UNIVERSITE DE STRASBOURG

ECOLE DOCTORALE 414 – SCIENCES DE LA VIE ET DE LA SANTE

CNRS UPR9022 - Modèles insectes de l'Immunité Innée

Inserm U1257 - Réponses immunitaires des moustiques vecteurs

THÈSE présentée par :

Emily GREEN

soutenue le **17 Décembre, 2019**

pour obtenir le grade de : **Docteur de l'Université de Strasbourg**

Discipline/ Spécialité : Aspects moléculaires et cellulaires de la biologie

Development of genetic control strategies for insect pests using CRISPR/Cas9

THÈSE dirigée par :
M Eric MAROIS — Chargé de Recherche, Université de Strasbourg

RAPPORTEURS EXTERNES:
Mme. Anna-Bella FAILLOUX — Directeur de recherches, Institut Pasteur de Paris
M. Jérémy BOUYER — Directeur de recherches, Centre de Coopération Internationale en Recherche Agronomique pour le Développement

EXAMINATEUR INTERNE :
M. João TRINDADE MARQUES — Professeur, Université de Strasbourg

INVITES :
M. Tony NOLAN — Senior Lecturer, Liverpool School of Tropical Medicine
M. Nikolai WINDBICHLER — Lecturer, Imperial College London

Acknowledgements

My sincere thanks to Eric Marois for taking me on as first a Master's student, then a PhD and for introducing me to the challenging, creative, and sometimes unexpected world of insect transgenesis and gene drive.

A big thanks to everyone who contributed to this project, those who I had the fortune to meet and those who I encountered only in the traces they left in the lab. To Tuyen and Fuatstine, who got the *suzukii* project going. A Irène, la chef cuisinière des mouches. A Laurianne, Johanna, et Philippe, un grand remercîment d'avoir contribué votre expertise protéomique. A Amandine, pour ta patience, ta assiduité, et surtout tout le tempes que tu a consacré aux souris. Finally, I must thank the two fantastic students I had the good fortune to work with during my PhD, Marrion Boeglin and Mallory Kastner.

To all the members of the Mosquito Team, past and present, my deepest thanks and appreciation. To Stéphanie Blandin, whose knowledge and attention to detail is inspiring. To our Post-Docs, Katharina, Nicholas, and especially Dennis, for his help and advice with the bite-back experiments, and to Roenick, who provided invaluable help and support for transgenesis and cloning. To my fellow PhDs, Albert, Alice-Anne, Brice, Vrushali, Antinéa and especially Raquel, who has been my constant lab-mate, bench-mate, and writing pal.

Our lab is lucky to have the knowledge and support of our lab manager, Noemie Jelly, who is an invaluable source for all things logistical and scientific in the lab. Et surtout nos ingénieurs et techniciennes, qui sont les fondations de notre laboratoire et un source de soutien et motivation quotidien: Cyril, Mélanie, Myrriam, et surtout Nathalie. Et finalement, un grande merci aux autres membres de l'UPR pour leur soutien et aide.

Thanks to the friends who have made the hard moments easier and the joyful moments sweeter during this six-year adventure in France. Anais, Magdalène, Sarah, Charles, Illaria, Akira, le meilleur résultat de master était notre amitié. My fellow PhDs, especially Catalina, Metchild, and Josipa, for all the support and distraction. And to those friends abroad, whose short messages made me feel loved from afar.

The path that took me to France began in Mme Mary Laglois classroom. I am deeply appreciative of the creativity and passion she brought to teaching, bringing not only the French language, but French culture, history, and, always, food into the lives of her students. I am also grateful to the French Professors at Western Washington University, particularly Christina Keppie, who first told me about the teaching assistant program in France and taught me how to gargle my French 'R'. I would not have come to Strasbourg without the help of Dr. Barb Quick, who put me in touch with her French colleagues and taught me so much about working with *Drosophila*.

And finally and most deeply, to my family, especially my parents, who have supported me and loved me no matter how far from home I have traveled.

Abbreviations

aa	Amino acid
ACTs	Artemisinin-based combination therapy
AMPs	Antimicrobial Peptides
ApoLp-I	ApoLipophorin-I
ApoLp-II	ApoLipopohrin-II
AsCas12a	Cas12a from *Acidaminococcus*
bp	Base Pair
Cas	CRISPR-associated
CI	Cytoplasmic Incompatibility
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeat
CRISPR-GD	CRISPR-Gene Drive
crRNA	CRISPR RNA
CSP	Circumsporozoite Protien
DDT	Dichlorodiphenyltrichloroethane
DSB	Double-Strand Break
dsDNA	Double-strand DNA
eSpCas9	Enhanced Specifity Cas9 from *Streptococcus pyogenes*
GD	Gene Drive
gRNA	guide RNA
HA	Homology Arm
HEG	Homing Endonuclease Gene
HR	Homologous Recombination
indel	Insertion Deletion
kb	Kilobase
kDa	Kilo Dalton
LbCas12a	Cas12a from *Lachnospiraceae*
Lp	Lipophorin
Lp-GD	Lipophoin Gene Drive
Medea	Maternal effect dominant embryonic arrest
NES	Nuclear Export Signal
NHEJ	Non-Homologous End Joining
NLS	Nuclear Localization Signal
nt	Nucleotide
PAM	Protospacer Adjacent Motif
RIDL	Release of Insects carrying a Dominant Lethal
RMCE	Recombinase-Mediated Cassette Exchange
RNAi	RNA interference
scFv	single chain Fragment variable
SGE	Selfish Genetic Element
sgRNA	Single guide RNA
SIT	Sterile Insect Technique
SpCas9	Cas9 from *Stre*
ssDNA	Cas9 from *Streptococcus pyogenes*
stil	*stand still*

TALEN	Transcription Activator-Like Effector Nuclease
TE	Transposable Element
tracrRNA	Trans-Activating crRNA
WHO	World Health Organization
WT	Wild Type
yl	*yolkless*
ZFN	Zinc Finger Nuclease

Table of Contents

List of Figures

List of Tables

Chapter I

General Introduction

Earth is home to approximately 1 million described insect species, and estimates of the total number of species range from 5 to 30 million (Stork et al., 2015). Found in virtually all terrestrial and freshwater ecosystems, insects play important ecological roles as predators, food sources, detritivores, and pollinators. Most insects are beneficial to humans, however a small number do cause harm, the most severe being damage to crops and natural resources by agricultural pests and transmission of diseases by insect vectors. Damage caused by agricultural pests leads to economic losses and potential shortages of food and other resources, while insect-borne diseases account for approximately 17% of all infectious diseases affecting humans, as well as causing illnesses in non-human animals (World Health Organization, 2017b).

Given this, effective control of insect pests is essential for human health and well-being. Current control strategies rely heavily on the use of chemical pesticides, however the emergence and spread of insecticide resistance coupled with a growing awareness of the negative environmental impact of these chemicals is lessening the effectiveness of current chemical control strategies. Alternative control strategies exist including environmental modification, agricultural practices such as post-harvest sanitation and crop rotation, the use of traps and barriers, and biological control agents. However these strategies are not always available for a given pest, or their implementation may be cost-prohibitive. At the same time, increasing movement of humans and goods around the globe, and a changing climate are leading to the emergence and spread of insect pests.

All these factors highlight the need to develop new control strategies. Genetic-based strategies are particularly attractive, as they are highly species-specific, thereby reducing the potential for negative off-target effects. While theoretical proposals for genetic control date back more than 90 years, movement from lab to field has been slow. Currently the only widely used genetic-based strategy used is the sterile-insect technique (SIT). Several researchers independently proposed the strategy in the 1930s, but it was not until the 1950s that it was fully developed and implemented to control the New World screwworm fly, *Cochliomyia hominivorax* (Scott et al., 2014). SIT involves the mass rearing insects and exposing them to gamma or x-rays prior to release. Irradiation causes dominant lethal mutations in germ cells, rendering the adults sterile. When released in large numbers, these sterile insects – preferably male – reduce the size of the target population. This strategy was used to eradicate *C. hominivorax* from North and Central America, and later applied to other species including the Mexican fruit fly, the tsetse fly, and the Mediterranean fruit fly (Scott et al., 2014).

While SIT has been successful in controlling some insect pests, it has several significant drawbacks. First, the intervention requires the ability to rear and release insects *en masse*. Second, the fitness of the released adults can be reduced by the radiation treatment, or during transportation and release, therefore hindering their ability to mate in the wild. SIT is also most efficient when only males are released, which means that techniques to efficiently sort males and females need to be developed. Successful interventions may also require that the target population be at a low level prior to release, either naturally or through pesticide treatment, and continuous releases are necessary to prevent re-establishment once the pest has been eradicated.

Despite its limitations, the sterile insect technique illustrates that genetic-based techniques can be effective alternative strategies. As genetic engineering tools developed, researchers sought to develop more precise strategies based on genetic modification. One example currently in use is the RIDL (Release of Insects carrying a Dominant Lethal) system. RIDL is similar to SIT, as it involves the mass rearing and release of insects for the purpose of population reduction. However instead of being sterilized prior to release, the insects carry a dominant lethal gene under the control of a conditional repression system (usually the tetracycline-repressible transactivator fusion protein system) (Thomas et al., 2000). Expression of the lethal gene is blocked in the presence of tetracycline, allowing fit and fertile insects to be reared for release. Once the repressor is removed, the gene will be expressed in the progeny of the RIDL insects. Depending on the promoter system chosen, the gene is expressed in one or both sexes, leading to death of all or some of the RIDL progeny.

The Oxitec Company has developed RIDL strains for a variety of pest insects and they show promise for local control of pest populations. However, the effects are limited to a few generations, so repeated releases are necessary to maintain population suppression. Strategies that persist longer would improve the efficiency genetic modification interventions. Additionally, in some cases it is potentially more desirable to modify target populations rather than eliminate them.

Given the limitations of current insect control techniques, there is an ongoing effort to develop new strategies. This study was initiated with the aim of contributing to the development of genetic-based strategies by applying a new strategy—CRISPR-based gene drives (CRISPR-GD)—in two important pest species. Various gene drive systems have been proposed for genetic control of insects for decades with limited success, however CRISPR-

GDs appear to be a viable intervention strategy that could finally realize the decades-long search for effective genetic control. This chapter provides an overview of the history of gene drives, the different strategies that have been tried over the decades, and the current state of CRISPR-GD.

Selfish-genetic elements and gene drives

While the principles of Mendelian genetics underpin our current understanding of inheritance, scientists have long been aware of genes that flout this biological rule. Called selfish genetic elements (SGE), these are genetic systems that are able to bias inheritance in their favor, increasing in frequency despite fitness costs to their hosts (Figure 1.1). SGEs have been found in most eukaryotic species, but their effects are most visible in species with rapid generation time and large numbers of offspring, such as some insects, plants, and rodents.

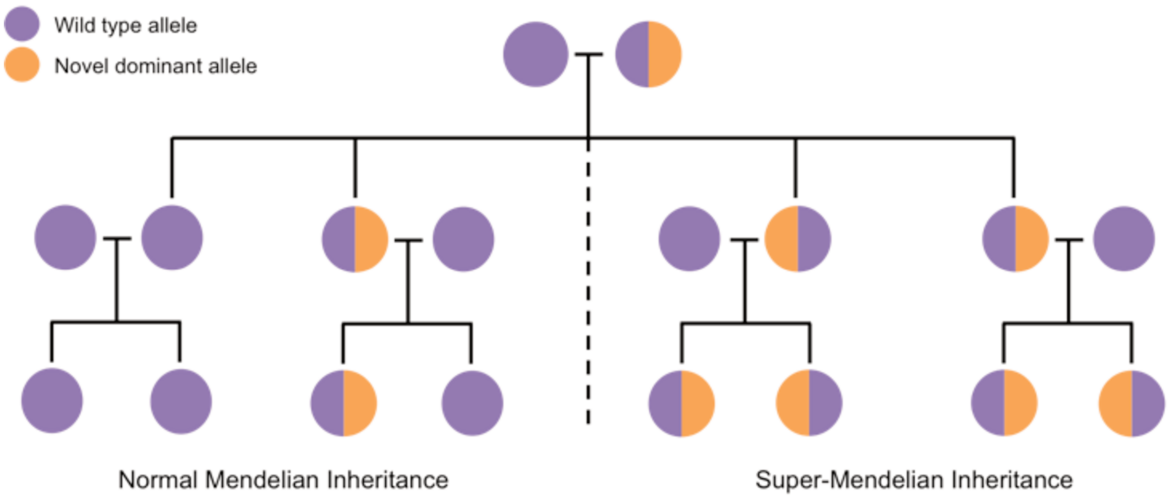


Figure 1.1: Spread of selfish genetic elements. Under normal Mendelian inheritance, a heterozygous individual will transmit a given allele to only 50% of its progeny. As a result, new alleles spread slowly in a population, relying on fitness benefits or chance to increase in frequency. In the case of selfish genetic elements, their ability to bias their inheritance means that a heterozygote will transmit the SGE-allele to >50% of its progeny. This allows an SGE to rapidly increase in frequency, even in the face of fitness costs.

The first known observations of selfish-genetic elements came from the identification of sex-distortion systems in *Drosophila* species. The first description of this phenomenon came from Morgan and colleagues in 1925, when they reported a pattern of sex bias in *D. affinis*, but it was not until work by Gershenson in 1928 that a fuller exploration was offered. Working with lines derived from wild-caught *D. obscura* females, Gershenson observed that some crosses gave rise to majority female progeny. This bias was only apparent when males from the lines were crossed to wild-type females, indicating that the trait was X-linked (Gershenson, 1928). This hypothesis was later proven and similar sex-ratio distortion systems, as they came to be known, have subsequently been documented in a number of *Drosophila* species (Capillon & Atlan, 1999; Jaenike, 1996; Wallace, 1948).

From this first observation of sex-distortion systems, further examples of other selfish genetic elements were identified, including meiotic drive systems, B chromosomes, and transposable elements. In 1988, the first comprehensive review on the topic defined the term "selfish genetic element" (Werren et al., 1988). Under this definition, selfish genetic elements (SGE) are described as systems with characteristics capable of enhancing their own transmission relative to the rest of the genome, and which are either neutral or detrimental to the host organism. Because of this mal-adaptive behavior, many SGEs are ultimately inactivated within their host, either via the emergence of resistance mechanisms such as germline silencing or acquisition of mutations after reaching fixation. SGEs escape this fate by moving horizontally to a new host species. Occasionally, SGEs can be beneficial to a host, generating new traits via their insertion or being coopted for new functions. For example, the molecular mechanism of yeast mate type switching evolved from a domesticated SGE called a homing endonuclease gene (Koufopanou & Burt, 2005).

Today, the umbrella of selfish genetic elements encompasses a range of systems with diverse molecular mechanisms and phenotypic effects. Some systems copy themselves from one DNA locus to the other, while others reduce the viability of non-SGE gametes or zygotes. Some have well-described molecular mechanisms, while others remain obscure. Many of the SGEs first identified came from insect species, leading researchers to propose that these systems could be coopted or mimicked as a method of pest control. Like SIT and RIDL, SGE-based strategies avoid the negative environmental and health effects of pesticides. However their ability to persist beyond one generation makes their use less costly and laborious. In addition to eradicating a population, SGEs could also be used to introduce a desirable trait, thereby expanding toolbox of insect control strategies. The advantages of such approaches were clear to early researchers, however the challenge was identifying SGE systems that could be coopted for pest control.

From SGEs to Gene Drives

A more recent and evocative term for SGEs is "Gene Drive" (GD). The name refers to the ability of these systems to "drive" in a population. Today, gene drive is used as both a noun and a verb, referring to the systems themselves and to their action within a population. While the terms SGE and gene drive are often used interchangeably in the literature, I will use SGE to refer to natural driving elements and gene drive to refer to coopted or artificial systems designed for the purpose of pest control.

Gene drive systems can be engineered for either population suppression or population modification (Figure 1.2). Various gene drive strategies have been tested and proposed – here I will review four systems with distinct modes of action, which illustrate the diversity of strategies available and also trace the history of the development of the field.

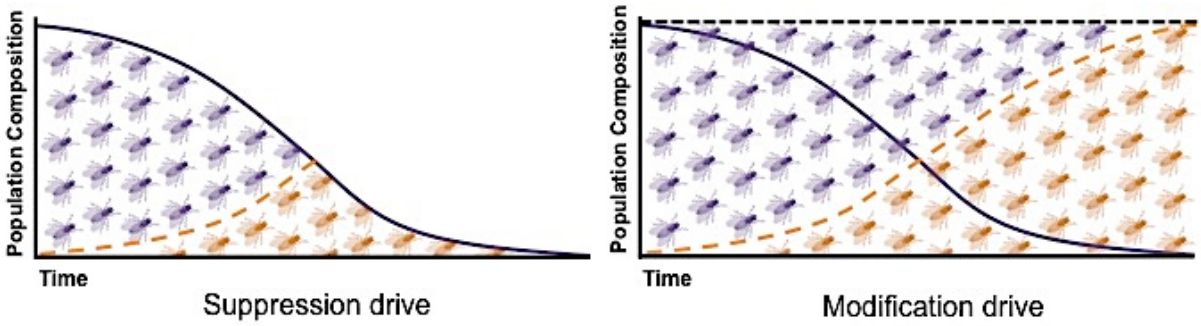


Figure 1.2: Gene drive applications. Gene drives can be engineered for two purposes. Suppression drives reduce the size of the target population by disrupting pathways essential for development or fertility. They can be considered an improvement on the SIT and RIDL strategy. Modification drives on the other hand introduce a novel trait while leaving the target population intact. Compared to other control strategies, modification drives are a novel approach, as the goal is not to remove the target species from its environment, but rather to remove its ability to cause harm to humans. The two approaches can also be combined. For example, a modification drive could be used to make the target species sensitive to a particular compound, which can then be used to reduce the population once the GD has spread above a certain threshold.

Underdominance systems: chromosomal translocations and toxin-antitoxin systems

Underdominance refers to a genetic trait that results in heterozygotes having a lower fitness compared to homozygotes. As a result, these systems exist in an unstable equilibrium, in which one allele will eventually dominate. Modeling and cage experiments of various underdominant systems show that the outcome depends on the initial frequencies of the alleles in the population and their respective fitness costs. In general, an introduced underdominant allele will spread to fixation provided it is released above a critical threshold.

The first underdominance-based gene drive system proposed for pest control was the use of reciprocal translocations. Reciprocal translocations occur when two non-homologous chromosomes exchange genetic material. Provided the translocation does not disrupt local gene function or expression, translocations can be tolerated in the population. However, translocation heterozygotes will have reduced fertility due to abnormal chromosomal segregation in a subset of their gametes. These gametes will be aneuploid for the

translocation loci and will produce unviable zytgotes. While not SGEs *per se*, these systems do exhibit distorted transmission and can therefore be used to create gene drives. The idea of using translocations was first proposed in 1940 by Alexander Sergeevitch Serebrovsky, however as he wrote and published exclusively in Russian, the idea did not spread to the wider scientific community. Almost 30 years later, Curtis proposed a similar approach, which was followed by the publication of a translation of Serebrovsky's original paper (Curtis, 1968; Serebrovsky, 1969).

Throughout the 1970s and 1980s, various translocation strains were identified and tested in several insect species, however the results were disappointing (Robinson, 1976). Laboratory trials using different *Drosophila melanogaster* translocation strains showed that some chromosomal re-arrangements could successfully invade a population (Fitz-Earle et al., 1973). Translocation strains of *Culex pipiens* and *Aedes aegypti* were also identified and eventually underwent short-term, small-scale field trials, during which the translocations were successfully incorporated into the population, however the effects were eventually lost (Laven et al., 1972; Rai et al., 1973). Trials in other insect species also occurred, however no strain capable of long-term, large-scale control was identified. Many of the chromosomal rearrangements, though apparently homozygous fit in laboratory studies, exhibited reduced fitness when introgressed into wild-type strains (Robinson 1976). These failures reduced interest in translocations for genetic control and spurred a search for new genetic control systems.

As genetic engineering progressed, new strategies to engineer underdominant systems were developed. One strategy is the use of two-locus toxin/antidote systems. First proposed by Davis and colleagues, underdominance is achieved by releasing two unlinked genetic

constructs (Davis et al., 2001). Each construct expresses its own toxin and an antidote to the toxin of the other locus. A proof-of-principle study was performed in *D. melanogaster* using an RNAi-based toxin system (Akbari et al., 2013). The team tested both single- and two-loci systems, all of which were able to invade a local population, though the two-locus system required a lower release threshold to invade.

It is now possible to make targeted translocations at specific genomic loci rather than relying on random rearrangements, thereby lessening the potential of generating unanticipated fitness costs due to genomic rearrangement. Recent work by Buchman and colleagues revived chromosomal translocations as a population control strategy (Buchman et al., 2018). Targeted chromosomal translocations were generated in *D. melanogaster* by incorporating artificial cassettes into two *att*P sites located on different chromosomes, and then expression of the I-*Sce*I endonuclease cleaved within the cassettes, generating translocations via homologous recombination between the two broken chromosomes. These translocation systems can become fixed in a population when released above a certain threshold frequency. It remains to be seen how similar systems in pest insect species behave.

Transposable elements

While the work of Buchman and colleagues described above has revived the idea of using chromosomal translocations for pest control, the failures experienced in the 1970s and 1980s lead researchers to abandon translocations at the time and search for a new drive system. The next system considered was a true SGE system – transposable elements. Transposable elements (TE) are SGEs that spread by moving from one genomic location to another in a process called "transposition". They are not fixed to a specific locus and therefore the copy number of a given TE in a genome can far exceed the limitations of Mendelian inheritance. Transposable elements are divided into two classes based on their mechanism of

transposition. Class 1 TEs or retrotransposons express an RNA that is then used as a template to produce a DNA copy that is inserted into a new location. Class 2 TEs or DNA transposons do not pass through an RNA-intermediate. Instead, they are excised from the genome and inserted at a new locus. At first glance, it is unclear how this mode of transposition allows Class 2 TEs to increase in copy number. This is possible when transposition occurs after DNA replication. Excision of the TE leaves behind a DNA break that can be repaired using the sister chromatid as a template, restoring the TE at the original locus (Engels et al., 1990).

TEs are the selfish genetic elements that have the most obvious effect on their hosts. Increase in TE copy number leads to expanded genome sizes. Among *Drosophila* species, TEs make up 10-40% of the genome, while over half of the *Ae. aegypti* genome is made up of transposable elements (Petersen et al., 2019). Insertion of a TE at a new locus alters local genome structure and can disrupt gene function. In response to TEs, organisms have evolved strategies that prevent their mobilization including epigenetic silencing and the expression of sequence-specific repressors. There is evidence that TEs themselves modulate their expression to avoid causing too much damage to their host. While these insertions can be mal-adaptive, TEs can also generate new traits. For example, TEs have been responsible for the appearance of insecticide resistance in several insect species including *D. melanogaster* and *Culex pipiens* (Aminetzach et al., 2005; Darboux et al., 2007). In the case of both *D. melanogaster* and *C. pipiens*, insertion of the TE disrupted an endogenous coding sequence, altering protein expression. In *D. melanogaster*, insertion of a TE in to the *CHKov1* gene results in the production of a truncated, apparently functional protein which alters choline metabolism, leading to an increase in resistance to organophosphates (Aminetzach et al., 2005). In *C. pipiens*, a TE insertion in the *cpm1* gene confers resistance to the binary toxin of

Bacillus sphaericus, a bacterium used as a bio-control agent for mosquitoes. *cpm1* is a membrane receptor that interacts with a prototoxin produced by *B. sphaericus* and insertion of a TE into the second exon of *cpm1* results in altered splicing that removes the prototoxin-interacting domains (Darboux et al., 2007).

Transposable elements can move horizontally as well as vertically. The most striking and best studied example of this ability comes from *D. melanogaster*. In the 1970s, drosophilists began observing high levels of mutations in the progeny of crosses between wild and laboratory strains. The phenomenon was eventually determined to be due to the presence of the P-element transposon. The transposon was likely acquired from *D. willistoni* via horizontal transfer sometime in the 20th century and spread rapidly (Engels, 1992). The exact mode of transfer is not known; one proposal is that the transposon was introduced by the mite *Proctolaelaps regalis* DeLeon, which feeds on *Drosophila* eggs and larvae, though the only piece of direct evidence for this proposal was the detection of free DNA corresponding to the P-element in the mite (Houck et al., 1991). Other potential vectors are viruses, other parasites/parasitoids, or endosymbiotic bacteria. Similar invasion and spread of P-elements has been seen in other drosophila species, in some cases likely via horizontal transfer (Serrato-capuchina et al., 2018; Yoshitake et al., 2018).

The rapid spread of P-elements generated excitement and interest in using transposons for insect control (Curtis & Graves, 1988; Kidwell & Ribeiro, 1992). Cage trials using *D. melanogaster* populations seeded with different initial frequencies of P-element showed that the system could invade a population, however there was strong variability in the rate of increase and the copy number of elements (Carareto et al., 1997). Additionally, autonomous transposons carrying their own source of transposase showed rapid breakdown of the P-

element. This and other trials highlighted the disadvantages of transposons as a control strategy. The inability to control the copy number and insertion site could lead to variable effects of payload genes. Transposons are also vulnerable to losing internal sequences during replication. Finally, P-elements proved nonfunctional in non-*Drosophilid* species and researchers were unable to identify other TEs with similarly high rates of transposition (O'Brochta et al., 2003). For this reason, TEs were abandoned as a control strategy.

Segregation distorters

Segregation distorters are selfish genetic elements that favor their own transmission by disabling gametes or zygotes that do not inherit the distorter. Examples of these systems have been identified in plants, animals, and fungi, including a number of well-described insect systems. In general, segregation distorters function by targeting a responder locus that is sensitive to its activity, however the precise modes of action are highly variable. The Segregation Distorter (SD) system found in *D. melanogaster* involves the interaction of two major loci on chromosome 2 – *Segregation distorter,* which arose from a truncated duplication of the *Ran*GAP gene, and the *responder locus*, which consists of a variable number of tandem repeats. The truncated RanGAP protein is mislocalized, and, through an unknown mechanism, this interferes with the correct chromatin condensation of spermatids containing a high number of *rsp* repeats (up to 2500 copies on highly sensitive chromosomes) (Larracuente et al., 2012). The SD locus is linked to an insensitive copy of *rsp,* protecting SD^+ sperm. Heterozygous males transmit SD to 95-99% of their progeny.

One particular type of segregation distorters are driving sex chromosomes, which result in a strong sex-bias in offspring. In *Drosophila*, several examples of X-linked sex distortion systems have been documented in different species, including the *sex-ratio (SR)* gene described by Gershenson (1928). The molecular mechanisms of many of these systems are

unknown; however most are associated with abnormal segregation of the Y chromosome. One *SR* system from *D. simulans*, the Paris system, leads to a failure of Y-chromosomes to segregate properly, resulting in an excess of X-containing sperm. A variant of a X-linked histone protein has recently been implicated in this system (Helleu et al., 2016). This protein, HP1D2, is highly enriched on the Y-chromosome in developing sperm. *SR* flies carry a dysfunctional copy of HP1D2 that is less abundant, presumably leading to abnormal organization of the Y-chromosome and the *SR* phenotype.

Another sex-distortion system is found in *Aedes* and *Culex* mosquitoes. In these species, sex is determined by an autosomal locus called the M locus. Females are mm and males are Mm. Sex-distorting M (M^D) alleles lead to an over-production of males, sometimes as high as 88-90% (Hickey & Craig, 1966). The cause of this distortion results from breaks in the *m* locus, which leads to an overproduction of *M* sperm (Sweeny & Barr, 1978). Different alleles of *m* are more or less sensitive to sex-distortion. The exact molecular mechanism by which M^D disrupts *m* loci is unknown.

Similarly to other SGE systems, segregation distorters were proposed as insect control strategies soon after their discovery (Hamilton, 1967). Cage trials were performed using a Y-linked *SD* system in *D. melanogaster* and M^D in *Ae. aegypti* (Hickey and Craig 1966; Lyttle 1977; Lyttle 1979). While the strains did generate sex-distortion and in some cases lead to population crashes, in many cases populations evolved resistance to the system. These results and documentation of natural resistance in the *m* locus lessened interest in using natural distorter systems (Hickey and Craig, 1966). Coupling distorter systems to a translocation was proposed as a strategy to improve efficiency and field tests showed that this dual system was

more efficient than simple translocations (Curtis, 1976). However these and other variations on natural sex-distortion systems were never implemented as full-scale control strategies.

In the 2000's, a team began re-investigating the M^D system in *Ae. aegypti*. The driving goal of this work is to molecularly characterize this system. To that end they isolated the T37 strain, which exhibits a strong, stable sex distortion (~15% female) (Mori et al., 2004). The team also proposed using the system to introduce effector genes, by linking them to either an insensitive *m* allele or to M^D. Cage trials showed that this system is able to consistently bias sex ratios over many generations and also to promote the spread of a payload gene, without the emergence of resistance (Cha et al., 2006). However, the system has not yet been further developed.

While work harnessing natural sex-distortion systems has faltered, a synthetic sex-distortion gene drive system was successfully engineered in *An. gambiae*. This system is based on the I-*Ppo*I endonuclease, which recognizes and cleaves a 29 nt sequence serendipitously present within the 28S ribosomal RNA gene. Ribosomal genes in eukaryotes are organized as clusters of tandem repeats called rDNA. In some species of *Anopheles* mosquitoes, rDNA is found only on the X chromosome (Collins et al., 1989). In 2008, Windbichler and colleagues engineered lines of *An. gambiae* that expressed I-*Ppo*I under the control of the male-germline *β2-tubulin* regulatory regions, with the goal of creating a synthetic X-shredder sex-distortion system (Windbichler et al., 2008). Surprisingly, transgenic males were sterile owing to zygotic activity of I-*Ppo*I cleaving the maternal X-chromosome, however PCR analysis showed that 90% of the embryos were male, indicating that the system successfully induced sex-bias. The system was improved by generating a less stable I-*Ppo*I variant which retained the cleavage ability of the native enzyme but did not persist in the sperm and zygote (Galizi et

al., 2014). This construct was successfully able to suppress small cage populations when introduced at 3x the amount of WT males. Large cage trials indicate that the system could be used as a local population control method, making this a potentially viable sex-distortion population control method (Facchinelli et al., 2019).

The segregation distorter systems described above are active during germline development and result in the generation of non-viable zygotes. There is another type of distorter systems that delay their activity until after fertilization. These systems cause the death of embryos that do not inherit them. The best described example of this is the Maternal Effect Dominant Embryonic Arrest (Medea) system, discovered in *Tribolium* beetles (Beeman et al., 1992). In this system, $Medea^{+}$ females express a toxin in their germline. $Medea^{+}$ embryos express the antidote and are therefore unharmed, while $Medea^{-}$ embryos are killed. Embryos can inherit Medea from either their mother or father. The distortion is seen only in females, progeny of $Medea^{+}$ males x WT females are all viable.

The precise mechanism of Medea and other post-segregation distorters is unknown, making direct exploitation of the systems for pest control difficult. However, synthetic toxin/antidote systems can be engineered which mimic theses systems. A synthetic system was engineered in *D. melanogaster* using an RNAi toxin/antitoxin system (Chen et al., 2007). The toxin was a microRNA that targets a gene essential for embryonic development, while the antidote was a resistant copy of the essential gene. Cage trials showed that introduction of the construct above a 25% threshold lead to complete population invasion within 10-12 generations. Various derivations of the Medea system have been proposed, though few have been engineered and tested in the lab. For example, the Medusa system uses a sex-linked

toxin/antitoxin system that produces a male bias, leading to population crash when the system is present above a certain threshold (Marshall & Hay, 2014).

Homing Endonuclease Genes

Many of the SGEs presented so far have highly complex modes of action. In contrast, homing endonuclease genes (HEGs) are elegantly compact and simple. HEGs are highly site and copy-number specific selfish genetic elements. At their simplest, they encode an endonuclease which recognizes a 14-40 bp target sequence which usually occurs only once in the host genome (Burt & Koufopanou, 2004). The HEG is inserted within its own target sequence, disrupting it and thereby rendering HEG^+ chromosomes resistant to HEG activity (Figure 1.3). When a HEG^+ chromosome comes into contact with a HEG^- chromosome, the endonuclease cleaves its target sequence on the HEG^- chromosome, producing a double-strand break. Repair of the break via homologous recombination (see below for a further description) copies the HEG onto the broken chromosome, converting it from HEG^- to HEG^+. This process is called 'homing'.

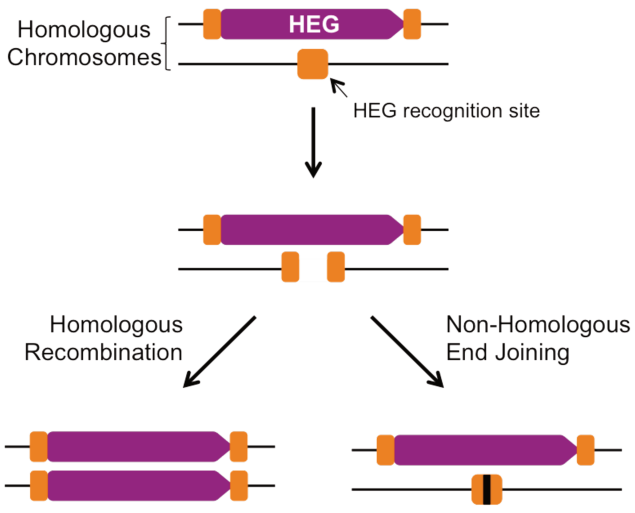


Figure 1.3: Homing endonuclease gene mode of action. A homing endonuclease gene is a highly site-specific selfish genetic element that is inserted within its target site (orange). When a HEG comes into contact with an intact recognition site, the endonuclease will cut the site, creating a double-strand break. Repair of the break by homologous recombination using the HEG^+ chromosome as a template copies the HEG to the second chromosome and disrupts the target site. If the break is repaired by non-homologous end joining, the intact target site may be regenerated, in which case the endonuclease can cut a second time, or the site will be mutated (black), potentially preventing the endonuclease from recognizing it.

HEGs have been described in a range of organisms including bacteria, fungi, protists, and some organelles. In nature, HEGs exhibit strong transmission bias, from 70%-99%. They are often found in self-splicing introns and inteins, which allows them to persist without significantly harming their host (Burt & Koufopanou, 2004).

In 2003, Burt proposed using homing endonucleases as gene drive systems (Burt, 2003). Compared to transposon-based systems, HEGs are restricted to one genomic locus, limiting their disruption of the genome. Unlike segregation distortion or underdominance systems, the molecular mechanism of HEGs is simple and well characterized, making it easier to engineer. Additionally, HEGs are predicted to require a much lower introduction frequency in order to spread. HEG-based interventions are dual-purpose: they can easily be designed either for population reduction or population modification. A HEG targeting a gene essential for development or fertility would reduce the size of a population. A HEG gene drive could also introduce a new trait into a population by disrupting an endogenous gene or by introducing a linked novel gene.

Burt also highlighted some of the challenges of applying HEG gene drive technology. The primary difficulty was the identification or ability to engineer the necessary site-specific endonucleases. The I-*Ppo*I X-shredder developed in *Anopheles* was possible because of naturally occurring endonuclease recognition sites at the desired location. However this was a fortuitous case. A second challenge was resistance to the gene drive. Burt identified several mechanisms that could lead to resistance. The first was mutations within the target site that abolish or reduce the ability of the HEG to cleave. A second possibility, in the case of suppression drives, was the evolution of compensatory mechanisms. Finally, evolution of systems that reduced or eliminated the HEG's ability to home, such as repression of HEG

expression or activity, would also prevent HEG spread. To overcome these challenges, Burt proposed targeting highly conserved amino acid-coding regions within target genes to avoid the generation of viable resistance mutations. He also suggested designing HEG gene drives which recognize multiple sites to either prevent the generation of resistance mutations and/or reduce the possibility of compensatory mutations arising. In the case of resistance to HEG activity, Burt concluded that, as HEGs are not naturally present in the likely target species, it is unlikely that natural resistance would already be present in target populations. If such resistance was to emerge, it could be overcome by using a different HEG with a different structure.

In 2011, two proof-of-concept HEG gene drives were engineered in *D. melanogaster* and *An. gambiae* (Chan et al., 2011; Windbichler et al., 2011). Both designs used the homing endonuclease I-*Sce*I and targeted a GFP marker. These studies provided several important pieces of information. Firstly, the HEG was active and able to home in both species. Secondly, as shown in *D. melanogaster*, the rate of homing is highly dependent on the promoter choice (Chan et al., 2011). Chan and colleagues tested various promoters active at different points in spermatogenesis and observed highly variable rates of homing, building upon previous work indicating that the predominant DNA break repair pathway varies over the course of gametogenesis (see below for a discussion of the major DNA repair pathways). However, subsequent improvements were observed by altering the 3' UTR regulatory regions of I-*Sce*I, which resulted in increased endonuclease expression, underscoring the importance of promoter and terminator choice when designing gene drives (Chan et al., 2013a). Finally, Windbichler et al (2011) showed in cage trials that their I-*Sce*I gene drive could rapidly invade a population over several generations.

While these proofs-of-concept were promising, full realization of HEG gene drives was hindered by the difficulty of finding or creating the necessary site-specific endonucleases. One strategy tested was identification and engineering natural HEGs. In one attempt, variants of the HEG I-*Onu*I were generated using site-directed mutagenesis and yeast-surface display was used to select those capable of cutting the desired target (a putative essential female fertility gene from *An. gambiae*) (Chan et al., 2013b). Further optimization was performed to improve the cleavage rate of the selected variants. However, the I-*Onu*I variants performed worse than a previously characterized I-*Sce*I gene drive (Chan et al., 2013b). Given the labor involved, the failure to recover an active HEG was a blow to further HEG-based gene drive development.

Fortunately, around this time other researchers began to develop strategies to create designer synthetic site-specific endonucleases. The first of these were ZFN, followed by TALENs, and finally the star of the current gene editing revolution: CRISPR systems. However, before discussing these systems, I will provide some background on the two main double-strand break repair pathways used by the cell.

Breaking DNA: the development of targeted double-strand break-based genome editing

The basis of all targeted DNA modification relies on the generation of precise double-strand breaks (DSB) at the site of interest. To maintain the integrity of the genome, cells have evolved robust DNA repair pathways. Detection of DNA damage leads to rapid mobilization of repair proteins. However, not all breaks are repaired using the same pathway. The pathway depends on the kind of DNA damage, cell type, and cell stage. In the context of genome editing, two major DSB repair pathways are important – the homologous recombination pathway and the non-homologous end-joining pathway (Figure 1.4).

Break repair via the homologous recombination pathway uses the homologous locus as a repair template. This pathway is favored during the S and G2 stages of the cell cycle, when the sister chromosome can serve as a repair template, and during meiosis. Upon sensing a double-strand break, the ends are processed to produce 3' single-stranded (ssDNA) ends which then invade the sister chromatid and anneal to homologous regions. Once this annealing occurs, the broken DNA strand is extended, using the homologous DNA as a template, and eventually ligated, forming a four-strand DNA structure called a Holliday Junction. The Holliday Junctions are resolved by nicking endonucleases, which separate the heteroduplex DNA strands. The final step is repair of the nicked DNA strand by a ligase.

