AGAP013178 was classified as H1 in 4 of the 5 samples with 4-9 marker SNPs, indicating strong signs of hybrid deposition. Only one transcript has been reported for this gene, and the *D. melanogaster* ortholog encodes a transcription factor (checkpoint suppressor 1-like) involved in the division of cardiac progenitor cells and a growth factor transcription (decapentaplegic) in testes (Gramates et al., 2022).

AGAP009572 was classified as H2 in 4 samples with 2-4 marker SNPs, suggesting potential hybrid deposition where the paternal contribution is diluted by the maternal deposition, which is also supported by the especially high deposition levels. This gene encodes a 40S ribosomal protein S15 (Amos et al., 2022) and is the ortholog of the *D. melanogaster* RpS15, whose transcript is also found in sperm (Fischer et al., 2012).

AGAP009341 was classified as H2 in 3 samples, with a very high number of marker SNPs (16-35). It encodes the protein phosphatase 2, and the *D. melanogaster* ortholog seems to be involved in metabolism and growth networks (Gramates et al., 2022).

AGAP009446 was also classified as H2 in 3 samples using a high number of marker SNPs (9-26). This gene codes for a sodium/hydrogen exchanger, which seems to be involved in potassium ion transmembrane transport and regulation of intracellular pH, according to the information available on the *D. melanogaster* ortholog (Gramates et al., 2022).

Finally, the only non-coding RNA was **AGAP013674**. It was identified as H1 in all 5 samples, indicating a strong signal for hybrid deposition. While the data from Taxiarchi et al. (2019) did not evaluate the expression of non-coding RNAs across the spermatogenesis process, other datasets show the expression of AGAP013674 in testes (Baker et al., 2011). It is a U2 spliceosomal RNA that seems to be involved in splicing of pre-mRNA (Amos et al., 2022; Karijolic and Yu, 2010).

7.4 Discussion

Finding regulatory sequences that induce paternal deposition would provide a very useful new tool for the development of alternative YLE strategies. Looking at transcripts present in the *Anopheles* embryos prior to zygotic genome activation in the describe approach allowed us to detect up to 29 transcripts with signs of paternal deposition. Based on deposition level and expression in the *An. gambiae* testes, 5 protein coding genes and 1 snRNA were selected as the candidates with a stronger pattern for paternal deposition.

The standard commercial kits for rRNA depletion optimised for human, mice and rat samples had been shown to efficiently remove rRNA in *Anopheles funestus* samples to the point where only ~3 - 12% of the RNA sequencing reads corresponded to rRNA sequences (Weedall et al., 2015). The enormous difference with our results, where over 95% of the reads corresponded to rRNA, casts doubt about the procedure with the rRNA depletion in our case. Nonetheless, the use of a customised probe pool showed to work exceptionally well.

The methodology applied was validated by the different tests performed. Firstly, they confirmed that there were no signs of zygotic genome activation in the samples, and therefore the transcripts found in the embryos had been deposited from the progenitors. But more importantly, the use of a spiked control sample, for which the allele frequencies were known in advance, confirmed that with this approach it was possible to detect those variants at the rates at which they were present in the sample, which is an essential aspect for the project and reinforces the credibility of the obtained results.

Besides the stringent filters used to detect transcripts with strong signs of paternal or hybrid deposition, a more lax approach was also used to identify transcripts with diluted signs of hybrid deposition. Transcripts classified as H2 could be explained by a considerably larger maternal contribution that diluted the signs for paternal deposition. Another explanation might be found in the lack of consistency in the deposition of these transcripts; similar to the detection of paternally deposited transcripts in only ~30% of the analysed *D. melanogaster* embryos (Fischer et al., 2012). In agreement with the transcripts present in the sperm of *D. melanogaster* (Fischer et al., 2012), the genes with

stronger paternal deposition signal seem to be involved in translation and splicing, in addition to the transmembrane ion transport.

It may be possible that some transcripts with paternal contribution have been undetected. Given that the variants were called as homozygous if one allele was present in ≥ 0.95 frequency, a transcript which paternal contribution is less than 5% compared to the maternal one would have been classified as maternally deposited. If the overall deposition levels of this hypothetical transcript were remarkably high, it could represent a good candidate for the use of its regulatory regions to induce paternal deposition of a transgene. However, the regulatory sequences of the selected candidates represent a priori better options for this purpose. Analysing the allelic imbalance of the deposited transcripts (i.e., number of *An. gambiae* reads divided the total coverage in that position) would give a better estimate of the actual paternal contribution per transcript.

An additional molecular verification of the paternal deposition of the transcripts from the selected candidates could be carried out by performing allele-specific quantitative PCR (qPCR). Obtaining new embryonic RNA samples from the hybrid progeny of an *An. gambiae* male and *An. arabiensis* female and using marker SNPs that are present across the species, it would be possible to amplify the *An. gambiae* deposited transcript without amplifying the *An. arabiensis* one, hence proving the presence of the paternal contribution. The following steps would involve the cloning of the regulatory sequences of one or more paternally deposited transcripts to attempt paternal deposition of a transgene. If this can also be achieved from a Y chromosome-linked construct, then a YLE strategy could be developed by inducing the deposition of a female-specific toxin. It has been shown that injecting the mRNA of the male-determining gene *Yob* in *An. gambiae* and *An. arabiensis* female embryos triggers lethality, while male embryos are unaffected (Krzywinska et al., 2016). Thus, depositing the *Yob* mRNA from a Y chromosome-linked construct would represent an efficient YLE strategy.

Finally, the data recovered in this experiment can be useful not only to find regulatory sequences that may cause paternal deposition, but also to find alternatives to cause maternal deposition, or both, which could be convenient for certain genetic control strategies. Moreover, besides the relevance for the development of genetic control tools, the data obtained might provide valuable information regarding the maternal-to-zygotic transition and the early stages of embryogenesis in mosquitoes.

7.5 Methods

7.5.1 Sample collection

A cage of ~600 wild-type *An. gambiae* males was set. Separately, a small cage was prepared with 20 wild-type *An. arabiensis* females that had been sexed at pupal stage. When the adults in both cages were 5-8 days-old, *An. arabiensis* females were added in groups of 4 to the cage containing the *An. gambiae* males. Following the same procedure as described in Methods 4.8.4.1, mating couples were trapped and isolated. Each couple was placed in a separate cage. Females were blood fed the following day and males were stored at -20°C for genomic DNA extraction.

Three days after blood feeding, the embryo collection was performed. For that, a wet filter paper was placed on a Petri dish. An *An. arabiensis* female was blown into one of the Petri dishes and rapidly trapped with a Petri dish cover. This was done to keep the female in touch with the moistened paper and stimulate oviposition. The covered Petri dishes were then placed in the dark and the time was registered. Every 25-35 minutes, the Petri dishes were checked to verify if the females had laid eggs. If that was not the case, they were placed in darkness again and the timer was reset. If the females had laid, they were removed with a mechanical aspirator and stored at -20°C for genomic DNA extraction. Embryos were then carefully collected with a thin paint brush and placed in microcentrifuge tubes with 50 µL of TRIzol (Invitrogen, ThermoFisher), where they were homogenised manually using pestles. Additional 450 µL of TRIzol were added after homogenisation. Samples were left directly on dry ice until they could be brought for storage to a -80°C freezer.

To avoid that rests of TRIzol in the paint brush could cause lethality in the remaining embryos, it was cleaned in ethanol and in two different glasses with water before collecting more embryos. A minimum of 15 embryos from each progeny were left in the Petri dish to allow them to hatch and verify that they were fertilised and viable eggs.

7.5.2 RNA extraction and RNA sequencing

To perform RNA extraction, 120 µL of chloroform were added to the samples in TRIzol. After centrifugation (4°C, ≥13200 rpm, 15 minutes), the aqueous phase was collected. RNA was then purified using Direct-zol RNA Microprep Kit (ZYMO RESEARCH).

The RNA samples obtained in the first attempt were sent for library preparation and RNA sequencing (GENEWIZ). At GENEWIZ, samples were processed with the standard ribosomal RNA (rRNA) depletion kit from QIAGEN, optimized for human, mouse and rat samples (QIAseq FastSelect -rRNA HMR Kit). Sequencing was performed with Illumina NovaSeq (2x150bp configuration) with an estimated data output of ~100 million paired-end reads per sample.

The RNA samples obtained in the second attempt were sent for library preparation and RNA sequencing to Imperial BRC Genomics Facility, where samples were processed with a customised rRNA depletion kit (the design process is explained below). Sequencing was performed with NextSeq2000 (2x150 bp configuration) with an estimated data output of ~100 million paired-end reads per sample.

Genomic DNA from the progenitors was extracted using the DNeasy Blood & Tissue Kit (QIAGEN). DNA samples were sent for whole genome sequencing to Imperial BRC Genomics Facility. Sequencing was performed with NextSeq2000 (2x150 bp configuration) with an estimated data output of ~30 million paired-end reads per sample.

7.5.3 Creation of a spiked RNA sample

Embryos from the *An. gambiae* colony and from the *An. arabiensis* strain were separately collected at ~12 hours post-oviposition. RNA extraction of each batch was performed as described above. Upon quantification of the RNA present in each sample, I calculated the necessary volumes to create an aliquot that would contain 90% of *An. arabiensis* embryos RNA and 10% of *An. gambiae* embryos RNA.

7.5.4 Design of the customised probe pool for ribosomal RNA depletion

A ribosomal DNA (rDNA) reference sequence was generated using data from nanopore sequencing of *An. gambiae* individuals from the G3 colony (Samantha O'Loughlin, personal communication). Galizi et al. (2016) analysed the rDNA of *An. gambiae* and designed sequences to target the ribosomal locus. These sequences were used to capture long-read sequences from the nanopore data that contained the whole ribosomal locus intact, which were then used to annotate the rDNA sequences.

The generated rDNA reference sequence was submitted to the NEB Custom RNA Depletion Design Tool (accessible at <https://depletion-design.neb.com/>), that automatically created a pool of probes for customized depletion with the NEBNext RNA Depletion Core Reagent Set.

7.5.5 Analysis of the WGS and RNA sequencing data

In order to identify the SNPs in the progenitors' genomes and in the embryonic RNA, the sequencing data was aligned to the *An. gambiae* PEST reference genome (Amos et al., 2022), using BWA software (Li and Durbin, 2009) and STAR software (Dobin et al., 2013), respectively. Variant calling was then performed. If a variant was present in a frequency ≥ 0.95 were classified as homozygous; otherwise, it was classified as a heterozygous position.

For a SNP to be selected as a marker, it was required that this position was covered in the RNA data in the embryos. This means that a variant found in an intron in both progenitors but absent in the transcript in the embryos would not be selected as a marker.

Conclusions and future directions

Developed self-sustaining genetic control strategies such as gene drives have shown great potential to suppress caged mosquito populations and represent one of the most promising tools for vector control. Nevertheless, vital aspects such as the control prospects and the international agreements needed for their deploy are retarding the use of gene drives in the field. On the other hand, existing self-limiting strategies are too inefficient to be incorporated in the malaria control programmes. Thus, the generation of efficient technologies that are geographically and/or temporally restricted are gathering attention both as an alternative to more invasive self-sustaining techniques and as an intermediate step in phased approaches towards the eventual use of gene drives.

This thesis described the development of a Y chromosome-linked genome editor (YLE) that induces sterility in the female offspring by transmitting a newly discovered female-specific dominant negative mutation in the *doublesex* gene. Mathematical modelling of such YLE has shown that it is a technology significantly more efficient than other self-limiting genetic control strategies, while allowing to suppress a target population in a more controlled manner than with the use of gene drives. Besides the potential of the YLE for vector control, its development proved the feasibility of generating genetic strategies linked to the Y chromosome of *An. gambiae*. Immediate follow up experiments for this technology were already described and include assessing the capability of this YLE to suppress caged populations. Following steps would require mechanisms to facilitate the scaling up the production of YLE males prior to releases, which could involve the addition of a fluorescent marker linked to the dominant mutation in a way that does not affect the transcription and translation of the mutated allele.