As repair via the HR pathway uses the duplicated sister chromatid as a template, the pathway usually restores the original DNA sequence, though errors can occur. For example, if the damaged sequence has multiple homologous regions in the genome (as in the case of repetitive sequences) genomic information can be lost or duplicated (Guirouilh-Barbat et al., 2014). Or if the homologous chromosome is used as a template, HR can result in a loss of heterozygosity at the broken locus. In yeast, HR is estimated to have an error rate 100-1000 times greater than that of DNA replication (Rattray et al., 2015). So while HR is theoretically the most faithful DNA repair pathway, it is not an error-free process.

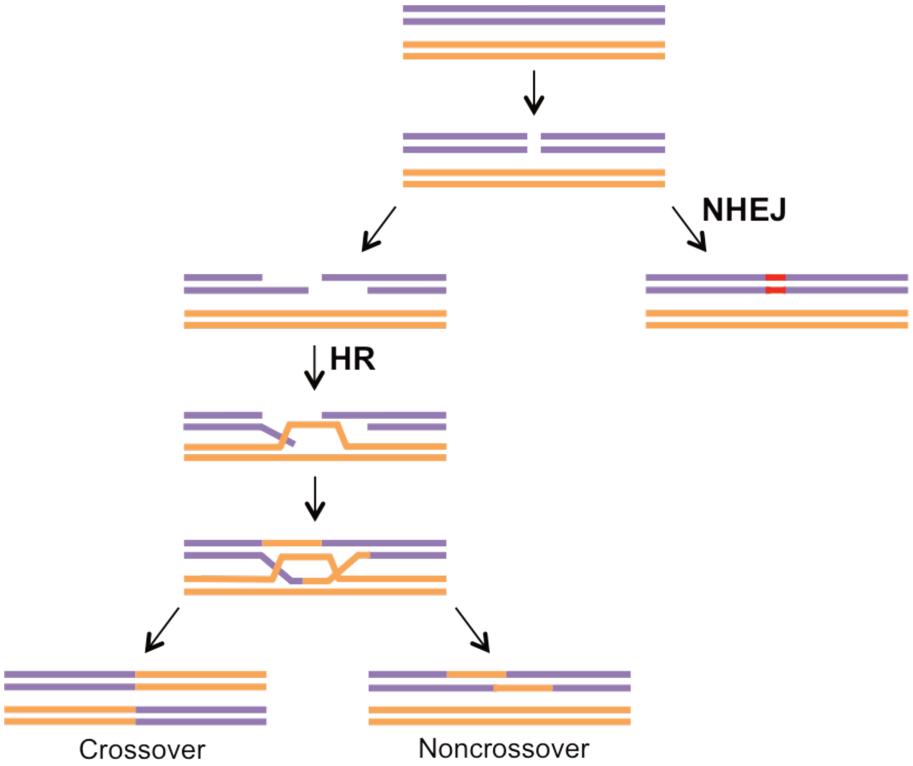


Figure 1.4: Major DNA double-strand break repair pathways. The two major DSB repair pathways used by the cell are homologous recombination (HR) and non-homologous end joining (NHEJ). HR uses the homologous chromosome (purple) as a repair template to fill in any missing DNA. Separation of the two chromosomes via Holliday Junctions can result in either crossing over of the chromosomes or non-crossing over. NHEJ is a simpler repair process in which the broken ends are ligated together, potentially introducing small mutations (red).

The other major DSB-repair pathway is the non-homologous end-joining pathway (NHEJ). This is the predominant pathway in the G1 phase of dividing cells and in G0 cells. In this pathway, the broken DNA ends are associated and ligated together (Figure 1.4). While NHEJ is usually considered more error-prone than HR, the outcome of repair depends on the type of DNA extremities generated by the break. If the broken ends of the DNA are perfectly compatible (i.e. blunt) and no internal sequences have been lost, NHEJ can restore the original sequences. However if the DNA ends require processing, small deletions or insertions (indels) can occur.

Early genetic studies relied on the serendipitous identification of mutants, or the use of mutagenic chemicals or radiation to generate mutations. However these resulted in random mutations. In the 1970s, researchers working with yeast discovered that exogenous DNA carrying a desired modification could be incorporated at a specific genomic locus by providing a plasmid containing homologous stretches of DNA (Hinnen et al., 1978). Subsequent work showed that using a restriction enzyme that cuts the desired modification site could increase the efficiency of these approaches (Rudin et al., 1989). While these early studies were done in yeast, similar approaches were developed in other model organisms.

More precise tools came with the development zinc-finger nucleases (ZFN) and later transcription activator-like effector nucleases (TALENs). These engineered site-specific endonucleases are composed of a DNA-binding domain linked to a DNA-cleavage domain (usually from the *Fok*I restriction enzyme) (Kim et al., 1996). The DNA-binding domain of ZFNs is comprised of 3 – 6 zinc finger binding domain repeats, with each binding domain recognizing 3 nucleotides (Urnov et al., 2010). TALENs are based on the DNA binding TALE domains from the plant pathogenic bacteria *Xanthomonas*, with each domain recognizing a single nucleic acid (Joung & Sander, 2013). A single TALEN domain can consist of up to 30 repeats.

Because the *Fok*I catalytic domain functions as a dimer, two ZFNs and TALENs are necessary to produce a double-strand break. ZFNs and TALEN monomers are designed to bind to target sites upstream and downstream of the break site on opposite DNA strands. ZFNs monomers are separated by a 5-7 bp spacer and TALENs by a 12-25 bp spacer. However, it is possible to design monomeric endonuclease by either targeting a palindromic sequence, or by using a different cleavage domain (Lin et al., 2015). In general, TALENs are

a more cost-effective and easier to use strategy compared to ZFNs, as they are based on a single-nucleotide code rather than triplets. They are also easily designed and assembled in a laboratory.

Modular synthetic gene drives: ZFN/TALENs

Similarly to other gene drive strategies, ZFN and TALEN-based gene drives were first tested in *D. melanogaster* (Simoni et al., 2014). Both systems performed better than the previous I-*Sce*I gene drive, exhibiting higher levels of homing into cleaved targets (49% for TALEN drives and 34% for ZFN drives). However, both systems also generated a large percentage of homed drives that were non-functional in subsequent crosses. PCR analysis of the endonucleases showed either loss of domains or loss of the entire endonuclease, indicating that the construct was either incompletely copied or part of the cassette was lost. As described above, DNA binding of both TALENs and ZFNs is based on repeated amino-acid domains. This means that the DNA sequences of both endonucleases contain closely linked repeats, making them vulnerable to recombination and sequence loss. TALENs were more vulnerable to domain loss during homing, likely explaining the better performance of ZFNs in cage trials. However, re-designing the TALEN drive to use a single monomeric TALEN for cleavage, thereby reducing the number of repeats, did improve the stability of the system (Simoni et al., 2014).

While ZFNs and TALENs can be used for homing-based gene drives, they have several drawbacks. As described above, the repetitive nature of the DNA binding-domains hinders the overall genetic stability of the systems. Additionally, the systems are expensive and time consuming to design, prepare, and test (more so for ZFNs than TALENs). For this reason, CRISPR systems have rapidly supplanted ZFNs and TALENs for most genetic modification applications.

A game-changing tool: RNA-guided endonucleases

CRISPR (Clustered Regularly Interspaced Short Palindromic Repeat) systems are an adaptive immune system found in some bacteria and archea. The system is based on two components: an endonuclease Cas (CRISPR-associated) protein or protein complex and a small RNA molecule called a guide-RNA (gRNA). Most Cas endonucleases cut DNA, though RNA endonucleases have also been identified. The gRNA directs the endonuclease to a specific DNA target, which the endonuclease cleaves, producing a double-strand break. In nature, gRNAs are expressed from CRISPR arrays that comprise repetitions of short palindromic repeats separated by non-repetitive inserts (protospacers). The protospacer sequences come from exogenous DNA and represent a memory of previous viral infections. New protospacers are added to the array during an infection in a process termed adaptation. The CRISPR array is then transcribed as a long precursor CRISPR RNA (crRNA), which is processed into individual crRNAs. Many crRNAs require an additional RNA, a trans-activating or tracrRNA, in order to be correctly processed and functional. The tracrRNA is homologous to the short palindromic repeat portion of the crRNA. A mature gRNA is made of a crRNA-tracrRNA complex.

Once the mature gRNA is expressed and in complex with the endonuclease, it is able to function as an immune system by cutting invading DNA in a process called interference. Interference involves the binding of the gRNA-Cas complex to the DNA at a region complementary to the gRNA protospacer sequence. In order to avoid cutting of the endogenous CRISPR array, most systems have a system to distinguish self from non-self called a PAM (Protospacer Adjacent Motif). The PAM is a short nucleotide sequence adjacent to the gRNA target site on the exogenous DNA. Sites without a PAM will not be cut.

CRISPR arrays were first described in 1987 and other components of the system were gradually uncovered in the decades that followed, culminating in a 2007 report that proved its hypothesized role in immunity (Ishino et al., 1987; Barrangou et al., 2007). Currently, CRISPR systems are classified into two classes with six types (Makarova et al., 2018). The classification is based on the number of proteins involved in the endonuclease activity. Interference in Class I systems is mediated by large, multi-protein complexes, while class 2 systems use only one protein. The best example of a type II system and current star of genetic modification is Cas9. Cas9 was the first CRISPR system to be used in heterologous species, first in another bacteria, *E. coli* (Sapranauskas et al, 2011) and then in human cells (Cong et al, 2013). Fusing of the crRNA-tracrRNA into a single RNA molecule (sgRNA) greatly facilitated subsequent genetic modification protocols (Jinek et al., 2012).

The advantage of CRISPR-based gene editing compared to previous approaches has led to a boom in gene editing and derived applications. To perform genome editing with CRISPR, all that is needed is expression of the Cas protein and the sgRNA. The components can be delivered as protein, RNA, or DNA, and can be expressed from a plasmid or transgenically. Once a modification system is in place, re-engineering the system is as simple as changing the sequence of the sgRNA. The system is much easier and less expensive than ZFNs and TALENs, and generally more efficient. The one major drawback of the system is the potential of off-target cutting by the endonuclease (discussed further below).

CRISPR/Cas9

Cas9 was the first CRISPR system adapted for genome editing and it remains the dominant system in use (Figure 1.5A). Most applications use Cas9 derived from *Streptococcus pyogenes* (SpCas9). SpCas9 is a 162 kDa protein which contains two nuclease domains, an HNH domain and a RuvC domain. The canonical gRNA target is 20 nt, but guides can be shortened to 18 nt without detectable reduction in cutting efficiency. The system recognizes the PAM sequence -NGG located directly downstream of the target sequence on the non-target strand. SpCas9 cuts 3 nt upstream of the PAM and produces a blunt DSB. Though SpCas9 is the predominant Cas9 orthologue used, other Cas9 proteins with differing size and PAM specificities have been identified in other species, and some have been used for genome editing. In addition to these natural variants, engineered versions of SpCas9 have been produced that recognize altered PAM sequences (Kleinstiver et al., 2015).

One concern with Cas9 and other CRISPR-systems is their potential to cause mutations at other sites in the genome - off-target mutations. The off-target rates reported in the literature are variable; however the delivery method, species, and cellular context (cell versus whole organism) of these studies are different, making it difficult to draw general conclusions or develop robust predictive software. Based on what is known about the mechanism of Cas9 *in vivo*, it seems clear that off-target mutations are more likely to sites in the genome where there is homology to the gRNA seed region. The seed region is defined as the 10-12 nt directly adjacent to the PAM within the target region (Jinek et al., 2012; Sternberg et al., 2014). Once associated with its guide RNA, Cas9 scans the DNA until it comes in contact with a PAM site (Sternberg et al., 2014). Association with the PAM triggers local unwinding of the DNA, allowing the gRNA to invade and form a DNA/RNA hybrid. DNA/RNA pairing begins from the PAM and increasing homology further stabilizes the structure, while

mismatches halt cleavage. Mismatches are increasingly tolerated the more distant they are from the PAM. Current recommendations for avoiding off-targets are therefore to choose guides without strong homology to non-target loci in the seed region. Engineered variants of Cas9 have lower off-target activity, such as eSpCas9 and SpCas9-HF1 (Kleinstiver et al., 2016a; Slaymaker et al., 2015). Both of these proteins have amino acid substitutions (3 in the case of eSpCas9 and 4 in the case of SpCas9-HF1), which were initially chosen to reduce protein/DNA interactions, making the stability of the complex more reliant on DNA/RNA pairing. However, work by Chen et al., (2017) indicated that the mutations instead alter the activation threshold of the HNH domain, leading to the development of another high-fidelity Cas9 variant, HypaCas9. It is also possible to use nickase-Cas9, in which one of the catalytic domains is inactivated, creating a Cas9 that cuts only one strand of DNA. In this case the design is analogous to TALENs and ZFNs, where two proteins binding to adjacent targets are used to produce a DSB, which reduces the impact of off-target cutting (Ran et al., 2013).

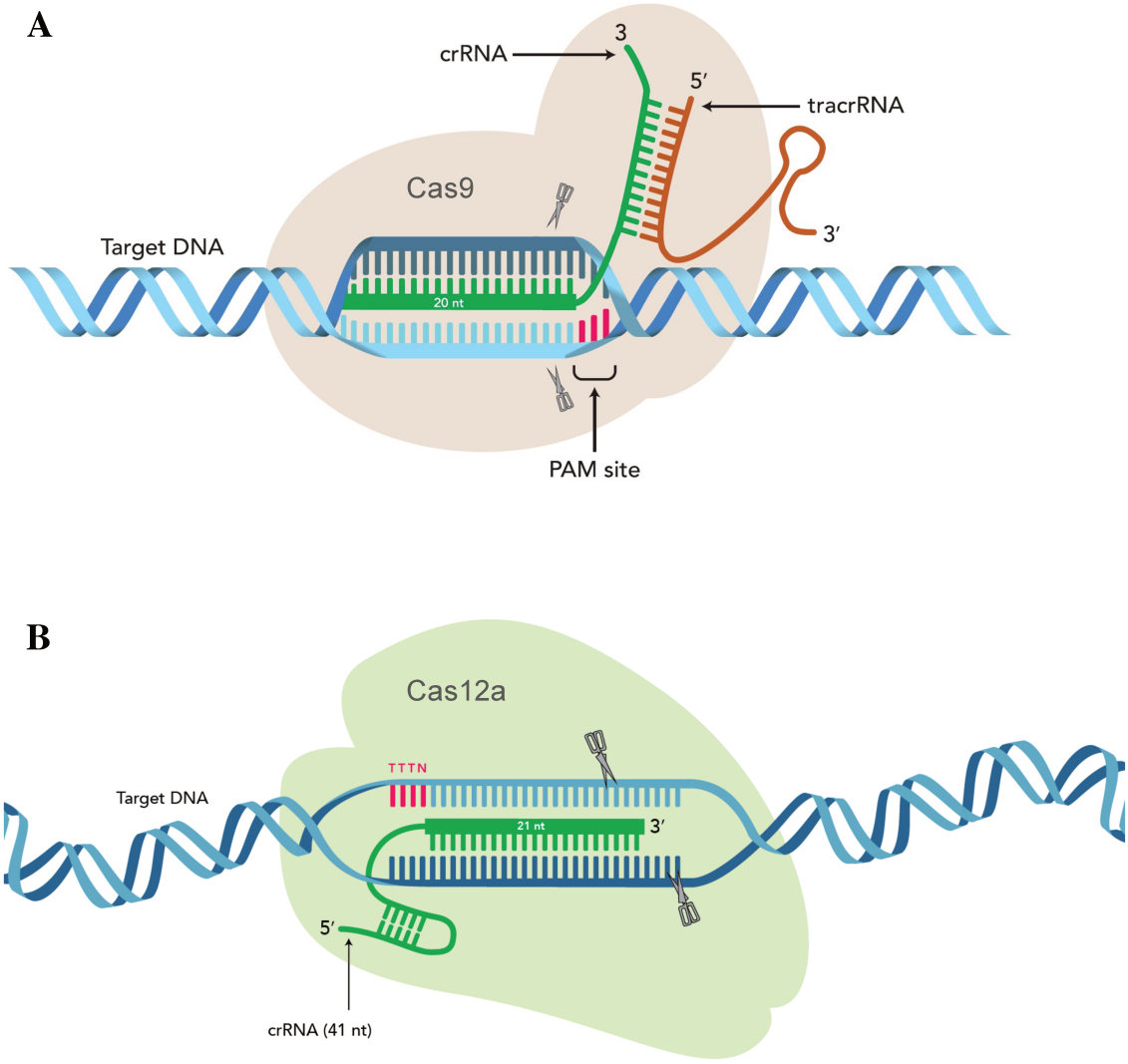


Figure 1.5: Type II CRISPR systems currently used for gene editing. **A**) Cas9 is guided by a two-molecule RNA complex comprised of the conserved tracrRNA (orange) and the semi-variable crRNA (green). The conserved region of the crRNA allows it to bind to the tracrRNA, forming an active guide RNA that Cas9 can bind too. The variable region of the crRNA is homologous to the DNA target site. Cas9 activity depends upon a short conserved PAM site located directly downstream of the target (pink). Once bound to the target site, Cas9 cuts 3 nucleotides upstream of the PAM, producing a blunt DSB. **B**) Cas12a is guided by a single semi-variable crRNA (green) and does not require a tracrRNA. Like Cas9, the variable region of the crRNA corresponds to the DNA target site. Cas12a requires a 4 nucleotide PAM upstream of the target site (pink) and creates a staggered cut 19-23 nucleotides downstream from the PAM. Image courtesy of IDT Technologies.

CRISPR/Cas12a

Recently, another type II CRISPR system was discovered, Cas12a (formerly Cpf1 – CRISPR from *Prevotella* and *Francisella*) (Zetsche et al., 2015) (Figure 1.5B). Cas12a is class II, type V CRISPR system. Like Cas9, the interfering machinery comprises a single protein and gRNA. Unlike Cas9, which requires RNAseIII activity to process the pre-crRNA, Cas12a is capable of cutting its own crRNA array, thanks to its dual RNase and DNase activities. Additionally, Cas12a does not require a tracrRNA. Its canonical PAM sequence is TTTN-, located upstream of the target site. Cas12a also has a different cutting profile compared to Cas9; it produces a staggered cut with 4-5 nt overhangs 18-23 nt downstream of the PAM site. The different motif means that the action of the NHEJ pathway usually results in large, 3-30 nt, deletions (Hu et al., 2017; Wang et al., 2017). The distance from the PAM site also reduces the likelihood that NHEJ repair will mutate the PAM site, meaning that Cas12a can potentially re-cleave the same target sequence.

In general, Cas12a appears less tolerant of mismatches than Cas9. Two adjacent mismatches within the guide sequence are sufficient to hinder its activity (Kleinstiver et al., 2016b). Similarly, Cas12a has been reported to have less off-target activity than Cas9 (Kim et al., 2016; Kleinstiver et al., 2016b). However, comparisons of gene editing efficiency between Cas9 and Cas12a have generated variable results, especially as there are at least three Cas12a orthologues currently in regular use for editing. The first reports of Cas12a came from work done in human cells, with both teams concluding that LbCas12a and AsCas12a were the most efficient variants for eukaryotic editing (Kleinstiver et al., 2016b; Zetsche et al., 2015). However, Zestche found that both Cas12a variants had similar efficiencies to SpCas9, whereas Kleinstiver concluded that SpCas9 was more efficient.

Based on these results, researchers began using Cas12a for genome editing in multicellular organisms. However, the results were perplexing: mutagenesis in mice was highly efficient, however results in *Drosophila,* plants, and zebrafish were inconsistent. In 2017, work by Moreno-Mateos and colleagues found that one of the commonly used Cas12a orthologues, AsCas12a, is highly sensitive to temperature variations. Comparing the efficiencies of AsCas12a and LbCas12a in *Xenopus* and zebrafish embryos, they found that LbCas12a was active at 28°C, but not AsCas12a (Moreno-Mateos et al., 2017). However, incubating embryos at 34°C after injection lead to activation of AsCas12a. In general, Cas12a appears to be more sensitive to temperature variations than Cas9, however its distinctive PAM and cutting motif makes it an interesting alternative. LbCas12a is currently being applied with success for plant editing. It is also possible that protein engineering will generate improved variants of Cas12a, as was the case for Cas9.

One of the most interesting features of Cas12a is its ability to process its own sgRNA, allowing researchers to easily express multiple guides as a single transcript. The short length of sgRNAs (~100 nt) means that individual guides must be expressed using a polymerase III promoter, such as the *H1* or *U6* promoters, which are specific for short RNAs. However, these promoters lack the cell/developmental specificity of many polymerase II promoters. While endonucleases are already expressed from a Pol II promoter, the ability to also limit guide expression spatially and temporally further increases the ability to precisely activate CRISPR activity. A pol II promoter can be used if multiple guides are expressed in the same transcript, a technique called guide multiplexing. Cas12a's inherent multiplexing ability has been demonstrated in rice, using six-multiplexed guides to simultaneously target three genes (Wang et al., 2017). While a similar inherent multiplexing strategy is not possible for Cas9,

synthetic alternatives exist including the use of ribozymes (Gao & Zhao, 2014) or tRNAs (Xie et al., 2015) (discussed further below).

Overview of CRISPR gene editing

To perform DNA editing using CRISPR systems, the first step is to select the guide(s). Many online programs have been developed to identify guide sites within a target locus, and to provide estimates about their activity and potential to induce off-targets. However, these algorithms have been shown to be biased by the data-set used in their construction, meaning the scores should be considered cautiously (Haeussler et al., 2016; Wilson et al., 2018). Similarly, off-target prediction is based on published reference genomes and therefore does not reflect the actual genetic variation in individual laboratory or field populations (Wilson et al., 2018). Additionally, most current models do not incorporate information about DNA-chromatin status. DNA accessibility can be constrained by histones and several studies have shown that CRISPR activity can be affected by DNA chromatin state (reviewed in Verkuijl and Rots 2019). As CRISPR editing continues and more robust models are developed, more specialized tools are likely to develop. For the moment these tools serve as a rapid way to identify guide sites within a target and provide a first indication of potential off-target sites, however their predictive power is limited. Guide selection is also constrained by the goal of the experiment. The choice of guides is necessarily different when the goal is to remove an entire coding region versus a specific exon versus knocking in a tag or marker.

Once the guide(s) have been selected, the next step is to determine how to deliver the endonuclease and guides. Endonucleases can be delivered as DNA, RNA or protein, while guides can be expressed from DNA or supplied as RNA (often pre-complexed with the protein). When performing knock-ins, a repair template is supplied as DNA. For smaller insertions (~50 bp), ssDNA can be used; larger insertions are delivered as dsDNA, usually in

plasmid form. Homology arm length is usually between 1-2 kb, though arms as short as 500 bp can be used for dsDNA templates, and 50 bp for ssDNA (Cong, n.d.). The start of the homology arms should be close to the cut site, ideally less than 10 bp. Once the components have been prepared, they are delivered using standard transformation techniques. In the case of insects, developing embryos are injected using a microinjector. Knockout mutations can also be generated in multicellular organisms by crossing protein and guide expressing transgenic lines.

While gene editing remains the predominant use of CRISPR systems, variants have been developed for other applications. The DNA-binding (and in some cases RNA-binding) ability of CRISPR proteins has been exploited to serve as platforms to assemble other protein domains. CRISPR systems now exist for gene activation or repression, cellular localization and imaging, RNA interference, and epigenetic modifications.

Gene drive in the CRISPR age

The ease and versatility of CRISPR compared to ZFNs, TALENs, or HEGs has rapidly led to its predominance in genome editing. Generating mutants is as simple as selecting guide targets, cloning guides, and expressing them – techniques all easily applied in standard laboratories. Given this, it is unsurprising that CRISPR has also galvanized gene drive development, for both HEG-based strategies and other systems.

Soon after CRISPR/Cas9 was first applied for genome editing in eukaryotic cells, an overview of gene drive strategies in the context of CRISPR was published, laying out the utility of CRISPR systems for gene drive development, an analysis of the likely long-term evolutionary stability of these systems, and their potential applications to combat vector-borne diseases, agricultural pests, and invasive species (Esvelt et al., 2014). Building on the work of

Burt, they highlighted the advantages of CRISPR compared to other site-specific endonucleases, including its simplicity, more efficient cutting rate, and the ability to target multiple loci with one endonuclease. Like all previously tested endonucleases, CRISPR-gene drives (CRISPR-GD) are vulnerable to resistance alleles, however this could be avoided by using multiple guides to target several sites, or by targeting an essential gene and providing an insensitive rescue copy within the GD. The sensitivities of the system to mismatch could also be used to target specific polymorphisms in a population, thereby acting only on a specific sub-population or allele. While pointing out the advantages of CRISPR for GD, Esvelt and colleagues also highlighted the biosafety and environmental concerns posed by such systems, especially in the case of eradicative gene drives, and proposed strategies to recall or limit a drive's spread (2014). Finally, the authors underlined the importance of thoughtful, open development of gene drive interventions that incorporate input from local, national, and international stakeholders. This last point is particularly important for gene drive development. The experience of GMOs in agriculture illustrates how negative public reaction to one application of a particular technology can damage its use in other contexts.

From theory to practice: the first proofs of concept

Unexpectedly, the first published CRISPR-GD was not conceived to be a gene drive *per se*. The goal was instead to develop a technique to efficiently generate homozygous mutants in *Drosophila* (Gantz and Bier, 2015). Gantz and Bier designed a simple Cas9 gene drive targeting the *yellow* gene. A *vasa*-Cas9, *U6:3*-sgRNA cassette was inserted into *yellow*, disrupting its function and generating an easily screened, visible recessive phenotype. Heterozygous flies transmitted the cassette to 95-100% of their progeny, which, in the context of *Drosophila* genetics, would decrease the time to generate homozygous mutants. While this approach was subsequently criticized in the context of *Drosophila*—given the invasive nature of CRISPR-GDs—the approach was recently used in *Candida albicans* to generate mutants,

illustrating that CRISPR-GDs can have a place in fundamental genetic research (Port et al., 2015; Shapiro et al., 2018).

The implications of Gantz and Bier's "mutagenic chain reaction" were immediately apparent for those working on GDs. The high levels of transmission observed in *Drosophila* underscored the need for those working on CRISPR-GDs to ensure that these systems remained confined in laboratory settings, particularly when developing GDs for eventual release (Scott et al., 2013). Molecular containment strategies include the use of split gene drives, where the 'driving components' (endonuclease and sgRNAs) are placed in separate loci, or targeting sequences found only in laboratory strains. Reproductive containment involves using laboratory strains that are unable to reproduce with wild mates. Barrier containment refers to all the physical precautions that can be put in place to prevent accidental escape, including how the GD organisms are housed, the location in which the organisms are housed and manipulated, and who has access to the strains. Finally, ecological or geographical containment involves performing experiments outside the habitable range of the species, or in an area where wild mates are not present. While no standard set of guidelines yet exists, researchers working with CRISPR-GD strains are recommended to employ at least two distinct containment strategies (Akbari et al., 2015; DiCarlo et al., 2015; NAS, 2016).

The next example of CRISPR-GD came from yeast and illustrated several confinement strategies as well as testing gene drive designs for specific applications (Akbari et al., 2015; DiCarlo et al., 2015). Using either a split system or a system targeting an artificial locus, DiCarlo and colleagues observed high levels of gene drive in yeast (over 99% inheritance). They also tested a recoding gene drive, designed to target a specific allele and replace it with

a recoded copy, and a reversal gene drive, which restores the original coding sequence. Each system tested showed similarly high levels of heritability.

Following these proofs-of-concept papers, the next two published CRISPR-GDs were designed for insect pest control. Both of these gene drives were designed with the purpose of controlling malaria, which is spread by *Anopheles* mosquitoes. As described above, GDs can be designed for one of two purposes. Suppression drives reduce the size of the population, while modification drives spread a particular trait. Using the CRISPR-Cas9 system, Gantz and colleagues designed a modification drive in *An. stephensi* to block malaria transmission. The 16 kb gene drive cassette contained Cas9 under control of the *An. stephensi vasa* regulatory elements, a single sgRNA targeting the *kynurenine hydroxylase*^{white} (*kh*^{w}) gene under control of an *An. stephensi* U6 promoter, a *3xP3-DsRed* marker, and two cargo effector genes, single-chain antibodies (scFv) sc2A10 and sc1C3, under the control of the *vitellogenin* and *carboxypeptidase* promoters from *An. stephensi*, respectively (Gantz et al., 2015). The combination of these two effector genes has previously been shown completely block malaria transmission (Isaacs et al., 2012) (for a further description of malaria transmission and single-chain antibodies, see Chapter 3). Disruption of *kh*^{w} results in mutant white-eyed adult mosquitoes. In initial crosses, both GD^{+} females and males exhibited strong super-Mendelian inheritance patterns (~99%), indicating that large CRISPR-GD cassettes can efficiently home in mosquitoes. However, by the fourth generation, differences between the lineages began to appear. While progeny of G2 males continued to show strong biased transmission of the GD cassette (98.5%), only 57.1% of the larval progeny of G2 females expressed DsRed, while at the adult stage, 41.2% of *white* adults did not express DsRed. PCR amplification and sequencing of the *kh*^{w} target site showed the presence of indels due to NHEJ repair of the DSB break. Additionally, the progeny of GD^{+} females frequently showed a mosaic rather than fully

white phenotype. The same phenomenon was observed in *Drosophila* CRISPR-GD experiments targeting *yellow* (Gantz et al., 2015). This is due to the presence of active Cas9 protein in the eggs of GD^{+} females, which can lead to cleavage of some of the paternal DNA in a subset of embryonic cells, resulting in a mosaic phenotype. Cleavage and NHEJ of the genetic material that gives rise to the germ tissue can also result in the generation of heritable resistant mutations.

Mosaicism was also observed in a CRISPR-GD designed to suppress *An. gambiae* populations. Again using *vasa*-Cas9 and a single sgRNA, Hammond and colleagues (2017), developed three gene drives targeting three different female fertility genes in *An. gambiae*, with the goal of population suppression. The cassettes were initially efficiently driven, with 91.4-99.6% heritability, however maternal-carryover of Cas9 again led to mosaic progeny and resistant allele formation. Additionally, as the GD targeted a fertility gene, mosaic heterozygous females also experienced reduced fertility. In cage studies, starting from a 50% prevalence, the GDs initially increased in prevalence for the first few generations, however by the 25^{th} generation, prevalence had fallen to 20%, and the majority of heterozygous females had a GD-resistant allele by the 20^{th} generation. A follow up cage experiment of the most promising candidate showed that the drive prevalence peaked at the 6^{th} generation, and also that functional resistant mutations were already present in the second generation (Hammond et al., 2017).

These first CRISPR-GD papers illustrated the superiority of CRISPR for homing gene drives compared to HEGs, ZFNs, or TALENS. CRISPR-GDs have a much higher cutting and homing efficiency than previously used endonucleases. They are easier to design and use, and can drive despite their large cassette size. However, as previous authors pointed out,

resistant alleles are the weak point of homing GDs (Burt, 2003; Deredec et al., 2008). Additionally, gene drives are spread by heterozygotes, so any fitness cost to these individuals will reduce the effectiveness of the gene drive. The mosaicism observed in *Anopheles* is an example of a heterozygous fitness cost that could hinder a drive's spread.

Refining gene drive design

In response to these first CRISPR-GD papers, several models were published illustrating that, at least for these first-generation gene drives, the evolution of resistance is highly likely and that resistant mutations will result in gene drive extinction, unless the resistant mutations themselves have a fitness cost on par with the gene drive (Burt, 2003; Deredec et al., 2008). Resistant alleles come from three different sources – natural existing genetic variation, *de novo* mutation, and NHEJ repair. Several strategies have been suggested to counteract the emergence of resistant alleles. These include pre-screening the target species or populations to select regions with low genetic variability, selecting sgRNA sites where mutations are likely to be mal-adaptive (highly conserved amino-acid residues, intron-exon splice junctions), using multiple guides or gene drives, selecting promoters which are tightly confined to the germline and during periods when HR is favored, modulating Cas9 activity, and suppressing non-homologous end joining (Drury et al., 2017; Unckless et al., 2015; Unckless et al., 2017).

Subsequent papers have demonstrated the utility of some of these proposals. Improvements in Cas9 expression was achieved in *D. melanogaster* by substituting the *vasa* promoter for *nanos* (Champer et al., 2017). In *An. gambiae*, the *zero-population growth* (*zpg*) promoter provides tighter germline expression than *vasa*, *nanos*, and *exuperantia* promoters (Hammond et al., 2018). A gene drive targeting a highly conserved guide site at an intron-exon junction of the *An. gambae doublesex* gene driven by *zpg*-Cas9 resulted in a highly invasive gene

drive, capable of reaching complete fixation within 7-11 generations without the appearance of functional resistant alleles (Kyrou et al., 2018). This gene drive results in complete serilization of females, making it a viable population control strategy. In addition to improving Cas9 expression using different promoters, two studies in yeast have also demonstrated that Cas9 gene drive activity can be modulated by the addition of nuclear localization and/or nuclear export signals, by modifying the time Cas9 spends in the nucleus (Goeckel et al., 2019; Roggenkamp et al., 2018).

Expression of multiple sgRNAs or multiplexing can be achieved by simply adding additional *U6*-sgRNA genes, or by using strategies to express multiple sgRNAs from the same promoter. Several techniques have been developed to multiplex Cas9 sgRNAs, the most widespread of which are the use of flanking ribozymes or tRNAs (Gao & Zhao 2014; Xie et al., 2015). Ribozymes are catalytic RNA molecules capable of carrying out enzymatic activity. Self-cleaving ribozymes such as the hammerhead and hepatitis delta virus are able to excise themselves from an RNA molecue. tRNAs on the other hand require external enzymes for their maturation. In eukaryotes, tRNAs are expressed as a pre-tRNA transcript, which often contains other tRNAs or non-coding RNA. Mature tRNAs are generated by enzymatic cleavage of their 5' and 3' ends, by RNase P and RNase Z respectivly (Xie et al., 2015). Flanking sgRNAs with either ribozymes or tRNAs has third benefits – first, it removes the nucleotide constraints of some pol III promoters; second, multiple sgRNAs can be expressed in the same transcript; third, these longer transcripts can be put under the control of pol II promoters, providing tighter control over sgRNA expression. Further multiplexing has been achieved by placing multiplexed sgRNAs within an artificial intron of Cas9 (Ding et al., 2018) or using an internal ribosome entry site (Yoshioka et al., 2015).

So far, only a few studies have tested muliplexing sgRNAs. In all of these cases, multiplexing reduced resistance allele formation, however the homing efficiencies varied depending on the number of guides. Use of two guides in yeast and *D. melanogaster* resulted in increased homing efficiency and fewer resistant alleles (Champer et al., 2018; Yan & Finnigan 2019). On the other hand, expression of four guides in *D. melanogaster* resulted in high levels of cleavage, but low levels of homing and instability of the gene drive cassette (Oberhofer et al., 2018). However, the guides used in this study were spread out over 2.2 or 8.8 kb of the target site, meaning that two of the guides were distant from the homology arms of the gene drive; different results could be observed by positioning the guides closer to the outermost cut sites, or by clustering all the guides at one particular locus.

Recent modeling by Champer and colleagues indicates that there is a balance between the number of guides, resistant allele formation, and drive efficiency (Champer et al., 2019). Increasing the number of guides does reduce resistant allele formation, but as the number of guides increases, the cleavage rate plateaus as Cas9 becomes the limiting factor. Based on their model and experimental results, Champer et al., recommend 2-3 guides for most applications.

If things go wrong: Gene Drive reversal strategies

While much attention has been paid to improving and enhancing homing gene drives ability to spread, the invasive nature of these systems has also raised alarm bells. Early modeling studies based on the first-generation gene drives showed that these systems are highly invasive, and that under some situations release of a single individual could lead to the GD becoming fixed in the population (Noble et al., 2018; Unckless et al., 2015). While resistant alleles could constrain a drives spread, a highly invasive drive can still reach a high frequency even in the face of fit resistant alleles (Noble et al., 2018). Other models incorporating spatial

and/or population structures illustrate some of the complexity of drive dynamics (Bull et al., 2019; Champer et al., 2019). While each model addresses a specific aspect of gene drive dynamics, the work done to date supports three general conclusions. First, drive spread is constrained by its intrinsic fitness cost and the presence/appearance of functional resistant mutations (Beaghton et al., 2017; Burt, 2003; Unckless et al., 2015). Second, structuring of the target population (physical isolation of sub-populations, mate choice) will reduce the effectiveness of a drive, potentially leaving time for resistance to evolve or leaving untouched pockets that can serve as a reservoir to for recolonization (Bull et al., 2019; Champer et al., 2019). Finally, and on the other hand, if there is any gene flow between populations, there is a high likelihood that the current generation of homing gene drives will invade that population (Noble et al., 2018; Unckless et al., 2015).

For this reason, at the same time as researchers have focused on enhancing drive efficiency, they have also proposed strategies to limit the spread of homing gene drives or recall a drive once it has been released. Many strategies have been proposed, however few have been demonstrate in the lab and none has been tested in a non-model organism. Limiting strategies are designed to reduce a drive's ability to spread in a population, with the goal of preventing it from becoming fixed while still having an effect on the target population. Proposed strategies include the daisy-chain and derived systems, in which a single gene drive is separated into multiple linked or unlinked loci, each one driving another, and tethered homing drives, in which Cas9 is linked to an underdominance system while the guides and any cargo genes are inserted into the homing locus (Dhole et al., 2019; Esvelt 2017; Min et al., 2017). However, none of these systems have been engineered and tested in any organism. The Esvelt team has declared that they plan to engineer their daisy systems in mice and *C. elegans*; given the

complexity of these systems it remains to be seen how easy they are to engineer and how reliably they behave in a population.

Reversal strategies are designed to stop the spread of an invading gene drive, either in the case of accidental release or unintended consequences, or, ominously, in the case of intentional release of a gene drive for bioterrorism purposes. Soon after the publication of the first gene drives, the US Military's research agency, the Defense Advanced Research Projects Agency (DARPA), established their Safe Genes Program in 2017, which provides funding for researchers developing and testing reversal strategies. A simple reversal measure is to release non-gene drive organisms with recoded sgRNA target site(s). This is most effective when the gene drive has a strong fitness cost, such as in the case of a suppression drive. However this may not be feasible depending on the organism or in the case of a modification gene drive.

Another option is to release a second gene drive that will replace the original drive. This strategy proved functional in yeast (Akbari et al., 2015). However this approach is undesirable when the goal is to remove a gene drive from the population. A third option is to use systems that are only active in the presence of the gene drive. These include the CATCHA, ERACER, and CHACR systems, which all are based on sgRNAs. The CATCHA construct incudes one or more sgRNAs which home into the Cas9 locus, disrupting its activity (Wu et al., 2016). This system was tested in *D. melanogaster*. A variant of CATCHA is the ERACER system, which again includes sgRNAs at the homing locus, but in this case the whole gene drive locus is removed (Gantz & Bier, 2016). The CHACR system also includes sgRNAs, this time targeting multiple loci at different locations, designed to inactivate the gene drive and also to kill the cell (Gantz & Bier, 2016). Neither the ERACER nor CHACR systems have yet been tested *in vivo*.

A final strategy is to use inhibitory anti-CRISPR proteins. In response to CRISPR systems, bacteriophages have evolved their own suite of counter measures, including proteins that bind to and inhibit some part of the CRISPR machinery. Two anti-CRISPR proteins, AcrII2A and AcrIIA4, which both inhibit Cas9, were shown to inhibit gene drive activity in yeast, however it remains to be seen how exactly these proteins would be introduced into a target population (Basgall et al., 2018). Independent modeling of these systems indicates that drive breaks such as CATCHA and ERACR can halt a drive spread, but on varying timescales and with varying dynamics, and provided that the drive break has a high enough fitness (Girardin et al., 2018; Vella et al., 2017). It is important to note that release of a generic reversal drive into the local environment (sgRNAs targeting Cas9, anti-CRISPR proteins) that could potentially interfere with other ongoing or future gene drive releases.

PhD Objectives

It has only been five years since the first CRISPR-GD papers were published, yet the field has advanced rapidly. While the technology remains controversial, approval has been given to pursue gene drives for insect control, particularly disease vectors. Gene drives can potentially fill a gap in pest control, either where current control strategies are not sufficient (the case of some disease vectors) or where they are lacking (for some new invasive species).

The goals of this PhD project were to develop gene drives for insect pest control in two species: *Drosophila suzukii* and *Anopheles gambiae*. *D suzukii* has recently emerged as a global, highly invasive pest of soft-skinned fruits, including many important agricultural crops. As *D. suzukii* is a recent arrival in most of its current range, we proposed to develop a suppression gene drive system to control its population. *An. gambiae*, on the other hand, is an endemic vector of malaria. For this species, we proposed to develop a gene drive that would disrupt the ability of the vector to transmit malaria, without removing it from its habitat.

In the next two chapters of this report I will describe each of these projects in detail, highlighting the successes and failures encountered when working with these two distinct species. I will begin by presenting our work with *D. suzukii* and my efforts to establish an effective transgenesis system, before passing to *An. gambiae* and the indirect gene drive system we have designed.

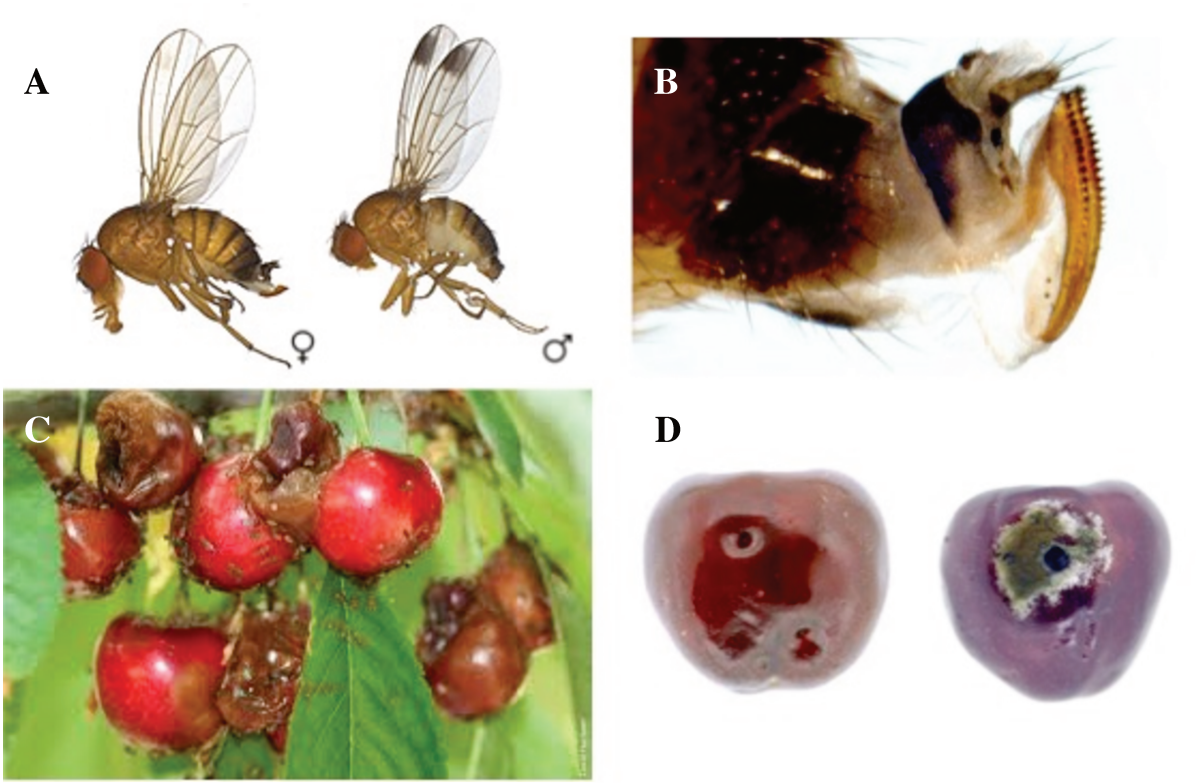


Figure 2.1: Biology of *D. suzukii*. A) *D. suzukii* females and males are morphologically similar to other *Drosophilids*, however females have a long serrated ovipositor. Males have a dark spot on their leading wing edge, but this takes several days to develop and is not necessarily diagnostic. **B)** A close up of a female ovipositor, showing the serrations that allow *D. suzukii* to pierce fruit skin. **C)** Cherry fruit infested with *D. suzukii*. **D)** Examples of primary damage (right) and secondary damage (left) caused by *D. suzukii*. The oviposition scar on the cherry marks the site where the eggs were laid. The damage also renders the fruit vulnerable to secondary infections, such as mold. (Photo credits: A) Eric LaGrasa, Washington State Department of Agriculture, B) Martin Hauser, California Department of Food and Agriculture, C) David Haviland, UC Davis, D) Martin Hauser).

Chapter II

Efforts to establish a suppression gene drive in *D. suzukii*

Introduction

Drosophila suzukii, also known as the Spotted Wing Drosophila or Cherry Drosophila, has recently and rapidly emerged as one of the most important global fruit pests. A member of the *Drosophilidae* family, *D. suzukii* is one of only two drosophilids that cause economic damage to crops (Walsh et al., 2011). Unlike the majority of *Drosophila* species, which are attracted to overripe fruit, *D. suzukii* has adapted to lay its eggs in ripening and ripe fruits. To do this, females have evolved a serrated ovipositor that allows them to pierce the skin of healthy fruits and deposit their eggs below the surface (Walsh et al., 2011) (Figure 2.1B). A single female can lay up to 60 eggs per day, and will deposit 1-3 eggs per oviposition site. Multiple females will also oviposit in the same fruit. As the larvae grow and develop, they consume the fruit, making it unfit for human consumption (Figure 2.1C and D). *D. suzukii* activity also leaves the fruit vulnerable to secondary infections from fungi, yeasts, bacteria, and other insects (Walsh et al., 2011).

D. suzukii is highly adaptable and is able to lay its eggs in a wide range of hosts. Its preference is for soft-skinned fruits such as strawberries, blueberries, raspberries, cherries, and peaches, but in the absence of these hosts it has been observed to use a variety of crop and non-crop species, including kiwis, tomatoes, grapes, wild rose, and dogwood (Asplen et al., 2015). A survey of 165 potential European hosts species from Italy, the Netherlands, and Switzerland found *D. suzukii* infestation in 84 of them (Kenis et al., 2016).

In addition to its broad host preferences, *D. suzukii* is also able to tolerate low temperatures. Cold exposure stimulates the development of a darker, winter morph adult form (Ryan et al., 2016; Toxopeus et al., 2016). Laboratory studies have shown that flies can survive a three-day exposure of -7.5°C, and the species is well established in Hokkaido, Japan, where winter temperatures can reach -12°C to -4° (Stockton et al., 2018; Walsh et al., 2011). In our lab, *D. suzukii* cultures were successfully stored over two months at 4°C without significant negative effect on the stock. The optimal temperature for *D. suzukii* is between 22°C – 28°C for egg-lay and development time, while survival decreases significantly at temperatures above 30°C (Ryan et al., 2016).

Until the 1980s, *D. suzukii* was restricted to its native range of SE Asia. The species was first described in Japan in 1931, though it is likely that the species originally came from China (Asplen et al., 2015; Walsh et al., 2011). In 1983, it was first reported on the Hawaiian Islands, where it has established itself as a crop pest (Figure 2.2). The current global invasion began in 2008, when the fly was simultaneously detected in North America and Europe. In North America, *D. suzukii* was first reported in 2008 in California, by 2009 it had been detected in all West Coast states and British Columbia, and by 2010 the fly had reached the East Coast. The invasion of North America was completed in 2011, when *D. suzukii* was detected in Mexico. In Europe, the invasion was first reported in Italy and Spain in 2008, though new data suggests that the pest could have been present in Southern France before this (Cini et al., 2014)(Asplen et al., 2015; Walsh et al., 2011). By 2012 it had reached Central and Northern Europe, including the UK. In 2013 it was detected in South America, and in 2016 in Turkey (Deprá et al., 2014; Orhan et al., 2016). Given ecological and environmental conditions, as well as global trade, there is a high potential that *D. suzukii* could also become established in Africa and Australia (dos Santos et al., 2017). *D. suzukii* has limited long-

range dispersal capacity, therefore the current invasion is almost certainly is due to human transportation and trade.

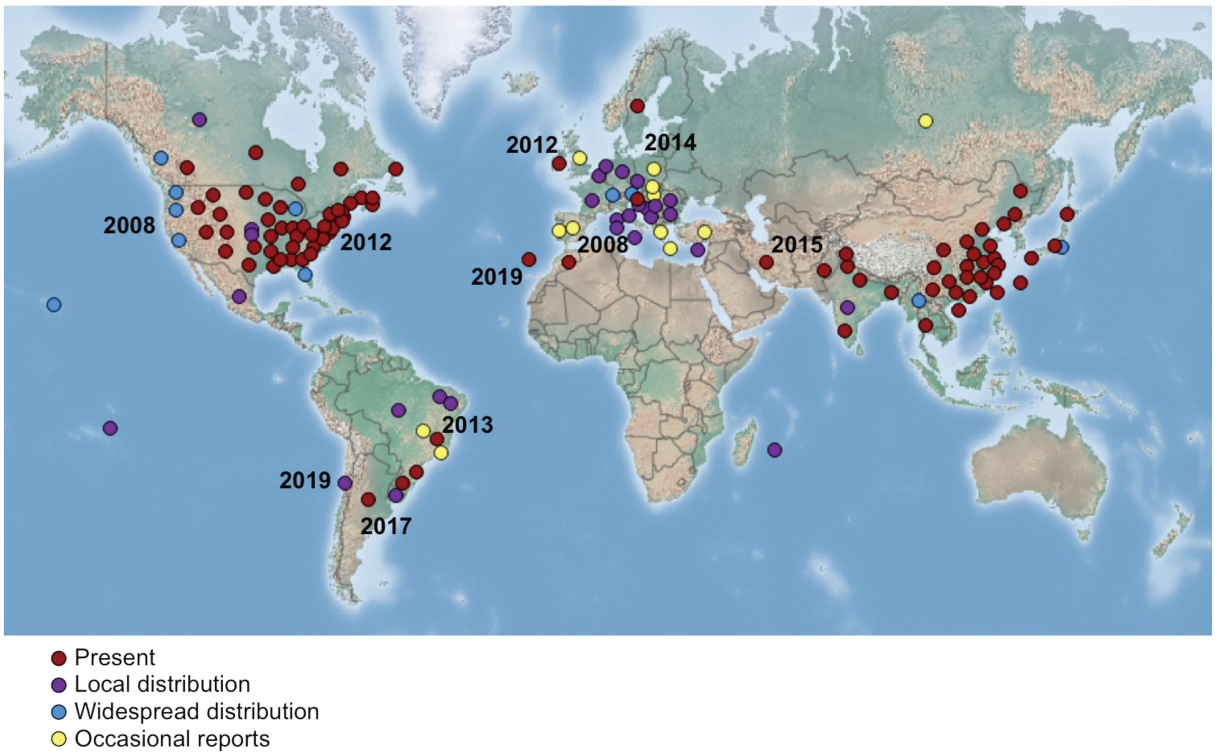


Figure 2.2: Global distribution of *D. suzukii* as of October 2019. *Drosophila suzukii* has been detected on all continents except Australia. Dates indicate when the pest was first detected. Map modified from the Center for Agricultural and Bioscience International (https://www.cabi.org/isc/datasheet/109283).

In the 11 years since *D. suzukii* began spreading across the globe, it has become one of the most significant agricultural pest species. To date, crop loss information and estimates of potential impact are limited, and are mostly based on US or European production. In various surveys of producers, reported loss due to *D. suzukii* ranges from 20–40% depending on the crop, but losses up to 80-100% have been reported. Estimates for economic losses are similarly variable, but often consequential. In one study from Trento Province in Italy, *D. suzukii* damage in 2011 lead to a loss of 3 million euros, while in 2017 raspberry farmers in Minnesota, USA, lost approximately 2.36 million USD (DiGiacomo et al., 2019; Ros et al.,

2012). In addition to direct revenue loss from crop damage, *D. suzukii* also increases producers' costs due to implementation of prevention and eradication strategies.

The particular life history of the pest limits the effectiveness of current control strategies. Once eggs are laid in the fruit, they are protected from most interventions. Chemicals are currently the primary control strategy, with broad-spectrum insecticides such as spinosyns, organophosphates, pyrethroids, and neonicotinoids being the most effective (Bruck et al., 2011; Van Timmeren & Isaacs, 2013). However, the close timing of infestation and harvesting imposes limits on pesticide application. Additionally, the current traps used for monitoring are not as effective as for other species, owing to the difficulty of identifying species-specific attractants. The difficulty of eradicating *D. suzukii* and its high potential for damage has lead many producers to increase or resume chemical applications. In US regions where the pest is established, many farmers currently begin spraying pesticides as soon as the fruit becomes attractive and apply the chemicals every 5-7 days throughout the growing season (Van Timmeren & Isaacs 2013). Such intensive insecticide can lead to the emergence of resistance. Resistance to spinosad was recently reported in California populations of *D. suzukii* and it is likely that other cases will be detected going forward (Gress & Zalom, 2019).

Beyond chemical applications, cultural control is currently the next most effective control strategy. Fine mesh netting can effectively prevent females from accessing the fruit (Leach et al., 2016). Increased crop sanitation such as clearing fallen fruit can also improve outcomes, however these strategies are not always economically viable. Many teams have focused on identifying potential predators for biological control. *D. suzukii* has a naturally strong constitutive hemocyte expression, limiting the ability of parasitoid wasps and other parasites to successfully develop within the fly (Kacsoh & Schlenke, 2012). So far, the most

successful potential bio-control agent identified is the parasitoid *Trichopria drosophilae*, which has performed well in both lab and field trials (Chabert et al., 2012; Rossi Stacconi et al., 2019).

Efforts are also underway to develop genetic-based strategies. Radiation limits for SIT have been determined; irradiation of pupae 24 hours before emergence with 75 Gy and 200 Gy resulted in sterility of females and males, respectively (Krüger et al., 2018). Additionally, sterile females showed reduced inclination to mate, meaning that released females are less likely to compete with wild females, while irradiated males were able to successfully mate with wild females. However, the rapid generation time and likely ongoing global invasion of the pest is a challenge for SIT.

Taking into account the limited control measures currently available, the recent timescale of its invasion, and its rapid generation time and limited dispersal capacity, *D. suzukii* is an excellent candidate for a gene drive control approach. Additionally, it is closely related to *D. melanogaster*, one of world's most comprehensively studied species. Finally, in 2013, the genome of *D. suzukii* was published, providing an important tool for genetic studies (Chiu et al., 2013). Analysis of the likely-protein coding genes indicated high levels of conservation between *D. suzukii* and *D. melanogaster*. As *D. suzukii* has only recently become established in the majority of its current range, it is unlikely that removal of this pest species will adversely harm local ecosystems, whereas the continued invasion poses both economic and environmental threats. Therefore, the goal of this part of my PhD was to develop a suppression drive targeting female fertility genes to reduce *D. suzukii* populations.

Results

Establishment of laboratory populations of D. suzukii

Prior to my arrival in the laboratory, a laboratory population of *D. suzukii* was established using wild-caught insects from Nancy and Strasbourg (Illkirch-Graffenstaden) France. Initially, the flies were reared on apple agar supplemented with yeast, however I was able to shift them to a standard simple cornmeal diet used in our unit for *D. melanogaster*.

In laboratory populations of *D. melanogaster*, wandering thirds migrate out of the food to pupate on the sides of fly vials. *D. suzukii* however pupates within the fruit it infests; in laboratory populations, this means that pupae remain in the food, where the humidity and action of younger larvae can submerge the pupae, preventing the adults from emerging. For this reason, I supplemented the food vials with a strip of Whatman paper, which serves as a support for flies and pupae and also absorbs extra humidity.

Selection of candidate genes and gene drive design

A former student, Thuy Tuyen Tran, compiled a list of candidate genes based on homology to essential female fertility genes in *D. melanogaster*. Three genes were ultimately selected as target genes: *cup, stand-still,* and *yolkless*. In *D. melanogaster, cup* is involved in the localization of mRNA transcripts within the oocyte and early embryo. *cup* null females produce abnormal eggs with an open-ended, cup shape, and cannot be fertilized. *stand still* (*stil*) controls the transcription of *out* and is essential for germline development. *stil* mutants have underdeveloped ovaries and are completely sterile. *yolkless* (*yl*) codes for a membrane receptor that is involved in vitellogenin uptake in late oocytes. *yl* null females produce eggs that lack a yolk and ultimately collapse. A summary of the selected targets, their putative *D. suzukii* orthologue, and the percent amino acid conservation is given in Table 2.1.

Table 2.1: Putative female essential genes used in this study

D. melanogaster gene	D. suzukii orthologue	% aa identity	% similarity
cup (FBgn0000392)	DS10_00001422	75%	83%
stand still (FBgn0003527)	DS10_00006669*	69%	76%
***Yolkless* (FBgn0004649)**	DS10_00007476	84%	91%

*In the current *D. suzukii* genome annotation (SpottedWingFlyBase, V1 assembly, October 2019), the putative *stil* CDS is fused to *ClC-b*

In *D. melanogaster*, *cup* and *stil* are located on the 2L and 2R chromosome arms, respectively, while *yolkless* is located on the X chromosome. While the current *D. suzukii* genome has not been assembled into chromosomes, given the high level of synteny between the species, it seems likely that the cytological location of these genes is conserved (Chiu et al., 2013). To avoid localizing a gene drive on the X chromosome, which would limit drive activity to females, and to maximize the effect of the gene drive on female fertility, E. Marois designed a gene drive to knock-out two putative fertility genes at the same time.

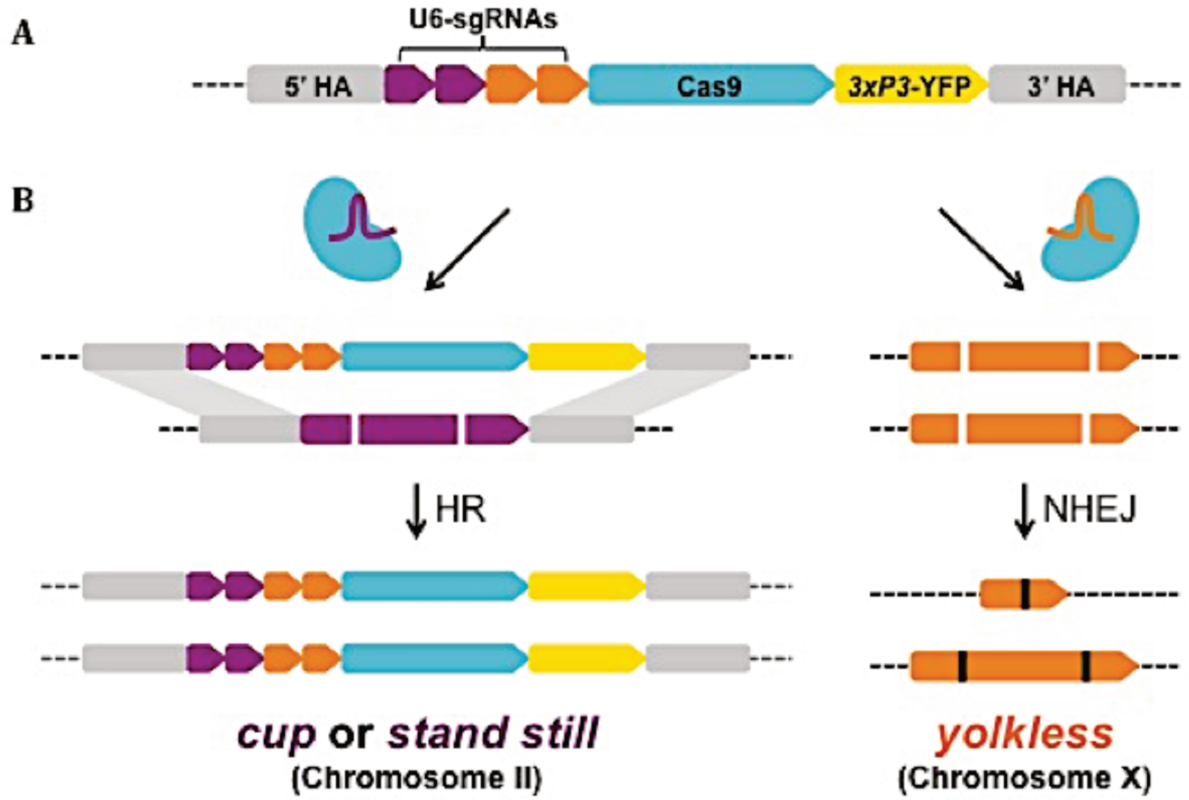


Figure 2.3: Suppression gene drive design for *D. suzukii*. **A)** The gene drive cassette comprises two ~2 kb 5' and 3' homology arms (HA), intended as a template for homologous recombination (HR). There are four individual *U6::*sgRNA expression genes, two targeting the homing locus (purple) and two targeting the deletion locus (orange). Cas9 is under the control of a germline promoter, either *nanos* or *β3-tubulin*. Finally, the *3xP3::*YFP marker identifies GD+ individuals. *3xP3* is a synthetic promoter that drives eye-specific expression in many insects. While the red pigment in *D. suzukii* eyes can obscure fluorescent expression, the promoter also drives expression in the ocelli (Schetelig & Handler, 2013). **B)** When Cas9 is expressed in the germline, the sgRNAs will direct the protein to cleave the WT homing locus and the deletion target. Repair of the homing target by HR copies the drive and disrupts the expression of the target gene, while NHEJ repair of the deletion target will result in either a large deletion or two indel mutations. Repair of the homing target by NHEJ will also likely disrupt gene function, thereby avoiding the emergence of functional resistant alleles, while repair of the deletion target by HR (possible only if the other chromosome is intact) will reconstitute the WT locus, which can then be cleaved again by Cas9 until mutated.

The proposed mode of action of the gene drive can be found in Figure 2.3. The gene drive would be inserted into one of the autosomal target loci (*cup* or *still*) and comprises four tandem sgRNAs each under the control of a putative *U6* promoter from *D. suzukii*, human-optimized Cas9 (Chiu et al., 2013) under the control of either the putative *D. suzukii* promoters *nanos* or *β3-tubulin*, and a *3xP3*-YFP marker. In *D. melanogaster*, the *nanos* promoter is used to drive expression of proteins, such as Cas9, in the female germline and embryo, while *β3-tubulin* is known to be expressed in both the male and female germline (Kimble et al., 1990; Port et al., 2014). Two of the sgRNAs target the homing target (*cup* or *stil*), while the other two target *yl*. Homing of the gene drive into its target locus disrupts the expression of either *cup* or *stil*. NHEJ of the DSB in *yl* will result in either one large deletion or smaller indel mutations, likely disrupting gene function. NHEJ repair of the homing targets will have the same outcome, meaning that mutations are likely to be non-functional, therefore guarding against the formation of functional resistant alleles.

The final gene drive cassettes were assembled via Golden Gate cloning. Three gene drive cassettes were prepared: p605 (pENTR[*U6*-sgRNAx4(*cup*),*tub85*-Cas9-sv40]); p606 (pENTR[*U6*-sgRNAx4(*cup*),*nos*-Cas9-*nos3'UTR*]); and p607 (pENTR[*U6*-sgRNAx4(*still*), *nos*-Cas9-*nos3'UTR*]).

Establishment of microinjection protocol

Microinjection of insect eggs begins with the collection and alignment of the eggs. When working with *D. melanogaster*, females are provided with an apple agar plate on which they lay and eggs are removed from the surface using a paintbrush. However, as described in the introduction, *D. suzukii* prefers to insert its eggs into the media. I tested several different egg collection strategies that would allow me to rapidly and easily collect the maximum number of eggs. The strategy I settled on is shown in Figure 2.4. Using this strategy, females are

provided with an attractive egg laying surface, but prevented from inserting their eggs by the Whatman paper. The eggs can then be removed from the paper using a paintbrush. Depending on the number of flies in the collection cage and the time of day, I was able to recover between 30-70 eggs in a 30 min period per collection chamber using this strategy.

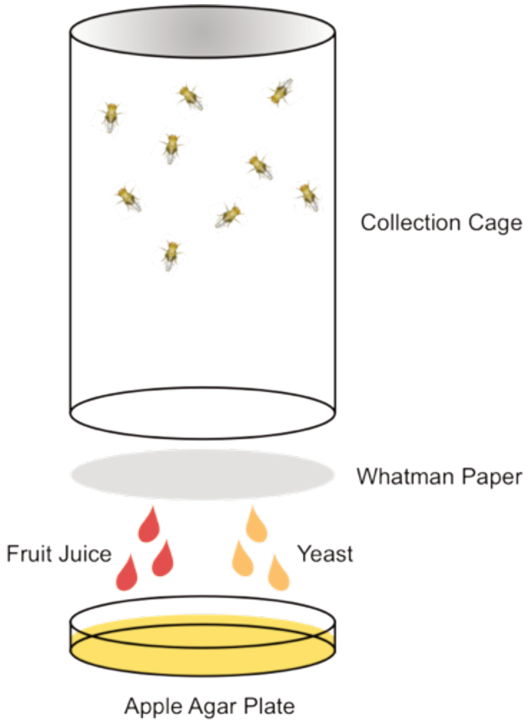


Figure 2.4: *D. suzukii* egg collection strategy. Adult male and female flies are transferred to the collection cage and females are supplied an egg laying support comprising an apple agar plate coated with fruit juice (cherry or strawberry) and yeast, which is covered with a circular Whatman paper. The juice and yeast serve to attract females to the plate, and lay their eggs in the paper. Eggs can be removed using a moistened paintbrush. The apple agar serves as a support for the paper and also keeps it humid.

Once the eggs have been collected and rinsed in water, they are aligned on a coverslip for injection. Initially, I followed a microinjection protocol for *D. melanogaster* used by our colleagues at the IBMC. Eggs are first aligned on a slice of agar and then a coverslip with a strip of semi-dry glue is pressed into the eggs. The eggs are covered with mineral oil prior to injection. Once injected, the oil is briefly drained and the coverslips are placed on apple agar plates. *D. suzukii* eggs are not dechorionated prior to injection.

While this microinjection strategy did allow me to successfully inject *D. suzukii* eggs, it was time consuming. Additionally, the viscosity of the mineral oil made it difficult for larvae to migrate off the coverslip. I eventually switched to using a protocol designed by N. Gompel's laboratory (Gompel & Schröde, 2015). In this protocol, eggs are aligned directly on the coverslip and briefly allowed to dry to the point at which they stick to the slide. They are then covered with olive oil, which can be removed after injection by briefly rinsing the coverslip in 95% ethanol, then water. The coverslip is then pushed into a food vial until the eggs touch the food. This protocol is much faster than the previous one and avoids larval mortality due to mineral oil.

Direct injection of gene drives

Initially, each gene drive plasmid was injected individually, at a concentration of 450 ng/μL. The GD plasmid served as a source of both Cas9/sgRNAs and as a DSB repair template. In total, 1987 eggs were injected with a survival rate of approximately 13% (Table 2.2). Surviving adults were outcrossed in batches of 2-4 to WT flies. I screened over 2000 G1 adult flies for each GD construct without recovering a positive transgenic.

Table 2.2: First injection of gene drive plasmids

Plasmid	Embryos Injected	Surviving Adults	G1 Screened	GD^+
p605	565	61	2193	0
p606	714	97	2257	0
p607	708	101	2401	0

I repeated the microinjections, this time including purified recombinant 240 ng/μL Cas9 protein in complex with synthetic sgRNAs targeting either *cup* or *stil*, and 1 μM Scr7, which was described as an inhibitor of *ligase IV* and therefore of NHEJ (Srivastava et al., 2012). Initial tests indicated that Scr7 could improve CRISPR-Cas9 knock-in efficiencies in human

cells (Chu *et al*., 2015)(V. T. Chu et al., 2015b). However more recent studies have found inconsistent effects, and highlighted discrepancies between the original published structure and some of the current commercially available products (Greco et al., 2016; Riesenberg & Maricic, 2018). In my hands, Scr7 did not have a detectable effect on Cas9 efficiency. In total, 1772 eggs were injected with a survival rate of approximately 15% (Table 2.3). Surviving adults were again outcrossed to WT and I screened over 1500 G1 adult flies for each GD construct without a positive transgenic.

Table 2.3: Second injection of gene drive plasmids with Cas9 protein and Scr7

Plasmid	Embryos Injected	Surviving Adults	G1 Screened	GD^+
p605	519	78	1708	0
p606	572	94	1822	0
p607	681	99	1836	0

G1 screen for large NHEJ deletions

To look for evidence of Cas9 activity, I screened a subset of the G1 adults from the second injection series by PCR for large deletions in the target region using primers flanking the Cas9 target sites within *cup* and *stil*. A WT copy of *cup* produces a 598 bp product and WT *stil* a 580 bp product. Precise deletion of the target region would produce products of 134 bp and 176 bp in *cup* and *stil* respectively. A total of 240 adult flies from each GD injection were screened. All PCR products were of the size predicted for the WT product, indicating that no large deletions had occurred at a detectable rate in these individuals (Figure 2.5). However, as the primers were designed to detect large deletions in the target region, it is not possible to rule out the presence of

small Cas9-induced indels within the target genes. Given the failure to detect any evidence of CRISPR/Cas9 activity, it is impossible to comment on the efficacy—or lack thereof—of Scr7.

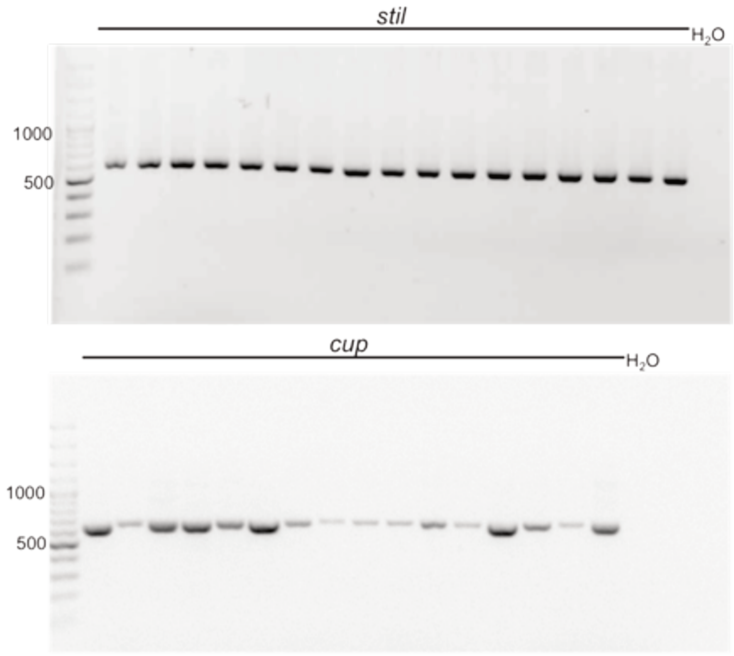


Figure 2.5: PCR amplification of *cup* and *stil* sgRNA target regions. Example of the results of the screen for Cas9-mediated deletions in *cup* or *stand-still*. No large deletions were detected in any of the individuals screened.

Design of transgenic tools cassettes

My failure to recover GD^+ transgenics from direct injection of the gene drive could have been due to the large size of the knock-in cassettes (~8.5 kb) or to a lack of activity of Cas9. Work with *D. melanogaster* has shown that gene editing using CRISPR can be improved by injecting into eggs that express Cas9 (Kondo & Ueda 2013; Port et al., 2014). Additionally, the efficiency of insertion of large cassettes can be improved using recombination-mediated cassette exchange (RMCE) (Zhang et al., 2014). The generation of transgenic gene drive *Anopheles* illustrates the difference between these two approaches. Gantz and colleagues screened 25,712 larvae to recover two GD^+ larvae, while Hammond and colleagues recovered 21 GD^+ larvae out of 10,000 using RMCE. (Gantz et al., 2015; Hammond et al., 2015)

Consequentially, I modified my knock-in strategy to use RMCE. The first step was to establish Cas9-expressing *D. suzukii* lines, which could be used to knock-in a small dual-*att*P cassette into the homing targets *cup* and *stil*. In a second step, I would use RMCE to integrate the full GD cassette (Figure 2.6). In addition to improving the efficiency of GD knock-in, this strategy would also allow me to verify the knock-out phenotype of the homing targets in *D. suzukii*.

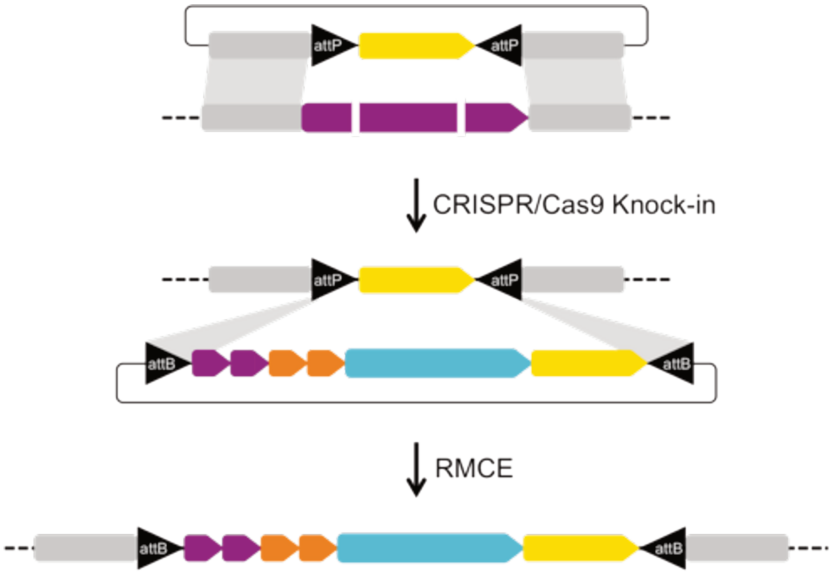


Figure 2.6: RMCE knock-in of CRISPR-GD. Schematic representation of knock-in of large cassettes using CRISPR/Cas9 and recombinase mediated cassette exchange. In a first step, CRISPR/Cas9 is used to knock-in a fluorescent marker flanked by reverse-complement *att*P sites. Once a dual-*att*P knock-in line is established and the integrity of the knock-in verified, Cre-recombinase catalyzes the exchange of the dual *att*P cassette for a dual *att*B cassette containing the full gene drive.

To establish transgenic lines in *D. suzukii*, I tested two transposon systems: piggyBac and P-element. I prepared three plasmids to test in *D. suzukii*: pENTR[*OpIE2*-GFP,*Hr5IE1*-DsRed] (piggyBac carrying two fluorescence markers), pENTR[*nos*-Cas9,*Ub*-DsRed] (piggyBac for Cas9 expression), and pENTR[*nos*-integrase,*Ub*-CFP] (P-element for ΦC31 integrase expression) (Figure 2.7A). Each plasmid contains the 5' and 3' piggyBac or P-Element terminal repeats, an *att*P site, and a transgene expression cassette flanked by *lox*P sites, which would allow the marked cassette to be excised, leaving behind an *att*P site that could be used

as a docking site. The piggyBac plasmids also include piggyBac transposase under the control of *hsp70* promoter from *D. melanogaster* in the backbone as a source of transposase. I also exchanged the eye-specific *3xP3* promoter for the strong ubiquitous promoters *OpIE2* and *Hr5IE1*, both from viruses, or *polyubiquitin (Ub)* from *D. suzukii*. pENTR[*OpIE2*-GFP,*Hr5IE1*-DsRed] was prepared to test the *OpIE2* and *Hr5IE1* promoters in addition to generating docking lines. The plasmids were co-injected with a final concentration of 100 ng/μL for each piggyBac plasmid, and 200 ng/μL for the P-Element plasmid along with 100 ng/μL of P-Element helper plasmid. To further improve my chances of finding a transgenic fly, I screened the surviving G0s for transient expression of the fluorescent markers (Figure 2.7B). Both CFP and GFP transient expression was visible at both larval and pupal stages, but RFP expression was never seen. This proved to be due to a mutation within the RFP coding sequence that was not detected before injection.