The discovery of a female-specific sterility-causing dominant mutation during this project was of immense relevance to increase the efficiency of the YLE strategy to the point where it can be a useful tool for population suppression. Moreover, the nature of the dominance encourages to search for other target genes where these mutations could be generated. Finding alternative genes will be required for preventing or reducing resistance emergence or to develop additional strategies with different targets to avoid interference

upon an eventual release in the field. The experiments performed during this thesis did not allow to verify the haplo-insufficiency of two candidate targets, although they proved that another one was not haplo-insufficient. Recent research in *D. melanogaster* discussed earlier provided an additional X chromosome-linked candidate to induce female lethality, which together with the putative dominant mutations that might be created in genes whose products form oligomers, offer numerous options to study alternative target genes. The identification of paternally deposited transcripts in the *Anopheles* embryos is also promising for the development of new YLE approaches based on the deposition of female-specific toxins in form of mRNAs or snRNAs. The next experiments to corroborate the deposition of the found transcripts were already described, and the evidence of lethality in the female embryos induced by the male-determining factor (Yob) mRNA provides a perfect toxin candidate. Future steps will need to assess whether enough transcript is deposited to cause lethality in the female embryos and ways to increase the deposited amount of toxin if it is not enough.

Efficient self-limiting genetic control strategies such as the here developed YLE are more likely to meet current international regulations and face fewer obstacles towards approval than the self-sustaining counterparts. Therefore, they constitute alternatives for a near future while scientists, regional and local communities, and governments address social and environmental concerns posed by more invasive and more efficient technologies. While enormous progress has been achieved in the genetic control field in relatively short time, malaria keeps causing over 600,000 deaths annually, almost all of them occurring in sub-Saharan Africa. The huge burden imposed by malaria in this region perpetuates the vicious circle between poverty and this disease. Hopefully, the development and use of novel genetic control strategies will prove helpful in the eradication of malaria.

Bibliography

- Achan, J., Talisuna, A.O., Erhart, A., Yeka, A., Tibenderana, J.K., Baliraine, F.N., Rosenthal, P.J., D'Alessandro, U., 2011. Quinine, an old anti-malarial drug in a modern world: role in the treatment of malaria. *Malaria Journal* 10, 144. <https://doi.org/10.1186/1475-2875-10-144>
- Adelman, Z., Akbari, O., Bauer, J., Bier, E., Bloss, C., Carter, S.R., Callender, C., Denis, A.C.-S., Cowhey, P., Dass, B., Delborne, J., Devereaux, M., Ellsworth, P., Friedman, R.M., Gantz, V., Gibson, C., Hay, B.A., Hoddle, M., James, A.A., James, S., Jorgenson, L., Kalichman, M., Marshall, J., McGinnis, W., Newman, J., Pearson, A., Quemada, H., Rudenko, L., Shelton, A., Vinetz, J.M., Weisman, J., Wong, B., Wozniak, C., 2017. Rules of the road for insect gene drive research and testing. *Nat Biotechnol* 35, 716–718. <https://doi.org/10.1038/nbt.3926>
- Adepoju, P., 2019. RTS,S malaria vaccine pilots in three African countries. *The Lancet* 393, 1685. [https://doi.org/10.1016/S0140-6736\(19\)30937-7](https://doi.org/10.1016/S0140-6736(19)30937-7)
- Akbari, O.S., Matzen, K.D., Marshall, J.M., Huang, H., Ward, C.M., Hay, B.A., 2013. A Synthetic Gene Drive System for Local, Reversible Modification and Suppression of Insect Populations. *Current Biology* 23, 671–677. <https://doi.org/10.1016/j.cub.2013.02.059>
- Alcalay, Y., Fuchs, S., Galizi, R., Bernardini, F., Haghighat-Khah, R.E., Rusch, D.B., Adrion, J.R., Hahn, M.W., Tortosa, P., Rotenberry, R., Papathanos, P.A., 2021. The Potential for a Released Autosomal X-Shredder Becoming a Driving-Y Chromosome and Invasively Suppressing Wild Populations of Malaria Mosquitoes. *Frontiers in Bioengineering and Biotechnology* 9.
- Alfonso-Parra, C., Ahmed-Braimah, Y.H., Degner, E.C., Avila, F.W., Villarreal, S.M., Pleiss, J.A., Wolfner, M.F., Harrington, L.C., 2016. Mating-Induced Transcriptome Changes in the Reproductive Tract of Female *Aedes aegypti*. *PLOS Neglected Tropical Diseases* 10, e0004451. <https://doi.org/10.1371/journal.pntd.0004451>
- Alphey, L., 2014. Genetic Control of Mosquitoes. *Annu. Rev. Entomol.* 59, 205–224. <https://doi.org/10.1146/annurev-ento-011613-162002>
- Alphey, L.S., Crisanti, A., Randazzo, F. (Fil), Akbari, O.S., 2020. Standardizing the definition of gene drive. *Proceedings of the National Academy of Sciences* 117, 30864–30867. <https://doi.org/10.1073/pnas.2020417117>
- Amos, B., Aurrecochea, C., Barba, M., Barreto, A., Basenko, E.Y., Bazant, W., Belnap, R., Blevins, A.S., Böhme, U., Brestelli, J., Brunk, B.P., Caddick, M., Callan, D., Campbell, L., Christensen, M.B., Christophides, G.K., Crouch, K., Davis, K., DeBarry, J., Doherty, R., Duan, Y., Dunn, M., Falke, D., Fisher, S., Flicek, P., Fox, B., Gajria, B., Giraldo-Calderón, G.I., Harb, O.S., Harper, E., Hertz-Fowler, C., Hickman, M.J., Howington, C., Hu, S., Humphrey, J., Iodice, J., Jones, A., Judkins, J., Kelly, S.A., Kissinger, J.C., Kwon, D.K., Lamoureux, K., Lawson, D., Li, W., Lies, K., Lodha, D., Long, J., MacCallum, R.M., Maslen, G., McDowell, M.A., Nabrzyski, J., Roos, D.S., Rund, S.S.C., Schulman, S.W., Shanmugasundram, A., Sitnik, V., Spruill, D., Starns, D., Stoeckert, C.J., Jr, Tomko, S.S.,

- Wang, H., Warrenfeltz, S., Wieck, R., Wilkinson, P.A., Xu, L., Zheng, J., 2022. VEuPathDB: the eukaryotic pathogen, vector and host bioinformatics resource center. *Nucleic Acids Research* 50, D898–D911. <https://doi.org/10.1093/nar/gkab929>
- 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, 3106–3111.
- Annas, G.J., Beisel, C.L., Clement, K., Crisanti, A., Francis, S., Galardini, M., Galizi, R., Grünewald, J., Immobile, G., Khalil, A.S., Müller, R., Pattanayak, V., Petri, K., Paul, L., Pinello, L., Simoni, A., Taxiarchi, C., Joung, J.K., 2021. A Code of Ethics for Gene Drive Research. *The CRISPR Journal* 4, 19–24. <https://doi.org/10.1089/crispr.2020.0096>
- 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., Sovannaroeth, 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., Phyto, 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., Tracking Resistance to Artemisinin Collaboration (TRAC), 2014. Spread of artemisinin resistance in *Plasmodium falciparum* malaria. *N Engl J Med* 371, 411–423. <https://doi.org/10.1056/NEJMoa1314981>
- Ashley, E.A., Phyto, A.P., 2018. Drugs in Development for Malaria. *Drugs* 78, 861–879. <https://doi.org/10.1007/s40265-018-0911-9>
- Ayala, D., Akone-Ella, O., Rahola, N., Kengne, P., Ngangue, M.F., Mezeme, F., Makanga, B.K., Nigg, M., Costantini, C., Simard, F., Prugnolle, F., Roche, B., Duron, O., Paupy, C., 2019. Natural *Wolbachia* infections are common in the major malaria vectors in Central Africa. *Evolutionary Applications* 12, 1583–1594. <https://doi.org/10.1111/eva.12804>
- Bachtrog, D., 2013. Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. *Nat Rev Genet* 14, 113–124. <https://doi.org/10.1038/nrg3366>
- 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, 296. <https://doi.org/10.1186/1471-2164-12-296>
- Baker, D.A., Russell, S., 2011. Role of Testis-Specific Gene Expression in Sex-Chromosome Evolution of *Anopheles gambiae*. *Genetics* 189, 1117–1120. <https://doi.org/10.1534/genetics.111.133157>

- Baker, R.H., Sakai, R.K., 1979. Triploids and male determination in the mosquito, *Anopheles culicifacies*. *Journal of Heredity* 70, 345–346. <https://doi.org/10.1093/oxfordjournals.jhered.a109271>
- Baldini, F., Rougé, J., Kreppel, K., Mkandawile, G., Mapua, S.A., Sikulu-Lord, M., Ferguson, H.M., Govella, N., Okumu, F.O., 2018. First report of natural *Wolbachia* infection in the malaria mosquito *Anopheles arabiensis* in Tanzania. *Parasites Vectors* 11, 635. <https://doi.org/10.1186/s13071-018-3249-y>
- Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., Horvath, P., 2007. CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes. *Science* 315, 1709–1712. <https://doi.org/10.1126/science.1138140>
- Barroso, G., Valdespin, C., Vega, E., Kershenovich, R., Avila, R., Avendaño, C., Oehninger, S., 2009. Developmental sperm contributions: fertilization and beyond. *Fertility and Sterility* 92, 835–848. <https://doi.org/10.1016/j.fertnstert.2009.06.030>
- BBC News, 2023. Ghana first to approve “world-changer” malaria vaccine. BBC News.
- Becker, N., Petrić, D., Zgomba, M., Boase, C., Madon, M.B., Dahl, C., Kaiser, A., 2020. *Mosquitoes: Identification, Ecology and Control*. Springer Nature.
- 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, 1037–1046. [https://doi.org/10.1016/0092-8674\(88\)90248-6](https://doi.org/10.1016/0092-8674(88)90248-6)
- Bernardini, F., Galizi, R., Menichelli, M., Papathanos, P.-A., Dritsou, V., Marois, E., Crisanti, A., Windbichler, N., 2014. Site-specific genetic engineering of the *Anopheles gambiae* Y chromosome. *Proceedings of the National Academy of Sciences* 111, 7600–7605. <https://doi.org/10.1073/pnas.1404996111>
- Bernardini, F., Galizi, R., Wunderlich, M., Taxiarchi, C., Kranjc, N., Kyrou, K., Hammond, A., Nolan, T., Lawniczak, M.N.K., Papathanos, P.A., Crisanti, A., Windbichler, N., 2017. Cross-Species Y Chromosome Function Between Malaria Vectors of the *Anopheles gambiae* Species Complex. *Genetics* 207, 729–740. <https://doi.org/10.1534/genetics.117.300221>
- Bernardini, F., Kriezis, A., Galizi, R., Nolan, T., Crisanti, A., 2019. Introgression of a synthetic sex ratio distortion system from *Anopheles gambiae* into *Anopheles arabiensis*. *Sci Rep* 9, 5158. <https://doi.org/10.1038/s41598-019-41646-8>
- 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.J., 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, 207–211. <https://doi.org/10.1038/nature15535>
- Bhaya, D., Davison, M., Barrangou, R., 2011. CRISPR-Cas Systems in Bacteria and Archaea: Versatile Small RNAs for Adaptive Defense and Regulation. *Annual Review of Genetics* 45, 273–297. <https://doi.org/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, 748–751.
<https://doi.org/10.1126/science.1236192>
- Bier, E., 2022. Gene drives gaining speed. Nat Rev Genet 23, 5–22.
<https://doi.org/10.1038/s41576-021-00386-0>
- Bilgo, E., Lovett, B., Fang, W., Bende, N., King, G.F., Diabate, A., St. Leger, R.J., 2017. Improved efficacy of an arthropod toxin expressing fungus against insecticide-resistant malaria-vector mosquitoes. Sci Rep 7, 3433. <https://doi.org/10.1038/s41598-017-03399-0>
- Blasco, B., Leroy, D., Fidock, D.A., 2017. Antimalarial drug resistance: linking Plasmodium falciparum parasite biology to the clinic. Nat Med 23, 917–928.
<https://doi.org/10.1038/nm.4381>
- Bonaccorsi, S., Santini, G., Gatti, M., Pimpinelli, S., Coluzzi, M., 1980. Intraspecific polymorphism of sex chromosome heterochromatin in two species of the Anopheles gambiae complex. Chromosoma 76, 57–64.
<https://doi.org/10.1007/BF00292226>
- 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, 403–415.
<https://doi.org/10.1101/gad.5.3.403>
- Brodersen, D.E., Nissen, P., 2005. The social life of ribosomal proteins. FEBS J 272, 2098–2108. <https://doi.org/10.1111/j.1742-4658.2005.04651.x>
- Buchman, A., Akbari, O.S., 2019. Site-specific transgenesis of the Drosophila melanogaster Y-chromosome using CRISPR/Cas9. Insect Molecular Biology 28, 65–73.
<https://doi.org/10.1111/imb.12528>
- Burt, A., 2014. Heritable strategies for controlling insect vectors of disease. Phil. Trans. R. Soc. B 369, 20130432. <https://doi.org/10.1098/rstb.2013.0432>
- Burt, A., 2003. Site-specific selfish genes as tools for the control and genetic engineering of natural populations. Proc. R. Soc. Lond. B 270, 921–928.
<https://doi.org/10.1098/rspb.2002.2319>
- Burt, A., Deredec, A., 2018. Self-limiting population genetic control with sex-linked genome editors 9.
- Burtis, K.C., Baker, B.S., 1989. Drosophila doublesex Gene Controls Somatic Sexual Differentiation by Producing Alternatively Spliced mRNAs Encoding Related Sex-Specific Polypeptides 14.
- Bushey, A.M., Dorman, E.R., Corces, V.G., 2008. Chromatin Insulators: Regulatory Mechanisms and Epigenetic Inheritance. Molecular Cell 32, 1–9.
<https://doi.org/10.1016/j.molcel.2008.08.017>
- Carballar-Lejarazu, R., Jasinskiene, N., James, A.A., 2013. Exogenous gypsy insulator sequences modulate transgene expression in the malaria vector mosquito, Anopheles stephensi. Proceedings of the National Academy of Sciences 110, 7176–7181. <https://doi.org/10.1073/pnas.1304722110>
- Carballar-Lejarazú, R., Ogaugwu, C., Tushar, T., Kelsey, A., Pham, T.B., Murphy, J., Schmidt, H., Lee, Y., Lanzaro, G.C., James, A.A., 2020. Next-generation gene drive for population modification of the malaria vector mosquito, Anopheles gambiae.

- Proceedings of the National Academy of Sciences 117, 22805–22814. <https://doi.org/10.1073/pnas.2010214117>
- Carter, S.R., Friedman, R.M., 2016. Policy and Regulatory Issues for Gene Drives in Insects.
- Carvalho, A.B., Vibranovski, M.D., Carlson, J.W., Celniker, S.E., Hoskins, R.A., Rubin, G.M., Sutton, G.G., Adams, M.D., Myers, E.W., Clark, A.G., 2003. Y Chromosome and Other Heterochromatic Sequences of the *Drosophila Melanogaster* Genome: How Far can we go? *Genetica* 117, 227–237. <https://doi.org/10.1023/A:1022900313650>
- 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, e0003864. <https://doi.org/10.1371/journal.pntd.0003864>
- Casas-Tintó, S., Ferrús, A., 2021. The haplolethality paradox of the wupA gene in *Drosophila*. *PLOS Genetics* 17, e1009108. <https://doi.org/10.1371/journal.pgen.1009108>
- Casas-Tintó, S., Maraver, A., Serrano, M., Ferrús, A., 2016. Troponin-I enhances and is required for oncogenic overgrowth. *Oncotarget* 7. <https://doi.org/10.18632/oncotarget.10616>
- Catteruccia, F., Nolan, T., Loukeris, T.G., Blass, C., Savakis, C., Kafatos, F.C., Crisanti, A., 2000. Stable germline transformation of the malaria mosquito *Anopheles stephensi*. *Nature* 405, 959–962. <https://doi.org/10.1038/35016096>
- Champer, J., Buchman, A., Akbari, O.S., 2016. Cheating evolution: engineering gene drives to manipulate the fate of wild populations. *Nat Rev Genet* 17, 146–159. <https://doi.org/10.1038/nrg.2015.34>
- Champer, J., Yang, E., Lee, E., Liu, J., Clark, A.G., Messer, P.W., 2020. A CRISPR homing gene drive targeting a haplolethal gene removes resistance alleles and successfully spreads through a cage population. *Proceedings of the National Academy of Sciences* 117, 24377–24383. <https://doi.org/10.1073/pnas.2004373117>
- Charlesworth, B., 1991. The evolution of sex chromosomes. *Science* 251, 1030–1033. <https://doi.org/10.1126/science.1998119>
- Chen, C.-H., Huang, H., Ward, C.M., Su, J.T., Schaeffer, L.V., Guo, M., Hay, B.A., 2007. A Synthetic Maternal-Effect Selfish Genetic Element Drives Population Replacement in *Drosophila*. *Science* 316, 597–600. <https://doi.org/10.1126/science.1138595>
- Chen, S., Corces, V.G., 2001. The gypsy insulator of *Drosophila* affects chromatin structure in a directional manner. *Genetics* 159, 1649–1658.
- Chevalier, B.S., Stoddard, B.L., 2001. Homing endonucleases: structural and functional insight into the catalysts of intron/intein mobility. *Nucleic Acids Research* 29, 3757–3774. <https://doi.org/10.1093/nar/29.18.3757>
- Cho, S., Wensink, P.C., 1998. Linkage between Oligomerization and DNA Binding in *Drosophila* Doublesex Proteins. *Biochemistry* 37, 11301–11308. <https://doi.org/10.1021/bi972916x>
- Cho, S., Wensink, P.C., 1997. DNA Binding by the Male and Female Doublesex Proteins of *Drosophila melanogaster*. *Journal of Biological Chemistry* 272, 3185–3189. <https://doi.org/10.1074/jbc.272.6.3185>

- 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, 224–226. <https://doi.org/10.1038/s41587-019-0032-3>
- Clements, A.N., 1992. *Biology of Mosquitoes : Development Nutrition and Reproduction*. Springer Netherlands.
- Concordet, J.-P., Haeussler, M., 2018. CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. *Nucleic Acids Research* 46, W242–W245. <https://doi.org/10.1093/nar/gky354>
- Connolly, J.B., Mumford, J.D., Fuchs, S., Turner, G., Beech, C., North, A.R., Burt, A., 2021. Systematic identification of plausible pathways to potential harm via problem formulation for investigational releases of a population suppression gene drive to control the human malaria vector *Anopheles gambiae* in West Africa. *Malar J* 20, 170. <https://doi.org/10.1186/s12936-021-03674-6>
- Connolly, J.B., Mumford, J.D., Glandorf, D.C.M., Hartley, S., Lewis, O.T., Evans, S.W., Turner, G., Beech, C., Sykes, N., Coulibaly, M.B., Romeis, J., Teem, J.L., Tonui, W., Lovett, B., Mankad, A., Mnzava, A., Fuchs, S., Hackett, T.D., Landis, W.G., Marshall, J.M., Aboagye-Antwi, F., 2022. Recommendations for environmental risk assessment of gene drive applications for malaria vector control. *Malaria Journal* 21, 152. <https://doi.org/10.1186/s12936-022-04183-w>
- Cook, R.K., Christensen, S.J., Deal, J.A., Coburn, R.A., Deal, M.E., Gresens, J.M., Kaufman, T.C., Cook, K.R., 2012. The generation of chromosomal deletions to provide extensive coverage and subdivision of the *Drosophila melanogaster* genome. *Genome Biol* 13, R21. <https://doi.org/10.1186/gb-2012-13-3-r21>
- Cortez, D., Marin, R., Toledo-Flores, D., Froidevaux, L., Liechti, A., Waters, P.D., Grützner, F., Kaessmann, H., 2014. Origins and functional evolution of Y chromosomes across mammals. *Nature* 508, 488–493. <https://doi.org/10.1038/nature13151>
- Coschigano, K.T., Wensink, P.C., 1993. Sex-specific transcriptional regulation by the male and female doublesex proteins of *Drosophila*. *Genes Dev* 7, 42–54. <https://doi.org/10.1101/gad.7.1.42>
- Crawford, J.E., Clarke, D.W., Criswell, V., Desnoyer, M., Cornel, D., Deegan, B., Gong, K., Hopkins, K.C., Howell, P., Hyde, J.S., Livni, J., Behling, C., Benza, R., Chen, W., Dobson, K.L., Eldershaw, C., Greeley, D., Han, Y., Hughes, B., Kakani, E., Karbowski, J., Kitchell, A., Lee, E., Lin, T., Liu, J., Lozano, M., MacDonald, W., Mains, J.W., Metlitz, M., Mitchell, S.N., Moore, D., Ohm, J.R., Parkes, K., Porshnikoff, A., Robuck, C., Sheridan, M., Sobecki, R., Smith, P., Stevenson, J., Sullivan, J., Wasson, B., Weakley, A.M., Wilhelm, M., Won, J., Yasunaga, A., Chan, W.C., Holeman, J., Snoad, N., Upson, L., Zha, T., Dobson, S.L., Mulligan, F.S., Massaro, P., White, B.J., 2020. Efficient production of male *Wolbachia*-infected *Aedes aegypti* mosquitoes enables large-scale suppression of wild populations. *Nat Biotechnol* 38, 482–492. <https://doi.org/10.1038/s41587-020-0471-x>
- 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, 433–441. <https://doi.org/10.1111/imb.12034>

- Criscione, F., Qi, Y., Tu, Z., 2016. GUY1 confers complete female lethality and is a strong candidate for a male-determining factor in *Anopheles stephensi*. *eLife* 5, e19281. <https://doi.org/10.7554/eLife.19281>
- Curtis, C.F., 1968. Possible Use of Translocations to fix Desirable Genes in Insect Pest Populations. *Nature* 218, 368–369. <https://doi.org/10.1038/218368a0>
- Dattoo, M.S., Natama, M.H., Somé, A., Traoré, O., Rouamba, T., Bellamy, D., Yameogo, P., Valia, D., Tegneri, M., Ouedraogo, F., Soma, R., Sawadogo, S., Sorgho, F., Derra, K., Rouamba, E., Orindi, B., Ramos Lopez, F., Flaxman, A., Cappuccini, F., Kailath, R., Elias, S., Mukhopadhyay, E., Noe, A., Cairns, M., Lawrie, A., Roberts, R., Valéa, I., Sorgho, H., Williams, N., Glenn, G., Fries, L., Reimer, J., Ewer, K.J., Shaligram, U., Hill, A.V.S., Tinto, H., 2021. Efficacy of a low-dose candidate malaria vaccine, R21 in adjuvant Matrix-M, with seasonal administration to children in Burkina Faso: a randomised controlled trial. *The Lancet* 397, 1809–1818. [https://doi.org/10.1016/S0140-6736\(21\)00943-0](https://doi.org/10.1016/S0140-6736(21)00943-0)
- Degner, E.C., Ahmed-Braimah, Y.H., Borziak, K., Wolfner, M.F., Harrington, L.C., Dorus, S., 2019. Proteins, Transcripts, and Genetic Architecture of Seminal Fluid and Sperm in the Mosquito *Aedes aegypti**. *Molecular & Cellular Proteomics* 18, S6–S22. <https://doi.org/10.1074/mcp.RA118.001067>
- Deredec, A., Burt, A., Godfray, H.C.J., 2008. The Population Genetics of Using Homing Endonuclease Genes in Vector and Pest Management. *Genetics* 179, 2013–2026. <https://doi.org/10.1534/genetics.108.089037>
- Deredec, A., Godfray, H.C.J., Burt, A., 2011. Requirements for effective malaria control with homing endonuclease genes. *Proceedings of the National Academy of Sciences* 108, E874–E880. <https://doi.org/10.1073/pnas.1110717108>
- Deutschbauer, A.M., Jaramillo, D.F., Proctor, M., Kumm, J., Hillenmeyer, M.E., Davis, R.W., Nislow, C., Giaever, G., 2005. Mechanisms of Haploinsufficiency Revealed by Genome-Wide Profiling in Yeast. *Genetics* 169, 1915–1925. <https://doi.org/10.1534/genetics.104.036871>
- Dietzl, G., Chen, D., Schnorrer, F., Su, K.-C., Barinova, Y., Fellner, M., Gasser, B., Kinsey, K., Oettel, S., Scheiblaue, S., Couto, A., Marra, V., Keleman, K., Dickson, B.J., 2007. A genome-wide transgenic RNAi library for conditional gene inactivation in *Drosophila*. *Nature* 448, 151–156. <https://doi.org/10.1038/nature05954>
- Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., Gingeras, T.R., 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21. <https://doi.org/10.1093/bioinformatics/bts635>
- Draptchinskaia, N., Gustavsson, P., Andersson, B., Pettersson, M., Willig, T.N., Dianzani, I., Ball, S., Tchernia, G., Klar, J., Matsson, H., Tentler, D., Mohandas, N., Carlsson, B., Dahl, N., 1999. The gene encoding ribosomal protein S19 is mutated in Diamond-Blackfan anaemia. *Nat Genet* 21, 169–175. <https://doi.org/10.1038/5951>
- Edgington, M.P., Alphey, L.S., 2018. Population dynamics of engineered underdominance and killer-rescue gene drives in the control of disease vectors. *PLOS Computational Biology* 14, e1006059. <https://doi.org/10.1371/journal.pcbi.1006059>