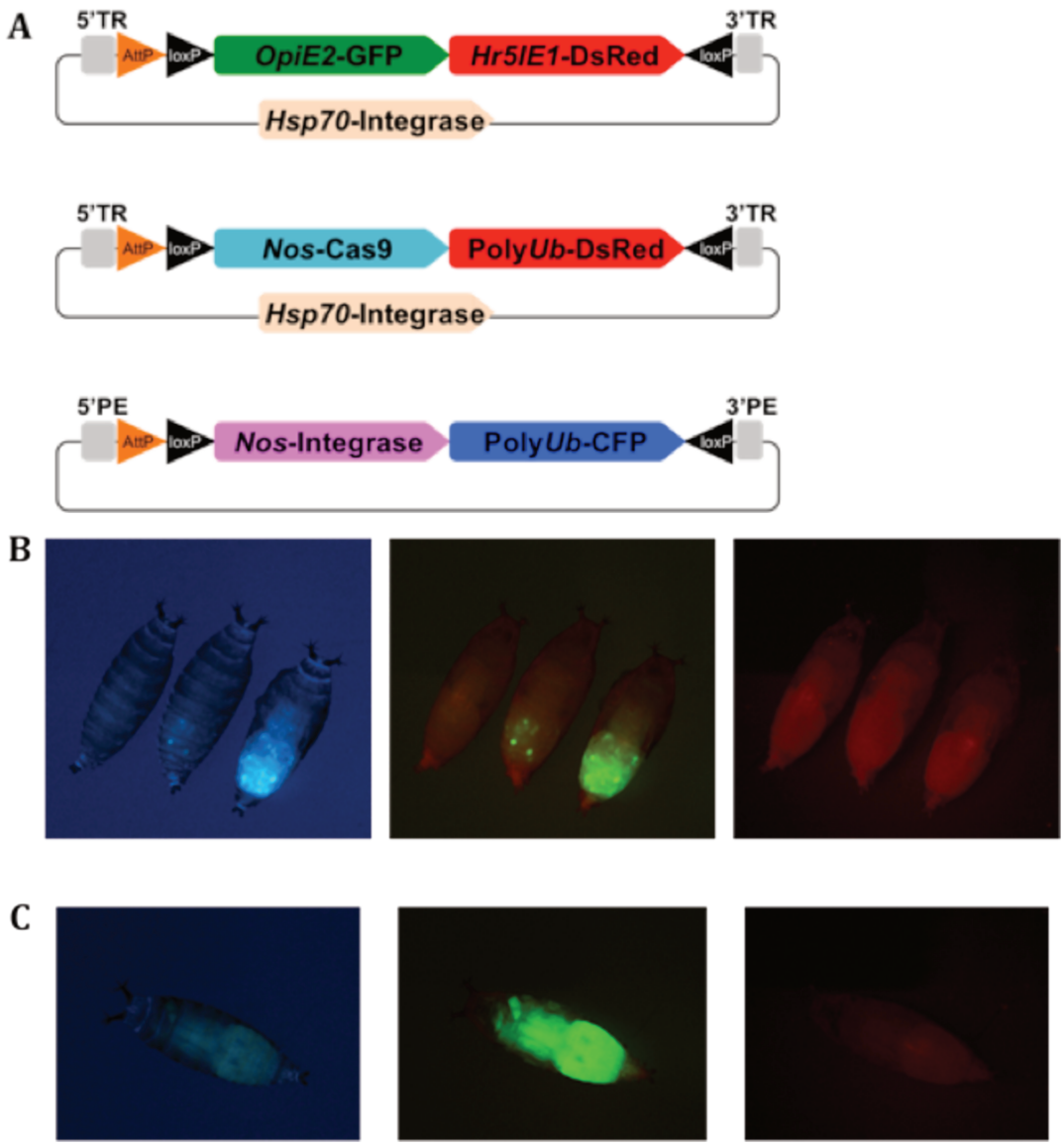


Figure 2.7: Transposon transgenesis in *D. suzukii*. A) Schematic representation of the three transposase plasmids tested in this project. Each plasmid has the same general layout: the cassettes are flanked by 5' and 3' terminal repeats (grey) from either piggyBac transposons or P-elements. The genes to be inserted are flanked by loxP sites, which can be excised to leave behind the *att*P site located directly after the 5' TR (orange). **B)** Examples of transient expression in G0 pupae. From left to right – no transient expression, moderate expression, and strong expression. Transient expression was never observed in the red channel. **C)** G1 transgenic pupae showing strong GFP expression. No fluorescence was ever observed in the blue or red channels.

After sorting the surviving G0 pupae by presence or absence of visible transient expression, I crossed transient-expressing pupae that in batches of 2-3 to WT flies, while non-transient expressing adults were self-crossed. From the first injection series (11/01-13/01), I found 2 GFP^+ G1s in a screen of 750 pupae, however as the pupae came from multiple crosses that were screened together on the same day, it was not possible to conclude if the transgenics occurred independently. For the second injection series (07/02), I screened each family individually. Out of 5 male crosses, one cross containing two G0 males gave two GFP^+ male G1s and out of 5 female crosses, one cross containing 4 G0 females produced one GFP^+ female G1.

Altogether, from these injections I recovered five transgenics that strongly expressed GFP, of which at least 4 represented different integration events (Figure 2.7C, Table 2.4). I never recovered any CFP-expressing transgenics from any of the crosses. Given the strong GFP and CFP transient expression observed in the G0s, the P-Element was either unable to integrate into *D. suzukii* or it was silenced in the G1s.

Table 2.4: Injection of transposon plasmids

Injection Date	Embryos Injected	Surviving Adults	G1 Screened	Transgenics
11/01/2017	86	27	~3000	2
12/01/2017	360	42		
13/01/2017	400	28		
07/02/2017	506	98	1315	3*

*2 confirmed independent lines

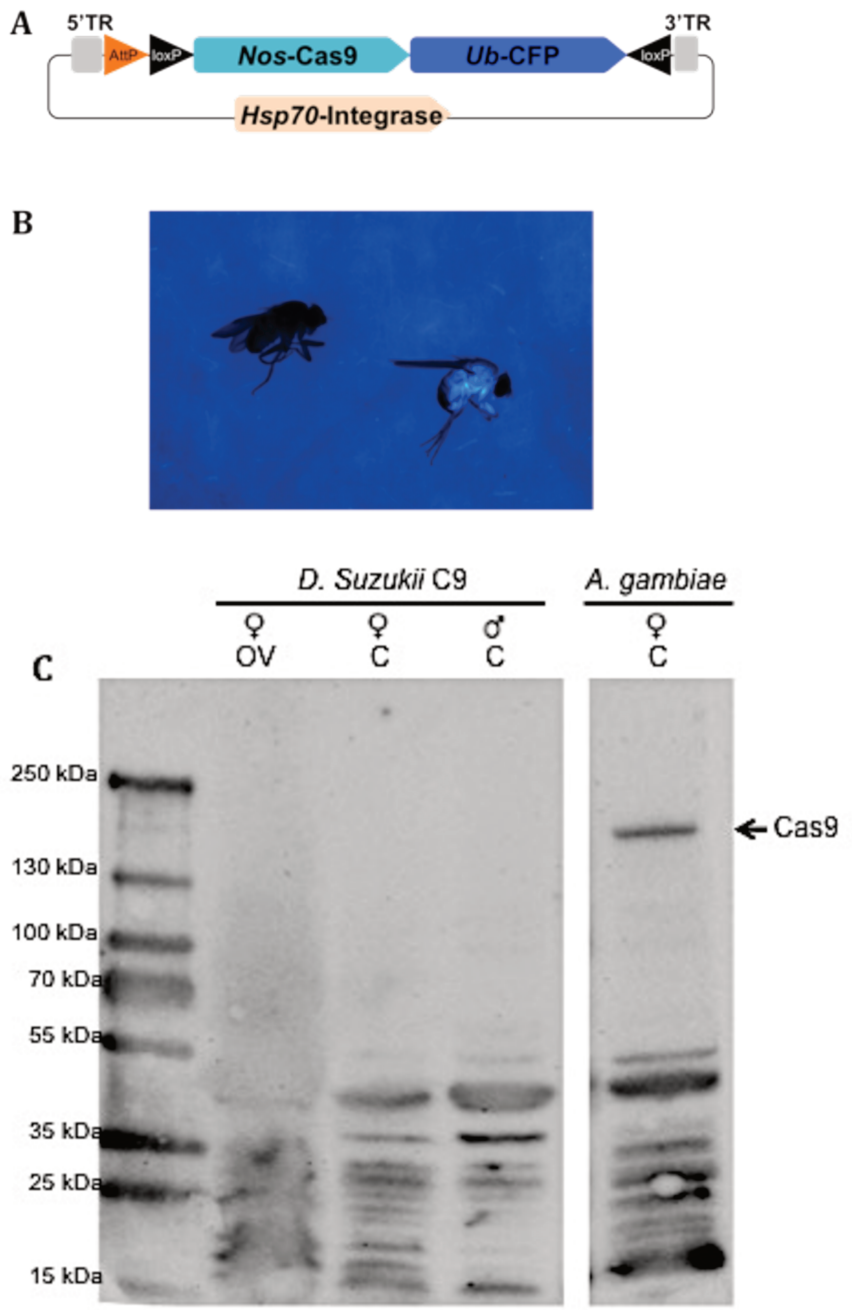


Figure 2.8: PiggyBac insertion of *nos*-Cas9 cassette. A) Schematic representation of the piggyBac plasmid pENTR-*nos*::Cas9-*Ub*::CFP used to try to create a Cas9 expressing line. **B)** WT male (left) and transgenic G1 male (right) expressing CFP. **C)** An example of western blot analysis of putative transgenic Cas9 *D. suzukii*. Cas9 is detected via an N-ter FLAG-tag. No evidence of Cas9 protein expression was seen in any of the transgenic lines. OV = ovaries, C = carcass

Design and injection of pENTR[nos-Cas9, Ub-CFP]

While I was not able to recover a Cas9 line from the first injections of transposons, I was able to determine that piggyBac transposons are able to integrate into *D. suzukii* and that both the *OpIE2* and *Ub* promoters could drive the expression of a fluorescent marker. With this in mind, I designed a new Cas9 piggyBac plasmid, containing *nos*-Cas9 and *Ub*-CFP (Figure 2.8A). I injected 394 eggs and recovered 157 adults, who were divided into 11 female crosses and 12 male crosses with 2-3 flies per tube. One female cross produced two CFP^+ males and one male cross produced one CFP^+ male (Figure 2.8B). The two CFP+ transgenics from the female cross showed X-linked inheritance and were recovered from the same cross, therefore were likely from the same G0 parent. Given this, overall I recovered two likely independent transgenic lines from 157 adults. I performed western blot analysis of CFP^+ F3 progeny using anti-Flag to detect Cas9, and using protein extracts from an established *An. gambiae* Cas9-expressing line as positive control (Figure 2.8C). None of the *D. suzukii* lines showed a band corresponding to Cas9.

After working on *D. suzukii* for over two years, it was disappointing not to obtain a *bona fide* Cas9 expressing line with which to advance the project. Development of transgenic tools in *D. suzukii* proceeded more slowly than anticipated. Additionally, at this time the lab moved into the new insectarium of the IBMC. This insectarium was designed for both mosquito and fruit fly research, however for the first year only the mosquito section was operational. The containment systems of this section were not sufficient for *D. suzukii*, as the primary exclusion nets covering the vents are not small enough to prevent a fly from escaping (though HEPA filters in the downstream circuit are designed to contain flies). Finally, the *Anopheles* project had reached a point that required full investment of time and energy. For these reasons, the *D. suzukii* project was placed on hold.

Discussion

Despite my failure to establish a gene drive system in *D. suzukii*, the work I performed on this insect provides a good baseline for future work with *D. suzukii* in our team. It also highlights several pitfalls that could be encountered when trying to establish gene drives in new species. During the two and a half years that I worked on this project, I established fly maintenance, egg collection, and transgenesis protocols in our laboratory. I also constructed a large library of plasmids that can serve as a basis for future *D. suzukii* genetic manipulation and successfully generated transgenic lines using the piggyBac system. Adoption of screening surviving G0s for transient reporter expression also allowed me to reduce the number of G1 pupae screened, therefore reducing the time and labor of G1 screening. While screening at the pupal stage does include a risk of missing small transient expression in the gonads, in four separate injections I never recovered transgenics from non-transient expressing G0s, indicating that this is a reasonable risk to take.

The first example of germline transformation in *D. suzukii* was published in 2013 using a piggyBac system (Schetelig & Handler, 2013). Compared to the transformation efficiencies reported their paper, my piggyBac injections yielded fewer transgenics. This was likely due to my lack of experience with microinjection, as the survival rates reported in their paper were higher than my survival rates. Improvements in piggyBac transgenesis in *D. suzukii* could also be made by using a stronger promoter to drive transposase (for example *Ub* from *D. suzukii*). Alternatively, a piggyBac transposase-expressing *D. suzukii* line was recently created using the *minos* transposon system (Chu et al., 2018). Injection of piggyBac plasmids into eggs expressing transposase improved transformation rates and also allowed the researchers to re-mobilize piggyBac plasmids.

Surprisingly, I was not able to detect any evidence of CRISPR/Cas9 activity in *D. suzukii*. As stated above, the large size of the gene drive cassette and the low numbers of G1s I was able to screen likely explains the failure to recover a transgenic knock-in. However it cannot be excluded that Cas9 expression from the injected plasmid was low or non-existent. The *nanos, β3-tubulin,* and *U6* promoters used in the designs were identified by searching the published *D. suzukii* genome database for putative orthologues to the *D. melanogaster* coding sequences, then amplifying likely promoter regions upstream of *D. suzukii* putative orthologues. Of note, transgenic *D. melanogaster* expressing sgRNAs from the cloned *D. suzukii U6* promoter successfully yielded mutant flies when crossed to Cas9-expressing flies (A. Acker, E. Marois, J.M. Reichhardt, unpublished), indicating that the *D. suzukii U6* promoter at least is functional. In the case of *nanos* and *β3-tubulin,* given the difficulty of identifying promoter regions, it is likely that the regions defined as "promoter" were incomplete. While I did screen for NHEJ activity in a subset of G1 individuals, the primers were designed to detect only large deletions. Depending on the individual activity of the sgRNAs, one sgRNA may have been more active than the other, favoring local repair of the break rather than large deletions.

So far, and rather surprisingly given the importance of the fruit pest, examples of CRISPR/Cas9 in *D. suzukii* have been limited. Injection of plasmids expressing Cas9 under the control of the *D. melanogaster vasa* promoter and sgRNAs expressed under the control of the *D. melanogaster* U6:3 promoter was used successfully to generate mutants in the *white* and *sex-lethal D. suzukii* orthologues (Li & Scott 2016). Knock-out mutations were also generated by injecting Cas9 protein and sgRNA again targeting *white* (Kalajdzic & Schetelig, 2017). Reported germline transmission for *white* mutations were 2.5% and 7.4-18.5% using plasmids and 4.5% using protein. CRISPR/Cas9 knock-in was also achieved by injecting

Cas9 protein and a repair template, as well as dsRNA against *lig4* (Li & Handler 2017). To date there is no published example of germline Cas9 expression in *D suzukii*.

RMCE has also been achieved in a follow up to the work of Schetelig and Handler (2013). Co-injection of a RMCE donor vector and *hsp70*-Cre helper plasmid led to complete exchange of cassettes flanked with heterospecific *lox* sites, with 20% germline transmission (Schetelig et al., 2019). Given the apparent ability of *D. melanogaster* promoters to drive expression in *D. suzukii*, a future gene drive could be constructed using previously validated *D. melanogaster* promoters, at least as a first step. Additionally, while *nanos* has been shown to better restrict GD activity in the germline compared to *vasa* in *D. melanogaster*, *nanos* was also found to be less efficient at driving homing in *Anopheles* compared to the *zpg* promoter (Champer et al., 2017; Hammond et al., 2018). If homing gene drives are pursued in *D. suzukii*, it would be worthwhile to test *zpg* and potentially other restricted germline promoters. To my knowledge, *β3-tubulin* has not yet been tested in a Cas9 homing gene drive.

Considering the labor involved in individually screening G1 flies, future efforts using CRISPR/Cas9 to target loci without a visible null phenotype should use strategies to identify those G0 most likely to yield transgenics. This could be achieved by including a fluorescent reporter and selecting for transient expressing-G0s and/or including a sgRNA targeting a visible marker such as *white*. When establishing knock-in lines to express Cas9 and other proteins for transgenesis, linking a fluorescent marker to the introduced gene via a self-cleaving 2A peptide could also speed up the screening process. Two 2A peptides were recently successfully tested in *D. suzukii* (Schwirz et al., 2019).

Because of the *D. suzukii* economic impact it is not surprising that research into new control strategies is currently underway. No homing gene drive has yet been published, but a MEDEA system was recently constructed using the same artificial construction used in *D. melanogaster* (Buchman et al., 2018). A piggyBac transposon was used to insert the MEDEA system comprising four miRNAs targeting the 5'UTR of *D. suzukii myd88* and a rescue copy of *myd88*. The miRNAs were expressed under the control of the putative *D. suzukii* female germline-specific *bicoid* promoter while the *myd88* rescue was expressed from the putative *D. suzukii* early-embryo promoter *bottleneck*. The system was able to invade a cage populations, however some variability in efficiency was seen when the system was tested in different genetic backgrounds. This was likely due to polymorphism at the miRNA target sites.

A temperature-sensitive sterile insect system was also recently developed (Li & Handler, 2017). In *D. melanogaster*, two missense point mutations in the *tra-2* gene produce a temperature-dependent sex-development phenotype. *D. melanogaster* develops normally bellow 20°C, but when raised at 26-29°C, females are masculinized and males are sterile. The same point mutations were introduced into the *D. suzukii tra-2* orthologue using CRISPR/Cas9. The mutants were non-viable when reared at 27-29°C, however when raised at lower temperatures then shifted to 29°C as pupae, XX individuals displayed an intersex phenotype, while XY individuals were sterile. XX pseudo-males were not interested in mating with WT females, but sterile XY males did successfully mate. This system is potentially an elegant basis for SIT as it eliminates the need to use radiation and separate sexes.

Finally, population control using *Wolbachia* has also recently been demonstrated in *D. suzukii*. *D. suzukii* is naturally infected by the *wSuz Wolbachia* strain (Cordaux et al., 2008). However, *wSuz* does not appear to impose any fitness costs. Notably, there is no evidence of cytoplasmic incompatibility (CI), which has been used in other species as a population reduction strategy (Cordaux et al., 2008). However, two strains from *D. simulans* were recently found to induce CI in *D. suzukii* and to suppress cage populations (Hamm et al., 2014). To be converted into an efficient pest control system, this approach would require developing a sex sorting system ensuring the release of only *Wolbachia*-infected males.

The history of genetic control of insects and the recent experience of CRISPR-GD in mosquitoes and *D. melanogaster* shows that, provided the optimal genetic components can be identified, genetic control strategies are viable approaches for insect pest control. Given the close relationship between *D. suzukii* and *D. melanogaster*, the barriers to development are much lower compared to other potential pests, where genetic targets and genome engineering techniques may be less developed. One GD system from *D. melanogaster* has already been successfully translated into *D. suzukii*. It is likely that homing CRISPR-GDs will also be developed. The strategy I attempted to develop in this part of my PhD project represents a potentially viable alternative strategy, though any future gene drive development should consider using molecular as well as barrier containment strategies, given the presence of *D. suzukii* in the Strasbourg area. Unfortunately, due to the challenges and logistical barriers discussed above, I was not able to able to carry through the project; however the tools and protocols are in place to continue at a later date.

Chapter III

Indirect gene drive in *An. gambiae*

Introduction

Mosquitoes belong to the insect order *Diptera*, family *Culicidae*, which is further sub-divided into three sub-families, *Anophelinae, Culicinae*, and *Toxorhynchitinae* (Clements, 1992). Currently, there are approximately 3500 described species of mosquitoes. Mosquitoes are found across the globe in all ecosystems, with the exception of permanently frozen areas, but are most present in tropical and sub-tropical regions. These insects split their lives between terrestrial and aquatic ecosystems. Eggs are laid on the surface of water and hatch into aquatic larvae that feed on algae and other organic material. The larvae undergo four molts, the final producing a pupa that does not feed, but remains at the water's surface as the adult tissues form. One to two days later an adult mosquito will emerge from the pupa. The total time to pupation depends on the species and on environmental conditions. As adults, mosquitoes feed on nectar, serving as pollinators. However female mosquitoes also require blood in order to produce eggs. Female mosquitoes bite a range of animals, including vertebrates like mammals, birds, reptiles, and even some fish, and some may also feed on invertebrates.

When a female mosquito takes a blood meal, her goal is to obtain proteins and other compounds required for egg development. Unfortunately, the saliva she uses to numb the bite area can contain pathogens—filarial worms, parasites, or viruses. Most mosquito species do not transmit diseases, but the handful that do have earned the mosquito the moniker "world's deadliest animal", because of the number of humans killed every year by mosquito-borne diseases. Among humans, members of three genera are capable of transmitting diseases;

Aedes, *Anopheles*, and *Culex*. Many of the diseases transmitted are viruses, such as dengue, yellow fever, West Nile fever, and Zika. However the disease that causes the most deaths by far is malaria, spread by *Anopheles* mosquitoes.

The Malaria-Mosquito relationship

Malaria is caused by unicellular protozoan parasites from the genus *Plasmodium*. There are over 120 *Plasmodium* species that can infect mammals, birds, or reptiles, four of which regularly infect humans: *P. falciparum, P. vivax, P. malariae,* and *P. ovale* (WHO, 2017). Humans may also become infected by species that primarily circulate in other primates, such as *P. knowlesi,* which primarily infects macaques, and *P. cynomolgi*, which is known to infect several species of macaques and the rhesus monkey (Ta et al., 2014; WHO, 2017). Though these cases are rare, reports of *P. knowlesi* infections in humans have become frequent enough to be a public health concern. Of all the *Plasmodium* species known to infect humans, *P. falciparum* and *P. vivax* are the most deadly. *P. falciparum* is considered the deadliest, due to its high prevalence in Sub-Saharan Africa, which bears the bulk of this disease burden. However, *P. vivax* is responsible for about half of the reported malaria cases outside of Africa and is the predominant *Plasmodium* species in the Americas. All *Plasmodium* parasites are transmitted by *Anopheles* mosquitoes, however each parasite has a limited host and vector range, meaning that a given parasite is only capable of infecting a limited number of vertebrate and insect species. Of over 500 currently described *Anopheles* species, only 30-40 are considered important malaria vectors (WHO 2017).

Plasmodium parasites have a complex life cycle that involves both sexual and asexual stages (Aly et al., 2009; Ashley et al., 2018). The malaria transmission cycle begins when a female *Anopheles* mosquito bites an infected human (Figure 3.1). Within the blood taken up by the female are gametocytes, the sexual stage of the parasite. Once inside the mosquito midgut,

the environmental change stimulates the gametocytes to develop. Male gametocytes undergo three cycles of replication to produce eight microgametes that fertilize female gametocytes, forming a zygote. The zygote develops into a mobile form called an ookinete, which crosses the peritrophic membrane and the midgut epithelium, settling in the basal side of the midgut to develop into a protected oocyst. Within the oocysts, the parasite undergoes many rounds of mitosis to form thousands of sporozoites, which are eventually released into the hemolymph. The sporozoites circulate within the hemolymph and preferentially bind to and invade the salivary glands. Once inside the sporozoites are ready to be transmitted to a new host the next time the female mosquito bites. The whole process takes approximately three weeks.

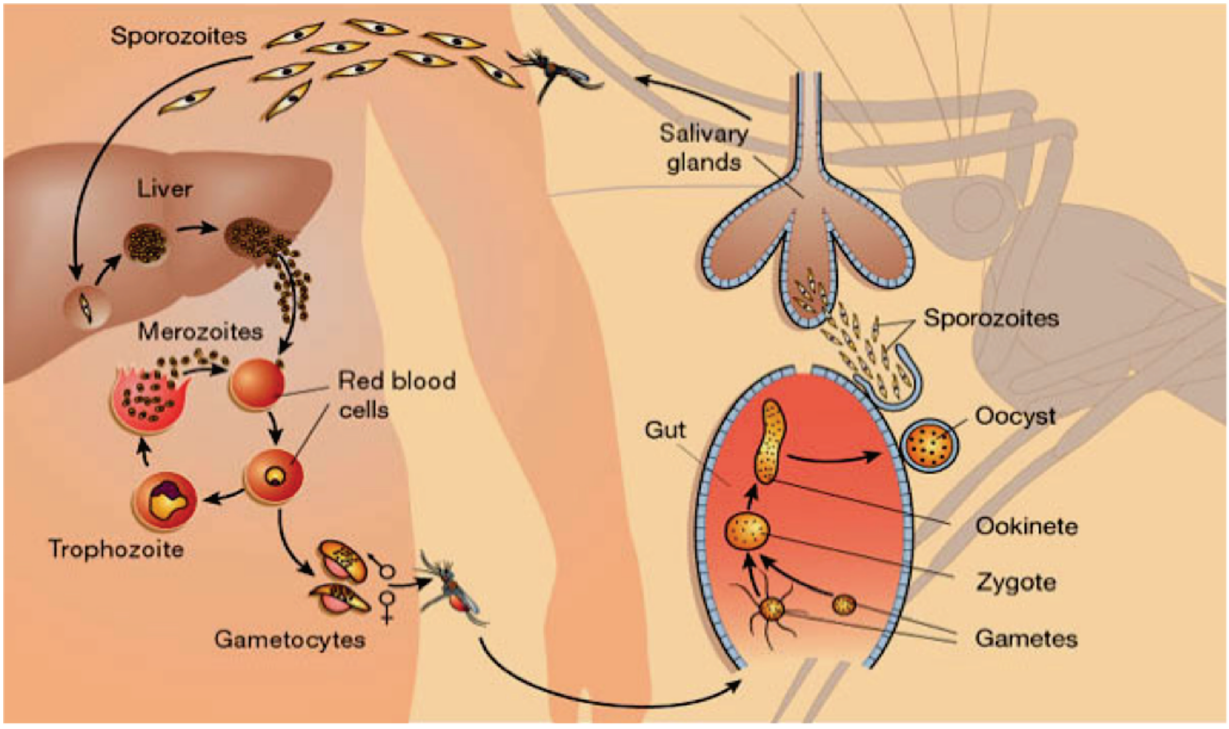


Figure 3.1: Malaria transmission cycle. The mosquito transmission cycle begins when female *Anopheles* ingest male and female *Plasmodium* gametocytes along with the blood meal (lower left). Once inside the insect, the parasites travel from the midgut through the hemocoel to the salivary glands, undergoing several developmental stages along the way. Once in the salivary glands, the sporozoite form of the parasite can be transmitted to the vertebrate host during the next blood meal, where it then infects first the liver, then the red blood cells. The blood stage infection is a cycle and is responsible for the majority of malarial symptoms. In response to environmental stimuli, some of the merozoites will develop into gametocytes, which when released into the blood stream can be taken up by a new *Anopheles* mosquito. Image from Wirth, 2002.

Once inside the human, the sporozoites migrate through the skin and blood vessels to the liver, where they invade the hepatocytes. *P. vivax, P. malariae,* and *P. ovale* can enter the dormant hypnozoite stage and remain present in the liver for months or years before becoming active again (Ashley et al., 2018). There they invade the hepatocytes and undergo asexual replication, producing merozoites that are released into the blood stream and invade erythrocytes, leading to the symptomatic stage of malaria in humans. Once inside the erythrocytes, the merozoites proceed through three developmental stages, rings, trophozoites, and schizonts, which produce more merozoites that burst from the erythrocyte and invade new cells, continuing the infection cycle. Some ring stage parasites will develop into gametocytes and enter the bloodstream, where they can be ingested by another mosquito, continuing the transmission cycle.

Malaria causes a variety of symptoms including acute recurrent fever, fatigue, head and muscle aches, nausea, and vomiting (Ashley et al., 2018). If untreated, malaria can progress to severe malaria, which includes anemia and potentially organ failure. Some forms of malaria can also infect the brain, causing cerebral malaria. Young children, pregnant women, and the elderly are particularly vulnerable to severe malaria, as are those with malnutrition or HIV. Malaria is also strongly linked to poverty; poorer countries are disproportionately burdened by malaria and within regions of disease transmission incidences of malaria are negatively correlated with income level (WHO 2018). Malaria imposes strong direct and indirect costs on the communities it affects – via increased healthcare costs, loss of economic production, lost time in school for children, and incalculable emotional and psychological trauma.

The history of malaria and the fight for eradication

The name malaria comes from the Italian "mal'aria" or "bad air", harkening back to the historical belief that the disease was caused by poisonous swamp air. The link between humid areas and malaria hints at the actual cause of the disease – female mosquitoes living in humid, damp environments. Malaria has likely afflicted humans since pre-historic times and there are written references to the disease going back to ancient China and Greece (Cox, 2010). However it was not until the 1880s that *Plasmodium* was identified as the causative agent and transmission linked to *Anopheles* mosquitoes.

Historically, the major achievements in malaria reduction have come from controlling mosquito populations. Land reclamation, drainage of breeding sites, and the improvements in sanitation that occurred as human society developed all contributed unintentionally to the decline of malaria. When the link between the disease and mosquitoes was discovered in 1897, the importance of controlling mosquito populations became clear. Initial efforts focused on killing the larval form by draining breeding sites, or applying oil or larvicides. Later, insecticides targeting the adult forms were developed. The first of these was DDT, a highly effective, long-lasting insecticide, which in combination with breeding site destruction allowed many western countries to successfully eradicate malaria in the 1950s–1970s.

In 1955, the WHO launched the Global Malaria Eradication Program, though the term global is a misnomer, as large parts of sub-Saharan Africa were not included in the campaign due to logistical challenges (Nájera et al., 2011). While the program fell short of the goal of global eradication, it did succeed in eliminating malaria from most of Europe, North America, the Caribbean, and parts of Asia and South and Central America. However the high reliance on DDT to fight mosquitoes and chloroquine to treat the disease in humans lead to the

appearance of resistance. This coupled with funding shortages and the negative environmental impacts of DDT lead to the programs suspension in 1969 (Nájera et al., 2011).

Despite this, individual countries continued eradication campaigns and in 1975 Europe was declared malaria free. In the years since, several new initiatives have been launched, and while malaria still remains a global challenge, until recently progress was steady. The current WHO Global Malaria Program has set the goal of reducing malaria cases and mortality by 90% and eliminating malaria from 35 countries by 2030 (WHO 2017).

In 2017, the WHO reported there were approximately 219 million cases of malaria and 435,000 deaths (WHO 2018). 3.4% of cases were caused by *P. vivax*, the rest were due to *P. falciparum*. 61% of deaths were children under the age of five. Africa continues to bear the highest burden of malaria; 92% of cases and 93% of deaths were in African countries in 2017. Alarmingly, progress in malaria control has stalled since 2015 and current forecasts indicate that the world is off course to meet the WHO's eradication objectives.

Current control strategies

Since the link between malaria and mosquitoes was discovered, malaria control has been based on both medical intervention and vector control. Recent global improvements are largely due to the development of artemisinin-based combination therapies (ACTs), on the medical side, and the adoption of insecticide treated bed nets and indoor residual spraying for vector control (Bhatt et al., 2015). ACTs involve the administration of two drugs, one a fast-acting artemisinin derivative, the other a longer acting drug with a distinct mode of action, thereby effectively treating the disease while limiting the appearance of resistance. This was the case until 2008, when resistance to artemisinin was reported in South East Asia (Noedl et

al., 2008). So far resistance has not spread outside of Asia, but its appearance in South East Asia has raised alarm bells for malaria control efforts.

For mosquito control, the most effective recent strategies have been the use of insecticide-treated bed nets and indoor residual spraying. Bednets serve the dual purpose of protecting humans and exposing mosquitoes to an insecticide. Indoor residual spraying involves applying long-lasting insecticides to areas where the female mosquitoes are likely to rest, thereby maximizing the effect of the application. Chemicals from five major insecticide classes are currently used for mosquito control: pyrethroids, organochlorides, organophosphates, carbamates, and neonicotinoids, however only pyrethroids are authorized for use in bednets (WHO, 2019). These two strategies have been effective when implemented correctly. Bednets in particular have been shown to have a significant effect on reducing malaria (Lengeler, 2004). However access to these interventions is not uniformly available. Additionally, resistance to commonly used insecticides has been detected in all major malaria vectors.

New and emerging strategies

Given the current limitations of effective control strategies, the emergence of resistance to both insecticides and medications, and the lack of newer tools coming to market, the world is off track to meet the WHO 2030 eradication objectives. Today there is a major push to develop new control strategies to continue the hard-won gains of the past few decades. New ACTs and drugs are being developed, and triple drug combinations are also being tested. Other new strategies include vaccine development, new insecticides, mass treatment with ivermectin (which can kill mosquitoes feeding on treated humans and animals), the sterile insect technique and RIDL strategies, and biocontrol agents such as modified bacteria (Barreaux et al., 2017; Hemingway et al., 2016; Mancini et al., 2016).

Gene drives are also a potentially valuable tool for malaria control. As described in chapter I, many of the first gene drive systems were tested in mosquitoes with the goal of vector control. Now in the CRISPR-GD age, both suppression and modification drives have been developed in two *Anopheles* species, and suppression systems are quickly advancing towards field trials (Gantz et al., 2015; Hammond et al., 2015; Ndiaga 2019). However, as many malaria vectors are endogenous species, modification drives could play a beneficial role in reducing disease transmission while avoiding potential negative environmental effects of eliminating a local species. Introduction of modified, pathogen-resistant strains into areas where malaria has already been reduced could provide a cost-effective way to prevent re-emergence from wild strains. Modification and suppression drives also provide a way to intervene in regions where traditional strategies are difficult.

In order to be transmitted to a new vertebrate host, the *Plasmodium* parasite must successfully travel from the interior of the mosquito midgut to the interior of the salivary glands. This requires the parasite to successfully identify and traverse the three tissue barriers – the peritrophic matrix, the midgut epithelium, and the salivary gland epithelium – at the same time avoiding destruction by the mosquito's immune system. While *Plasmodium* parasites do not cause a debilitating infection in the mosquito, they do represent a foreign body that causes damage to the mosquito (tissue damage, nutrient hijacking). The journey of the parasite involves numerous and complex interactions between parasite factors and mosquito host factors, and modulation or disruption of these interactions can influence the vectorial capacity of the mosquito.

Once researchers began to understand these interactions, they started searching for ways to exploit this knowledge to block transmission. One strategy is to modulate the overall mosquito immune system by either overexpressing *Plasmodium* antagonists or knocking out/down *Plasmodium* agonists. Though our understanding of the mosquito immune response to *Plasmodium* is far from complete, several genes and pathways have been identified to play a role. Extensive work in *D. melanogaster* has identified and characterized the immune pathways Toll, IMD, and JAK-STAT. Canonically, Toll and IMD respond to bacterial or fungal infection via the production of anti-microbial peptides (AMPs), while JAK-STAT plays a role in antiviral defense, though these distinct divisions are not as strict as is often described (Smith et al., 2014). Activation of these pathways involves signal cascades that lead to the expression of AMPs.

In *Anopheles*, all of these pathways have been implicated in the response to *Plasmodium*. Up-regulation of the Toll and IMD *Anopheles* transcription factors Rel1 and Rel2 can be achieved by silencing their negative regulators, respectively *cactus* and *caspar*. In both cases, parasite infection is reduced, but the effects are species specific: *P. berghei* infection is inhibited by Rel1 overexpression, while Rel2 activity appears more specific for *P. falciparum* (Frolet et al., 2006; Garver et al., 2009). Similarly, silencing of the STAT suppressor SOCS reduces infection of both *P. berghei* and *P. falciparum* in *Anopheles* (Gupta et al., 2009). Beyond the general involvement of these pathways, specific important anti-*Plasmodium* factors have been identified including Thioester-containing Protein 1 (TEP1) and fibrinogen domain-containing immunolectin 9 (FBN9). TEP1 is involved in the early-phase response to *Plasmodium* infection, specifically binding to ookinetes, leading to their destruction (Blandin et al., 2004). FBN9 is a member of the fibrinogen-related protein family, the largest pattern recognition receptor in *An. gambiae*, and has been implicated in control of *Plasmodium*, though the

precise mechanism is not known (Dong & Dimopoulos, 2009). RNAi knockdown of both TEP1 and FBN9 leads to an increase of both *P. berghei* and *P. falciparum* oocysts.

Identification of anti-*Plasmodium* immune factors and pathways led naturally to efforts to engineer resistant strains via over-expression of key immune genes, with mixed results. Direct overexpression of TEP1 had little effect on infection, while overexpression of FBN9 reduced infection of *P. berghei* but not *P. falciparum* (Simões et al., 2017; Volohonsky et al., 2017). Overexpression of Rel2 under the control of the *carboxypeptidase* or *vitellogenin* promoters significantly reduced *P. falciparum* infection levels, via general up regulation of immune genes including TEP1 (Dong et al., 2011). Subsequent analysis of these lines showed that the overall fitness of the mosquitoes was not impaired, though in the case of lines expressing Rel2 under the control of the midgut *carboxypeptidase* promoter, the mosquito microbiota was altered (Pike et al., 2017). Curiously, this distortion in the microbiota affected the mating preference of the transgenic males, who preferentially mated with wild type females, leading the transgene to increase in frequency over time. This effect appears to be linked specifically to the microbiota, as both treatment of the transgenic mosquitoes with antibiotics and addition of bacteria during rearing and sugar feeding reduced the strength of mating preference. It remains to be seen how these transgenic lines perform against other *Plasmodium* species and strains and how these mosquitoes would preform in field conditions, where a more complex environment could result in a fitness cost due to perturbed gut microbiota.

The IMD and Toll pathways kill pathogens via the expression of AMPs. Instead of modulating the overall mosquito immune system, resistant mosquitoes can be generated via overexpression of endogenous AMPs or introduction of exogenous toxins. AMPs are small,

usually positively charged 15-45 aa proteins, which are thought to function primarily by disrupting the cell membrane, though disruption of protein/RNA synthesis and the cell wall have also been described (Brogden, 2005). In the mosquito, three types of endogenous AMPs have been identified: defensins, cecropins, and gambicin. Expression of each of these has been documented in response to *Plasmodium* infection, though there are variations between mosquito and *Plasmodium* species. Upregulation of defensin in response to *P. berghei* was reported in *An. gambiae*, however an RNAi knock-down failed to detect an effect on infection (Blandin et al., 2002; Richman et al., 1997). Defensin has been reported to be effective at controlling P. *gallinaceum* in *Ae. aegypti* (Shin et al., 2003). Gambicin was shown to be active against *P. berghei in vitro* and silencing of gambicin increased *P. berghei* oocyst levels *in vivo*, but had no effect on *P. falciparum* (Dong et al., 2006). The most direct effect of an endogenous AMP on *Plasmodium* in *Anopheles* was demonstrated by ectopic expression of cecropin A in the midgut, which reduced *P. berghei* infection (Kim et al., 2004).

Exogenous AMPs with anti-*Plasmodium* properties have also been identified. Gomesin, Magainin, and Scorpine isolated from frogs, tarantulas, and scorpions, respectively, have been shown in different assays to inhibit *Plasmodium* development (Conde et al., 2000; Gwadz et al., 1989; Moreira et al., 2007). Finally, natural toxins such as Phospholipase A2 (PLA2) and melittin from bee venom, and synthetic toxins Vida3 and Shiva1 are also effective against *Plasmodium in vitro* (Meredith et al., 2011; Moreira et al., 2002; Rodriguez et al., 1995). Confirmation of many of these molecules' antiparasitic activity by transgenic expression in mosquitoes is still largly lacking. So far, transgenic mosquito lines expressing PLA2 and Vida3 have been engineered and successfully reduced *Plasmodium* development by 85-87% (Meredith et al., 2011; Moreira et al., 2002). In the case of melittin and Vida3, unpublished

work using transgenic mosquitoes has failed to find any effect *in vivo* against *P. berghei* (E. Marois, pers. comm.)