- Edgington, M.P., Harvey-Samuel, T., Alphey, L., 2020. Split drive killer-rescue provides a novel threshold-dependent gene drive. *Sci Rep* 10, 20520. <https://doi.org/10.1038/s41598-020-77544-7>
- Elgin, S.C.R., Reuter, G., 2013. Position-Effect Variegation, Heterochromatin Formation, and Gene Silencing in *Drosophila*. *Cold Spring Harbor Perspectives in Biology* 5, a017780–a017780. <https://doi.org/10.1101/cshperspect.a017780>
- Erdman, S. e., Burtis, K. c., 1993. The *Drosophila* doublesex proteins share a novel zinc finger related DNA binding domain. *The EMBO Journal* 12, 527–535. <https://doi.org/10.1002/j.1460-2075.1993.tb05684.x>
- 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, 1639–1652. <https://doi.org/10.1093/genetics/144.4.1639>
- Erickson, J.W., Quintero, J.J., 2007. Indirect Effects of Ploidy Suggest X Chromosome Dose, Not the X:A Ratio, Signals Sex in *Drosophila*. *PLOS Biology* 5, e332. <https://doi.org/10.1371/journal.pbio.0050332>
- Fischer, B.E., Wasbrough, E., Meadows, L.A., Randlet, O., Dorus, S., Karr, T.L., Russell, S., 2012. Conserved properties of *Drosophila* and human spermatozoal mRNA repertoires. *Proceedings of the Royal Society B: Biological Sciences* 279, 2636–2644. <https://doi.org/10.1098/rspb.2012.0153>
- Fuchs, S., Nolan, T., Crisanti, A., 2013. Mosquito Transgenic Technologies to Reduce *Plasmodium* Transmission, in: Ménard, R. (Ed.), *Malaria: Methods and Protocols, Methods in Molecular Biology*. Humana Press, Totowa, NJ, pp. 601–622. https://doi.org/10.1007/978-1-62703-026-7_41
- 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. <https://doi.org/10.1038/ncomms4977>
- 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, 31139. <https://doi.org/10.1038/srep31139>
- Gamez, S., Chaverra-Rodriguez, D., Buchman, A., Kandul, N.P., Mendez-Sanchez, S.C., Bennett, J.B., Sánchez C., H.M., Yang, T., Antoshechkin, I., Duque, J.E., Papathanos, P.A., Marshall, J.M., Akbari, O.S., 2021. Exploiting a Y chromosome-linked Cas9 for sex selection and gene drive. *Nat Commun* 12, 7202. <https://doi.org/10.1038/s41467-021-27333-1>
- Gantz, V.M., Jasinskiene, N., Tatarenkova, O., Fazekas, A., Macias, V.M., Bier, E., James, A.A., 2015. Highly efficient Cas9-mediated gene drive for population modification of the malaria vector mosquito *Anopheles stephensi*. *Proceedings of the National Academy of Sciences* 112, E6736–E6743. <https://doi.org/10.1073/pnas.1521077112>
- Garrett-Engle, C.M., Siegal, M.L., Manoli, D.S., Williams, B.C., Li, H., Baker, B.S., 2002. *intersex*, a gene required for female sexual development in *Drosophila*, is

- expressed in both sexes and functions together with doublesex to regulate terminal differentiation. *Development* 129, 4661–4675. <https://doi.org/10.1242/dev.129.20.4661>
- Garziera, L., Pedrosa, M.C., de Souza, F.A., Gómez, M., Moreira, M.B., Virginio, J.F., Capurro, M.L., Carvalho, D.O., 2017. Effect of interruption of over-flooding releases of transgenic mosquitoes over wild population of *Aedes aegypti* : two case studies in Brazil. *Entomol Exp Appl* 164, 327–339. <https://doi.org/10.1111/eea.12618>
- Geci, R., Willis, K., Burt, A., 2022. Gene drive designs for efficient and localisable population suppression using Y-linked editors. *PLOS Genetics* 18, e1010550. <https://doi.org/10.1371/journal.pgen.1010550>
- Goddard, M.R., Greig, D., Burt, A., 2001. Outcrossed sex allows a selfish gene to invade yeast populations. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 268, 2537–2542. <https://doi.org/10.1098/rspb.2001.1830>
- Gòdia, M., Swanson, G., Krawetz, S.A., 2018. A history of why fathers' RNA matters†. *Biology of Reproduction* 99, 147–159. <https://doi.org/10.1093/biolre/iroy007>
- Gramates, L.S., Agapite, J., Attrill, H., Calvi, B.R., Crosby, M.A., dos Santos, G., Goodman, J.L., Goutte-Gattat, D., Jenkins, V.K., Kaufman, T., Larkin, A., Matthews, B.B., Millburn, G., Strelets, V.B., the FlyBase Consortium, 2022. FlyBase: a guided tour of highlighted features. *Genetics* 220, iyac035. <https://doi.org/10.1093/genetics/iyac035>
- Grilli, S., Galizi, R., Taxiarchi, C., 2021. Genetic Technologies for Sustainable Management of Insect Pests and Disease Vectors. *Sustainability* 13, 5653. <https://doi.org/10.3390/su13105653>
- Grover, N., Rigby, J., Rigby, J., 2023. Ghana first to approve Oxford's malaria vaccine. Reuters.
- Haldane, J.B.S., 1922. Sex ratio and unisexual sterility in hybrid animals. *Journ. of Gen.* 12, 101–109. <https://doi.org/10.1007/BF02983075>
- Halder, K., Bhattacharjee, S., Safeukui, I., 2018. Drug resistance in Plasmodium. *Nat Rev Microbiol* 16, 156–170. <https://doi.org/10.1038/nrmicro.2017.161>
- 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, E2114–2123. <https://doi.org/10.1073/pnas.1525164113>
- 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, 273. <https://doi.org/10.1186/1471-2164-14-273>
- Hamilton, W.D., 1967. Extraordinary Sex Ratios. *Science* 156, 477–488. <https://doi.org/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, 78–83. <https://doi.org/10.1038/nbt.3439>
- Hammond, A., Karlsson, X., Morianou, I., Kyrou, K., Beaghton, A., Gribble, M., Kranjc, N., Galizi, R., Burt, A., Crisanti, A., Nolan, T., 2021a. Regulating the expression of gene drives is key to increasing their invasive potential and the mitigation of resistance. *PLOS Genetics* 17, e1009321. <https://doi.org/10.1371/journal.pgen.1009321>
- 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., 2021b. Gene-drive suppression of mosquito populations in large cages as a bridge between lab and field. *Nat Commun* 12, 4589. <https://doi.org/10.1038/s41467-021-24790-6>
- 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, e1007039. <https://doi.org/10.1371/journal.pgen.1007039>
- Han, B., Weiss, L.M., 2017. Microsporidia: Obligate Intracellular Pathogens Within the Fungal Kingdom. *Microbiology Spectrum* 5, 10.1128/microbiolspec.funk-0018–2016. <https://doi.org/10.1128/microbiolspec.funk-0018-2016>
- Hanboonkunupakarn, B., White, N.J., 2022. Advances and roadblocks in the treatment of malaria. *Br J Clin Pharmacol* 88, 374–382. <https://doi.org/10.1111/bcp.14474>
- 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, 828–830. <https://doi.org/10.1038/nbt.2350>
- 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 Medicine* 7, e1000209. <https://doi.org/10.1371/journal.pmed.1000209>
- He, F., Jacobson, A., 2015. Nonsense-Mediated mRNA Decay: Degradation of Defective Transcripts Is Only Part of the Story. *Annu Rev Genet* 49, 339–366. <https://doi.org/10.1146/annurev-genet-112414-054639>
- 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, 2187. <https://doi.org/10.1038/s41467-020-16121-y>
- Herskowitz, I., 1987. Functional inactivation of genes by dominant negative mutations. *Nature* 329, 219–222. <https://doi.org/10.1038/329219a0>
- Hildreth, P.E., 1965. Doublesex, a Recessive Gene That Transforms Both Males and Females of *Drosophila* into Intersexes. *Genetics* 51, 659–678.

- 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, 454–457. <https://doi.org/10.1038/nature10356>
- Holt, R.A., Subramanian, G.M., Halpern, A., Sutton, G.G., Charlab, R., Nusskern, D.R., Wincker, P., Clark, A.G., Ribeiro, José M.C., 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., Flanagan, 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., Venter, J.C., Weissenbach, J., Kafatos, F.C., Collins, F.H., Hoffman, S.L., 2002. The Genome Sequence of the Malaria Mosquito *Anopheles gambiae*. *Science* 298, 129–149. <https://doi.org/10.1126/science.1076181>
- Hosken, D.J., Hodgson, D.J., 2014. Why do sperm carry RNA? Relatedness, conflict, and control. *Trends in Ecology & Evolution* 29, 451–455. <https://doi.org/10.1016/j.tree.2014.05.006>
- Huestis, D.L., Dao, A., Diallo, M., Sanogo, Z.L., Samake, D., Yaro, A.S., Ousman, Y., Linton, Y.-M., Krishna, A., Veru, L., Krajacich, B.J., Faiman, R., Florio, J., Chapman, J.W., Reynolds, D.R., Weetman, D., Mitchell, R., Donnelly, M.J., Talamas, E., Chamorro, L., Strobach, E., Lehmann, T., 2019. Windborne long-distance migration of malaria mosquitoes in the Sahel. *Nature* 574, 404–408. <https://doi.org/10.1038/s41586-019-1622-4>
- Hug, N., Longman, D., Cáceres, J.F., 2016. Mechanism and regulation of the nonsense-mediated decay pathway. *Nucleic Acids Research* 44, 1483–1495. <https://doi.org/10.1093/nar/gkw010>
- 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, e1002043. <https://doi.org/10.1371/journal.ppat.1002043>
- Hughes, G.L., Rivero, A., Rasgon, J.L., 2014. *Wolbachia* Can Enhance *Plasmodium* Infection in Mosquitoes: Implications for Malaria Control? *PLOS Pathogens* 10, e1004182. <https://doi.org/10.1371/journal.ppat.1004182>

- Hughes, G.L., Vega-Rodriguez, J., Xue, P., Rasgon, J.L., 2012. Wolbachia Strain wAlbB Enhances Infection by the Rodent Malaria Parasite *Plasmodium berghei* in *Anopheles gambiae* Mosquitoes. *Appl. Environ. Microbiol.* 78, 1491–1495. <https://doi.org/10.1128/AEM.06751-11>
- Inoue, K., Hoshijima, K., Higuchi, I., Sakamoto, H., Shimura, Y., 1992. Binding of the *Drosophila* transformer and transformer-2 proteins to the regulatory elements of doublesex primary transcript for sex-specific RNA processing. *Proceedings of the National Academy of Sciences* 89, 8092–8096. <https://doi.org/10.1073/pnas.89.17.8092>
- 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. *The American Journal of Tropical Medicine and Hygiene* 98, 1–49. <https://doi.org/10.4269/ajtmh.18-0083>
- James, S.L., Dass, B., Quemada, H., 2023. Regulatory and policy considerations for the implementation of gene drive-modified mosquitoes to prevent malaria transmission. *Transgenic Res* 32, 17–32. <https://doi.org/10.1007/s11248-023-00335-z>
- Jeffries, C.L., Cansado-Utrilla, C., Beavogui, A.H., Stica, C., Lama, E.K., Kristan, M., Irish, S.R., Walker, T., 2021. Evidence for natural hybridization and novel Wolbachia strain superinfections in the *Anopheles gambiae* complex from Guinea. *Royal Society Open Science* 8, 202032. <https://doi.org/10.1098/rsos.202032>
- 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. <https://doi.org/10.12688/wellcomeopenres.14765.2>
- Jiang, X., Biedler, J.K., Qi, Y., Hall, A.B., Tu, Z., 2015. Complete Dosage Compensation in *Anopheles stephensi* and the Evolution of Sex-Biased Genes in Mosquitoes. *Genome Biol Evol* 7, 1914–1924. <https://doi.org/10.1093/gbe/evv115>
- 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, 816–821. <https://doi.org/10.1126/science.1225829>
- Johnson, A.F., Nguyen, H.T., Veitia, R.A., 2019. Causes and effects of haploinsufficiency. *Biological Reviews* 94, 1774–1785. <https://doi.org/10.1111/brv.12527>
- Kandul, N.P., Liu, J., Sanchez C., H.M., Wu, S.L., Marshall, J.M., Akbari, O.S., 2019. Transforming insect population control with precision guided sterile males with demonstration in flies. *Nat Commun* 10, 84. <https://doi.org/10.1038/s41467-018-07964-7>
- Karijolich, J., Yu, Y.-T., 2010. Spliceosomal snRNA modifications and their function. *RNA Biology* 7, 192–204. <https://doi.org/10.4161/rna.7.2.11207>

- Karousis, E.D., Nasif, S., Mühlemann, O., 2016. Nonsense-mediated mRNA decay: novel mechanistic insights and biological impact. *Wiley Interdiscip Rev RNA* 7, 661–682. <https://doi.org/10.1002/wrna.1357>
- Kleinschmidt, I., Bradley, J., Knox, T.B., Mnzava, A.P., Kafy, H.T., Mbogo, C., Ismail, B.A., Bigoga, J.D., Adechoubou, A., Raghavendra, K., Cook, J., Malik, E.M., Nkuni, Z.J., Macdonald, M., Bayoh, N., Ochomo, E., Fondjo, E., Awono-Ambene, H.P., Etang, J., Akogbeto, M., Bhatt, R.M., Chourasia, M.K., Swain, D.K., Kinyari, T., Subramaniam, K., Massougboji, A., Okê-Sopoh, M., Ogouyemi-Hounto, A., Kouambeng, C., Abdin, M.S., West, P., Elmardi, K., Cornelie, S., Corbel, V., Valecha, N., Mathenge, E., Kamau, L., Lines, J., Donnelly, M.J., 2018. Implications of insecticide resistance for malaria vector control with long-lasting insecticidal nets: a WHO-coordinated, prospective, international, observational cohort study. *The Lancet Infectious Diseases* 18, 640–649. [https://doi.org/10.1016/S1473-3099\(18\)30172-5](https://doi.org/10.1016/S1473-3099(18)30172-5)
- Knipling, E.F., 1959. Sterile-male method of population control 130, 3.
- Koonin, E.V., 2005. Orthologs, Paralogs, and Evolutionary Genomics. *Annu. Rev. Genet.* 39, 309–338. <https://doi.org/10.1146/annurev.genet.39.073003.114725>
- Kranjc, N., Crisanti, A., Nolan, T., Bernardini, F., 2021. *Anopheles gambiae* Genome Conservation as a Resource for Rational Gene Drive Target Site Selection. *Insects* 12, 97. <https://doi.org/10.3390/insects12020097>
- Kranjc, N., Thomas, C., 2023. nkran/guido: v0.1.3. <https://doi.org/10.5281/zenodo.8056051>
- Krawetz, S.A., 2005. Paternal contribution: new insights and future challenges. *Nat Rev Genet* 6, 633–642. <https://doi.org/10.1038/nrg1654>
- Krzywinska, E., Dennison, N.J., Lycett, G.J., Krzywinski, J., 2016. A maleness gene in the malaria mosquito *Anopheles gambiae*. *Science* 353, 67–69. <https://doi.org/10.1126/science.aaf5605>
- Krzywinska, E., Ferretti, L., Li, J., Li, J.-C., Chen, C.-H., Krzywinski, J., 2021. femaleless Controls Sex Determination and Dosage Compensation Pathways in Females of *Anopheles* Mosquitoes. *Current Biology* S0960982220318340. <https://doi.org/10.1016/j.cub.2020.12.014>
- Krzywinski, J., Nusskern, D.R., Kern, M.K., Besansky, N.J., 2004. Isolation and Characterization of Y Chromosome Sequences From the African Malaria Mosquito *Anopheles gambiae*. *Genetics* 166, 1291–1302. <https://doi.org/10.1534/genetics.166.3.1291>
- 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, 1062–1066. <https://doi.org/10.1038/nbt.4245>
- Labbé, G.M.C., Scaife, S., Morgan, S.A., Curtis, Z.H., Alphey, L., 2012. Female-Specific Flightless (fsRIDL) Phenotype for Control of *Aedes albopictus*. *PLOS Neglected Tropical Diseases* 6, e1724. <https://doi.org/10.1371/journal.pntd.0001724>
- Lalancette, C., Miller, D., Li, Y., Krawetz, S.A., 2008. Paternal contributions: New functional insights for spermatozoal RNA. *J. Cell. Biochem.* 104, 1570–1579. <https://doi.org/10.1002/jcb.21756>

- Lantz, V., Ambrosio, L., Schedl, P., 1992. The *Drosophila orb* gene is predicted to encode sex-specific germline RNA-binding proteins and has localized transcripts in ovaries and early embryos. *Development* 115, 75–88. <https://doi.org/10.1242/dev.115.1.75>
- Laurens, M.B., 2018. The Promise of a Malaria Vaccine—Are We Closer? *Annual Review of Microbiology* 72, 273–292. <https://doi.org/10.1146/annurev-micro-090817-062427>
- Lawler, C.D., Hernandez, N., Parra Nunez, A.K., Bhidé, S., Baxter, S., Robin, C., 2023. The haplolethal gene *wupA* of *Drosophila* exhibits potential as a target for an X-poisoning gene drive (preprint). *Genetics*. <https://doi.org/10.1101/2023.06.23.546292>
- Lawlor, M.A., Cao, W., Ellison, C.E., 2021. A transposon expression burst accompanies the activation of Y-chromosome fertility genes during *Drosophila* spermatogenesis. *Nat Commun* 12, 6854. <https://doi.org/10.1038/s41467-021-27136-4>
- Laxminarayan, R., 2004. Does reducing malaria improve household living standards? *Tropical Medicine & International Health* 9, 267–272. <https://doi.org/10.1046/j.1365-3156.2003.01171.x>
- Lee, M.T., Bonneau, A.R., Giraldez, A.J., 2014. Zygotic Genome Activation During the Maternal-to-Zygotic Transition. *Annual Review of Cell and Developmental Biology* 30, 581–613. <https://doi.org/10.1146/annurev-cellbio-100913-013027>
- Lemos, B., Branco, A.T., Hartl, D.L., 2010. Epigenetic effects of polymorphic Y chromosomes modulate chromatin components, immune response, and sexual conflict. *Proceedings of the National Academy of Sciences* 107, 15826–15831. <https://doi.org/10.1073/pnas.1010383107>
- Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>
- Li, M., Yang, T., Bui, M., Gamez, S., Wise, T., Kandul, N.P., Liu, J., Alcantara, L., Lee, H., Edula, J.R., Raban, R., Zhan, Y., Wang, Y., DeBeaubien, N., Chen, J., Sánchez C., H.M., Bennett, J.B., Antoshechkin, I., Montell, C., Marshall, J.M., Akbari, O.S., 2021. Suppressing mosquito populations with precision guided sterile males. *Nat Commun* 12, 5374. <https://doi.org/10.1038/s41467-021-25421-w>
- Liang, H.-L., Nien, C.-Y., Liu, H.-Y., Metzstein, M.M., Kirov, N., Rushlow, C., 2008. The zinc-finger protein Zelda is a key activator of the early zygotic genome in *Drosophila*. *Nature* 456, 400–403. <https://doi.org/10.1038/nature07388>
- Lindquist, D.A., Abusowa, M., Hall, M.J.R., 1992. The New World screwworm fly in Libya: a review of its introduction and eradication. *Med Vet Entomol* 6, 2–8. <https://doi.org/10.1111/j.1365-2915.1992.tb00027.x>
- Lloyd, V.K., Sinclair, D.A., Grigliatti, T.A., 1999. Genomic imprinting and position-effect variegation in *Drosophila melanogaster*. *Genetics* 151, 1503–1516.
- Long, K.C., Alphey, L., Annas, G.J., Bloss, C.S., Campbell, K.J., Champer, J., Chen, C.-H., Choudhary, A., Church, G.M., Collins, J.P., Cooper, K.L., Delborne, J.A., Edwards, O.R., Emerson, C.I., Esvelt, K., Evans, S.W., Friedman, R.M., Gantz, V.M., Gould, F., Hartley, S., Heitman, E., Hemingway, J., Kanuka, H., Kuzma, J., Lavery, J.V., Lee, Y., Lorenzen,

- M., Lunshof, J.E., Marshall, J.M., Messer, P.W., Montell, C., Oye, K.A., Palmer, M.J., Papathanos, P.A., Paradkar, P.N., Piaggio, A.J., Rasgon, J.L., Rašić, G., Rudenko, L., Saah, J.R., Scott, M.J., Sutton, J.T., Vorsino, A.E., Akbari, O.S., 2020. Core commitments for field trials of gene drive organisms. *Science* 370, 1417–1419. <https://doi.org/10.1126/science.abd1908>
- Lovett, B., Bilgo, E., Millogo, S.A., Ouattarra, A.K., Sare, I., Gnambani, E.J., Dabire, R.K., Diabate, A., St. Leger, R.J., 2019. Transgenic *Metarhizium* rapidly kills mosquitoes in a malaria-endemic region of Burkina Faso. *Science* 364, 894–897. <https://doi.org/10.1126/science.aaw8737>
- Lucchesi, J.C., 1973. Dosage Compensation in *Drosophila*. *Annual Review of Genetics* 7, 225–237. <https://doi.org/10.1146/annurev.ge.07.120173.001301>
- Lynch, K.W., Maniatis, T., 1995. Synergistic interactions between two distinct elements of a regulated splicing enhancer. *Genes Dev.* 9, 284–293. <https://doi.org/10.1101/gad.9.3.284>
- Magnusson, K., Lycett, G.J., Mendes, A.M., Lynd, A., Papathanos, P.-A., Crisanti, A., Windbichler, N., 2012. Demasculinization of the *Anopheles gambiae* X chromosome. *BMC Evol Biol* 12, 69. <https://doi.org/10.1186/1471-2148-12-69>
- 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, 460. <https://doi.org/10.1186/1475-2875-13-460>
- Marchi, A., Mezzanotte, R., 1990. Inter- and intraspecific heterochromatin variation detected by restriction endonuclease digestion in two sibling species of the *Anopheles maculipennis* complex. *Heredity* 65, 135–142. <https://doi.org/10.1038/hdy.1990.80>
- Marinotti, O., Jasinskiene, N., Fazekas, A., Scaife, S., Fu, G., Mattingly, S.T., Chow, K., Brown, D.M., Alphey, L., James, A.A., 2013. Development of a population suppression strain of the human malaria vector mosquito, *Anopheles stephensi*. *Malaria Journal* 12, 142. <https://doi.org/10.1186/1475-2875-12-142>
- Markstein, M., Pitsouli, C., Villalta, C., Celniker, S.E., Perrimon, N., 2008. Exploiting position effects and the gypsy retrovirus insulator to engineer precisely expressed transgenes. *Nat Genet* 40, 476–483. <https://doi.org/10.1038/ng.101>
- Marois, E., Scali, C., Soichot, J., Kappler, C., Levashina, E.A., Catteruccia, F., 2012. High-throughput sorting of mosquito larvae for laboratory studies and for future vector control interventions. *Malaria Journal* 11, 302. <https://doi.org/10.1186/1475-2875-11-302>
- Marshall, J.M., 2010. The Cartagena Protocol and genetically modified mosquitoes. *Nat Biotechnol* 28, 896–897. <https://doi.org/10.1038/nbt0910-896>
- Marygold, S.J., Roote, J., Reuter, G., Lambertsson, A., Ashburner, M., Millburn, G.H., Harrison, P.M., Yu, Z., Kenmochi, N., Kaufman, T.C., Leever, S.J., Cook, K.R., 2007. The ribosomal protein genes and Minute loci of *Drosophila melanogaster*. *Genome Biol* 8, R216. <https://doi.org/10.1186/gb-2007-8-10-r216>