The final transgenic transmission-blocking strategy is to block parasite or mosquito residues that are important for parasite development within and traversal of the mosquito. Parasite recognition of mosquito tissues is mediated by receptor ligand recognition and disruption of these interactions by either knocking-out mosquito ligands or blocking parasite binding can reduce transmission. Known pro-*Plasmodium* mosquito genes include *Frep1, CSP-BP,* and *PRS1* (Dong et al., 2018; Ghosh et al., 2009; J. Wang et al., 2013). Frep1 is involved in midgut invasion via anchoring of the parasite to the peritrophic matrix. A CRISPR/Cas9 knockout of *Frep1* was viable, and resulted in reduced oocysts and sporozoite levels, but exhibited reduced developmental time, lifespan, fertility, and blood-feeding (Dong et al., 2018). CSP-BP and Saglin are salivary gland proteins that are ligands for, respectively, sporozoite surface proteins CSP and TRAP (Ghosh et al., 2009; Wang et al., 2013). In the case of *saglin*, recent work has challenged its role in *Plasmodium* infection (O'Brochta et al., 2019). However unpublished results from our group does show that sporozoite load is reduced in the salivary glands of *saglin* loss-of-function mosquitoes (E. Marois, pers. comm.).

A key tool for disrupting mosquito/parasite interactions is the use of monoclonal antibodies (mAb) and derived single-chain variable fragment antibodies (scFv) (Figure 3.2). The discovery that antibodies could block transmission came in the 1970s and grew out of efforts to develop transmission-blocking vaccines against gametocytes. Researchers discovered that mAbs raised against gametocytes or midgut tissue could block transmission when fed to mosquitoes along with infectious blood (Yoshida et al., 1999). Full-length mAbs are too large to be expressed transgenically, however smaller synthetic fusions of the epitope binding

domain, scFvs, can be engineered for expression in mosquitos (Figure 3.2) (Raag & Whitlow, 1995). Yoshida and colleagues (1999) were the first to show that a scFv could recapitulate the transmission-blocking capacities of a mAb using a feeding assay, engineering a single-chain antibody against the *P. berghei* ookinete surface protein Pbs21. Other parasite protein targets of scFvs include chitinase, which is required for parasite traversal of the peritrophic matrix, the ookinete surface protein Pfs25, and the sporozoite surface protein CSP (Isaacs et al., 2011). The rare non-scFv example of an exogenous receptor-blocking factor is the SM1 peptide (Ghosh et al., 2001). SM1 was the first blocking factor to be engineered for transgenic expression in an *Anopheles* mosquito, reducing *P. berghei* oocysts levels by in *An. stephensi* (Ito et al., 2002). It also led to the identification of Saglin as the SM1 target and as a putative receptor for the *Plasmodium* sporozoite surface protein TRAP (Ghosh et al., 2009).

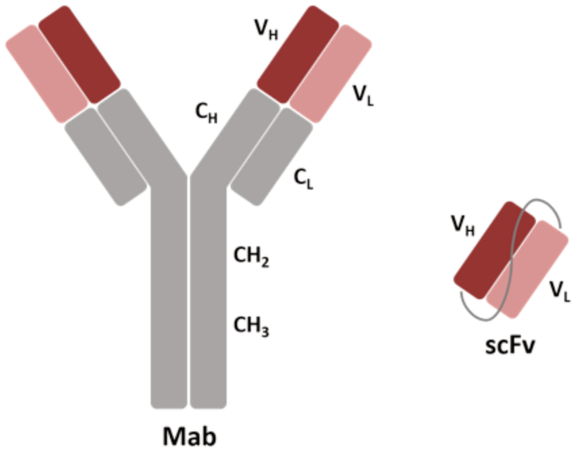


Figure 3.2: mAbs and scFvs. Single chain variable fragment antibodies (scFv) are synthetic peptides generated by linking the light- and heavy- variable regions of a full antibody (mAb) with a short polypeptide linker (10-25 aa). scFvs are the smallest minimal structure which retains the full epitope binding capacity of the full-length mAb. Their small size allows them to be expressed transgenically.

While all of the above strategies have been effective at reducing transmission, no strategy has achieved complete blockage. For this reason, many engineered refractory transgenic lines involve the expression of two or more synergistic transmission blocking factors. Targeting multiple parasite-mosquito factors has the added benefit of reducing the risk of resistance arising in the parasite, analogous to combination therapies. For example, expression of both

defensin and cecropin in *Ae. aegypti* rendered the mosquitoes completely resistant to *P. gallinaceum* (Kokoza et al., 2010). Another approach is to express several scFvs to target the parasite at different stages. Transgenic expression of two distinct scFvs targeting chitinase and CSP in *An. stephensi* completely blocked *P. falciparum* transmission (Isaacs et al., 2012).

Researchers continue to search for additional genes involved in *Plasmodium* transmission and new transmission-blocking factors. However, the diverse panel of factors already identified is primed to be exploited for new vector control strategies. As stated in Chapter I, the goal of this part of my PhD project was to develop a modification drive strategy to block *P. falciparum* transmission by *An. gambiae*. Over the course of my PhD, I worked on different strategies, one of which, indirect gene drive, will be presented in full here. The other strategies will be presented briefly at the end.

Indirect gene drive

As described in Chapter I, CRISPR-GDs will rapidly invade a population, provided they do not encounter or produce functional resistant mutations. In the case of suppression drives targeting essential genes, careful selection of sgRNA target sits and/or increasing the number of guides can overcome this challenge. However, modification drives designed to introduce novel genes are not constrained to a specific locus. The first modification drive published in mosquitoes was inserted into the kh^{w} locus in order to generate a visible marker (Gantz et al., 2015). While useful in laboratory settings, such mutations may not be desirable when developing gene drives for release. The modification drives could be inserted into neutral loci. Thanks to docking lines generated by random insertions of transposons, loci with no significant fitness costs in laboratory settings have been identified and could serve as sites for GD cassettes. However the absence of a fitness cost means that NHEJ mutations are less likely to be lost from the population.

One suggested strategy is to link the modification drive to an essential gene. The gene drive would home into an essential gene and carry along a rescue copy with mutated sgRNA sites. This strategy was successfully demonstrated in yeast using a homing gene drive (DiCarlo et al., 2015). The gene drive "cleave and rescue" system recently engineered in *D. melanogaster* also involves targeting an essential gene with Cas9 and rescuing mutants with a recoded gene copy, though in this case the system is not a homing gene drive and targets an essential gene *in trans* (Oberhofer et al., 2019).

The strategy proposed in this project, indirect gene drive, separates the effector and driving components of a classic modification drive, and involves the release of two transgenic lines (Figure 3.3). One is a suppression CRISPR-GD targeting an essential gene, while the other contains the effector gene linked to the recoded rescue copy. The gene drive will remove wild-type copies of the essential gene, favoring the spread of the effector-rescue construct. This strategy exploits the propensity of suppression GDs to stimulate the spread of functional GD resistant alleles and has several advantages compared to an "all-in-one" modification drive. First, any fitness costs of the gene drive (for example off-targets) or drive instability will not affect the effector gene. Secondly, the driving components will not remain indefinitely in the population, as they are also a loss-of-function allele of the essential gene. Thirdly, the rescue gene will not be able to serve as a repair template for DSB repair, which could otherwise lead to the formation of a functional resistant allele. Finally, the effector line can be assessed in the laboratory and field prior to the release of the full gene drive system. This strategy was briefly mentioned by Burt in his 2003 paper proposing HEG-gene drives, and was more explicitly described and modeled by Beaghton and colleagues (2017). However it has yet to be engineered and tested in a population.

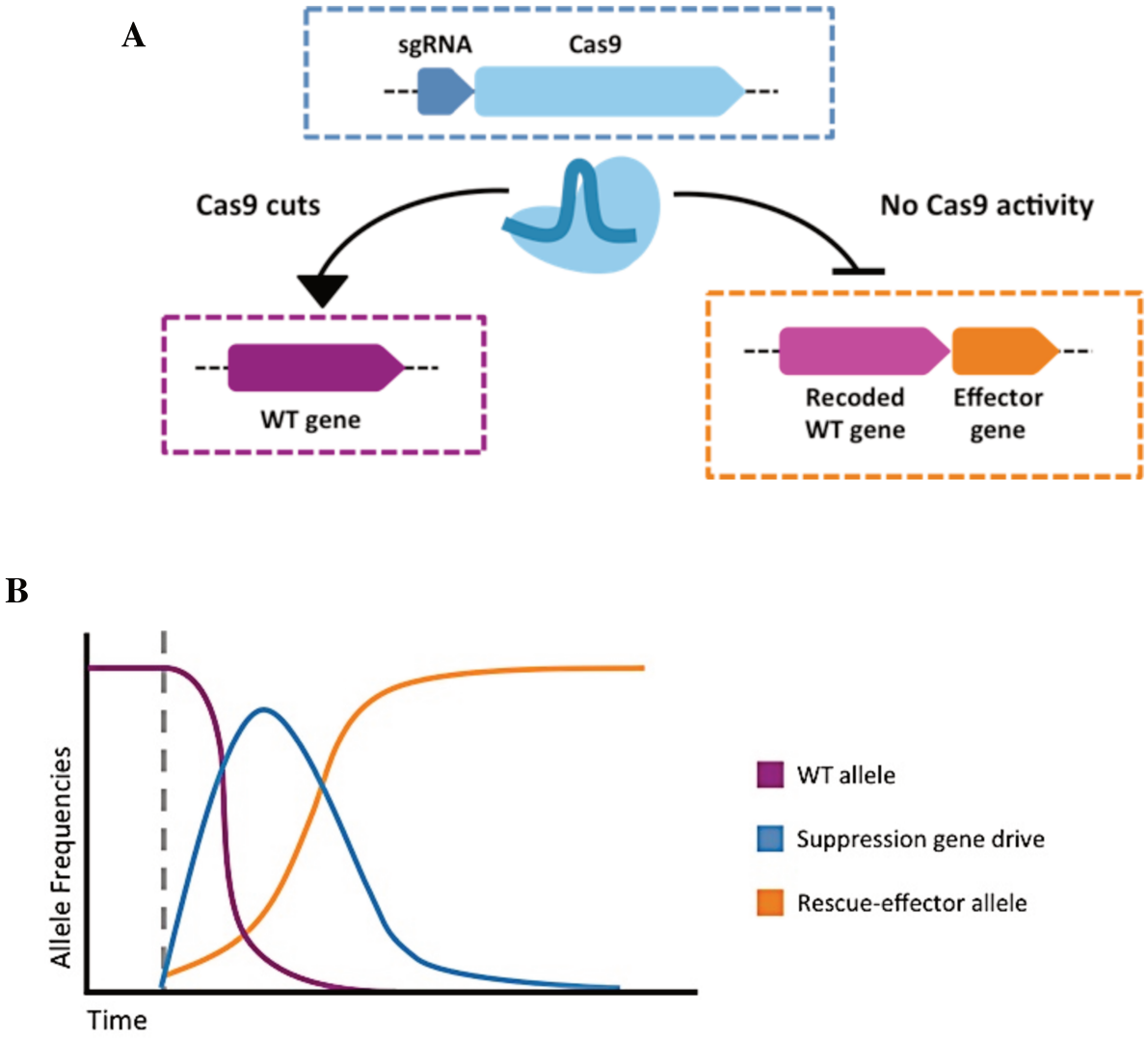


Figure 3.3. Indirect gene drive mode of action. **A**) An indirect gene drive system comprises three distinct genetic elements: the wild-type gene (purple), the CRISPR-GD (blue) and the desired effector allele (pink) linked to a GD resistant copy of the essential gene (orange). **B**) Initially, only the WT allele is present in the system. When the indirect gene drive components are introduced, the CRISPR-GD will initially increase in frequency, to the detriment of the WT allele. However, as the GD homes into and disrupts an essential gene, it will eventually be lost from the population, leaving only the rescue-effector allele. Figure adapted from model by Beaghton et al, 2017.

The indirect gene drive I primarily focused on is designed to introduce a single-chain antibody (scFv) that targets the sporozoite Circumzporozoite Protein (CSP). CSP is the most abundant protein on the surface of the sporozoite and is involved in both parasite transmission and infection in the vertebrate host. The protein is highly immunogenic and many of the natural anti-*Plasmodium* antibodies found in humans react to CSP. It is also the basis for many of the vaccines currently in development.

In this project, I use two anti-CSP scFvs, sc2A10 and sc125. 2A10 was one of the earliest transmission-blocking antibodies developed. It was generated from mice immunized with *P. falciparum* sporozoites and played an important role in the discovery and characterization of CSP (Nardin et al., 1982). 2A10 inhibits sporozoites ability to infect hepatocytes *in vitro* and reduces infection when pre-incubated with sporozoites (Hollingdale et al., 1984; Nardin et al., 1982). Several teams have tested the transmission-blocking ability of scFV 2A10 (Isaacs et al., 2011; Sumitani et al., 2013). Isaacs and colleagues found that expression of sc2A10 under the control of the vitellogenin promoter reduced sporozoite load and prevalence in the salivary glands, while Sumitani and colleagues found that expression of sc2A10 in the salivary glands using the *aapp* promoter significantly reduced transmission to mice. sc2A10 was also used in the modification gene drive developed in *A. stephensi* (Gantz et al., 2015).

sc125 was isolated by the laboratory of H. Wardemann using blood samples from Gabon and tested in transgenic mosquitoes in E. Levashina's lab in collaboration with our group (Triller et al., 2017). Like 2A10, transgenic expression in salivary glands significantly reduced transmission to mice and also resulted in a two-day delay in infection. Both sc2A10 and sc125 bind to the NANP repeat region of CSP.

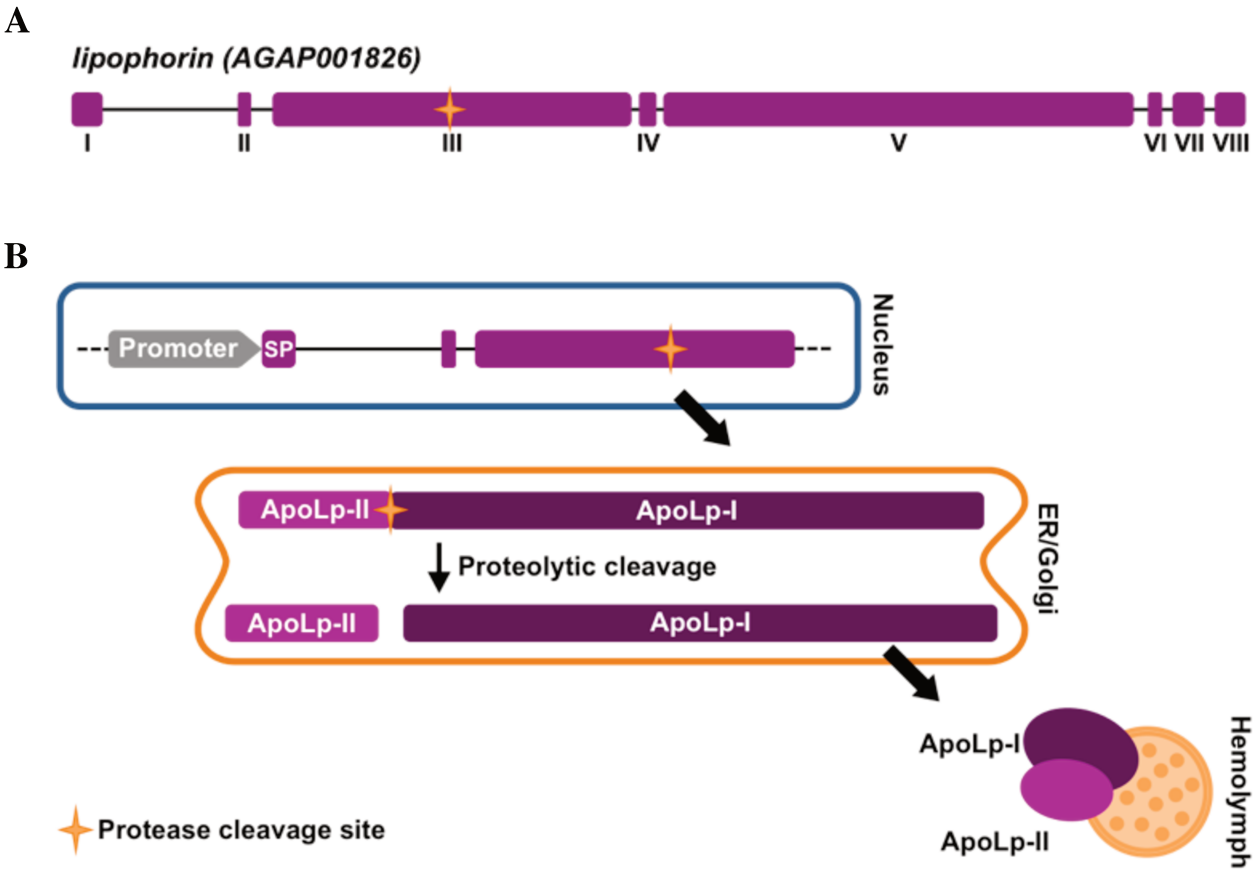


Figure 3.4. Lipophorin in *Anopheles*. **A)** The *lipophorin* locus in *Anopheles gambiae* is located on the right arm of chromosome two. It consists of seven exons. The first exon codes for the signal peptide and is followed by a large, 1.7 kb intron. The protease cleavage site is located in the third exon. **B)** After translation, the signal peptide directs the Lipophorin polypeptide into the ER/Golgi secretion pathway. Once inside the Golgi, a furin protease cleaves the polypeptide, producing the mature subunits ApoLp-I and ApoLp-II. The subunits are then secreted into the hemolymph, where they act in concert to shuttle lipids and other molecules between tissues.

The other component of an indirect gene drive is the essential target gene. For this project, the indirect gene drive targets *lipophorin* (Lp). Lipophorin forms the protein component of insect hemolymph lipoprotein complexes, which transport lipids and other insoluble compounds such as cuticular hydrocarbons, steroids, and pheromones between insect tissues via the hemolymph (Blacklock & Ryan, 1994). Lipophorin is essential for both development and fertility, transporting lipids from the midgut to fat body for storage and tissues and eggs for catabolism. It has also been shown to play a role in *Plasmodium* infection; knocking down Lp expression reduces oocysts loads (Rono et al., 2010; Vlachou et al., 2005). While it is common to speak of 'lipophorin', the lipoprotein complex is actually made up of two major proteins – apoipophorin-I (ApoLp-I) and apolipophorin-II (ApoLp-II). A third protein, apolipophorin-III, can further bind to lipoprotein complexes to provide additional stability.

In *An. gambiae*, ApoLp-I is 260 kDa and ApoLp-II is 74 kDa (Atella et al., 2006). Both proteins are transcribed from the *lipophorin* gene as a single polypeptide within the fat body, which is then cleaved and secreted via the ER/Golgi secretory pathway (Weers et al., 1993) (Figure 3.4B). The *lipophorin* gene (AGAP001826) is located on the right arm of chromosome 2 (11116762:11163371). The 10507 bp gene contains 7 exons. A short first exon of 75 bp comprising the secretory signal peptide is followed by a large 1.7 kb intron (Figure 3.4A). To date, there is no evidence of the presence of coding or non-coding genes within this intron. A second short exon and short intron are followed by the third exon, which contains the ApoLpII/ApoLpI protease cleavage site.

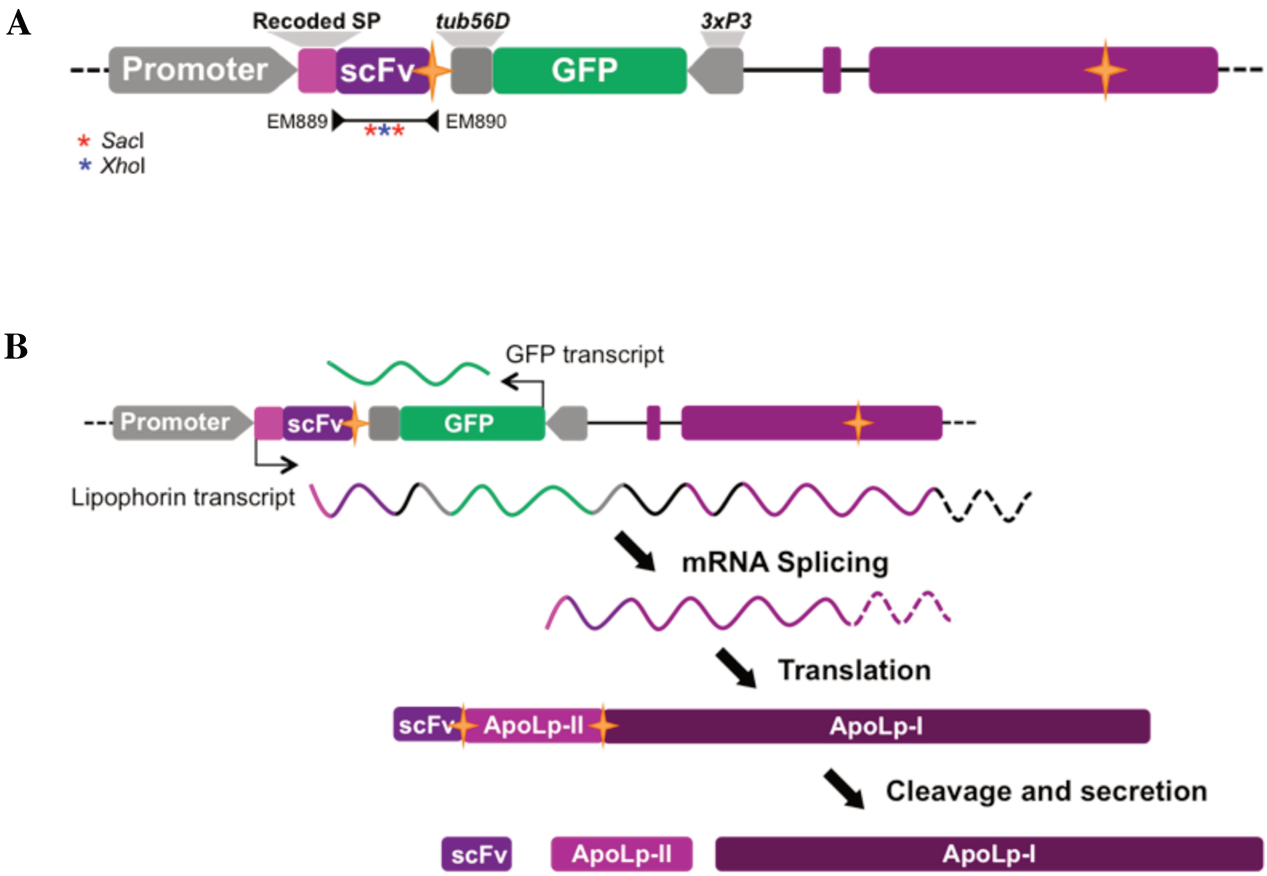


Figure 3.5: Diagram of the Lp::scFv locus. **A)** The knock-in cassette comprises the recoded first exon of *lipophorin* (pink), the scFV (purple), followed by the *lipophorin* protease cleavage peptide. The *3xP3*-GFP marker is inserted into the first intron, preserving the endogenous splicing signals. **B) Expression of the scFV**. The GFP marker is expressed from its own *3xP3* promoter in the eyes. The scFV is expressed under the control of the endogenous *lipophorin* promoter as part of the full-length mRNA. The GFP marker is removed from the transcript during mRNA splicing. The scFV is translated as a fusion with the Lipophorin subunits. The full polypeptide is then cleaved in the ER/Golgi and the three separated proteins are secreted into the hemolymph.

Results

Establishment of Lp::scFv effector lines

The Lp::scFv effector lines were produced knocking-in either sc125 or sc2A10 behind the first exon of lipophorin (Figure 3.5). The majority of the Cas9 sgRNA target sites within exon 1 and the beginning of intron 1 were also re-coded or removed to allow for maximal selection of guides when designing a *lipophorin* gene drive. A GFP marker under the control of the artificial *3xP3* promoter and the *D. melanogaster Tubulin56D* terminator was inserted into the first intron of lipophorin, with care taken to preserve the splice junctions, and in reverse orientation relative to *Lp* so that the transcription terminator would not perturb *Lp* transcription. Transcription of the *Lp::scFv* gene is therefore expected to produce a polypeptide containing the endogenous signal peptide followed by the scFv, ApoLpII, and ApoLpI. The endogenous proteolytic cleavage site of Lipophorin was duplicated between the scFv and ApoLpII to allow the separation of the scFv from the polypeptide during maturation.

The Lp::scFv knock-in lines were generated by co-injection of the knock-in plasmids pENTR-Lp::sc2A10 and pENTR-Lp::sc125 into *vasa*-eSpCas9 eggs. The knock-in plasmids contain three sgRNAs targeting the first exon of *lipophorin* under the control of the *An. gambiae U6* pol III promoter (AGAP013557), ~1 kB of 5' homology region, the knock-in region, and ~1 kB 3' homology arm. The target site of one of the guides is added at the 3' extremity of the 3' homology arm to allow the plasmid to be linearized by Cas9 upon injection. Positive G1s were identified by GFP expression in the eyes (Figure 3.6A). Single female families were established and genotyped by PCR amplifying the knock-in region, followed by a diagnostic digest using *Xho*I and *Sac*I, which cut specifically in sc2A10. Ultimately, two independent lines were established for each scFv.

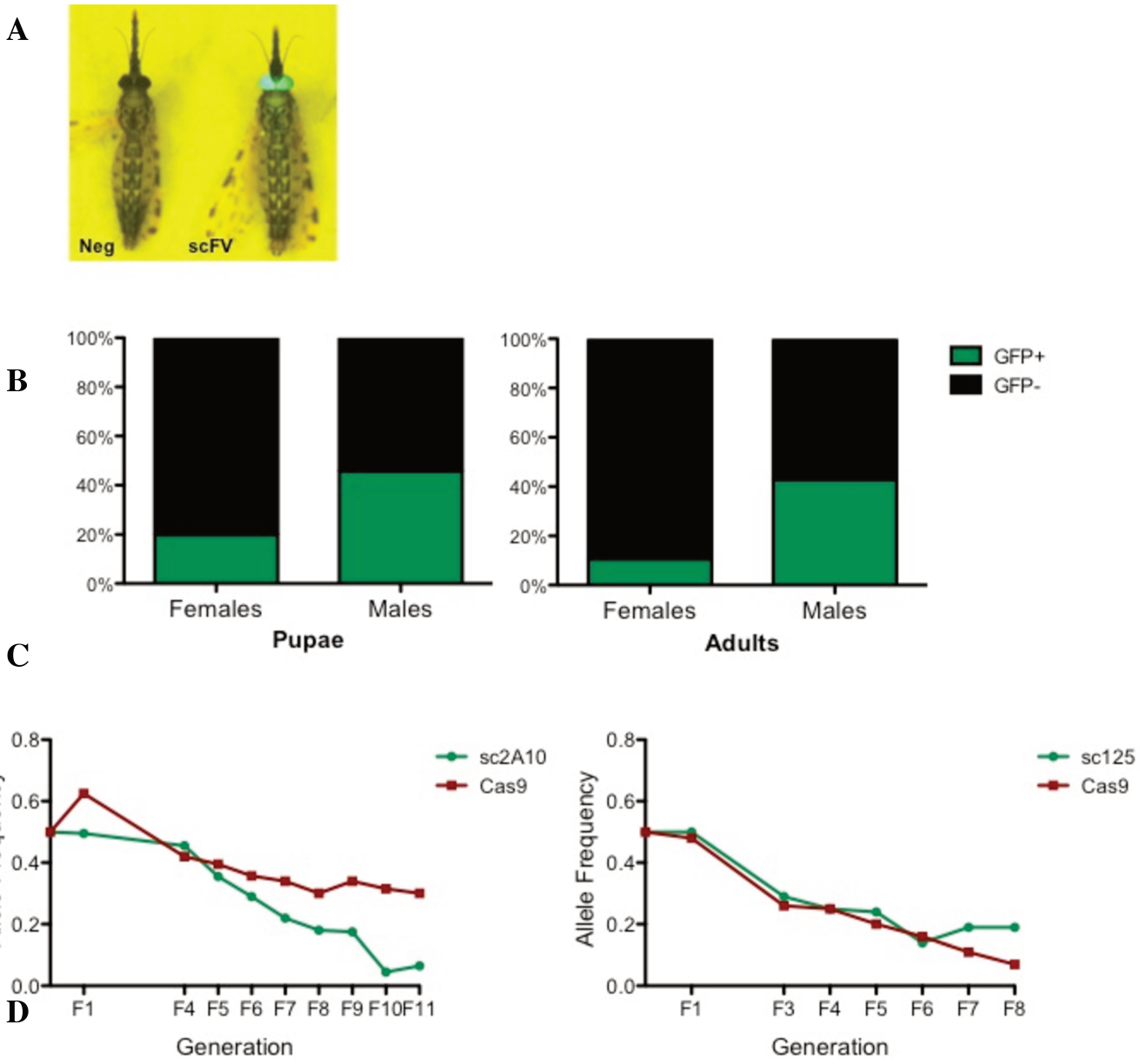


Figure 3.6: Characterization of *Lp::scFV* lines. A) Negative (Left) and transgenic Lp::scFV (right) adult *An. gambiae*). **B)** The sex ratios of mosquitoes expressing Lp::sc125 are strongly biased towards males. This bias is apparent at the pupal stage and becomes more pronounced at the adult stage. Bar graphs represent the pooled averages of four independent experiments. **C)** Over time, both scFvs are lost from the mosquito population. Mixed cages of Lp::scFV and eSpCas9 were tracked over several generations. The initial frequency of each population was calculated using the COPAS and allele frequencies were calculated using the Hardy-Weinberg equation. During this period, the insectary experienced several dramatic temperature and humidity decreases, which influenced mosquito viability and egg batch quality. The Lp::sc125 cage was lost during a significant temperature fluctuation at the 8th generation, while Lp::2A10 was tracked to the 11th generation before the cross was lost.

Fitness costs of Lp::scFv knock-in

Homozygous lines of Lp::sc125 and Lp::sc2A10 were established. In the case of Lp::sc2A10, there is no obvious fitness costs: the lines are fit and can be maintained as a homologous population. However in the case of Lp::sc125, males are fertile but homozygous females are completely sterile. Additionally, Lp::sc125 homozygotes exhibit a significant sex-bias during development (Table 3.1 and 3.2; Figure 3.6B). To compare the effects of the transgene on female survival, I selected equal numbers (200 – 300) neonate homozygous GFP^+ larvae and their negative siblings from the progeny of Lp::sc125 heterozygotes. The larvae were raised in the same pan and the numbers of males and females were counted at the pupa and adult stages. At the pupal stage, 20±6% of females are GFP^+ instead of the expected 50% of females. By the adult stage the percentage of GFP^+ females has dropped to 11±3%. This striking difference is not seen for GFP^+ males, indicating that sc125 has a female-specific fitness cost.

Table 3.1: Observed numbers of GFP^+ and GFP^- pupae from four independent experiments compared by sex.

	Female		**Male**	
Pan	GFP^+	GFP^-	GFP^+	GFP^-
1	24 (23%)	80 (77%)	95 (53%)	83 (47%)
2	14 (18%)	66 (83%)	82 (45%)	99 (55%)
3	27 (11%)	218 (89%)	62 (36%)	111 (64%)
4	36 (28%)	94 (72%)	89 (48%)	90 (54%)
Total	101 (20±6%)	458 (80%±6%)	328 (46±7%)	388 (54%±7%)

Table 3.2: Observed numbers of GFP^+ and GFP^- adults from four independent experiments compared by sex.

	Female		**Male**	
Pan	GFP^+	GFP^-	GFP^+	GFP^-
1	9 (12%)	66 (88%)	60 (40%)	89 (60%)
2	4 (9%)	42 (91%)	53 (42%)	74 (58%)
3	7 (8%)	76 (92%)	68 (53%)	61 (47%)
4	17 (14%)	101 (86%)	93 (49%)	97 (51%)
Total	37 (11±3%)	285 (89±3%)	274 (46±6%)	321 (54±6%)

The total number of individuals observed in each replicate pan is presented. Percentages were calculated using the total of females or males, to compare the numbers of GFP^+ and GFP^-. A significant ($p<0.0001$) difference between groups, calculated using Fisher's exact test with two-sided p-value

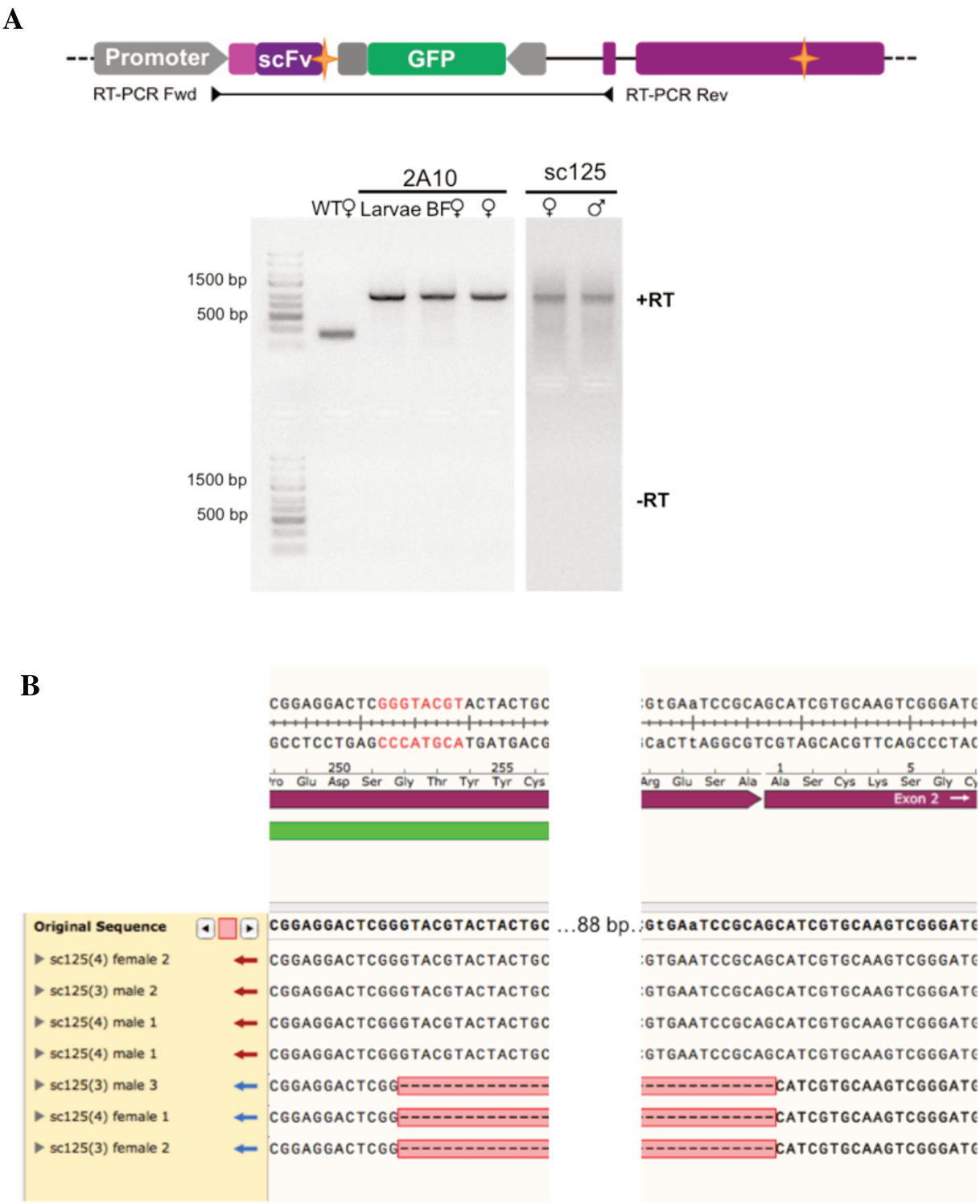
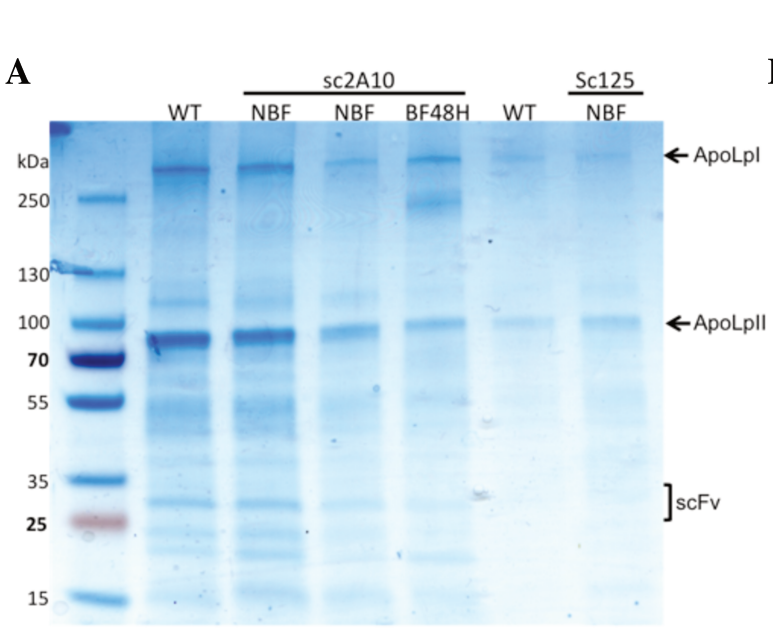


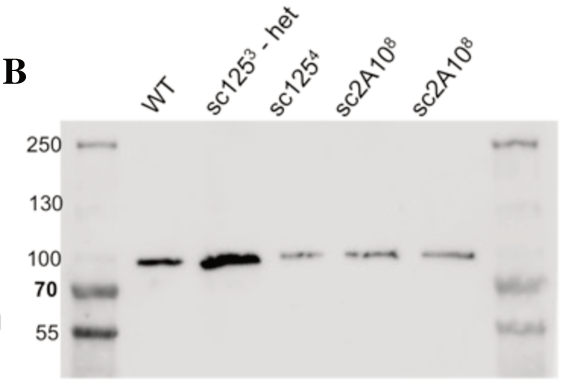
Figure 3.7: RT-PCR analysis of *Lp::scFv* lines. A) RT-PCR was performed using primers that span the first exon and intron. All lines tested showed the expected PCR product size corresponding to the *Lp::scFv* insertion. **B)** Alignment of individual mRNA transcripts from *Lp::sc125*. The first four transcripts correspond with the anticipated mRNA sequence, while the last three transcripts have been spliced at an upstream alternative splicing site located within sc125 (red).

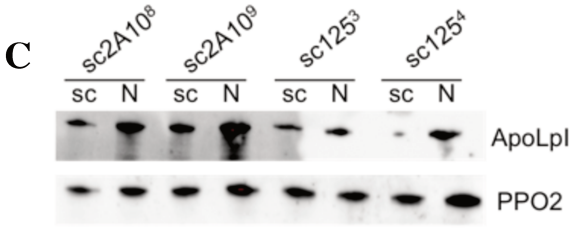
To evaluate the overall stability of the Lp::scFV transgenes in a population, the frequencies of the transgenes were tracked over time by COPAS analysis starting from a heterozygous population (Figure 3.6C). Starting from an initial 50% allele frequency, the frequency of both Lp::sc125 and Lp-2A10 steadily decreases in the population. Given the fitness consequences to homozygous Lp::sc125 females, the decrease of sc125 is unsurprising. The decline of sc2A10 is initially less dramatic compared to Lp::sc125, however it too decreases. eSpCas9 marked with DsRed was included as a reference 'neutral' locus, however this transgene does not behave consistently between the two populations. In the sc2A10 cage, Lp::sc2A10 decreases more rapidly than eSpCas9, while in sc125, both transgenes follow a similar trajectory. However, during the time that these cages were followed, they experienced several bottlenecks due to several dramatic decreases in temperature and/or humidity in the insectarium.

Transcription of scFvs and intron excision

The 3xP-GFP marker that identifies the Lp::scFV lines was inserted into the first intron of *lipophorin*. To verify that insertion of the scFV and/or the GFP marker had not disrupted mRNA processing, I performed RT-PCR on total RNA extracted from Lp::sc2A10 or Lp::sc125 homozygotes and their negative siblings, using primers spanning the first exon (Figure 3.7). Correctly spliced WT mRNA produces a 199 bp product and correctly spliced Lp::scFv produces a 970 bp product. All lines show the PCR product corresponding to a correctly spliced transcript (Figure 3.7). The PCR products were purified and sequenced, to compare the predicted transcript sequences to the actual sequences. In the case of sc2A10, sequencing of all the RT-PCR products showed the expected sequence. However, for sc125 the sequencing results of the 3' end of the transcript were consistently poor. The RT-PCR was performed using pooled RNA from 5 adult mosquitoes, meaning that variation within







D

Sample	ApoLpI	ApoLpII	scfv
Neg sc2A10	2075	486	0
sc2A10^{8}	1627	405	11
sc2A10^{8} (het)	1557	351	2
sc2A10^{9}	1002	267	7
sc2A10^{9} (het)	2290	664	2
Neg sc125	3218	768	0
sc125	1893	487	18

E

Sample	ApoLpI	ApoLpII	scfv
Neg sc2A10	5.8×10^{-3}	5.1×10^{-3}	0
sc2A10^{8}	5.1×10^{-3}	4.8×10^{-3}	0.4×10^{-3}
sc2A10^{8} (het)	10.1×10^{-3}	8.7×10^{-3}	0.1×10^{-3}
sc2A10^{9}	4.4×10^{-3}	4.4×10^{-3}	0.3×10^{-3}
sc2A10^{9} (het)	6×10^{-3}	6.7×10^{-3}	0.1×10^{-3}
Neg sc125	10×10^{-3}	9.3×10^{-3}	0
sc125	7.1×10^{-3}	7.0×10^{-3}	0.7×10^{-3}

Figure 3.8: Analysis of hemolymph from Lp::scFv lines. A) Example Coomassie staining of hemolymph collected from adult females that were either non-blood fed (NBF) or received a blood meal 48h before collection (BF48H). No specific band or varying band intensity was detected in the region where the scFvs would be expected to migrate. B) ApoLpII western blot on hemolymph from 10 mosquitos collected 48 hours after blood feeding. No detectable shift in size was observed which would correspond to a fusion between the scFv and ApoLPII that would be indicative of failed separation of scFv and ApoLpII. All Lp::scFv females were homozygous for the Lp::scFv, except for sc125^{3} which were heterozygous. C) ApoLpI/PPO2 Western blot using hemolymph from 6-7 female mosquitos 48 hours after blood feeding. sc = Lp::scFv homozygotes, N = negative siblings. D) Total number of peptide spectra corresponding to ApoLpI, ApoLpII, and sc2A10 or sc125 detected in hemolymph from Lp::scFv lines or negative siblings. Het = heterozygous. E) Relative abundance of ApoLpI, ApoLpII, and sc2A10 or sc125 in mass spectrometry samples. A normalization coefficient was calculated by averaging the total number of spectra from the abundant hemolymph proteins ApoLpIII, APLC1, LRIM1, Nimrod, Phenoloxidase, and TEP1 across all samples. The ApoLpI, ApoLpII, and scFv spectral counts were normalized by multiplying the total number of spectra by the normalization coefficient and dividing by the total number of abundant protein spectra for each sample. The relative abundance was calculated by dividing the normalized spectral counts by the molecular weight in daltons.

individuals could account for the poor sequencing quality. I cloned the RT-PCR product into the pJet plasmid and sent 2 – 3 colonies to sequence for each RT-PCR product. The results showed that in many of the transcripts the RNA was spliced using an alternative splice junction within the sc125 coding sequence (Figure 3.7 B). This alternative splicing maintains the reading frame and results in a truncation of sc125 that lacks 25 C-terminal amino acids as well as the protease cleavage site, therefore the truncated sc125 will remain fused to ApoLpII. Given the frequency with which I observed this alternative splicing, this appears to occur often in the Lp::sc125 lines and could account for the unanticipated fitness costs and female sterility I observed.

The scFvs are present in the hemolymph

The scFvs used for this project were not tagged in order to minimize risks of decreased activity, so it is not possible to detect them by western blot. Initially, I tried running hemolymph samples on a polyacrylamide gel followed by Coomassie staining, to look for additional bands within the expected size range of the scFvs (Figure 3.8A). However I was unable to detect an additional band or a difference in band intensity between the Lp::scFv lines and negative controls.

I therefore turned to mass spectrometry to try to detect the scFvs. The mass spectrometry analysis was performed by the IBMC proteomics platform using whole hemolymph samples collected from female mosquitoes. For all samples analyzed, peptides corresponding to the scFvs were found only in the Lp::scFv lines (Figure 3.8D). Additionally, the mass spectrometry analysis showed that the signal peptide was correctly cleaved from the polypeptide. The mass spectrometry results did not allow me to confirm that the scFvs were correctly cleaved from ApoLpII, as the region is rich in lysine and arginine residues, and the peptides generated by trypsin digestion in this region are too small to be detected. However,

comparing the relative abundance of ApoLp I, ApoLp II, and the scFv shows that ApoLp I and ApoLP II are present in a roughly 1:1 ratio, while the scFv is roughly 10x less abundant (Figure 3.8E). This indicates that the scFv is following a separate pathway within the hemolymph and is therefore likely cleaved from ApoLpII. Additionally, a western blot on ApoLpII did not uncover a shift in the size of ApoLpII from Lp::scFv lines (Figure 3.8B). All of these observations indicate that the scFvs are correctly cleaved from ApoLpII at the ectopic protease cleavage site, at least in the case of the Lipophorin circulating in the adult hemolymph.

I performed a western blot on ApoLpI using hemolymph samples from 6-7 homozygous Lp::scFv females and their negative siblings to see if insertion of the scFv affected the expression levels of Lipophorin. After transfer and blocking, I cut the membrane at the 130 kDa mark and probed the upper molecular weight proteins for ApoLpI (approximately 260 kDa) and the lower molecular weight proteins for PPO2 (aproximatly 78 kDa) as a loading control (Figure 3.8C). Though the quality of the blot is not optimal, ApoLpI appears to be less abundant in Lp::sc125 lines.

sc2A10 significantly reduces transmission of Pb-PfCSP to naïve mice

Once I had confirmed that the scFvs were expressed as expected and present in the hemolymph, the next step was to assess their ability to block transmission. As stated in the introduction, these scFVs have previously been tested and are known to reduce transmission, albeit expressed under the control of different promoters, from a different genomic locus, and in different tissues. To assess the transmission-blocking capacity of the scFVs, I performed bite-back experiments using a rodent model. *P. berghei* is a standard laboratory model for the study of malaria, as its life-cycle is highly similar to that of human parasites and it can be transmitted by *Anopheles gambiae* and other vectors of human malaria. Additionally, *P.*

berghei can be genetically engineered to express proteins from *P. falciparum* and other human parasites. As the scFVs used in this study specifically recognize CSP from *P. falciparum*, I used a *P. berghei* strain in which the CSP of *P. berghei* had been replaced with the CSP from *P. falciparum*. Initially, the *Pb-PfCSP* strain was marked with *eif1a*::GFP (Triller et al., 2017). While this GFP marker is visible in oocysts, the fluorescence levels were not high enough to allow me to detect blood-stage parasites using the current FACS set-up in the lab. In order to exploit the FACS, E. Marois generated a *Pb-PfCSP* strain marked with *hsp7*::GFP, which allowed me to track the evolution of infection in mice in over time using FACS.

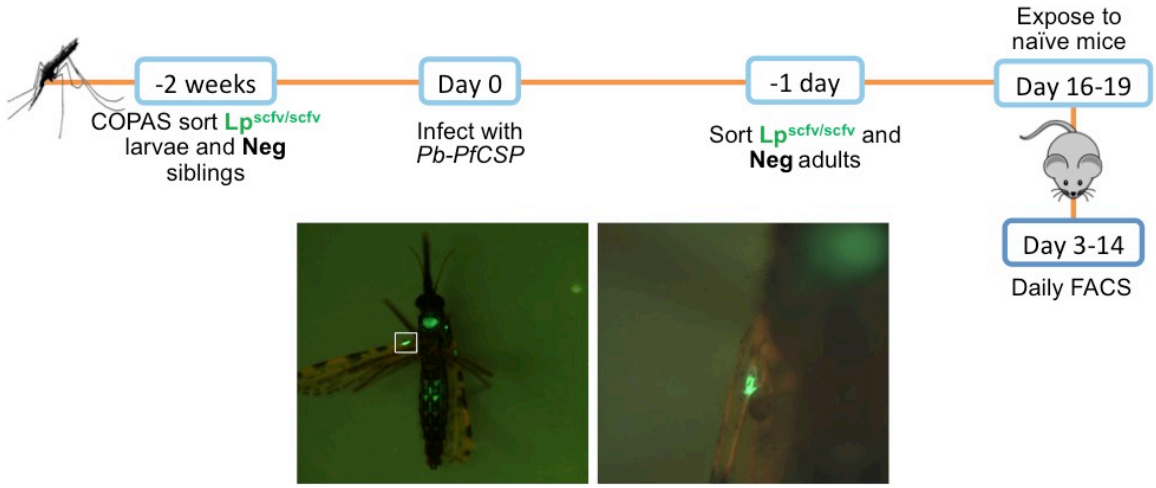


Figure 3.9: Schematic representation of the bite-back protocol. Images show GFP expression from *Pb-PfCSP* oocysts in the abdomen and sporozoites in the wing. Wing sporozoites typically accumulate in the hinge region.

A schematic of the bite-back protocol can be found in Figure 3.9. Briefly, experimental (GFP^+) larvae and negative control siblings were COPAS sorted from the progeny of Lp::scFV heterozygotes. Experimental and control mosquitoes were raised and infected together by blood-feeding from the same mouse. 16-20 days after infection, the Lp::scFV females were separated from negative controls. Mosquitoes were first sorted as transgenic (GFP^+) or non-transgenic (GFP^-) and then *Plasmodium-carrying* mosquitoes were selected based on the presence of midgut oocysts and sporozoites in the wing visible by GFP

fluorescence through the cuticule of live, cold anesthetized females (Figure 3.9). Naïve mice were then exposed to either Lp::scFV or negative control mosquitoes. Two independent lines of Lp::sc2A10 were used for bite-backs (Lp::sc2A10^{8} and Lp::sc2A10^{9}), while only one line of Lp::sc125 (Lp::sc125^{4}) was tested. When possible, the mosquitoes were dissected following the bite-back to count oocysts numbers and look for sporozoites in the salivary glands. Mouse blood samples were analyzed daily by FACS for two weeks starting from 3-4 days after bite-back.

Confirming previous reports, sc2A10 significantly reduced transmission of *Pb-PfCSP*. Only 29.7% of mice exposed to Lp::sc2A10 *An. gambiae* became infected, compared to 97.1% exposed to control mosquitoes (Figure 3.10A). Additionally, those mice that did become infected experienced a 1-day delay in the development of infection (Figure 3.10C). No difference was observed between the biological replicated lines Lp::sc2A10^{8} and Lp::sc2A10^{9}. In contrast, sc125 exhibited no transmission blocking activity or effect on the infection (Figure 3.10A). However the bite-back experiments were performed using *Lp::sc125* heterozygotes. Given this and the alternative splicing that appears to have occurred with high frequency, functional sc125 was presumably much less abundant than sc2A10. I performed a preliminary bite-back experiment using Lp::sc2A10 heterozygotes, however transmission by control mosquitoes was abnormally poor in one of the replicates (Table 3.3).

Table 3.3: Total number of mice infected or uninfected after exposure to Lp::sc2A10 heterozygotes from either line 8 or line 9.

	Lp::sc2A10^{8}	**Neg^{8}**	**Lp::sc2A10^{9}**	**Neg^{9}**
Infected/Uninfected	0/6	1/5	0/5	4/1

Results from one experiment

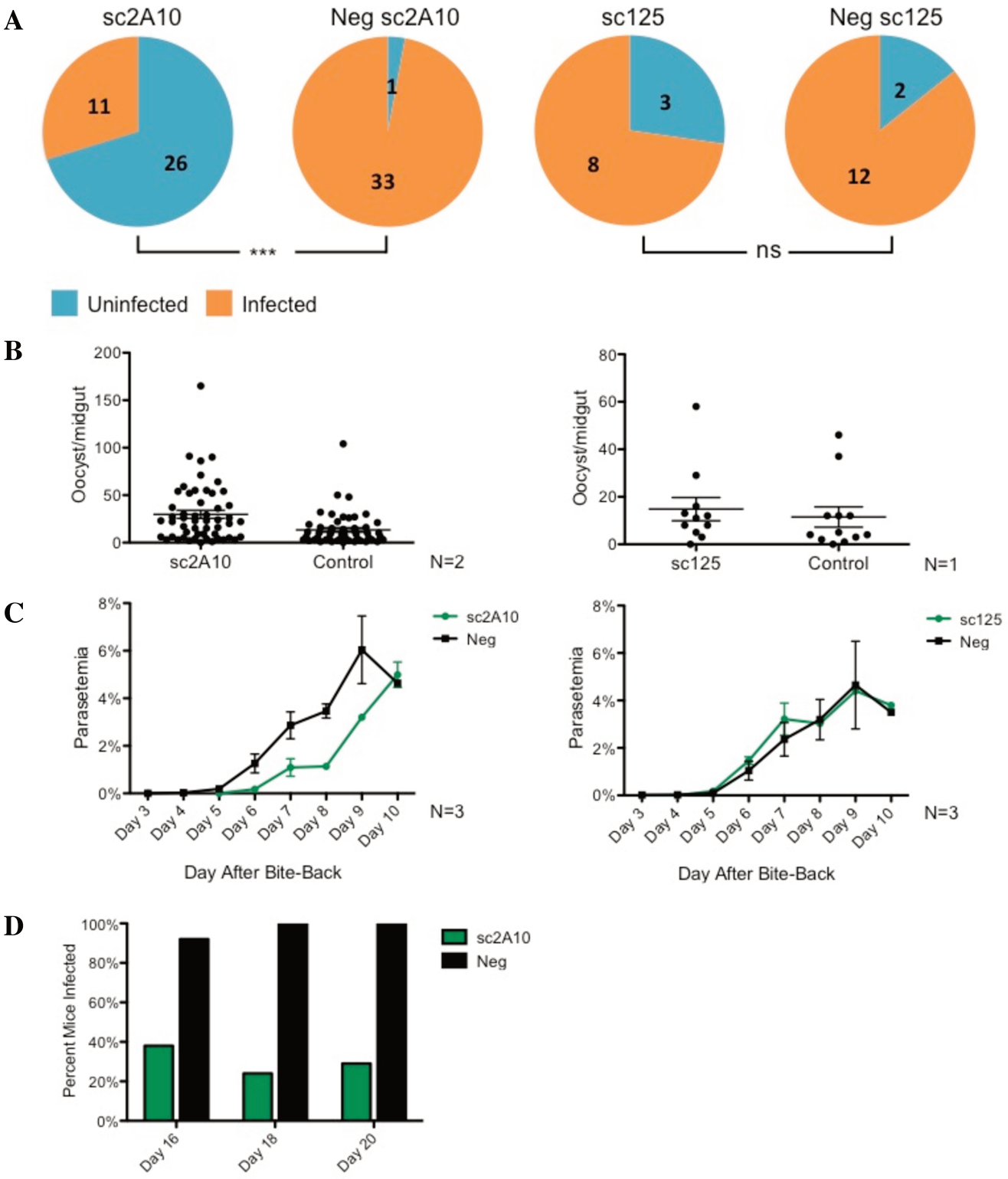


Figure 3.10: Bite-back results. A) Total number of infected (orange) and uninfected (blue) mice after exposure to *Lp::scFv* females or their negative siblings. Aggregated data from six (sc2A10) and four (sc125) independent experiments were analyzed using Fisher's exact test with a two-tailed p-value. *** $p<0.0001$, ns $p=0.6232$. **B**) Total number of oocysts per midgut from a subset of mosquitoes used for bite-back. Mosquitoes were dissected after bite-back, on day 18-22. The higher oocysts levels in the *Lp::scFv* lines is likely due to bias introduced during mosquito sorting before the bite-back. **C**) Evolution of parasitemia in mice that became infected after exposure to *Pb-PfCSP*-infected mosquitoes. Mice exposed to *Lp::sc2A10* mosquitoes experience a one-day delay in the onset of infection. Only experiments were both scFv and neg mice became infected were included in this analysis. **D**) Comparison of infectivity of *Lp::sc2A10* mosquitoes exposed to mice on different days. There is no evidence of difference in outcome based on when the infection occurred. The percentages were calculated from the aggregated data from 5 (Day 16), 6 (Day 18), or 3 (Day 20). Total number of mice: sc2A10 – Day 16: 13, Day 18: 17, Day 20: 7; Neg – Day 16: 12, Day 18: 14, Day 20: 7

Previous work on transmission blocking in the team has shown that the infectivity of transmission-impaired mosquitoes can increase over time. The number of *P. berghei* sporozoites in the salivary glands of *saglin* null mosquitoes was observed to increase from day 18 to day 22 (E. marois, unpublished). In the case of sc125 and sc2A10, preliminary results do not indicate that infectivity of the Lp::scFv lines generated in this study increases with time (Figure 3.10D). In terms of oocysts load, the transgenic mosquitoes dissected after bite-back appear to be slightly more infected than controls (Figure 3.10D). However this is likely a sorting bias, as the GFP fluorescence from the Lp::scFV transgenic marker can obscure GFP expression from the parasites in the wing, leading to the selection the most highly infected transgenic mosquitoes. The reduced transmission of these mosquitoes is therefore all the more striking.

To verify that the transmission-blocking phenotype of Lp::sc2A10 was indeed specific to the CSP from *P. falciparum*, I performed parallel bite-backs using *Pb-PfCSP* and *Pb-GOMO14*, a GFP-marked *P. berghei* strain expressing endogenous *PbCSP* and commonly used for infections in our laboratory (Manzoni et al., 2014). Both Lp::sc2A10 and negative mosquitoes transmitted *Pb-GOMO14* to mice with equal efficiency, indicating that the transmission blocking capacity of sc2A10 is specific (Table 3.4).

Table 3.4: Number of mice infected or uninfected after exposure to mosquitoes infected with *Pb-PfCSP* or *Pb-GOMO14*.

	Pb-PfCSP		**Pb-GOMO14**	
	Lp::sc2A10	WT	Lp::sc2A10	WT
Infected/Uninfected	0/4	3/0	5/0	4/0

Pooled results of one experiment using two biological replicates

First attempt to establish a gene drive targeting lipophorin

Based on its transmission-blocking abilities, Lp::sc2A10 is a good candidate for an indirect gene drive. As described in the introduction, an indirect gene drive has two components: the resistant effector line and a suppression gene drive line. In the previous section, I described the Lp::scFV effector line; now I will turn to my efforts to establish a gene drive targeting *lipophorin*.

The *lipophorin* gene drive (Lp-GD) design is a standard suppression CRISPR-GD (Figure 3.11A). It comprises four sgRNAs, eSpCas9 under the control of the *zpg* germline specific promoter and terminator regions (Hammond et al., 2018), and a fluorescent marker. To express the sgRNAs, I used the tRNA-multiplexing strategy developed in rice by Xie and colleagues (2015), in which multiple sgRNAs are expressed as a single transcript of tandem tRNA-sgRNA repeats (Figure 3.11B). Processing of the tRNAs by the endogenous tRNA maturation system liberates the sgRNAs. This strategy has been adapted to *D. melanogaster* (Port & Bullock, 2016). It has also successfully been used in our laboratory for tissue-specific CRISPR knock-out (R. Mela-Lopez, unpublished data).

Initially, I designed a Lp-GD containing three sgRNAs marked with *3xP3*-CFP. I injected this plasmid into *vasa*-eSpCas9 eggs, however I failed to recover any transgenics, despite observing transient expression of CFP in injected G0 larvae. As *lipophorin* is essential for development and fertility, one explanation is that Cas9 cleaved too efficiently and produced an excess of homozygous mutant cells, either killing G0 larvae or sterilizing them. Alternatively, the sgRNAs used were not optimal.

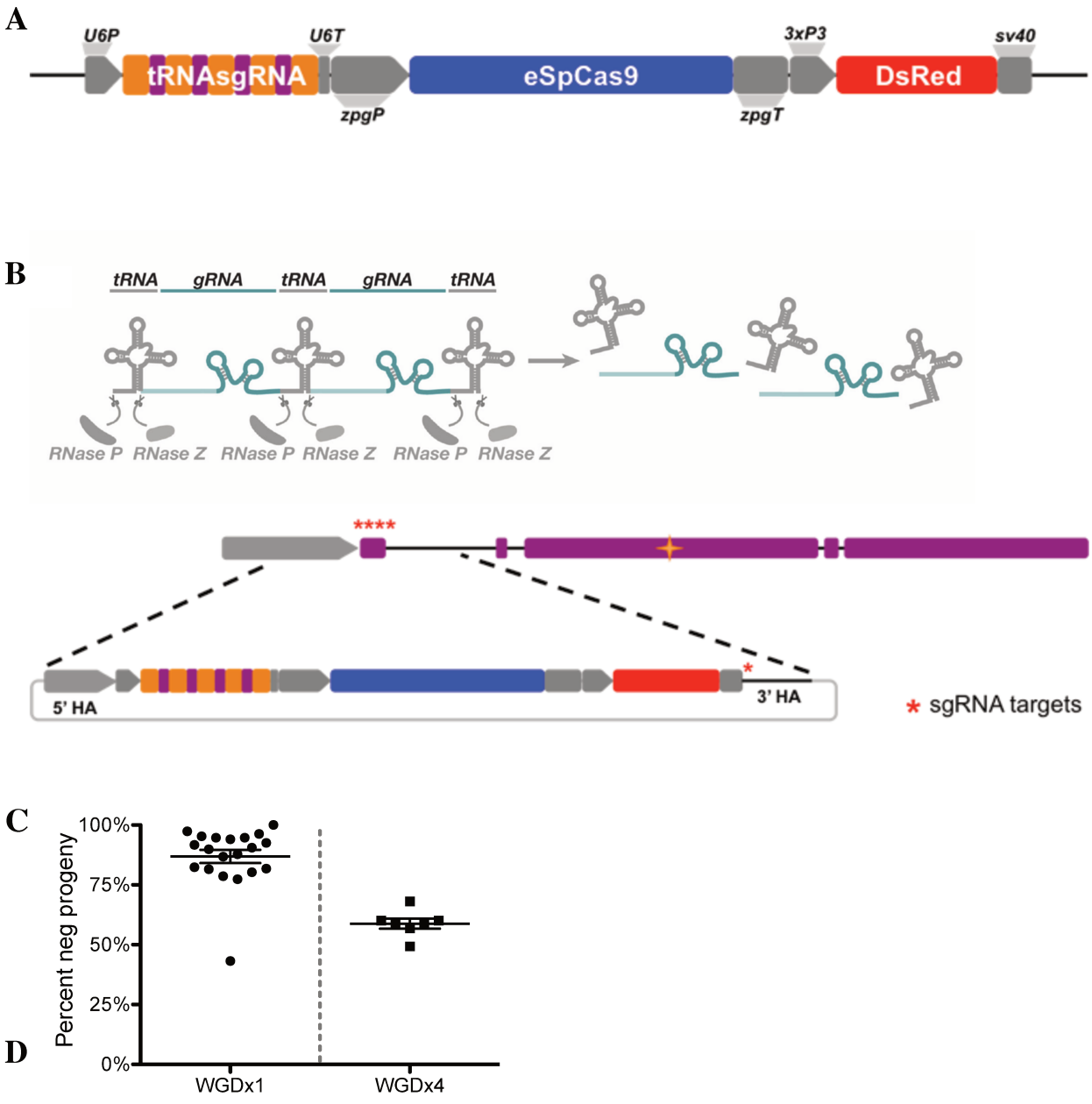


Figure 3.11: Establishment of a CRISPR-GD targeting *lipophorin*. A) Schematic of the CRISPR/Cas9 gene drive targeting *lipophorin*. The final gene drive design included four sgRNAs expressed from a tRNA array under the control of the U6 promoter, eSpCas9 under control of the *zpg* promoter, and a *3xP3*::DsRednls. **B)** sgRNAs can be expressed from the same transcript by flanking each sgRNA with a tRNA. Cleavage of the tRNAs by the endogenous processing enzymes RNase P and RNase Z produces mature sgRNAs. From Port and Bullock, 2016. **C)** Schematic of the LpGD knockin into *lipophorin*. The four sgRNAs target the first exon and start of the first intron of *liphophorin* and are used for both gene drive knock-in and subsequent gene drive activity. By error, the sgRNA target site within the first intron was included at the start of the 3' homology arm, allowing Cas9 to cleave the plasmid. **D)** Transmission bias of split gene drives expressing one (WGDx1) or four (WGDx4) sgRNAs under the control of the *U6* promoter and targeting a *3xP3*-YFP YFP marker. Females heterozygous for the gene drive and expressing *vasa*::eSpCas9 from an unlinked locus were crossed to negative males and the percentage of YFP^{-} progeny was calculated for single female families. A higher percentage of progeny from WGDx1 females lost YFP expression compared to progeny from WGDx4.

I re-designed the gene drive, this time including sgRNAs used for the initial Lp::scFv knock-in and at the same time replacing the *3xP3*-CFP marker with *3xP3*-DsRednls. The rationale for this substitution was to facilitate COPAS analysis of mixed Lp-GD (DsRed) and Lp::scFv (GFP) populations. I also injected the new gene drive in eggs heterozygous for Lp::sc2A10, thereby ensuring that only one copy of *lipophorin* would be disrupted. The injected eggs also expressed a *vasa*-Cas9 transgene (marked with *3xP3*-YFP) to ensure the presence of sufficient Cas9 protein in the embryo, in addition to any Cas9 expressed from the injected GD plasmid.

Several dozens GFP/RFP G1 larvae were recovered from the G0 parents and outcrossed to wild-type partners. As the two transgenes should be located at the same locus on homologous chromosomes, they should segregate independently. Surprisingly, the majority of GFP/RFP outcrosses gave rise to 50% GFP/RFP, 50% negative progeny, while a minority gave rise to larvae that expressed either GFP or RFP, as expected. Preliminary analysis of the GFP/RFP lines indicates that the Lp-GD was inserted upstream of Lp::sc2A10, following a rearrangement between homologous chromosomes in addition to integration of the injected plasmid. When I redesigned the Lp-GD, I failed to realize that there was an sgRNA target site in the plasmid at the 5' end of the 3' homology arm (Figure 3.11C). Cas9 could therefore cleave the injected plasmid, separating the 3' homology region. GFP/RFP homozygotes are viable and fertile, indicating that the *lipophorin* promoter is still functional. Further PCR analyses will provide additional insight into the exact configuration of this locus.

Though the RFP population was the minority of the recovered G1 transgenics, enough were recovered to create several lines. PCR analysis indicates that the gene drive was incorporated as expected into the *lipophorin* locus. In these lines, the sgRNA target site in the plasmid shows a small indel mutation just upstream of the PAM, indicating NHEJ repair of the

synthetic construct. Analysis of the progeny of the Lp-GD lines showed Mendelian-like inheritance, indicating that the gene drive is inactive. The RFP line has been crossed to a *vasa*-eSpCas9 line and to a control sgRNA line to check the activity of the guides and Cas9, respectively. Preliminary results suggest that the sgRNA-tRNA array is functional, whereas *zpg*-eSpCas9 has no detectable activity (E. Marois, pers. comm.).

The initial hypothesis for the failure of the Lp-GD to drive was that the sgRNAs were not expressed properly. This has previously been observed in our lab using the tRNA multiplexing system under the control of the U6 promoter. Early in my PhD, I designed several split-gene drives expressing one or four sgRNAs. The primary goal of this project was to compare the homing efficiencies of Cas9 and Cas12a by targeting an artificial YFP locus in one of the *An. gambiae* lines in our lab. However, the Cas12a variant chosen, AsCas12a, proved to be the highly temperature sensitive variant with optimal activity at 37°C, so this part of the project was abandoned (Moreno-Mateos et al., 2017). However, by the time Moreno-Mateos and colleagues published their observations on the temperature constraints of AsCas12a, I had already generated Cas9 split drives expressing one (WGDx1) sgRNA or four (WGDx4) sgRNAs multiplexed using the tRNA system. A summer intern, Mallory Kastner, used these lines to compare the driving efficiencies of these two constructs using *vasa*-eSpCas9 at the X1 locus as a source of Cas9. Surprisingly, the single-guide construct was much more efficient than the multi-guide construct (Figure 3.11D). 89% of the progeny of WGDx1/YFP females lost YFP expression, compared to 59% of progeny from WGDx4/YFP females. 96% of the progeny of WGDx1/YFP males inherited a disrupted copy of YFP, compared to 73% of WGDx4/YFP males. Based on these results and the lack of activity by the Lp-GD, it seems that the *U6*-tRNAsgRNA is expressed less efficiently in the

germline, despite being functional in somatic tissue under the control of Pol II promoters and for generating CRISPR/Cas9 knock-ins when injected as naked plasmid DNA.

To confirm that the knock-in of the Lp-GD into *lipophorin* is lethal as expected, E. Marois crossed the Lp-GD (RFP) line to Lp::sc2A10 (GFP), with the goal of obtaining heterozygotes to self-cross. However, the majority of Lp-GD/Lp::sc2A10 die during development, indicating that Lp::sc2A10 hemizygotes have a fitness cost that was not readily apparent in the presence of a wild-type chromosome, but is evident when crossed to a loss-of-function Lp allele. Additional preliminary experiments to characterize the loss-of-function phenotype of the Lp-GD transgene show an elevated death rate of heterozygous larvae during development in spite of the presence of a WT copy of *lipophorin*. This suggests that *lipophorin* loss-of-function is partially haploinsufficient, meaning that it will be difficult to establish an efficient gene drive in the *lipophorin* locus.

Other indirect gene drive candidate loci

While the bulk of my PhD thesis focuses on the Lp::scFv project, I also worked on other gene drive projects over the course of my PhD. Prior to beginning working on the Lp::scFv lines, I attempted to construct a modification drive to simultaneously disrupt two salivary gland proteins, *saglin* and *csp-bp*, which are known to be important for sporozoite invasion of the salivary glands. However this design made use of AsCas12a and was therefore abandoned when it sensitivity to temperature was discovered. Prior to this, I generated a dual-*att*P knock-in into the *saglin* gene, generating a loss-of-function mutant. This line is homozygous viable and shows no obvious fitness costs, and could be used to host future GD constructs, such as a new gene drive targeting *lipophorin* to promote the spread of the Lp::sc2A10 construct.

I also performed a screen to identify essential genes that could be exploited for modification drives. I selected 8 putative essential genes based on homology to *D. melanogaster* genes. Candidate genes were selected based on the following criteria: size (150 aa or less), knock-out phenotype (homozygous recessive lethal, early development essential), homology to *An. gambiae* genes, quality of annotation (well-defined molecular function and *in vivo* evidence in *D. melanogaster*) and no physical overlap with other mapped genes in *An. gambiae*. From this list, I chose 8 candidate genes and constructed dual-*att*P knock-in plasmids to disrupt the *An. gambiae* orthologues (Table 3.5). Four of these plasmids used sgRNAs for Cas12a and therefore never gave transgenics when injected into *vasa*-AsCas12a eggs. The other four used Cas9 sgRNAs for knock-ins. From co-injections of knock-out plasmids targeting the essential genes, I recovered transgenic G1 larvae expressing YFP, corresponding to a knockout of the *An. gambiae* orthologue of *roadblock* (robl, AGAP003360). In *D. melanogaster, roadblock* (FBgn0024196) is a dynein-associated protein that plays a role in intracellular transport. In *An. gambiae,* knock-out mutants of *robl* are heterozygous viable but homozygotes die early during larval development. *robl* is therefore another candidate locus that could be exploited to create a modification gene drive in *An. gambiae*.

Table 3.5: Candidate essential genes selected for screening in *An. gambiae*

D. melanogaster gene	An. gambiae orthologue	Size bp (aa)	Biological Process
cyclope (FBgn0015031)	AGAP007768	542 (78)	ATP synthesis
***roadblock* (FBgn0024196)**	AGAP003360	883 (97)	Intracellular Transport
Spase 12-subunit (FBgn0040623)	AGAP004296	483 (96)	Part of ER membrane
adaptor protein complex 2, σ subunit (FBgn0043012)	AGAP001703	718 (142)	Endocytosis
transcription-factor IIA-S (FBgn0013347)	AGAP004370	769 (112)	Transcription
***mago nashi* (FBgn0002736)**	AGAP010755	718 (148)	mRNA splicing
Ribosomal protein L23 (FBgn0010078)	AGAP010252	664 (140)	Translation
Replication protein A3 (FBgn0266421)	AGAP010177	494 (120)	Replication

Knock-in plasmids for the genes in grey were constructed using sgRNAs for Cas12a.

Discussion

While I was not able to successfully test the indirect gene drive design during my PhD, I was able to characterize the Lp::scFv knock-in lines, as well as identify an additional locus, *roadblock*, which could serve as a basis for a future indirect gene drive design. I showed that insertion of sc2A10 into the *lipophorin* locus retained the transmission blocking abilities previously reported. One advantage of expressing a sporozoite effector molecule fused to native *lipophorin* is that the molecule is constitutively expressed; as opposed to blood-meal inducible scFv transgenic lines that required repeated feeding to maintain scFv expression (Isaacs et al., 2011). Additionally, embedding the scFv in the coding sequence of *lipophorin,* under the control of its own promoter and signal peptide, should strongly protect the transgene from spontaneous loss by mutation.

Based on the difference in relative abundance between the scFvs and ApoLpI:ApoLpII observed in the mass spectrometry analysis and the lack of visible shift in the size of ApoLpII, both scFvs appear to be cleaved from the polypeptide, provided the protease cleavage site is intact. While the relative abundance allowed me to compare the ratios of ApoLpI/ApoLpII/scFv within samples, the variability in protein quantity between hemolymph samples makes comparison between samples impossible. Better quantification with mass spectrometry could be obtained by labeling the scFv or spiking the sample with a synthetic peptide. However, based on the preliminary western blot of ApoLpI/PPO2, Lipophorin appears to be less abundant in *Lp::sc125* samples, though this experiment does need to be replicated to confirm the results. Taking into account the alternative splicing in a subset of *Lp::sc125* transcripts, it is possible that truncated sc125::ApoLpII fusion proteins are more rapidly degraded.

The fitness costs of sc125 were surprising and unanticipated, as well as its poor transmission-blocking phenotype. sc125 had previously been tested in transgenic *An. gambiae* mosquitoes under the control of the salivary gland promoter *aapp*, where it displayed strong transmission blocking ability (Triller et al., 2017). However, the alternative splicing of the *Lp::sc125* mRNA likely explains the poor performance of sc125 and its fitness costs. This is an important phenomenon to take into account when designing fusion proteins in the future. When *sc125* was designed, obvious splice junctions were searched for and removed, however, as my results indicate, cryptic splice junctions can be missed. Given how often I found alternative splicing of the transcript, a high proportion of sc125 was likely expressed as a non-functional truncated fusion to ApoLpII. This would interfere with sc125's ability to bind to sporozoites and also likely perturb the function of ApoLpII. It is possible that the fusion

specifically interferes with ApoLpII's ability to interact with pathways important for female development, explaining the female-specific phenotype of *Lp::sc125*.

Lipophorin itself is known to play a role in infection. Knock-down of *lipophorin* levels reduces the number of oocysts in the midgut and also reduces transmission of both *P. berghei* and *P. falciparum* (Costa et al., 2017; Rono et al., 2010). However, the Lp::sc2A10 lines are able to transmit the wild-type *P. berghei* strain *PbGOMO14*, indicating that any reduction of Lipophorin protein in these lines is not contributing substantially to the transmission blocking of *Pb-PfCSP*.

My attempt to assess the transmission-blocking capacity of Lp::sc2A10 heterozygotes was inconclusive, due to the small scale of the experiment and abnormal results in one of the controls. This is an important experiment to replicate in order to model the transmission-blocking potential of this line in a mixed population. Beyond quantification of Lipophorin, more in-depth fitness studies should be performed on Lp::sc2A10. While neither Eric Marois nor I observed significant differences in developmental time, egg-lay, or survival, these observations are based on our experience rearing the experimental populations for bite-backs and maintaining the lines. The fertility costs of *Lp::sc125* were detected as soon as the lines were established and the sex-bias clear when I began performing bite-back experiments. However, as the crosses to Lp-GD illustrate, there could be additional fitness problems. Given that the ultimate goal of this project is to develop transgenic lines that could be released in the field, precise measurements of fitness are important to model how an indirect gene drive based on this construct might spread. Over time, both transgenes were steadily lost from caged populations, though in each case only one replicate cage was followed and both

cages experienced bottlenecks during the move from the old to new insectarium and during adjustment to the new insectarium.

The failure to generate a functional gene drive targeting *lipopohrin* is disappointing. Experiments are ongoing to understand which components of the drive are not functioning as desired. It is important to note that the U6 promoter I cloned in the Lp-GD is not the same U6 promoter that was used in conjunction with *zpg* by Hammond and colleagues (2018). While U6 promoters are canonically described as ubiquitous, both published reports (Port et al., 2014) and observations in our team demonstrate that this *U6* promoter, which is used routinely in our group to generate mutant and knock-ins, is not active in all tissues (Raquel Mela-Lopez, unpublished). Additionally, given the results of crosses between LpGD and Lp::sc2A10, the utility of *lipophorin* as an indirect gene drive homing locus is diminished. It would be possible to drive the Lp::sc2A10 construct from an additional locus, for example *saglin* or *roadblock*. This would also open the door to include additional transmission-blocking strategies, via disruption of *saglin* and/or introduction of additional effector molecules in *roadblock*.