- Metzstein, M.M., Krasnow, M.A., 2006. Functions of the Nonsense-Mediated mRNA Decay Pathway in *Drosophila* Development. *PLOS Genetics* 2, e180. <https://doi.org/10.1371/journal.pgen.0020180>
- Miller, D., Ostermeier, G.C., 2006. Spermatozoal RNA: why is it there and what does it do? *Gynécologie Obstétrique & Fertilité, Onzièmes Journées nationales de la FFER* (Paris, 11-13 octobre 2006) 34, 840–846. <https://doi.org/10.1016/j.gyobfe.2006.07.013>
- Miller, D., Ostermeier, G.C., Krawetz, S.A., 2005. The controversy, potential and roles of spermatozoal RNA. *Trends in Molecular Medicine* 11, 156–163. <https://doi.org/10.1016/j.molmed.2005.02.006>
- Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., Steinegger, M., 2022. ColabFold: making protein folding accessible to all. *Nat Methods* 19, 679–682. <https://doi.org/10.1038/s41592-022-01488-1>
- Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G.A., Sonnhammer, E.L.L., Tosatto, S.C.E., Paladin, L., Raj, S., Richardson, L.J., Finn, R.D., Bateman, A., 2021. Pfam: The protein families database in 2021. *Nucleic Acids Research* 49, D412–D419. <https://doi.org/10.1093/nar/gkaa913>
- Mitchell, S.E., Seawright, J.A., 1989. Recombination between the X and Y Chromosomes in *Anopheles quadrimaculatus* Species A. *Journal of Heredity* 80, 496–499. <https://doi.org/10.1093/oxfordjournals.jhered.a110911>
- Mnzava, A., Monroe, A.C., Okumu, F., 2022. *Anopheles stephensi* in Africa requires a more integrated response. *Malar J* 21, 156. <https://doi.org/10.1186/s12936-022-04197-4>
- 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, 1268–1278. <https://doi.org/10.1016/j.cell.2009.11.042>
- Mouillaud, T., Berger, A., Buysse, M., Rahola, N., Daron, J., Agbor, J.-P., Sango, S.N., Neafsey, D.E., Duron, O., Ayala, D., 2023. No *Wolbachia* protection against *Plasmodium falciparum* infection in the major malaria mosquito *Anopheles moucheti*. <https://doi.org/10.1101/2023.05.28.542560>
- Murdock, C.C., Blanford, S., Hughes, G.L., Rasgon, J.L., Thomas, M.B., 2015. Temperature alters *Plasmodium* blocking by *Wolbachia*. *Sci Rep* 4, 3932. <https://doi.org/10.1038/srep03932>
- Nagoshi, R.N., Baker, B.S., 1990. Regulation of sex-specific RNA splicing at the *Drosophila* doublesex gene: cis-acting mutations in exon sequences alter sex-specific RNA splicing patterns. *Genes & Development* 4, 89–97. <https://doi.org/10.1101/gad.4.1.89>
- Nattoh, G., Maina, T., Makhulu, E.E., Mbaisi, L., Mararo, E., Otieno, F.G., Bukhari, T., Onchuru, T.O., Teal, E., Paredes, J., Bargul, J.L., Mburu, D.M., Onyango, E.A., Magoma, G., Sinkins, S.P., Herren, J.K., 2021. Horizontal Transmission of the Symbiont Microsporidia MB in *Anopheles arabiensis*. *Frontiers in Microbiology* 12.

- Nattoh, G., Onyango, B., Omoke, D., Makhulu, E., Mbaisi, L., Kamau, L., Gesuge, M.M., Ochomo, E., Herren, J.K., 2023. Microsporidia MB in the primary malaria vector *Anopheles gambiae* s.s. is avirulent and undergoes maternal and horizontal transmission (preprint). In Review. <https://doi.org/10.21203/rs.3.rs-2664771/v1>
- Navarro-Payá, D., Flis, I., Anderson, M.A.E., Hawes, P., Li, M., Akbari, O.S., Basu, S., Alphey, L., 2020. Targeting female flight for genetic control of mosquitoes. *PLOS Neglected Tropical Diseases* 14, e0008876. <https://doi.org/10.1371/journal.pntd.0008876>
- Nazni, W.A., Hoffmann, A.A., NoorAfizah, A., Cheong, Y.L., Mancini, M.V., Golding, N., Kamarul, G.M.R., Arif, M.A.K., Thohir, H., NurSyamimi, H., ZatilAqmar, M.Z., NurRuqqayah, M., NorSyazwani, A., Faiz, A., Irfan, F.-R.M.N., Rubaaini, S., Nuradila, N., Nizam, N.M.N., Irwan, S.M., Endersby-Harshman, N.M., White, V.L., Ant, T.H., Herd, C.S., Hasnor, A.H., AbuBakar, R., Hapsah, D.M., Khadijah, K., Kamilan, D., Lee, S.C., Paid, Y.M., Fadzilah, K., Topek, O., Gill, B.S., Lee, H.L., Sinkins, S.P., 2019. Establishment of Wolbachia Strain wAlbB in Malaysian Populations of *Aedes aegypti* for Dengue Control. *Current Biology* 29, 4241-4248.e5. <https://doi.org/10.1016/j.cub.2019.11.007>
- Neuhaus, C.P., Caplan, A.L., 2017. Ethical lessons from a tale of two genetically modified insects. *Nat Biotechnol* 35, 713–716. <https://doi.org/10.1038/nbt.3927>
- Nguyen, T.H., Nguyen, H.L., Nguyen, T.Y., Vu, S.N., Tran, N.D., Le, T.N., Vien, Q.M., Bui, T.C., Le, H.T., Kutcher, S., Hurst, T.P., Duong, T.T.H., Jeffery, J.A.L., Darbro, J.M., Kay, B.H., Iturbe-Ormaetxe, I., Popovici, J., Montgomery, B.L., Turley, A.P., Zigterman, F., Cook, H., Cook, P.E., Johnson, P.H., Ryan, P.A., Paton, C.J., Ritchie, S.A., Simmons, C.P., O'Neill, S.L., Hoffmann, A.A., 2015. Field evaluation of the establishment potential of wMelpop Wolbachia in Australia and Vietnam for dengue control. *Parasites Vectors* 8, 563. <https://doi.org/10.1186/s13071-015-1174-x>
- Nickless, A., Bailis, J.M., You, Z., 2017. Control of gene expression through the nonsense-mediated RNA decay pathway. *Cell & Bioscience* 7, 26. <https://doi.org/10.1186/s13578-017-0153-7>
- Noedl, H., Se, Y., Schaecher, K., Smith, B.L., Socheat, D., Fukuda, M.M., 2008. Evidence of Artemisinin-Resistant Malaria in Western Cambodia. *New England Journal of Medicine* 359, 2619–2620. <https://doi.org/10.1056/NEJMc0805011>
- North, A.R., Burt, A., Godfray, H.C.J., 2020. Modelling the suppression of a malaria vector using a CRISPR-Cas9 gene drive to reduce female fertility. *BMC Biology* 18, 98. <https://doi.org/10.1186/s12915-020-00834-z>
- North, A.R., Burt, A., Godfray, H.C.J., 2019. Modelling the potential of genetic control of malaria mosquitoes at national scale. *BMC Biol* 17, 26. <https://doi.org/10.1186/s12915-019-0645-5>
- Nöthiger, R., Leuthold, M., Andersen, N., Gerschwiler, P., Grüter, A., Keller, W., Leist, C., Roost, M., Schmid, H., 1987. Genetic and developmental analysis of the sex-determining gene 'double sex' (*dsx*) of *Drosophila melanogaster*. *Genet. Res.* 50, 113–123. <https://doi.org/10.1017/S001667230002351X>
- Oberhofer, G., Ivy, T., Hay, B.A., 2019. Cleave and Rescue, a novel selfish genetic element and general strategy for gene drive. *Proceedings of the National Academy of Sciences* 116, 6250–6259. <https://doi.org/10.1073/pnas.1816928116>

- O'Leary, S., Adelman, Z.N., 2020. CRISPR/Cas9 knockout of female-biased genes *AeAct-4* or *myo-fem* in *Ae. aegypti* results in a flightless phenotype in female, but not male mosquitoes. *PLOS Neglected Tropical Diseases* 14, e0008971. <https://doi.org/10.1371/journal.pntd.0008971>
- Ostera, G.R., 2011. Biosafety Concerns Involving Genetically Modified Mosquitoes to Combat Malaria and Dengue in Developing Countries. *JAMA* 305, 930. <https://doi.org/10.1001/jama.2011.246>
- Ostermeier, G.C., Miller, D., Huntriss, J.D., Diamond, M.P., Krawetz, S.A., 2004. Delivering spermatozoan RNA to the oocyte. *Nature* 429, 154–154. <https://doi.org/10.1038/429154a>
- Oye, K.A., Esvelt, K., Appleton, E., Catteruccia, F., Church, G., Kuiken, T., Lightfoot, S.B.-Y., McNamara, J., Smidler, A., Collins, J.P., 2014. Regulating gene drives. *Science* 345, 626–628. <https://doi.org/10.1126/science.1254287>
- Papathanos, P.A., Bossin, H.C., Benedict, M.Q., Catteruccia, F., Malcolm, C.A., Alphey, L., Crisanti, A., 2009. Sex separation strategies: past experience and new approaches. *Malaria Journal* 8, S5. <https://doi.org/10.1186/1475-2875-8-S2-S5>
- 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, 65. <https://doi.org/10.1186/1471-2199-10-65>
- Parks, A.L., Cook, K.R., Belvin, M., Dompe, N.A., Fawcett, R., Huppert, K., Tan, L.R., Winter, C.G., Bogart, K.P., Deal, J.E., Deal-Herr, M.E., Grant, D., Marcinko, M., Miyazaki, W.Y., Robertson, S., Shaw, K.J., Tabios, M., Vysotskaia, V., Zhao, L., Andrade, R.S., Edgar, K.A., Howie, E., Killpack, K., Milash, B., Norton, A., Thao, D., Whittaker, K., Winner, M.A., Friedman, L., Margolis, J., Singer, M.A., Kopczynski, C., Curtis, D., Kaufman, T.C., Plowman, G.D., Duyk, G., Francis-Lang, H.L., 2004. Systematic generation of high-resolution deletion coverage of the *Drosophila melanogaster* genome. *Nat Genet* 36, 288–292. <https://doi.org/10.1038/ng1312>
- Paysan-Lafosse, T., Blum, M., Chuguransky, S., Grego, T., Pinto, B.L., Salazar, G.A., Bileschi, M.L., Bork, P., Bridge, A., Colwell, L., Gough, J., Haft, D.H., Letunić, I., Marchler-Bauer, A., Mi, H., Natale, D.A., Orengo, C.A., Pandurangan, A.P., Rivoire, C., Sigrist, C.J.A., Sillitoe, I., Thanki, N., Thomas, P.D., Tosatto, S.C.E., Wu, C.H., Bateman, A., 2023. InterPro in 2022. *Nucleic Acids Research* 51, D418–D427. <https://doi.org/10.1093/nar/gkac993>
- Pessia, E., Makino, T., Bailly-Bechet, M., McLysaght, A., Marais, G.A.B., 2012. Mammalian X chromosome inactivation evolved as a dosage-compensation mechanism for dosage-sensitive genes on the X chromosome. *Proceedings of the National Academy of Sciences* 109, 5346–5351. <https://doi.org/10.1073/pnas.1116763109>
- 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, 11. <https://doi.org/10.1186/1741-7007-5-11>

- Pollegioni, P., Persampieri, T., Minuz, R.L., Bucci, A., Trusso, A., Martino, S.D., Leo, C., Bruttini, M., Ciolfi, M., Waldvogel, A.-M., Tripet, F., Simoni, A., Crisanti, A., Müller, R., 2023. Introgression of a synthetic sex ratio distortion transgene into different genetic backgrounds of *Anopheles coluzzii*. *Insect Molecular Biology* 32, 56–68. <https://doi.org/10.1111/imb.12813>
- Price, D.C., Egizi, A., Fonseca, D.M., 2015. The ubiquity and ancestry of insect doublesex. *Sci Rep* 5, 13068. <https://doi.org/10.1038/srep13068>
- Pritchard, D.K., Schubiger, G., 1996. Activation of transcription in *Drosophila* embryos is a gradual process mediated by the nucleocytoplasmic ratio. *Genes Dev.* 10, 1131–1142. <https://doi.org/10.1101/gad.10.9.1131>
- Ranson, H., N'Guessan, R., Lines, J., Moiroux, N., Nkuni, Z., Corbel, V., 2011. Pyrethroid resistance in African anopheline mosquitoes: what are the implications for malaria control? *Trends in Parasitology* 27, 91–98. <https://doi.org/10.1016/j.pt.2010.08.004>
- Reuters, 2023. Nigeria regulator grants approval to Oxford's malaria vaccine. Reuters.
- Rice, W.R., 1996. Evolution of the Y Sex Chromosome in Animals: Y chromosomes evolve through the degeneration of autosomes. *BioScience* 46, 331–343. <https://doi.org/10.2307/1312947>
- Rogers, D.W., Whitten, M.M.A., Thailayil, J., Soichot, J., Levashina, E.A., Catteruccia, F., 2008. Molecular and cellular components of the mating machinery in *Anopheles gambiae* females. *Proceedings of the National Academy of Sciences* 105, 19390–19395. <https://doi.org/10.1073/pnas.0809723105>
- Rose, G., Krzywinska, E., Kim, J., Revuelta, L., Ferretti, L., Krzywinski, J., 2016. Dosage compensation in the African malaria mosquito *Anopheles gambiae*. *Genome Biol Evol* evw004. <https://doi.org/10.1093/gbe/evw004>
- Roseman, R. r., Pirrotta, V., Geyer, P. k., 1993. The su(Hw) protein insulates expression of the *Drosophila melanogaster* white gene from chromosomal position-effects. *The EMBO Journal* 12, 435–442. <https://doi.org/10.1002/j.1460-2075.1993.tb05675.x>
- Ruggero, D., Pandolfi, P.P., 2003. Does the ribosome translate cancer? *Nat Rev Cancer* 3, 179–192. <https://doi.org/10.1038/nrc1015>
- Ryder, E., Ashburner, M., Bautista-Llacer, R., Drummond, J., Webster, J., Johnson, G., Morley, T., Chan, Y.S., Blows, F., Coulson, D., Reuter, G., Baisch, H., Apelt, C., Kauk, A., Rudolph, T., Kube, M., Klimm, M., Nickel, C., Szidonya, J., Maróy, P., Pal, M., Rasmuson-Lestander, A., Ekström, K., Stocker, H., Hugentobler, C., Hafen, E., Gubb, D., Pflugfelder, G., Dorner, C., Mechler, B., Schenkel, H., Marhold, J., Serras, F., Corominas, M., Punset, A., Roote, J., Russell, S., 2007. The DrosDel deletion collection: a *Drosophila* genomewide chromosomal deficiency resource. *Genetics* 177, 615–629. <https://doi.org/10.1534/genetics.107.076216>
- Ryder, E., Blows, F., Ashburner, M., Bautista-Llacer, R., Coulson, D., Drummond, J., Webster, J., Gubb, D., Gunton, N., Johnson, G., O'Kane, C.J., Huen, D., Sharma, P., Asztalos, Z., Baisch, H., Schulze, J., Kube, M., Kittlaus, K., Reuter, G., Maroy, P., Szidonya, J., Rasmuson-Lestander, A., Ekström, K., Dickson, B., Hugentobler, C., Stocker, H., Hafen, E., Lepesant, J.A., Pflugfelder, G., Heisenberg, M., Mechler, B., Serras, F.,

- Corominas, M., Schneuwly, S., Preat, T., Roote, J., Russell, S., 2004. The DrosDel collection: a set of P-element insertions for generating custom chromosomal aberrations in *Drosophila melanogaster*. *Genetics* 167, 797–813. <https://doi.org/10.1534/genetics.104.026658>
- Sachs, J., Malaney, P., 2002. The economic and social burden of malaria. *Nature* 415, 680–685. <https://doi.org/10.1038/415680a>
- Sackton, T.B., Montenegro, H., Hartl, D.L., Lemos, B., 2011. Interspecific Y chromosome introgressions disrupt testis-specific gene expression and male reproductive phenotypes in *Drosophila*. *Proceedings of the National Academy of Sciences* 108, 17046–17051. <https://doi.org/10.1073/pnas.1114690108>
- Sakai, R.K., Baker, R.H., Raana, K., Hassan, M., 1979. Crossing-over in the long arm of the X and Y chromosomes in *Anopheles culicifacies*. *Chromosoma* 74, 209–218. <https://doi.org/10.1007/BF00292273>
- Scali, C., Catteruccia, F., Li, Q., Crisanti, A., 2005. Identification of sex-specific transcripts of the *Anopheles gambiae* doublesex gene. *Journal of Experimental Biology* 208, 3701–3709. <https://doi.org/10.1242/jeb.01819>
- Schier, A.F., 2007. The Maternal-Zygotic Transition: Death and Birth of RNAs. *Science* 316, 406–407. <https://doi.org/10.1126/science.1140693>
- Scholte, E.-J., Ng'habi, K., Kihonda, J., Takken, W., Paaijmans, K., Abdulla, S., Killeen, G.F., Knols, B.G.J., 2005. An Entomopathogenic Fungus for Control of Adult African Malaria Mosquitoes. *Science* 308, 1641–1642. <https://doi.org/10.1126/science.1108639>
- 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. <https://doi.org/10.1038/ncomms11772>
- She, W., Lin, W., Zhu, Y., Chen, Y., Jin, W., Yang, Y., Han, N., Bian, H., Zhu, M., Wang, J., 2010. The gypsy Insulator of *Drosophila melanogaster*, Together With Its Binding Protein Suppressor of Hairy-Wing, Facilitate High and Precise Expression of Transgenes in *Arabidopsis thaliana*. *Genetics* 185, 1141–1150. <https://doi.org/10.1534/genetics.110.117960>
- Simoni, A., Hammond, A.M., Beaghton, A.K., Galizi, R., Taxiarchi, C., Kyrou, K., Meacci, D., Gribble, M., Morselli, G., Burt, A., Nolan, T., Crisanti, A., 2020. A male-biased sex-distorter gene drive for the human malaria vector *Anopheles gambiae*. *Nat Biotechnol*. <https://doi.org/10.1038/s41587-020-0508-1>
- 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. *Parasites Vectors* 3, 117. <https://doi.org/10.1186/1756-3305-3-117>
- Sinka, M.E., Bangs, M.J., Manguin, S., Rubio-Palis, Y., Chareonviriyaphap, T., Coetzee, M., Mbogo, C.M., Hemingway, J., Patil, A.P., Temperley, W.H., Gething, P.W., Kabaria, C.W.,

- Burkot, T.R., Harbach, R.E., Hay, S.I., 2012. A global map of dominant malaria vectors. *Parasites & Vectors* 5, 69. <https://doi.org/10.1186/1756-3305-5-69>
- Sinka, M.E., Pironon, S., Massey, N.C., Longbottom, J., Hemingway, J., Moyes, C.L., Willis, K.J., 2020. A new malaria vector in Africa: Predicting the expansion range of *Anopheles stephensi* and identifying the urban populations at risk. *Proceedings of the National Academy of Sciences* 117, 24900–24908. <https://doi.org/10.1073/pnas.2003976117>
- Skaletsky, H., Kuroda-Kawaguchi, T., Minx, P.J., Cordum, H.S., Hillier, L., Brown, L.G., Repping, S., Pyntikova, T., Ali, J., Bieri, T., Chinwalla, A., Delehaunty, A., Delehaunty, K., Du, H., Fewell, G., Fulton, L., Fulton, R., Graves, T., Hou, S.-F., Latrielle, P., Leonard, S., Mardis, E., Maupin, R., McPherson, J., Miner, T., Nash, W., Nguyen, C., Ozersky, P., Pepin, K., Rock, S., Rohlfig, T., Scott, K., Schultz, B., Strong, C., Tin-Wollam, A., Yang, S.-P., Waterston, R.H., Wilson, R.K., Rozen, S., Page, D.C., 2003. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* 423, 825–837. <https://doi.org/10.1038/nature01722>
- Soh, Y.Q.S., Alföldi, J., Pyntikova, T., Brown, L.G., Graves, T., Minx, P.J., Fulton, R.S., Kremitzki, C., Koutseva, N., Mueller, J.L., Rozen, S., Hughes, J.F., Owens, E., Womack, J.E., Murphy, W.J., Cao, Q., de Jong, P., Warren, W.C., Wilson, R.K., Skaletsky, H., Page, D.C., 2014. Sequencing the Mouse Y Chromosome Reveals Convergent Gene Acquisition and Amplification on Both Sex Chromosomes. *Cell* 159, 800–813. <https://doi.org/10.1016/j.cell.2014.09.052>
- Spinner, S.A.M., Barnes, Z.H., Puinean, A.M., Gray, P., Dafa'alla, T., Phillips, C.E., Nascimento de Souza, C., Frazon, T.F., Ercit, K., Collado, A., Naish, N., Sulston, E., Ll. Phillips, G.C., Greene, K.K., Poletto, M., Sperry, B.D., Warner, S.A., Rose, N.R., Frandsen, G.K., Verza, N.C., Gorman, K.J., Matzen, K.J., 2022. New self-sexing *Aedes aegypti* strain eliminates barriers to scalable and sustainable vector control for governments and communities in dengue-prone environments. *Frontiers in Bioengineering and Biotechnology* 10.
- Stoeckius, M., Grün, D., Rajewsky, N., 2014. Paternal RNA contributions in the *Caenorhabditis elegans* zygote. *EMBO J* 33, 1740–1750. <https://doi.org/10.15252/embj.201488117>
- Tadros, W., Lipshitz, H.D., 2009. The maternal-to-zygotic transition: a play in two acts. *Development* 136, 3033–3042. <https://doi.org/10.1242/dev.033183>
- Tautz, D., Lehmann, R., Schnürch, H., Schuh, R., Seifert, E., Kienlin, A., Jones, K., Jäckle, H., 1987. Finger protein of novel structure encoded by hunchback, a second member of the gap class of *Drosophila* segmentation genes. *Nature* 327, 383–389. <https://doi.org/10.1038/327383a0>
- Taxiarchi, C., Beaghton, A., Don, N.I., Kyrou, K., Gribble, M., Shittu, D., Collins, S.P., Beisel, C.L., Galizi, R., Crisanti, A., 2021. A genetically encoded anti-CRISPR protein constrains gene drive spread and prevents population suppression. *Nat Commun* 12, 3977. <https://doi.org/10.1038/s41467-021-24214-5>
- Taxiarchi, C., Kranjc, N., Kriezis, A., Kyrou, K., Bernardini, F., Russell, S., Nolan, T., Crisanti, A., Galizi, R., 2019. High-resolution transcriptional profiling of *Anopheles gambiae* spermatogenesis reveals mechanisms of sex chromosome regulation. *Sci Rep* 9, 14841. <https://doi.org/10.1038/s41598-019-51181-1>