Chapter IV

General Discussion

Over the course of my PhD, I worked on establishing gene drives in two distinct pest species. In the case of *An. gambiae*, I was able to characterize the transmission-blocking capacities of two scFvs inserted into the *lipophorin* gene. One of these insertions, *Lp::sc2A10* is able to significantly reduce transmission. These results show that scFvs can be expressed from endogenous promoters without significantly perturbing local gene expression, provided cryptic splice junctions are not present, as was the case for sc125. Gene expression is a complex process involving not only promoter and terminator sequences, but also interactions of enhancers and repressor elements that can be located further away from the gene locus. An advantage of using endogenous promoters is that it avoids the difficulty of identifying full functional promoter regions and alterations in expression patterns that can occur when a promoter is introduced into a new genomic locus. Given the small size of scFvs, it could be possible to express multiple scFvs from the same locus, thereby targeting several parasite antigens to limit the appearance of resistance in the parasite.

It is important to note that the levels of infection observed under laboratory conditions are much higher than what is usually observed in the field (Whitten et al., 2006), therefore the transmission-blocking capacity of *Lp::sc2A10* and other modified refractory lines is likely to be higher under field conditions. Conversely, transmission-blocking lines need to be tested against field *P. falciparum* isolates prior to release, to ensure that their activity is conserved. Finally, it is crucial that transgenic mosquitoes released to combat one mosquitoes-borne disease do not have increased susceptibility to other diseases, or increase resistance to other mosquito control strategies. Only a few of the transgenic lines published have so far been assessed in this way. Pike and Dimopoulos tested ectopic-Rel2 mosquito lines against the

o'nyong'nyong virus, four insecticides, and different *P. falciparum* isolates (2018). They found no increased susceptibility to the virus, resistance to insecticides, or difference in transmission-blocking between the different *P. falciparum* strains. Mumford et al. tested the infectivity of *An. gambiae* lines expressing the I-*Ppo*I x-shredder transgene towards the o'nyong'nyong virus and one *P. falciparum* strain (2019). Again, no increased susceptibility to the virus or parasite was found between transgenic and non-transgenic mosquitoes. Of note, the specificity of scFvs to a specific pathogen-antigen makes this class of effectors less likely to alter the mosquito's response to non-target pathogens than modification of a general immune pathway.

The indirect gene drive design proposed in this paper and modeled by Beaghton et al, (2017) remains to be tested in mosquito populations. Theoretically, this strategy represents an efficient way to introduce a novel gene in a population while guarding against the appearance of resistant mutations. It can also be advantageous that components of the CRISPR/Cas9 system are only transiently present in the wild population. It is an interesting strategy that I hope will be explored in living mosquitoes. At the moment, the theoretical gene drive literature is much more extensive than engineered examples; many of the approaches are intriguing, but they must be engineered and tested in living organisms in order to contribute substantially to the development of applied gene drives.

My work in *An. gambiae* benefited from the well-established protocols already in place in our team for generating CRISPR/Cas9 transgenics. In the case of *D. suzukii*, I was not able to build upon the wealth of prior transgenics experience in the lab or an extensive transgenic literature. In retrospect, attempting to directly knock-in a large gene drive was perhaps too ambitious an approach and more success could have been achieved by beginning with

optimizing a knock-in protocol of smaller cassettes before building up to a full gene drive. This is a consideration in future efforts to establish gene drives in non-laboratory organisms.

The advent of targeted genetic engineering has revolutionized many fields of science and insect control is no different. The development of CRISPR-GD in insects has been rapid, particularly in *Anopheles* mosquitoes. This is exciting, as gene drives represent a way to exploit decades of research on malaria transmission by mosquitoes to achieve the goal of disease reduction. However, the rapidity of gene drive development is also outstripping current regulations, raising urgent questions about how and when these systems should be employed. Since the first CRISPR-GD papers were published, there have been numerous national and international discussions and hearings in an effort to develop guidelines for the use of this technology. While some researchers feel that the risks of CRISPR-GD merit a complete moratorium, so far no official authority has approved a ban. Most public research bodies and agencies have given tempered approval to GD development, citing in particular the potential benefits to human health that the technology could provide (Callaway, 2018; NAS, 2016).

The public response to CIRSPR-GD has been similar, though extensive surveys are limited. A recent survey in the US found more support for GD to control disease vectors compared to other potential applications such as invasive species eradication or agricultural pest control (Jones, Delborne, Elsensohn, Mitchell, & Brown, 2019). Those surveyed were also more positive towards GD development and deployment by public sector organizations such as universities and the US Department of Agriculture, compared to private sector organizations. In terms of the potential side-effects of gene drives, respondents were most concerned about the effects of a gene drive on human health and the potential environmental costs due to

removal of a pest species. Similar concerns were raised in problem formulation workshops held at several locations in Africa, where respondents raised concerns that released mosquitoes could have a negative impact on human health and the environment (Teem et al., 2019).

These are important concerns when developing gene drives and deploying them in the field. The history of insect control provided examples of successful programs abandoned due to public backlash or economic limitations. Fears of bioterrorism ended a trials of SIT and translocation mosquitoes in India, while control of the Sheep Blowfly in Australia was ended due to lack of investment by local industry partners (Gould & Schliekelman, 2004). The recent backlash in Brazil following reports of DNA from Ocitec RIDL strains being detected in local *Ae. aegypti* highlights how tenuous public acceptance of GMO insects can be (Evans et al., 2019; Servick, 2019). Although the claims in the paper have been disputed and their conclusions critiqued as overly alarmist, the initial alarm generated by the report has already done damage.

There are currently no comprehensive laws or regulations that sufficiently govern gene drives. This places enormous responsibility on those involved in gene drive research – researchers, institutions, and funding agencies – to ensure a sufficient level of rigor and reflection. Many of the researchers leading the way on gene drives have been active in calling for stronger guidance and emphasizing the need for openness in order to sustain public trust (Oye et al, 2014, regulating gene drives; Kofler, 2018). They have also made efforts to include other involved organizations, governments, and local communities in their projects. The WHO in collaboration with researchers and other organizations has proposed a plan for phased testing for field trials, and several groups are laying the ground for field trials. These trials will be

valuable not only in continuing to develop gene drives, but also in developing the regulatory structure that will be necessary as the field develops. It remains to be seen how CRISPR-GDs preform in the field, but given the current state of research, they could be a valuable tool for improving human health and prosperity. However, when it comes to genetic modification, who is doing the modifying and how they are being monitored will matter as much or more than the ultimate purpose of the modification. Once outside of the lab, the application of CIRSPR-GDs will become a matter of risk, reward, and public trust.

Methodology

Insect rearing conditions

All insect work and experiments using mice were performed in the insectary of the Institut de Biologie Moléculaire et Cellulaire at the University of Strasbourg. The facilities and mice protocols have been certified by the regional veterinary services (authorization N° F67-4822) and by the national ethics committee in animal experimentation (authorization for project APAFIS-20562-2019050313288887 v3).

Anopheles gambiae maintenance

Mosquito colonies were reared at 27 ± 2°C with a 75% relative humidity and a 12 hour day/night cycle. Females were blood fed for 15 minutes on CD1 mice anaesthetized by i.p. injection of 8.5 mg/kg of xylazine (Rompun) and 42.5 mg/kg of a mix of tiletamine and zolazepam (w/w 1:1, Zoletil) diluted in saline solution (85 ul of the mix injected per 10 g). Two days later, an egg-laying dish containing a wet filter was placed in the cage. Newly hatched larvae were collected two days later and placed into container with deionized water. Larvae were fed two times a day on powdered fish food (TetraMin). Pupae were collected from the larval pans and transferred to mosquito cages prior to emergence. Adult mosquitoes were fed a 10% sucrose solution from either cotton pads placed on the top of cages or from homemade feeding dispensers inside the cage.

Drosophila suzukii maintenance

Wild caught *D. suzukii* collected in June 2015 from cherry trees and blackberry bushes in Nancy and Illkirch-Graffenstaden, France were used to establish a laboratory colony at the IBMC in Strasbourg. The flies were raised at room temperature on a standard simple

cornflower medium used for *D. melanogaster*. As an added confinement, vials containing flies were stored in plastic containers with a mesh lid and flies were anesthetized with CO_2 prior to transfer or manipulation.

Preparation of plasmids for microinjection

Multi-gene cassettes for microinjection were prepared using the Golden Gate assembly technique (Engler et al., 2009). This technique allows multiple DNA fragments excised from donor plasmids to be simultaneously cloned into the same destination plasmid. It is based on the activity of type IIS restriction enzymes, which cut outside their recognition sequence. The protocol used in this project the *Bsa*I enzyme, which recognizes the sequence 5'-GGTCTC(1/5)^, producing 4 nt overhangs. The order of the inserts in the final plasmid is determined by the sequence of the overhangs.

Donor plasmid preparation

Individual DNA inserts were first PCR amplified from genomic or plasmid DNA using Phusion High-Fidelity Polymerase (Thermo) using primers with the required *Bsa*I site and four-nucleotide overhang at their 5' end. PCR products were run on a 0.8% TAE gel and purified using either a homemade spin column for rough extraction or the NucleoSpin Gel and PCR Clean-Up Kit (Machery-Nagel) for a clean preparation. Rough extraction was performed as follows: a hole was poked into the end of a 0.5 mL Eppendorf tube and a small piece of sterile polypropylene fiber packed at the bottom. The agarose gel slice containing the PCR product was placed in the tube. The 0.5 mL tube was placed in a 1.5 mL Eppendorf tube, and the tubes were centrifuged at maximum speed for 1.5-3 min, until all the liquid had been completely separated from the agarose.

The purified PCR product was cloned into either the *Sma*I or *Eco*RV site of a modified pBluescriptKS plasmid that lacks a *Bsa*I site (pKSB$^-$). Cloning was performed using a restriction-ligation reaction consisting of 1-2 µL PCR product, 10 ng pKSB$^-$, 0.6 µL restriction enzyme, 0.4 µL (2U) T4 DNA ligase, 1 µL 10x enzyme buffer, and 1 mM ATP, with a final volume of 10 µL. The reaction was incubated for 2-16 hours at room temperature, followed by inactivation of the enzymes at 65°C for 20 minutes. After inactivation, an additional 0.5 µL of restriction enzyme was added to re-open empty pKSB$^-$ vectors. All enzymes were purchased from ThermoFisher.

For several inserts, the PCR product was cloned into the pJet or pTOPO plasmids using the CloneJet PCR cloning kit (ThermoFishcer) or TOPO XL Cloning Kit (Invitrogen).

A complete list of the primers and donor plasmids used in the project can be found in the Annex.

sgRNA expression plasmid preparation

Generic U6-sgRNA or U6-tRNA-sgRNA donor plasmids were ordered as synthetic gBlocks (Integrated DNA Technologies) and cloned into pKSB$^-$ using the above protocol. Derived modules with altered *Bsa*I overhangs were prepared by PCR amplification of the insert, followed by cloning into pKSB$^-$. In these sgRNA-expression scaffolds, two inverted *Bbs*I restriction sites allow the cloning of target-specific nucleotides provided as short linkers (see below). Full sequences of the *An. gambiae* and *D. suzukii* sgRNA expression modules are provided in the Annex.

sgRNA target site selection and plasmid preparation

sgRNA target sites were identified manually using SeqBuilder or the online tool CRISPOR (Version 4.0 or later: http://crispor.tefor.net/). The sgRNAs were ordered as two reverse complement oligonucleotides (Integrated DNA Technologies) with 4 nt overhangs. A 10 μM stock linker solution was prepared by mixing 5 μL of each oligo of a pair (100 μM) in a final volume of 50 μL. The mixture was denatured at 95°C for 5 min in a heat block, then cooled to room temperature. The annealed oligos were cloned into the *Bbs*I sites of sgRNA expression donor modules under the control of a U6 promoter from either *An. gambiae* (AGAP013557) or *D. suzukii*. 1 μL of annealed oligo was mixed with 10 ng of linear, pre-*Bbs*I digested plasmid, 1 μL T4 Ligase (Thermo Scientific), and 1 μL 10x T4 ligase buffer (Thermo Scientific), in a final volume of 10 μL. The reaction was incubated for 30 minutes at room temperature.

A complete list of oligonucleotides used to make the sgRNA linkers can be found in the Annex.

Bacterial transformation

2.5 μL of restriction-ligation product was used to transform 25 μL of chemically competent DSH5α *E. coli* prepared using either calcium chloride or the Mix and Go competent cell preparation kit (Zymo Research). Transformed *E. coli* were plated on LB + Ampicillin plates with the addition of 20 μL of IPTG and 40 μL of X-Gal, and incubated overnight at 37°C. Putative positive colonies were identified by blue-white screening and colonies with the correct insert were identified by colony PCR using GoTaq Green Master Mix (Promega). Positive colonies were grown overnight at 37°C in 5 mL of LB+Amp and plasmids were purified using the NucleoSpin Plasmid EasyPure Kit (Machery-Nagel). Plasmids were digested with *Bsa*I or *Bbs*I to confirm the presence of an insert prior to sequencing by GATC.

Golden Gate reaction

40 fmol/µL stock solutions of donor plasmids and the destination plasmid pENTR-ATCC-LacZ-GCTT were prepared using the formula: ([DNA in ng/μl] x 1520)/plasmid size in base pairs. Golden Gate reactions were performed in a final volume of 20 µL with 2 mM ATP, 2 µL enzyme buffer (NEB), 1 µL *Bsa*I (NEB) or *Eco31*I (ThermoFisher), 1 μl T4 DNA Ligase (Thermo Scientific), 1 µL pENTR-ATCC-LacZ-GCTT, and 1 µL of each donor plasmid. The assembly reaction was performed using a thermocycler with the following program: five cycles of 37°C, 10 min, 20 °C 10 min; 50 min at 20°C, 20 min at 70°C, 12°C hold. After the reaction, 0.5 µL BsaI, 0.5 µL enzyme buffer in a final volume of 5 µL was added to digest any reconstituted empty destination plasmid.

In cases where the assembly of the desired cassette was too complex to assemble directly, the reaction was divided into two and a two-step process was used. If this proved unsuccessful, sub-modules were prepared by PCR amplifying multiple inserts from partial Golden Gate reactions and cloning into pKSB$^-$ as described above.

5 µL of Golden Gate product was used to transform 50 μl of competent DH5α following the protocol described above, with the following changes: kanamycin was used as a selection marker and colony PCR was performed using primers that spanned two or more DNA inserts. A diagnostic digestion of plasmids was performed prior to sequencing. Positive plasmids were re-grown in 50-100 mL LB+Kana and purified using the NucleoBond Xtra Midi EF Kit (Machery-Nagel)

A complete list of the donor plasmids used to make the multi-cassette plasmids used in this PhD project can be found in the Annex.

Transgenesis

Selection of candidate female fertility genes for D. suzukii

A complete genome of *D. suzukii* was published in 2013 and can be accessed via the website http://spottedwingflybase.oregonstate.edu/ (Chiu *et al*., 2013). A list of candidate genes was compiled by searching the *D. melanogaster* genome database FlyBase (version FB2015_05 or later) for genes whose disruption produces a recessive female sterile phenotype. Candidate gene sequences were then used to search the *D. suzukii* database (version 1.0) to identify putative homologues. The sgRNA target regions were PCR amplified and sequenced to check for polymorphisms within the sgRNA targets. The following primers were used:

cup	ctcgtgataggaactcctcgt gctgctgcttgtgctgc
stil	cgttctcgaaagctagaaggtt cagctcgtggctttcgt
ylk region 1	gcaagaagcccaaggtga ccgtccgtcgcactcc
ylk region 2	ggaccaggacgctggtctcgt caaactgcgaggcgaagaagtgtc

Collection and alignment of D. suzukii eggs

The day before egg collection, *D. suzukii* adults were transferred to a collection cage and supplied with an apple agar plate supplemented with yeast paste. 30 minutes before egg collection, the flies were briefly anesthetized by cold and the agar plate was replaced with a fresh plate supplemented with cherry juice and yeast paste, and covered with a Whatman paper circle (No. 1). After 30 min, the collection plate was exchanged and the eggs were removed from the paper using a paintbrush. The eggs were briefly rinsed in 1xPBS and aligned on a coverslip. The eggs were allowed to dry to the point at which they stuck to the glass slip, and then covered with organic olive oil.

Collection and alignment of An. gambiae eggs

Two to three days before injection, the G0 *An. gambiae* cage was blood-fed. On the day of injection, a fresh egg dish was placed and left until eggs began to darken. Eggs were aligned against a nitrocellulose membrane held in place with filter paper humidified with demineralized water. The genetic background of all mosquito lines used in this study is an N'Gousso strain that has been selected for high susceptibility to *Plasmodium* (S1High, *Blandin* et al., unpublished)

Microinjection of An. gambiae and D. suzukii eggs

Eggs were injected in the posterior pole using a FemtoJet 4X injector (Eppendorf). Eggs were injected with 400-500 ng/μL of plasmid DNA in endonuclease-free H_2O. For knock-in experiments, the injection mix was supplemented with 1 μM of Scr7, a putative inhibitor of Ligase IV (Chu *et al*., 2015). For injections in *D. suzukii*, the knock-in plasmid served as a source of Cas9, with or without 240 ng/μL of recombinant Cas9 protein (DNA Bio Inc) was included in the injection mix with 120 ng/μL of sgRNA (IDT). In *An. gambiae*, knock-in plasmids were injected in to *vasa*-eSpCas9 eggs, with the plasmid serving as a source of sgRNA.

Eggs were injected with homemade quartz needles (0.5 cm, WPI) pulled on a P-2000 micropipette puller (Sutter Instruments). Needles were beveled using a BV-10 micropipette beveller (Sutter Instruments), at an angle of 22.5° for *An. gambiae* and 45° for *D. suzukii*.

Egg treatment and transgenic screening for D. suzukii

After injection, the coverslip containing injected *D. suzukii* eggs was drained of oil. The coverslip was either placed into an apple agar plate or pushed into a tube containing cornmeal media until the eggs were near to the media. Surviving G1 was screened as larvae or pupae for transient expression of the fluorescent reporter. Larvae that showed transient expression were out-crossed to wild-type flies in small batches (2-3 transgenic flies per vial). Non-transient expressing larvae were self-crossed. G1 larvae or pupae were screened for fluorescent marker expression.

PCR screens for NHEJ deletions in *cup* and *stil* were performed using the Phire Animal Tissue Direct PCR Kit (Thermofisher) following the manufacturer's instructions. Adult G1 *D. suzukii* were anesthetized by cold prior to DNA extraction and were screened in pools of eight. The following primers were used:

cup	tgtcgttcttcgctcgtgatagga
	ggtgaactgcagtacagccgtgat
stil	tccttttccttcgttctcg
	gctggaacaccgttttgatt

Egg treatment and transgenic screening for An. gambiae

Slides containing injected *An. gambiae* eggs were placed at an angle in a small pan of water as illustrated in Volohonsky et al., (2015), with the end of the filter most distant from the eggs in contact with the water to ensure the eggs remained moist. Two days after injection, the eggs were washed into the water and hatched. Surviving injected G0 adults were outcrossed and G1 were manually screened as young L1 larvae for expression of the fluorescent transgenesis marker.

COPAS sorting

Un-fed, newly hatched L1 *An. gambiae* larvae were analyzed and sorted on a COPAS SELECT (Union Biometrica) using the Biosort 5295 software. For population analysis, total larvae for each genotype were either counted directly on the Biosort read-out or using the WinMDI software.

Plasmodium berghei infections and bite-back

Parasite strains

A *Plasmodium berghei* strain in which the endogenous *CSP* gene was exchanged for that of *P. falciparum*, *Pb-PfCSP*, *eif1*-GFP (Triller et al., 2017), was provided by Shahid Khan and Chris Janse (University of Leiden, The Netherlands). To generate a Pb-*Pf*CSP strain with stronger GFP expression, a mouse was co-infected with *Pb-GOMO14* (Manzoni et al., 2014) and *Pb-PfCSP*, *eif1a*-GFP at a 1:20 ratio. Naïve mosquitoes fed on the mouse to allow the hybridization of the two parasite strains. 17 days later, the infected mosquitoes carrying strongly GFP positive sporozoites in the wings were selected to bite a new mouse. Infected mouse blood was collected in PBS when the parasitemia reached 0.1% and 3000 red blood cells containing strong GFP expressing parasites were purified by FACS (IGBMC cell sorting platform) and injected into a new mouse. Eleven days later, the mouse became positive and its blood was passaged to a new mouse. Infected mouse blood with a parasitemia of 0.4% was diluted in PBS to contain approximately 1 parasite per 150 μl. For parasite cloning, 11 naïve mice were each injected with 150 μl. All mice became positive with strongly GFP-expressing parasites, and frozen parasite stocks were prepared and genotyped by PCR using primers specific for the *Pf*CSP allele. Three putative *Pb-Pf*CSP, *PbHSP70*-GFP parasite clones were determined to harbor only the *P. falciparum* CSP gene and were used for subsequent bite-back infections.

Mosquito infection

The mosquito genetic background used in all infection experiments is N'Gousso from Cameroon. The parental line for all infections is a derived N'Gousso line selected for high susceptibility to *P. berghei*, S1High (S. Blandin et al., unpublished). Mixed batches of homozygous transgenic mosquitoes and their negative siblings were selected by COPAS and raised together. Mixed cages of 4-7 day old mosquitoes were allowed to feed on infected mice. Infected cages were kept at 21°C, 60% relative humidity on a 12/12 day/night cycle. Non-blood fed females and males were removed one day after infection.

Bite-Back

One-day before the bite-back, transgenic and non-transgenic females were separated and placed into cups in batches of 10-12. Mosquitoes infected with *Pf*CSP-*hsp*-GFP *P. berghei* were additionally sorted visually by the presence of GFP fluorescent oocysts in the abdomen and sporozoites in the wing. On the day of the bite-back, one anesthetized mouse was placed per cup and the mosquitoes were allowed to feed for 15-30 minutes. The cups were regularly checked to verify that the mosquitoes were feeding and the mice were moved to promote probing. Mice infection was monitored starting from day 4-5 by either smear or FACS for two weeks or until the mouse became infected, at which point it was sacrificed.

FACS analysis

1-5 µL of blood was taken from anesthetized mice and mixed with 5 µL of heparin. The samples were diluted in 1 mL 1xPBS, then a 200 µL aliquot was diluted again in 1 mL 1x PBS. The samples were analyzed on an Accuri FACS system (BD).

Molecular biology

Separation of Lp::sc125 and Lp::sc2A10

DNA from single family-founder females was extracted by crushing mosquitoes in a home-made grinding buffer consisting of 0.1 M NaCl, 0.2 M Sucrose, 0.1 M TRIS (pH 9.2), 0.05 M EDTA, and 0.5% SDS. After crushing, the homogenate was incubated for 30 mintues at 68°C, followed by addition of 7 μL 8 M KAc, and incubation for 30 minutes on ice. The mixture was centrifuged at maximum speed for 10 minutes and 1 μL of supernatant was diluted in 29 μL of water. 1 μL of DNA was used in a PCR reaction using the 2xGoTaq green mix and the following primers to amplify the scFv insert:

ggtctctagaaacgaccacaggagtcttctgcctcccagcacccacatcgttccgtgtcg
ggtctcattctcgtcagcttggtcctcatccagtccgtcagcgcagcccagatccagctggtgc

The PCR reaction was digested with *Sac*I and *Xho*I, which cut specifically in sc2A10, and the total reaction was run on a gel.

RT-PCR

Larvae or adult *An. gambiae* were collected directly in 600 μL Trizol Reagent (Invitrogen) and crushed with ceramic beads in a Precellys 24 homogenizer (Bertin Technologies) using the following parameters: 5,500 rpm for 2 x 25 seconds with a 10 second interval. RNA extraction was performed using the Direct-zol RNA MiniPrep kit (Zymo Research). Final RNA concentration was measured on a NanoDrop One system (Thermo Scientific). DNA was removed using the RapidOut DNA removal kit (Thermo Scientific) and samples were diluted to a final concentration of 50 ng/μL. 8 μL of DNase-treated RNA was used for reverse transcription. Reverse transcription was performed using the RevertAid H minus Reverse Transcriptase with random primers and RiboLock RNase Inhibitor (Thermo).

PCR amplification of the Lp::scFv transcript was done using the GoTaq Green Master Mix (Promega) with 1 μL of cDNA, and primers spanning the exon 1-exon 2 junction:

ccatgttgaactgtaaggtctagt
ccatgttgaactgtaaggtctagtgaacagaaca

Total PCR product was run on a gel and the PCR products were cut out and purified using the NucleoSpin Gel and PCR Clean-Up Kit (Machery-Nagel). The PCR products were sequenced by GATC Biotech using the same primers used for amplification. Sequencing results were analyzed using Sequencher.

Hemolymph collection

Adult female *An. gambiae* were anesthetized with CO_2 and transferred to ice. Their proboscises were cut and forceps were used to gently squeeze the abdomen, forcing out a drop of hemolymph, which was collected from the proboscis directly into 1x Laemmli buffer. Protein samples were boiled at 95°C for 5 minutes and stored at -20°C until use.

Coomassie gels

Hemolymph samples from 25 mosquitoes in 1x Laemmli buffer were run on pre-cast Mini-PROTEAN TGX Stain-Free Precast gels, 4-15% (BioRad) with PageRuler Plus Prestained Protein Ladder (Thermo Scientific). The gels were stained with Coomassie dye (Invitrogen) for 1 hour, then destained in water.

Lipophorin western blots

Hemolymph samples from 6-10 mosquitoes in 1x Laemmli buffer were run on pre-cast Mini-PROTEAN TGX Stain-Free Precast gels, 4-15% (BioRad) with PageRuler Plus Prestained Protein Ladder (Thermo Scientific). The proteins were transferred to a methanol-activated

PVDF membrane (GE Healthcare) using a Pierce G2 Fast Blotter (Thermo Scientific) and blotting paper soaked in 1-Step Transfer Buffer (Thermo Scientific). Membranes were blocked in Pierce Protein-Free T20 (PBS) Blocking Buffer (Thermo Scientific) for one hour at RT and incubated for 1 hour at RT or overnight with primary antibody diluted in blocking buffer. Membranes were washed in PBS-T, and then incubated with secondary antibody fused to horseradish peroxidase for 1 hour at room temperature. The following primary antibodies were used: mouse anti-ApoIILp 2H5 (1:4000), mouse-anti ApoILp 2C6 (1:4000) (Rono et al., 2010), rabbit anti-PPO2 (1:10000) (Fraiture et al., 2009). For ApoLpI/PPO2 western blots, the membrane was cut at the 130 kDa marker after blotting and the upper and lower molecular weight regions were probed independently with ApoLpI and PPO2, respectively. Membranes were washed and visualized using the SuperSignal West Pico PLUS kit (Thermofisher) and imaged using a ChemiDoc Imaging System (Biorad).

Mass spectrometry analysis of hemolymph

Hemolymph from 25-30 female mosquitoes was collected directly into 1x Laemmli buffer 48 hours post-blood feeding using the proboscis clipping method. The Proteomics Platform at the IBMC in Strasbourg performed the mass spectrometry sample preparation and analysis. Samples were precipitated and digested with trypsin, and 1/5 of the digestion product was analyzed on a Q Exactive Plus Mass Spectrometer coupled to an Easy-nanoLC1000 (Thermo). The acquired data was searched against the *Anopheles* UniProt database plus the sc125 and sc2A10 sequences using Mascot. The total number of spectra (spectrum counting values) were normalized using the combined abundance of common hemolymph proteins APL1C, Apolipophorin III, LRIM1, Nimrod, Prophenoloxidase, and TEP1 and divided by the protein molecular weight to estimate protein abundance.

Résumé en Français

Introduction Générale

Parmi les millions d'espèces d'insectes sur notre planète, seulement une petite fraction cause des dommages aux humains. Les ravageurs agricoles détruisent les cultures, provoquant des pertes économiques et contribuant à l'insécurité alimentaire. Les vecteurs de maladies transmettent des agents pathogènes à l'homme et aux animaux. Prises ensemble, ces espèces ravageuses représentent une énorme charge pour l'humanité.

Actuellement, la plupart de ces insectes sont contrôlés par des pesticides. Mais l'utilisation élevée de ces produits chimiques a entrainé l'apparition et la propagation de résistance. De plus, la prise de conscience croissante de l'impact négatif des insecticides sur l'environnement défavorise de plus en plus leur utilisation. Enfin, les mouvements humains et le changement climatique contribuent à l'émergence de nouveaux insectes ravageurs, pour lesquels la lutte est actuellement limitée.

Tous ces faits soulignent l'importance de développer de nouveaux moyens de lutte contre les insectes. Une technique intéressante est l'utilisation des outils génétiques. Contrairement aux pesticides, ces outils sont spécifiques à l'espèce ciblée et leur effet peut durer plusieurs générations. Ils peuvent en outre être utilisés soit pour diminuer la taille de la population, comme des pesticides, soit pour la modification spécifique d'une population, par exemple l'introduction de gènes qui bloqueront la transmission d'une maladie. Ce dernier type d'outil permettrait de garder une espèce native dans son écosystème naturel en réduisant son impact négatif pour l'homme.

Le forçage génétique est un phénomène naturel qui permet à des gènes de surmonter les lois de l'hérédité de Mendel et être transmis à la plupart de leurs descendants. Il existe plusieurs mécanismes de forçage génétique, mais l'un des plus simples et mieux connus est la stratégie utilisée par les endonucléases de *homing* (homing endonuclease genes - HEG). Ce sont des gènes qui expriment une endonucléase qui coupe l'ADN double brin à un endroit précis sur le chromosome homologue. Lors de la création d'une coupure double brin, des mécanismes de réparation naturelle de la cellule s'activent. Si la cassure est réparée par la voie de recombinaison homologue, le HEG sera copié sur le chromosome coupé. L'insertion du HEG interrompt le site de reconnaissance de l'endonucléase, rendant le chromosome qui porte le HEG résistant à cette enzyme. Par ce mécanisme, une cellule hétérozygote pour le HEG devient une cellule homozygote. En revanche, la cassure peut aussi être réparée par la voie de réparation par jonction d'extrémités non-homologues. Dans ce cas, les deux extrémités sont jointes ensemble, provoquant potentiellement de petites mutations qui peuvent modifier le site de reconnaissance du HEG et rendent le chromosome insensible à l'action de l'endonucléase.

Le système de modification génétique CRISPR/Cas9 peut être utilisé pour créer des gènes de forçage HEG artificiels. En 2015 le premier gène de forçage basé sur CRISPR/Cas9 a été publié, suivi rapidement par d'autres exemples. Ces exemples prouvent que le forçage génétique par CRISPR/Cas9 fonctionne chez les eucaryotes, y compris les insectes, mais des difficultés sont apparues avec cette stratégie, notamment la formation d'allèles résistants à Cas9. Ce sont des mutations dans la région d'ADN reconnue par Cas9 qui trouvent leur origine soit dans la diversité génétique préexistante, soit créées par l'activité de Cas9 elle-même.

L'objectif de mon projet de thèse était de créer des systèmes de forçage génétique par CRISPR/Cas9 dans deux espèces. L'une, la mouche *Drosophila suzukii,* est une espèce invasive actuellement présente en Amérique du Nord et du Sud et en Europe, y compris en France. Cette espèce infeste les baies et fruits à peau fine et constitue une menace potentielle pour l'agriculture. Mon but était de créer des forçages génétiques capables de diminuer la taille de sa population en ciblant des gènes essentiels à la fertilité des femelles. L'autre espèce, *Anopheles gambiae*, est le vecteur majeur du paludisme en Afrique. Le paludisme est provoqué par des parasites du genre *Plasmodium* qui sont transmis par des moustiques femelles lors d'une piqûre. En 2017, il y avait environ 219.000.000 de cas de paludisme, dont 90% en Afrique, et 435.000 morts, dont 61% d'enfants de moins de 5 ans. L'objectif de ce projet était de créer des gènes de forçage qui bloquent la transmission de ces parasites par les moustiques.

Vers la création d'un système de forçage génétique à but d'élimination chez *D. suzukii*

Introduction

D. suzukii représente une énorme menace potentielle pour l'agriculture. Les statistiques actuelles sur les coûts sont limitées, mais les pertes de récoltes se situent généralement entre 20 et 40 %, bien que des pertes allant jusqu'à 100 % aient également été signalées. En plus des pertes de revenus dues aux pertes de récoltes, *D. suzukii* peut imposer des coûts supplémentaires en raison de la nécessité de mettre en œuvre des stratégies de lutte intensives supplémentaires. Actuellement, les techniques de contrôle efficaces contre *D. suzukii* sont limitées. Contrairement à la plupart des autres espèces de drosophiles, les femelles pondent dans les fruits mûrs. Une fois dans le fruit, la larve est protégée contre certains outils de contrôle, tels que les pesticides. De plus, la courte période entre la ponte et la récolte impose

des restrictions à l'application de pesticides. D'autres interventions sont en cours de développement, tel que contrôla lutte biologique avec le la guêpe parasitoïde *Trichopria drosophilae* et la technique de l'insecte stérile. Cependant, un système de forçage génétique CRISPR/Cas9 pourrait constituer une stratégie d'intervention supplémentaire puissante.

Résultats et discussion

Pour concevoir un système de forçage génétique qui rendrait les femelles stériles, nous avons sélectionné trois gènes cibles candidats en nous basant sur l'homologie avec des gènes connus pour être essentiels à la fertilité femelle chez *D. melanogaster*. Nous avons conçu un mécanisme de forçage génétique pour perturber simultanément deux gènes à la fois - l'un par l'insertion du mécanisme de forçage génétique dans le gène (voie de recombinaison homologue) et l'autre par des délétions dans le gène cible à distance (voie de jonction d'extrémités non-homologues).

Initialement, nous avons essayé d'insérer directement le système de forçage génétique dans le génome de la mouche *suzukii* avec CRISPR/Cas9, mais nous n'avons pas obtenu de mouches transgéniques, probablement à cause de la grande taille de ces cassettes (8.5 kb), en plus de 4 kb de brins d'homologie, faisant de l'insertion un évènement trop rare. En utilisant une stratégie semblable, Gantz et ses collègues ont dû cribler plus de 25.000 larves d'*Anopheles stephensi* pour trouver 2 larves transgéniques (Gantz et al, 2016). Le taux de fécondité plus faible chez *D. suzukii* rend un tel criblage difficile.

Nous avons donc décidé de modifier notre stratégie et d'insérer les systèmes de forçage génétique en utilisant l'approche d'*échange de cassettes médié par la recombinase* en deux étapes - d'abord en insérant une cassette à double site *att*P en utilisant CRISPR/Cas9 et ensuite, dans une deuxième étape, en insérant la cassette de forçage génétique flanquée de sites *att*B. Chez la mouche *D. melanogaster*, l'efficacité de la transgénèse par CRISPR/Cas9

était augmentée en utilisant des lignées de mouches qui expriment la protéine Cas9 dans leurs œufs. Nous avons décidé de tester cette stratégie pour *D. suzukii* et commencé par créer des mouches *D. suzukii* qui expriment Cas9 sous le control d'un promoteur maternel. Pour y parvenir, nous avons utilisé un transposon piggyBac contenant Cas9 sous le contrôle du promoteur *nanos* et un marquer fluorescent sous le contrôle du promoteur fort *polyubiquitin* (*Ubi*). Ces deux promoteurs venaient de *D. suzukii*. Après l'injection de ce plasmide, nous avons isolé 3 lignées transgéniques qui expriment fortement la protéine fluorescente. Cependant, aucune de ces lignées n'a montré de signes d'expression de Cas9, peut-être à cause de la sélection d'une mauvaise région promotrice.

Après avoir travaillé sur *D. suzukii* pendant plus de deux ans, il était décevant de ne pas obtenir une véritable lignée d'expression de Cas9 avec laquelle faire avancer le projet. Le développement d'outils transgéniques chez *D. suzukii* s'est déroulé plus lentement que prévu. De plus, à ce moment, le laboratoire a déménagé dans le nouvel insectarium de l'IBMC. Les systèmes de confinement de cette section n'étaient pas suffisants pour *D. suzukii*, car les filets d'exclusion primaires couvrant les évents ne sont pas assez petits pour empêcher une mouche de s'échapper (bien que les filtres HEPA dans le circuit en aval soient conçus pour contenir les mouches). Enfin, le projet *Anopheles* avait atteint un point qui nécessitait un investissement complet de temps et d'énergie. Pour ces raisons, le projet *D. suzukii* a été mis en attente.

Mon travail avec *D. suzukii* illustre les difficultés qui peuvent être rencontrées lors de l'établissement de la transgénèse chez une nouvelle espèce, même une espèce proche d'un organisme modèle de laboratoire. L'identification de promoteurs endogènes fonctionnels est un challenge important pour ces études. Mais malgré la non-obtention d'un forçage génétique

chez *D. suzukii*, en partant d'une espèce nouvellement domestiquée dans notre laboratoire, j'ai réussi à développer un protocole de transgénèse qui a produit plusieurs lignées transgéniques. De plus, quelques plasmides que j'ai créés pour ce projet ont contribué à d'autres projets dans le laboratoire. A l'heure actuelle, il y a très peu d'exemples de modification génétique chez *D. suzukii* par CRISPR/Cas9 et tous les mutants publiés ont été générés par injection de Cas9 sous forme d'ADN ou de protéine. Pour le moment, les outils génétiques et les stratégies de contrôle pour *D. suzukii* sont toujours limités, alors que la mouche continue d'envahir des nouvelles régions.

Le forçage génétique indirect chez *A. gambiae*

Introduction

En 2017, l'OMS a rapporté environ 219 millions de cas de paludisme et 435 000 décès (OMS 2018). 61 % des décès étaient des enfants de moins de cinq ans. L'Afrique continue de supporter le fardeau le plus lourd du paludisme ; 92 % des cas et 93 % des décès se sont produits dans les pays africains en 2017. Il est alarmant de constater que les progrès dans la lutte contre le paludisme sont au point mort depuis 2015 et les prévisions actuelles indiquent que le monde n'est pas sur la bonne voie pour atteindre les objectifs d'éradication de l'OMS.

Le paludisme est causé par des protozoaires parasites unicellulaires du genre *Plasmodium* qui sont transmis par les moustiques femelles Anophèle. Le cycle de transmission commence lorsqu'un moustique pique un humain infectieux. Le moustique ingère les parasites présents dans le sang. Une fois dans le moustique, les parasites passent par plusieurs stades de développement et migrent de l'intestin aux glandes salivaires. Lorsqu'ils arrivent dans les

glandes salivaires, les parasites peuvent être transmis à un autre humain lors de la prochaine piqûre par le moustique.

Depuis que le lien entre le paludisme et les moustiques a été identifié, le contrôle du vecteur est une partie importante des programmes de lutte et d'éradication du paludisme. Les premiers efforts ont porté sur la destruction de la forme larvaire par le drainage des sites de reproduction ou l'application d'huile ou de larvicides. Par la suite, des insecticides ciblant les formes adultes ont été découverts. Le premier d'entre eux a été le DDT, un insecticide très efficace et de longue durée, qui, combiné à la destruction des sites de reproduction, a permis à de nombreux pays occidentaux d'éradiquer avec succès le paludisme dans les années 1950-1970. Mais l'apparition d'une résistance au DDT et à d'autres insecticides limite l'efficacité à long terme des stratégies de lutte chimique.