- Teklehaimanot and Paola Mejia, Awash, 2008. Malaria and Poverty. *Annals of the New York Academy of Sciences* 1136, 32–37. <https://doi.org/10.1196/annals.1425.037>
- Terradas, G., Buchman, A.B., Bennett, J.B., Shriner, I., Marshall, J.M., Akbari, O.S., Bier, E., 2021. Inherently confinable split-drive systems in *Drosophila*. *Nat Commun* 12, 1480. <https://doi.org/10.1038/s41467-021-21771-7>
- Thawornwattana, Y., Dalquen, D., Yang, Z., 2018. Coalescent Analysis of Phylogenomic Data Confidently Resolves the Species Relationships in the *Anopheles gambiae* Species Complex. *Molecular Biology and Evolution* 35, 2512–2527. <https://doi.org/10.1093/molbev/msy158>
- The *Anopheles gambiae* 1000 Genomes Consortium, 2017. Genetic diversity of the African malaria vector *Anopheles gambiae*. *Nature* 552, 96–100. <https://doi.org/10.1038/nature24995>
- Thomas, D.D., Donnelly, C.A., Wood, R.J., Alphey, L., 2000. Insect Population Control Using a Dominant, Repressible, Lethal Genetic System. *Science* 287, 2474–2476. <https://doi.org/10.1126/science.287.5462.2474>
- Turner, J.M.A., 2007. Meiotic sex chromosome inactivation. *Development* 134, 1823–1831. <https://doi.org/10.1242/dev.000018>
- Utarini, A., Indriani, C., Ahmad, R.A., Tantowijoyo, W., Arguni, E., Ansari, M.R., Supriyati, E., Wardana, D.S., Meitika, Y., Ernesia, I., Nurhayati, I., Prabowo, E., Andari, B., Green, B.R., Hodgson, L., Cutcher, Z., Rancès, E., Ryan, P.A., O'Neill, S.L., Dufault, S.M., Tanamas, S.K., Jewell, N.P., Anders, K.L., Simmons, C.P., 2021. Efficacy of Wolbachia-Infected Mosquito Deployments for the Control of Dengue. *New England Journal of Medicine* 384, 2177–2186. <https://doi.org/10.1056/NEJMoa2030243>
- 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., Arie, 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* 26, 1602–1608. <https://doi.org/10.1038/s41591-020-1005-2>
- Valcárcel, 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, 171–175. <https://doi.org/10.1038/362171a0>
- Valenzuela, L., Kamakaka, R.T., 2006. Chromatin Insulators. *Annual Review of Genetics* 40, 107–138. <https://doi.org/10.1146/annurev.genet.39.073003.113546>
- Veitia, R. a., Caburet, S., Birchler, J. a., 2018. Mechanisms of Mendelian dominance. *Clinical Genetics* 93, 419–428. <https://doi.org/10.1111/cge.13107>
- Veitia, R.A., 2002. Exploring the etiology of haploinsufficiency. *BioEssays* 24, 175–184. <https://doi.org/10.1002/bies.10023>
- Verhulst, E.C., van de Zande, L., 2015. Double nexus—Doublesex is the connecting element in sex determination. *Briefings in Functional Genomics* 14, 396–406. <https://doi.org/10.1093/bfgp/elv005>
- Vitale, M., Leo, C., Courty, T., Kranjc, N., Connolly, J.B., Morselli, G., Bamikole, C., Haghighat-Khah, R.E., Bernardini, F., Fuchs, S., 2023. Comprehensive characterization of a

- transgene insertion in a highly repetitive, centromeric region of *Anopheles* mosquitoes. *Pathog Glob Health* 117, 273–283. <https://doi.org/10.1080/20477724.2022.2100192>
- 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. *Journal of Economic Entomology* 93, 123–135. <https://doi.org/10.1603/0022-0493.93.1.123>
- 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, 450–453. <https://doi.org/10.1038/nature10355>
- Wallrath, L.L., Elgin, S.C., 1995. Position effect variegation in *Drosophila* is associated with an altered chromatin structure. *Genes & Development* 9, 1263–1277. <https://doi.org/10.1101/gad.9.10.1263>
- Weedall, G.D., Irving, H., Hughes, M.A., Wondji, C.S., 2015. Molecular tools for studying the major malaria vector *Anopheles funestus*: improving the utility of the genome using a comparative poly(A) and Ribo-Zero RNAseq analysis. *BMC Genomics* 16, 931. <https://doi.org/10.1186/s12864-015-2114-z>
- Werren, J.H., Baldo, L., Clark, M.E., 2008. *Wolbachia*: master manipulators of invertebrate biology. *Nat Rev Microbiol* 6, 741–751. <https://doi.org/10.1038/nrmicro1969>
- West, A.G., Gaszner, M., Felsenfeld, G., 2002. Insulators: many functions, many mechanisms. *Genes Dev.* 16, 271–288. <https://doi.org/10.1101/gad.954702>
- White, B.J., Collins, F.H., Besansky, N.J., 2011. Evolution of *Anopheles gambiae* in Relation to Humans and Malaria. *Annual Review of Ecology, Evolution, and Systematics* 42, 111–132. <https://doi.org/10.1146/annurev-ecolsys-102710-145028>
- White, M.T., Verity, R., Griffin, J.T., Asante, K.P., Owusu-Agyei, S., Greenwood, B., Drakeley, C., Gesase, S., Lusingu, J., Ansong, D., Adjei, S., Agbenyega, T., Ogutu, B., Otieno, L., Otieno, W., Agnandji, S.T., Lell, B., Kriesner, P., Hoffman, I., Martinson, F., Kamthunzu, P., Tinto, H., Valea, I., Sorgho, H., Onoko, M., Otieno, K., Hamel, M.J., Salim, N., Mtoro, A., Abdulla, S., Aide, P., Sacarlal, J., Aponte, J.J., Njuguna, P., Marsh, K., Bejon, P., Riley, E.M., Ghani, A.C., 2015. Immunogenicity of the RTS,S/AS01 malaria vaccine and implications for duration of vaccine efficacy: secondary analysis of data from a phase 3 randomised controlled trial. *The Lancet Infectious Diseases* 15, 1450–1458. [https://doi.org/10.1016/S1473-3099\(15\)00239-X](https://doi.org/10.1016/S1473-3099(15)00239-X)
- Wiedenheft, B., Sternberg, S.H., Doudna, J.A., 2012. RNA-guided genetic silencing systems in bacteria and archaea. *Nature* 482, 331–338. <https://doi.org/10.1038/nature10886>
- Wilson, A.L., Courtenay, O., Kelly-Hope, L.A., Scott, T.W., Takken, W., Torr, S.J., Lindsay, S.W., 2020. The importance of vector control for the control and elimination of vector-borne diseases. *PLoS Negl Trop Dis* 14, e0007831. <https://doi.org/10.1371/journal.pntd.0007831>

- Windbichler, N., Menichelli, M., Papathanos, P.A., Thyme, S.B., Li, H., Ulge, U.Y., Hovde, B.T., Baker, D., Monnat, R.J., Burt, A., Crisanti, A., 2011. A synthetic homing endonuclease-based gene drive system in the human malaria mosquito. *Nature* 473, 212–215. <https://doi.org/10.1038/nature09937>
- Windbichler, N., Papathanos, P.A., Catteruccia, F., Ranson, H., Burt, A., Crisanti, A., 2007. Homing endonuclease mediated gene targeting in *Anopheles gambiae* cells and embryos. *Nucleic Acids Research* 35, 5922–5933. <https://doi.org/10.1093/nar/gkm632>
- 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, e1000291. <https://doi.org/10.1371/journal.pgen.1000291>
- Wongsrichanalai, C., Pickard, A.L., Wernsdorfer, W.H., Meshnick, S.R., 2002. Epidemiology of drug-resistant malaria. *The Lancet Infectious Diseases* 2, 209–218. [https://doi.org/10.1016/S1473-3099\(02\)00239-6](https://doi.org/10.1016/S1473-3099(02)00239-6)
- World Health Organization, 2022. World malaria report 2022.
- World Health Organization, 2015. Guidelines for the treatment of malaria, 3rd ed. ed. World Health Organization, Geneva.
- World Health Organization, 2012. Global plan for insecticide resistance management in malaria vectors 131.
- Wu, Y., Hu, W., Biedler, J.K., Chen, X.-G., Tu, Z.J., 2018. Pure early zygotic genes in the Asian malaria mosquito *Anopheles stephensi*. *Parasites Vectors* 11, 652. <https://doi.org/10.1186/s13071-018-3220-y>
- Wyss, J.H., 2000. Screwworm Eradication in the Americas. *Annals of the New York Academy of Sciences* 916, 186–193. <https://doi.org/10.1111/j.1749-6632.2000.tb05289.x>
- Yadav, A.K., Butler, C., Yamamoto, A., Patil, A.A., Lloyd, A.L., Scott, M.J., 2023. CRISPR/Cas9-based split homing gene drive targeting *doublesex* for population suppression of the global fruit pest *Drosophila suzukii*. *Proc. Natl. Acad. Sci. U.S.A.* 120, e2301525120. <https://doi.org/10.1073/pnas.2301525120>
- Yang, Y., Zhang, W., Bayrer, J.R., Weiss, M.A., 2008. Doublesex and the Regulation of Sexual Dimorphism in *Drosophila melanogaster*: STRUCTURE, FUNCTION, AND MUTAGENESIS OF A FEMALE-SPECIFIC DOMAIN*. *Journal of Biological Chemistry* 283, 7280–7292. <https://doi.org/10.1074/jbc.M708742200>
- 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. *Malaria Journal* 7, 258. <https://doi.org/10.1186/1475-2875-7-258>
- Zélé, F., Nicot, A., Berthomieu, A., Weill, M., Duron, O., Rivero, A., 2014. Wolbachia increases susceptibility to Plasmodium infection in a natural system. *Proc. R. Soc. B* 281, 20132837. <https://doi.org/10.1098/rspb.2013.2837>

Appendix

Licences and Permissions

Page	Name of work	Source of work	Copyright holder	Permission required for its use	Permission note
28	Figure 1.1 – Distribution of the most dominant malaria vectors and Plasmodium falciparum incidence rate in Africa. (Left)	Sinka, M.E., Bangs, M.J., Manguin, S. et al. A global map of dominant malaria vectors. <i>Parasites Vectors</i> 5, 69 (2012). https://doi.org/10.1186/1756-3305-5-69	Copyright © 2012, Sinka et al; licensee BioMed Central Ltd. Publisher: Springer Nature	No	Creative Commons Attribution 2.0 Generic (CC BY 2.0)
28	Figure 1.1 – Distribution of the most dominant malaria vectors and Plasmodium falciparum incidence rate in Africa. (Right)	Malaria Atlas Project (https://malariaatlas.org/)	Malaria Atlas Project	No	Creative Commons Attribution 3.0 Unported License
30	Figure 1.2 – The life cycle of the mosquito.	Wikimedia Commons https://commons.wikimedia.org/wiki/File:Culex_mosquito_life_cycle_en.svg	Credit: Mariana Ruiz Villarreal LadyofHats MRV@LadyofHats.com	No	Creative Commons Public Domain
40	Figure 1.4 – The CRISPR/Cas9 system.	Redman M, King A, Watson C, et al. What is CRISPR/Cas9? <i>Archives of Disease in Childhood - Education and Practice</i> 2016;101:213-215.	© 2016, BMJ Publishing Group Ltd. Redman et al., 2016 Open Access article distributed under the terms of CC BY	No	Creative Commons Attribution (CC BY 4.0)
42	Figure 1.5 – Classification of genetic control strategies.	Grilli, Silvia, Roberto Galizi, and Chrysanthi Taxiarchi. 2021. "Genetic Technologies for Sustainable Management of Insect Pests and Disease Vectors" <i>Sustainability</i> 13, no. 10: 5653. https://doi.org/10.3390/su13105653	© 2021 Licensee MDPI, Grilli et al., 2021 Open Access article distributed under the terms of CC BY	No	Creative Commons Attribution (CC BY 4.0)
61	Figure 2.2 – Modelling on population control with different self-limiting strategies.	Burt Austin and Deredec Anne, 2018. Self-limiting population genetic control with sex-linked genome editors <i>Proc. R. Soc. B</i> 2852018077620180776 http://doi.org/10.1098/rspb.2018.0776	© 2018 Burt and Deredec 2018, Published by the Royal Society under the terms of the Creative Commons Attribution License	No	Creative Commons Attribution (CC BY 4.0)