Une autre stratégie de contrôle consiste à bloquer la transmission du parasite par le moustique. Pour être transmis à un nouvel hôte vertébré, le parasite Plasmodium doit réussir à voyager de l'intérieur de l'intestin moyen du moustique à l'intérieur des glandes salivaires. Pour cela, le parasite doit traverser trois barrières tissulaires - la matrice péritrophique, l'épithélium de l'intestin moyen et l'épithélium des glandes salivaires - tout en évitant d'être détruit par le système immunitaire du moustique. Le voyage du parasite implique des interactions nombreuses et complexes entre les facteurs du parasite et les facteurs de l'hôte du moustique, et la modulation ou la perturbation de ces interactions peut influencer la capacité vectorielle du moustique.

Une fois que les chercheurs ont commencé à comprendre ces interactions, ils se sont mis à chercher des moyens d’exploiter ces connaissances pour bloquer la transmission. Une stratégie consiste à moduler l'ensemble du système immunitaire des moustiques, soit en

surexprimant les antagonistes de Plasmodium, soit en réduisant ou bloquant l'expression des agonistes. Une autre stratégie consiste à exprimer de petits peptides antimicrobiens qui tuent les agents pathogènes. La dernière stratégie est d'exprimer des protéines qui bloquent des facteurs du parasites ou des moustiques importants pour le développement du parasite et sa migration dans le moustique.

Un outil clé pour perturber les interactions moustiques/parasites est l'utilisation d'anticorps monoclonaux (mAb) et les anticorps à fragment variable à chaîne unique dérivés (scFv). La découverte que les anticorps pouvaient bloquer la transmission est survenue dans les années 1970. Les chercheurs ont découvert que les anticorps dirigés contre les parasites ou les tissus de l'intestin pourraient bloquer la transmission lorsqu'ils sont administrés aux moustiques avec le sang infectieux. Les mAbs sont trop grands pour être exprimés transgéniquement. Par contre, leurs dérivés synthétiques, les scFvs, sont des protéines qui gardent la spécificité et la capacité de liaison des mAbs, et peuvent être insérés dans le génome de moustique. Plusieurs scFvs ont été développés qui bloquent la transmission de parasites chez les moustiques Anophèles. Dans ce projet de thèse, nous avons développé un système de forçage génétique basé sur l'activité de deux scFvs, sc125 et sc2A10. Ces deux scFvs reconnaissent la protéine CSP, qui est une protéine de surface du stade sporozoite de *Plasmodium falciparum*. Ces scFvs ont été déjà testés dans le moustique et démontrent une capacité importante de blocage de transmission.

Le forçage génétique indirect

Afin d'introduire ces scFvs dans une population des moustiques Anophèle, nous avons décidé de tester une stratégie que nous avons appellée «forçage génétique indirect ». Ce système implique deux constructions génétiques. L'une est un simple gène de forçage CRISPR/Cas9 qui cible un gène essentiel. L'autre contient les gènes effecteurs liés à une copie du gène

essentiel dans lequel les sites-cibles des sgRNAs ont été recodés pour ne pas être reconnus par Cas9. Lors de l'introduction des deux constructions dans une population, le système de forçage génétique réduira la fréquence du gène essentiel sauvage dans la population, ce qui favorisera la dissémination du gène effecteur. Comparé avec un système de forçage génétique « simple », où le gène effecteur est positionné dans la cassette avec Cas9 et les guides, le forçage génétique indirect à plusieurs avantages. Tout d'abord, les coûts de vigueur du système de forçage génétique n'affecteront pas le gène effecteur. Deuxièmement, les guides et Cas9 ne resteront pas indéfiniment dans la population, car ils représentent un allèle de perte de fonction du gène essentiel. Troisièmement, le gène de sauvetage ne sera jamais en situation de servir de modèle de réparation pour la réparation des DSB, ce qui pourrait autrement mener à la formation d'un allèle fonctionnel résistant. Enfin, la lignée de l'effecteur peut être évaluée en laboratoire et sur le terrain avant la libération du système complet de forçage génétique. Cette stratégie a été déjà proposée par des chercheurs, mais n'a encore jamais été testée dans une espèce.

Notre système de forçage génétique indirect est inséré dans le gène essentiel de la lipophorine, qui joue un rôle dans le transport des lipides et d'autres molécules dans l'hémolymphe des insectes. Le locus code pour deux protéines, qui sont exprimées sur le même précurseur polypeptidique. Les deux sous-unités sont par la suite clivées et secrétées dans l'hémolymphe. La séquence codant pour un scFv (2A10 ou sc125) a été insérée dans la *lipophorine* directement après le premier exon et la séquence endogène de clivage de la *lipophorine* a été dupliquée après le scFv (Figure 3A). De cette façon, le scFV sera exprimé dans le polypeptide et suivra la même voie de maturation et de sécrétion. Un marqueur GFP dans le premier intron du gène permet l'identification des moustiques transgéniques.

Résultats et Discussion

Caractérisation moléculaire des lignées Lp::sc2A10 et Lp::sc125

Les lignées transgéniques Lp::sc2A10 ou Lp::sc125 ont été établies et l'expression de scFv a été vérifiée par RT-PCR et spectrométrie de masse. Toutes les lignées sont viables et celles qui expriment 2A10 sont fertiles. En revanche, les femelles homozygotes pour sc125 sont stériles. De plus, il y a une différence significative des *sex ratios* aux stades pupe et adulte entre les mâles et les femelles homozygotes pour Lp::sc125. Les résultats du séquençage des produits de RT-PCR ont révélé un site d'épissage alternatif cryptique dans la séquence codant sc125 en amont du premier intron. L'épissage alternatif utilisant ce site enlève la partie C-terminale du sc125 et crée une fusion entre le début du sc125 et la première sous-unité de la lipophorin. Ce phénomène peut potentiellement expliquer le phénotype inattendu du sc125. Alternativement, ce phénotype peuvt être dû à une activité non-spécifique du sc125.

Tests de blocage de transmission

Pour évaluer l'effet du scFv, nous avons faits des tests de blocage de transmission. Ces essais consistent à infecter les moustiques avec *P. berghei*, parasite de rongeur, puis laisser les moustiques piquer des souris 16 à 20 jours plus tard. Ces souris sont suivies pendant deux semaines après piqûre afin de suivre le développement ou non de maladie. Puisque les scFvs utilisé dans cette étude reconnaissent spécifiquement la protéine CSP de *P. falciparum*, nous avons utilisé une souche du *P. berghei* qui exprime le CSP de *P. falciparum* (*Pb-PfCSP*). Cette souche exprime aussi la GFP, qui nous a permis d'identifier les moustiques infectieux grâce à la présence des sporozoites GFP+ dans l'aile des moustiques.

Confirmant les études précédentes, sc2A10 a réduit significativement la transmission de *Pb-PfCSP*. Seulement 29,7 % des souris exposées à *A. gambiae* Lp::sc2A10 sont devenues

malades, comparativement à 97,1 % des souris exposées aux moustiques contrôles. De plus, les souris qui sont devenues malades suite aux piqûres des moustiques *Lp::sc2A10* ont montré un retard d'un jour dans le développement de l'infection. Enfin, l'activité de sc2A10 est bien spécifique à la CSP de *P. falciparum*, car il n'y a pas de différence de transmission d'une souche *P. berghei* sauvage entre les moustiques *Lp::sc2A10* et les moustiques contrôles.

En revanche, sc125 n'a présenté aucune activité de blocage de la transmission ni aucun effet sur le développement de maladie. Cependant, les expériences sur la lignée Lp::sc125 ont été réalisées en utilisant des moustiques hétérozygotes. Compte-tenu de cela et de l'épissage alternatif qui semble avoir eu lieu à haute fréquence, la protéine sc125 fonctionnelle était probablement beaucoup moins abondante que la sc2A10.

Création d'un système de forçage génétique ciblant la lipophorine

Au vu de ses capacités de blocage de la transmission, Lp::sc2A10 est un bon candidat pour un système de forçage indirect. Comme décrit dans l'introduction, un système de forçage indirect comporte deux composantes : la lignée effectrice et un système de forçage indirect à but de suppression. La lignée *Lp::sc2A10* est une bonne candidate comme lignée effectrice, mais pour tester le système complet il était nécessaire de créer un système de forçage génétique qui cible la lipophorine.

La système de forçage génétique ciblant la *liphoroine* (Lp-GD) comprends quatre sgRNAs exprimés en utilisant le système de multiplex des tRNAs, la protéine eSpCas9 sous le contrôle du promoteur et terminateur du gène *zpg*, et une marqueur fluorescent rouge. Cette cassette a été injectée dans des œufs hétérozygotes pour Lp::sc2A10, afin de préserver une copie fonctionnelle de la lipophorine. A partir de ces injections, nous avons retrouvé deux

populations d'animaux transgéniques – l'une n'exprimant que le marqueur RFP et l'autre les marqueurs RFP et GFP. L'analyse de la deuxième population a montré que la cassette de forçage génétique s'est insérée avec le plasmide tout entier en amont du gène Lp::s2A10. L'autre population a incorporé le système de forçage génétique comme attendu, mais la descendance de ces individus ne montre aucune évidence d'héritabilité préférentielle. Afin de comprendre pourquoi le système ne fonctionne pas comme prévu, nous avons réalisé des croisements avec d'autres lignées transgéniques exprimant soit des sgRNAs (pour tester la fonctionnalité du *zpg*-eSpCas9), soit *vasa*-Cas9 (pour tester l'activité des sgRNAs dirigés contre la lipophorine).

Le système de forçage indirect est basé sur l'inactivation d'un gène essentiel et la présence d'un allèle de sauvetage. Pour un fonctionnement optimal, ces deux allèles en présence l'un de l'autre doivent donner un moustique viable et fertile. Malheureusement, la descendance d'un croisement entre la lignée Lp::sc2A10 et Lp-GD montre un phenotype sublétal : la majorité de la descendance n'atteint pas la stade adulte et la minorité qui y arrive souffre d'un défaut de vigueur. Ceci indique que les individus hémizygotes pour Lp::sc2A10 ont un coût de vigueur qui n'est pas évident en présence d'un allèle sauvage. De plus, d'autres expériences préliminaires visant à caractériser le phénotype de perte de fonction de la lignée Lp-GD présente un taux de mortalité élevé des larves hétérozygotes au cours du développement malgré la présence d'une copie WT de la lipophorine. Cela suggère que la perte de la lipophorin est partiellement haploinsuffisante, ce qui signifie que la lipophorine n'est pas un locus optimal pour un système de forçage génétique indirect.

Perspectives

Malheureusement, nous n'avons pas pu tester notre concept d'un système de forçage génétique indirect dans la *lipophorine*. Mais nous avons pu générer et caractériser des lignées Lp::scFV, dont l'une, Lp::sc2A10, est capable de bloquer significativement la transmission des parasites. En plus, nous avons identifié un locus supplémentaire, *roadblock*, qui est un gène essentiel qui pourrait être utilisé pour tester le forçage génétique indirect. Par ailleurs, nous envisageons de favoriser l'héritabilité de Lp::sc2A10 à partir d'un autre locus. Pour cette approche, nous utiliserons et inactiverons le gène de la s*agline* du moustique, connu pour favoriser l'infection.

Conclusion générale

Au cours de ce projet, nous avons travaillé sur le développement des systèmes de forçage génétique chez deux espèces distinctes, *D. suzukii* et *A. gambiae*. Malgré la non-obtention d'un système de forçage génétique chez *D. suzukii*, ce travail a produit des nombreux outils qui sont actuellement utilisé dans notre équipe et par nos collègues. Ce projet montre aussi les difficultés qui peuvent être rencontrées lors du développement de la transgénèse chez de nouvelles espèces. Dans le cas d'*A. gambiae*, nous avons développé des lignées qui expriment des scFvs capable de bloquer la transmission, qui serviront pour tester de nouveaux designs de forçage génétique.

Le forçage génétique par CRISPR/Cas9 reste une technologie nouvelle et comme la plupart des avancées, elle offre des opportunités et présente des risques. Des essais préliminaires sur le terrain sont en cours de préparation. De nombreuses questions restent ouvertes, mais dans l'état actuel de la recherche, le forçage génétique continue de représenter un outil potentiellement très important pour l'amélioration de la santé publique et de la prospérité

humaine. Cependant, en ce qui concerne la modification génétique, l'identité de l'opérateur des modifications et leur surveillance et régulation auront au moins autant d'importance que le but ultime de la modification. L'application de systèmes de forçage génétique sur le terrain sera une question de balance bénéfice / risques, et de confiance du public.

Annex I: Primer and Plasmid List

*Bsa*I sites are marked in bold, four nucleotide overhangs in uppercase

D. suzukii **project**

Gene drive:

Primers

5' homology arm *cup*	**ggtctcg**ATCCgcattcgaccaaattcac
	ggtctcaTTACggttcatgtgagccgctggtt
5' homology arm *stil*	**ggtctcc**ATCCggatcccaacacaccact
	ggtctcaTTACcaagttgtcggacagtttcagtctc
DsuzU6-sgRNA2	**ggtctcc**ATACatcaagagtagaaaaagcgc
	ggtctctTTCCtcaccctcaggtttgcaa
DsuzU6-sgRNA3	**ggtctcc**GGAAatcaagagtagaaaaagcgc
	ggtctctCTCTtcaccctcaggtttgcaa
DsuzU6-sgRNA4	**ggtctcc**AGAGatcaagagtagaaaaagcgc
	ggtctctATGGtcaccctcaggtttgcaa
DsuzβTub2 Promoter	**ggtctct**CCATtaatttcttggctctttcgatgg
	ggtctcgCATTttgctggtggtgatttg
Cas9	**ggtctcg**AATGgactataaggaccacgacgga
	ggtctcaGATAcattgatgagtttggacaaaccac
3xP3-YFP	**ggtctca**TATCtaattcaattagagactaattca
	ggtctccGCGTatcgataagcttta
3' homology arm *cup*	**ggtctct**ACGCggacatggatcgcgaag
	ggtctcaAAGCgtggcgattaattctcctcg
3' homology arm *stil*	**ggtctcg**ACGCcggcgtttttctggacatc
	ggtctcaAAGCcgaattactgggcgctcctc
sgRNA1 *cup*	CCTTgctgccatgaccctggttca
	AAACtgaaccagggtcatggcagc
sgRNA2 *cup*	CCTTgcatggtctactggtgcagt
	AAACactgcaccagtagaccatgc
sgRNA1 *stil*	CCTTgtccgacaacttgaacggtt
	AAACaaccgttcaagttgtcggac
sgRNA2 *stil*	CCTTgcagctgcgcagagtaattc
	AAACgaattactctgcgcagctgc
sgRNA1 *yl*	CCTTggcattctcacagccgtggg
	AAACcccacggctgtgagaatgcc
sgRNA2 *yl*	CCTTgtagatgtgactgccatcgt
	AAACacgatggcagtcacatctac
DsuzNanos Promoter	**ggtctca**CCATcccgcgcaagggcagctat
	ggtctcgCATTgcgaaagtacggctcgaaagtaacc
DsuzNanos 3'UTR	**ggtctct**TAAGaacacatccggcaggagcagag
	ggtctctGATAcatcttcctggccctttcga

DsuzU6-sgRNA1 (gBlock):

ggtctccGTAAatcaagagtagaaaaagcgccactagtttaaatttggaacatcatgaaacaccaccgctagaggtcgctaggagtcacgtactt
ttataattcccaactgctttttctgaatggagctagtatatatacgtccttttcgatactaaatcgtccttgggtcttcgaattcgaagacctgttttagagct
agaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgctttttttgcaaacctgagggtgAATA**caga**
gacc

Plasmids for Golden Gate Assembly

pENTR[*U6*-sgRNAx4(*cup*),*tub85*-Cas9-sv40]
pENTR R4 ATCC LacZ GCTT
pKSB-5'Cupflk ATCC, GTAA
pKSB-U6dsuz-sgRNA1-Cup1 GTAA, ATAC
pJET-U6dsuz-sgRNA2-Cup2 ATAC, GGAA
pJET-U6dsuz-sgRNA3-Ylk1 GGAA, AGAG
pKSB-U6dsuz-sgRNA4-Ylk2 AGAG, CCAT
pJET1.2-TubPdsuz CCAT, AATG
pKSB-Cas9-sv40 AATG, TATC
pJET1.2-3xP3-YFP TATC, ACGC
pJET1.2-3'Cupflk ACGC, GCTT

pENTR[*U6*-sgRNAx4(*cup*),*nos*-Cas9-*nos3'UTR*]
pENTR R4 ATCC LacZ GCTT
pKSB-5'Cupflk ATCC, GTAA
pKSB-U6dsuz-sgRNA1-Cup1 GTAA, ATAC
pJET-U6dsuz-sgRNA2-Cup2 ATAC, GGAA
pJET-U6dsuz-sgRNA3-Ylk1 GGAA, AGAG
pKSB-U6dsuz-sgRNA4-Ylk2 AGAG, CCAT
pJET1.2-NanosPdsuz CCAT, AATG
pKSB-Cas9 AATG, TAAG
pKSB-Nos 3'UTR dsuz TAAG, TATC
pJET1.2-3xP3-YFP TATC, ACGC
pJET1.2-3'Cupflk ACGC, GCTT

pENTR[*U6*-sgRNAx4(*still*), *nos*-Cas9-*nos3'UTR*]
pENTR R4 ATCC LacZ GCTT
pJET1.2-5'Stilflk ATCC, GTAA
pKSB-U6dsuz-sgRNA1-Stil1 GTAA, ATAC
pJET1.2-U6dsuz-sgRNA2-Stil2 ATAC, GGAA
pJET-U6dsuz-sgRNA3-Ylk1 GGAA, AGAG
pKSB-U6dsuz-sgRNA4-Ylk2 AGAG, CCAT
pJET1.2-NanosPdsuz CCAT, AATG
pKSB-Cas9 AATG, TAAG
pKSB-Nos 3'UTR dsuz TAAG, TATC
pJET1.2-3xP3-YFP TATC, ACGC
pJET1.2-3'Stilflk ACGC, GCTT

Transposons:

Primers

P-Element 5'TR	**ggtctcc**ATCCcatgatgaaataacataaggtgg
	ggtctcaTTACgatatcgctgctgctctaaacgac
Integrase	**ggtctca**AATGggccgatgcgcagcatg
	ggtctcaCTTActacgccgctacgtcttccg
DsuzPub Promoter	**ggtctcc**TATCggcttgctgttcttcgc
	ggtctctGAACtttggattattctgcgggtag
mTurquoise	**ggtctct**GTTCaccatggtgagcaaggg
	ggtctcaAAGCgatacattgatgagtttgga
P-Element 3'TR	**ggtctct**TAAGgatatctgcgtactcgcaaattattaaaa
	ggtctccAAGCcatgatgaaataacataaggt

OpIE2 promoter	**ggtctcg**ATCCatgatgataaacaatgtatggtgc
	ggtctcaCATTgtggccctcctatagtgagtc
Hr5IE1 promoter	**ggtctcg**TAAGcgcccgcgtaaaacacaat
	ggtctctGAACgtcgttcgcgggcgcaa
DsRedsv40	**ggtctca**AAGCgatacattgatgagtttgga
	ggtctcaGTTCaccatggtgcgctcctccaagaa

dsDNA linkers

dsDNA linker CATG, CCAT	**CCTG**ataacttcgtataatgtatgctatacgaagttat
	ATGGataacttcgtatagcatacattatacgaagttat
dsDNA linker ATCC, CCAT	**ATCC**gcctgccattcaggctcgaactgcagggccaa
	ATGGttggccctgcagttcgagcctgaatggcaggc

Plasmids for Golden Gate Assembly

pENTR[*OpIE2*-GFP,*Hr5IE1*-DsRed]
pENTR piggyBac ATCC lacZ GCTT loxAttP
pKSB-OpiE2Promoter ATCC, AATG
pKSB-GFP-sv40 AATG, TAAG
pKSB-Hr5IE1 promoter TAAG, GTTC
pKSB-DsRed-sv40 GTTC, GCTT

pENTR[*nos*-integrase,*Ub*-CFP]
pENTR R4-P5'TR ATCC LacZ GCTT P3'TR
pKSB-attP site ATCC, CATG
dsDNA linker1 CATG, CCAT
pJET1.2-NanosPdsuz CCAT, AATG
pJet-Integrase AATG, TAAG
pKSB-Nos 3'UTR dsuz TAAG, TATC
pKSB-UbiPdsuz TATC,GTTC
pKSB-mTurq-sv40 GTTC, GCTT

pENTR[*nos*-Cas9,*Ub*-DsRed]
pENTR piggyBac ATCC lacZ GCTT loxAttP
dsDNA linker ATCC, CCAT
pJET1.2-NanosPdsuz CCAT, AATG
pKSB-Cas9 AATG, TAAG
pKSB-Nos 3'UTR dsuz TAAG, TATC
pKSB-UbiPdsuz TATC, GTTC
pKSB-DsRed-sv40 GTTC, GCTT

pENTR[*nos*-Cas9,*Ub*-mTurquoise]
pENTR tTpiggyBac ATCC LacZ GCTT loxAttP
dsDNA linker ATCC, CCAT
pJET1.2-NanosPdsuz CCAT, AATG
pKSB-Cas9 AATG, TAAG
pKSB-Nos 3'UTR dsuz TAAG, TATC
pKSB-UbiPdsuz TATC, GTTC
pKSB- mTurquoise-sv40 GTTC, GCTT

An. gambiae project

Lp::scFv Knock-In:

Primers

5' homology arm *Lp*	**ggtctcg**AACAgttcattcccgattgagg
	ggtctctAGAAacgaccacaggagtcttctgcctcccagcacccacatcgttccgtgtcg
scFv	**ggtctca**TTCTcgtcagcttggtcctcatccagtccgtcagcgcagcccagatccagctggtgca
	ggtctcaCCTTggcgcgcttgatctccagcttc
3xP3-GFP	**ggtctca**CGAAtttacttgtacagctcgtccatgcc
	ggtctctGGGGatctaattcaattagagac
3' homology arm *Lp*	**ggtctca**CCCCgatggaagagatggcgaaggttctc
	ggtctctAAGCtcggagttaagacctccttcttttttgtc
sgRNA1-Lp	AAACggtgtcgctagtgctgattc
	CCTTgaatcagcactagcgacacc
sgRNA2-Lp	AAACtccctgacgatatttcacgc
	CCTTgcgtgaaatatcgtcaggga
sgRNA3-Lp	AAACccgaggacccacatcgttcc
	CCTTggaacgatgtgggtcctcgg
dsDNA linker	GCTTgcgtgaaatatcgtcagggacg
	AAGCcgtccctgacgatatttcacgc

DmTubterm (gBlock):

ggtctcgAAGGaacgtttccgtcgcggaattcgtgaatccgcaggtatgttcctatacgaaaccccaacaaaaaccataattgtttagacttgtgaa
caaaattggatccgactttattgattacgttgttaagagaacaaatcttttacaactgaattcatttgttctcgtttcatttttttcgcaaaacattgatcgaga
attcgattgatttccgatTCGA**atgagacc**

sgRNA expression modules

pKSB- U6agam-sgRNA1 ATCC, GGAA
pKSB-U6agam-sgRNA2 GGAA, AGAG
pKSB-Ugagam-sgRNA3 AGAG, AACA

Plasmids for Golden Gate Assembly

Lp-sc125

pENTR R4 ATCC LacZ GCTT
pKSB- U6agam-sgRNA(Lp1) ATCC, GGAA
pKSB-U6agam-sgRNA(Lp2) GGAA, AGAG
pKSB-Ugagam-sgRNA(Lp3) AGAG, AACA
pKSB-Lp5'flk AACA,TTCT
pKSB-sc125 TTCT, AAGG
pKSB-DmTubterm AAGG, GCAA
pKSB-3xP3GFPNoTerm CGAA, CCCC
pKSB-Lp3'flk CCCC, GCTT

Lp-sc2A10

pENTR R4 ATCC LacZ GCTT
pKSB- U6agam-sgRNA(Lp1) ATCC, GGAA
pKSB-U6agam-sgRNA(Lp2) GGAA, AGAG
pKSB-Ugagam-sgRNA(Lp3) AGAG, AACA
pKSB-Lp5'flk AACA,TTCT
pKSB-2A10 antibody TTCT, AAGG

pKSB-DmTubterm AAGG, GCAA
pKSB-3xP3GFPNoTerm CGAA, CCCC
pKSB-Lp3'flk CCCC, GCTT

Lp-GD

Primers

First Lp-GD (CFP)

5' homology arm *Lp*	**ggtctcc**ATCCaacagttcattcccgattgagg
	ggtctccATAAacccacatcgttccgtgt
ZPG promoter	**ggtctct**ATTCgctggcggtggggac
	ggtctccCATTctcgatgctgtatttgttgttgggctgtttgtta
3xP3-CFP	**ggtctca**TATCtaattcaattagagactaattca
	ggtctccGCGTatcgataagcttta
3' homology arm *Lp*	**ggtctcc**ACGCctagtgctgattcaaagtgtgt
	ggtctccAAGCcctccttcttttttgtctactattcttcc
tRNAsgRNA	**ggtctcg**AGAGagcatcggtggttcagtggtag
	ggtctcgTTCCagcaccgactcggtgcc
sgRNA1-Lp	ATGCaggaacgatgtgggtcctcgg
	AAACccgaggacccacatcgttcct
sgRNA2-Lp	ATGCagaatcagcactagcgacacc
	AAACggtgtcgctagtgctgattct
sgRNA3-Lp	ATGCcaagagcagcctccttccaccg
	AAACcggtggaaggaggctgctctt

second Lp-GD (DsRed)

3xP3 DsRed TATC, ACGC	**ggtctca**TATCtaattcaattagagactaattca
	ggtctcaGCGTtaagatacattgatgagtttggacaa
sgRNA(Lp1)	ATGCacacggaacgatgtgggtcct
	AAACaggacccacatcgttccgtgt
sgRNA(Lp2)	ATGCagaggctgctctggagcttcc
	AAACggaagctccagagcagcctct
sgRNA(Lp3)	ATGCactgattcaaagtgtgtccgc
	AAACgcggacacactttgaatcagt
sgRNA(Lp4)	ATGCagcgtgaaatatcgtcaggga
	AAACtccctgacgatatttcacgct

sgRNA expression modules

pKSB-tRNAgly-sgRNA1 GTAA, AGAG
pKSB-tRNAgly-sgRNA2 AGAG, GGAA
pKSB-tRNAgly-sgRNA3 GGAA, ATTC
pKSB-tRNAsgRNA1 GTAA, AGAG
pKSB-tRNAsgRNA2 AGAG, CCAT
pKSB-tRNAsgRNA3 CCAT, GGAA
pKSB-tRNA-sgRNA4-tRNA-U6terminator GGAA, ATTC

Plasmids for Golden Gate Assembly

First Lp-GD (CFP)

pENTR R4 ATCC LacZ GCTT
pKSB-Lp 5' Homology Arm ATCC, TTAT
pKSB-U6agam promoter TTAT, GTAA
pKSB-tRNAgly-sgRNA(Lp1) GTAA, AGAG
pKSB-tRNAgly-sgRNA(Lp2) AGAG, GGAA
pKSB-tRNAgly-sgRNA(Lp3) GGAA, ATTC
pKSB-ZPG promoter ATTC, AATG
pKSB-eSpCas9 AATG, AATT
pKSB-ZPG terminator AATT, TATC
pKSB-3xP3-mTurq TATC, ACGC
pKSB-Lp 3' Homology Arm ACGC, GCTT

Second Lp-GD (DsRed)

pENTR R4 ATCC LacZ GCTT
pKSB-Lp 5' Homology Arm ATCC TTAT
pKSB-U6agam promoter TTAT, GTAA
pKSB-tRNAsgRNA(Lp1) GTAA, AGAG
pKSB-tRNAsgRNA(Lp2) AGAG, CCAT
pKSB-tRNAsgRNA(Lp3) CCAT, GGAA
pKSB-tRNA-sgRNA(Lp4)-tRNA-U6terminator GGAA, ATTC
pKSB-ZPG promoter ATTC, AATG
pKSB-eSpCas9 AATG, AATT
pKSB-ZPG terminator AATT, TATC
pKSB-3xP3-DsRednls TATC, ACGC
pKSB-Lp 3' Homology Arm ACGC, GCTT

Essential Genes Knock-Out

Primers

Spase12 sgRNA1	CCTTgtgacaggggcagtctagtt
	AAACaactagactgcccctgtcac
Spase12 sgRNA2	CCTTggactttgagggtcagggta
	AAACtaccctgaccctcaaagtcc
Spase12 sgRNA3	CCTTgcggaaccgtcggtctagtgt
	AAACacactagaccgacggttccgc
5' HA Spase12	**ggtctcc**AACAatagcatatcggcagtctcaa
	ggtctctTTAGgcgagaatggcactattcaa
3' HA Spase12	**ggtctct**AATTagcagttttcgcaaacggtgt
	ggtctccAAGCccagcgtcgatggatagatatac
robl sgRNA1	CCTTgagcaattttccccgtcgcgg
	AAACccgcgacggggaaaattgctc
robl sgRNA2	CCTTgttgacgacggtcgcaccttt
	AAACaaaggtgcgaccgtcgtcaa
robl sgRNA3	CCTTgaggcgtcgtaggaacgattg
	AAACcaatcgttcctacgacgcctc
5' Homology Arm robl	**ggtctcc**AACAgctaatcaaccacttgtgttgt
	ggtctcgTTAGgactgcgattcgttttgttttga
3' Homology Arm robl	**ggtctcg**AATTggtcaataatgaaggtacgt
	ggtctccAAGCaagcaagagcaagacaactg

mago sgRNA1	CCTTgcgctactatgtgggccaca
	AAACtgtggcccacatagtagcgc
mago sgRNA2	CCTTgcatactcgactcggaaatta
	AAACtaatttccgagtcgagtatgc
mago sgRNA3	CCTTgaacgttactcccgaccaatc
	AAACgattggtcgggagtaacgttc
5' Homology Arm mago	**ggtctct**AACActatatccttccgcaccac
	ggtctctTTAGggtagaaatcttccgtgctcg
3' Homology Arm mago	**ggtctcc**AATTatccggtcaccattttgttgt
	ggtctctAAGCaccggttgttgagacagcga
AP-2σ sgRNA1	CCTTgtcggccgtttaggaatgttt
	AAACaaacattcctaaacggccgac
AP-2σ sgRNA2	CCTTgatcagtatgaagcgaatctg
	AAACcagattcgcttcatactgatc
AP-2σ sgRNA3	CCTTgagaagtacagccccgcgtac
	AAACgtacgcggggctgtacttctc
5' Homology Arm AP-2σ	**ggtctct**AACAtgcagccggtgctgcaagat
	ggtctctTTAGagggaagagagcacgtcgt
Lox-3xP3 GGGG TCGA	**ggtctca**GGGGcgaagacacggggataact
	ggtctctTCGAcccgattgtttagcttgt
3' Homology Arm AP-2σ	**ggtctcg**AATTggctgtacttctgcatct
	ggtctctAAGCcgatggggttcgctcgtta
Cas12a guide-linker1 against agam cype	agattcgattcgtgcagagcaaagccg**taatttctactcttgtagat**
	tgat**atctacaagagtagaaatta**cggctttgctctgcacgaatcga
Cas12a guide-linker2 against agam cype	ttcttgaagatgtctgaagtcgc**taatttctactcttgtagat**
	agtg**atctacaagagtagaaatta**gcgacttcagacatcttca
Cas12a guide-linker3 against agam cype	cactcacttgtagaactcggcgt**taatttctactcttgtagat**
	aaac**atctacaagagtagaaatta**acgccgagttctacaagtg
cype 5' HA	**ggtctcg**AACAgtgccagacgttgctgcatt
	ggtctccTTAGcaacactttggcccaccca
cype 3' HA	**ggtctcg**AATTcaaaatggtgctgcga
	ggtctcaAAGCccagcattcactcctaaact
Cas12a guide-linker1 against agam RPA323	agatctcccgcggatcctccacgtcct**taatttctactcttgtagat**
	tgtg**atctacaagagtagaaatta**aggacgtggaggatccgcgggag
Cas12a guide-linker2 against agam RPA323	cacatgggtcaatgatacggggc**taatttctactcttgtagat**
	caca**atctacaagagtagaaatta**gccccgtatcattgaccca
Cas12a guide-linker3 against agam RPA323	tgtgaaagacacacttgattttg**taatttctactcttgtagat**
	aaac**atctacaagagtagaaatta**caaaatcaagtgtgtctttcaca
5' Homology Arm RPA323	**ggtctca**AACAcgaacaatgtcacatgt
	ggtctctTTAGaaatccgattgccggaa
lox-mScarlet TCGA AATT	**ggtctcg**TCGAgataacttcgtatagc
	ggtctcgAATTcctaggtaagatacattg
3' Homology Arm RPA323	**ggtctcg**AATTcgtcatcaactgtgctgacaa
	ggtctcaAAGCcccaaacgtaccggatccaatt
Cas12a guide-linker1 against agam TFIIAS	agatatcgtaaaatcatatcaacacga**taatttctactcttgtagat**
	cctg**atctacaagagtagaaatta**tcgtgttgatatgattttacgat
Cas12a guide-linker2 against agam TFIIAS	caggagagtttggacgagctgat**taatttctactcttgtagat**
	tcga**atctacaagagtagaaatta**atcagctcgtccaaactct
Cas12a guide-linker3 against agam TFIIAS	tcgaactgcaccaaaactcgcac**taatttctactcttgtagat**
	aaac**atctacaagagtagaaatta**gtgcgagttttggtgcagt
5' Homology Arm TFIIAS	**ggtctcg**AACAtaccatgaacctcaccttgccgaat
	ggtctctTTAGttacgatgaaattgtagatgatgaagctc
3xP3 infusion	ctagaactagtggatcccccggtctccggggatctaattcaa
	tagacaccatccgattgtttagcttgttcagctgc
mNeonGreen module for infusion	aaacaatcggatggtgtctaagggagaagagg
	tgctagcttacgggtcctccaccttccg
sv40 module for infusion	ggaggacccgtaagctagcataaaatcagcca
	atatcgaattcctgcagcccggtctcatctttaagatacattga

3' Homology Arm TFIIAS	**ggtctcg**AATTaacacgtaccgattctgcgacaacgtct
	ggtctcaAAGCctcgaacggatgacgaaccagaatgaa
Cas12a guide-linker1 against agam Rep Prot	agattcgattcgtgcagagcaaagccg**taatttctactcttgtagat**
	tgat**atctacaagagtagaaatta**cggctttgctctgcacgaatcga
Cas12a guide-linker2 against agam Rep Prot	atcagcgctccgttcacgatcgt**taatttctactcttgtagat**
	tgag**atctacaagagtagaaatta**acgatcgtgaacggagcgc
Cas12a guide-linker3 against agam Rep Prot	ctcaggatgggtcgaagtgatag**taatttctactcttgtagat**
	aaac**atctacaagagtagaaatta**ctatcacttcgacccatcc
5' HARep Prot	**ggtctct**AACAcgaggagaacggaacagaatacagat
	ggtctccTTAGcgtgaatcttcctgcgtttggtg
add NLS to 3xP3-CFP module	aaggtggaggacccgtaaaatcagccataccacat
	ccgtttcttcttgggtgccttgtacagctcgtccat
3' Rep Prot	**ggtctcc**AATTtaggcattgcggctcccaat
	ggtctcaAAGCtcccgtcagctgtggctgatt

sgRNA expression plasmids

pKSB-U6agam-sgRNA1 ATCC, GGAA
pKSB-U6agam-sgRNA2 GGAA, AGAG
pKSB-Ugagam-sgRNA3 AGAG, AACA
pKSB-sgRNACas12a ATCC, AACA

Plasmids for Golden Gate Assembly

Spase12
pENTR R4 ATCC LacZ GCTT
pKSB-U6agam-sgRNA-Spase12-1 ATCC, GGAA
pKSB-U6agam-sgRNA-Spase12-2 GGAA, AGAG
pKSB-U6agam-sgRNA-Spase12-3 AGAG, AACA
pKSB-Spase12 5' Homology Arm AACA, CTAA
pKSB-attP-3xP3-CFPnls-sv40-attP CTAA, AATT
pKSB-Spase12 3' Homology Arm AATT, GCTT

robl
pENTR R4 ATCC LacZ GCTT
pKSB-U6agam-sgRNA-robl-1 ATCC, GGAA
pKSB-U6agam-sgRNA-robl-2 GGAA, AGAG
pKSB-Ugagam-sgRNA-robl-3 AGAG, AACA
pKSB-robl 5' Homology Arm AACA, CTAA
pKSB-attP-3xP3-YFPsv40-attP CTAA, AATT
pKSB-robl 3' Homology Arm AATT, GCTT

mago
pENTR R4 ATCC LacZ GCTT
pKSB- U6agam-sgRNA-mago-1 ATCC, GGAA
pKSB-U6agam-sgRNA-mago-2 GGAA, AGAG
pKSB-U6agam-sgRNA-mago-3 AGAG, AACA
pKSB-mago 5' Homology Arm AACA, CTAA
pKSB-attP-3xP3-GFPsv40-attP CTAA, AATT
pKSB-mago 3' Homology Arm AATT, GCTT

AP-2σ
pENTR R4 ATCC LacZ GCTT
pKSB-U6agam-sgRNA-AP-2σ-1 ATCC, GGAA
pKSB-U6agam-sgRNA-AP-2σ-2 GGAA, AGAG
pKSB-U6agam-sgRNA-AP-2σ-3 AGAG, AACA
pKSB-AP-2σ 5' Homology Arm AACA, CTAA
pKSB-attP-3xP-mNeonGreen-attP CTAA, AATT
pKSB-AP-2σ 3' Homology Arm AATT, GCTT

cype
pENTR R4 ATCC LacZ GCTT
pKSB-U6agam-cypeRNAXas12ax3 ATCC, AACA
pKSB-cype 5' Homology Arm AACA, CTAA
pKSB-attP-3xP-YFPnls-attP CTAA, AATT
pKSB-cype 3' Homology Arm AATT, GCTT

RepProt
pENTR R4 ATCC LacZ GCTT
pKSB-U6agam-sgRNACas12ax3 RepProt ATCC, AACA
pKSB-ReplicationProtein 5' Homology Arm AACA, CTAA
pKSB-attP-3xP-mNeonGreen-attP CTAA, AATT
pKSB-RepProt 3' Homology Arm AATT, GCTT

RPA323
pENTR R4 ATCC LacZ GCTT
pKSB-U6agam-RPA3sgRNACas12ax3 ATCC, AACA
pKSB-RPA323 5' Homology Arm AACA, CTAA
pKSB-attP-3xP-CFP-attP CTAA, AATT
pKSB-RPA323 3' Homology Arm AATT, GCTT

TFIIAS
pENTR R4 ATCC LacZ GCTT
pKSB-U6agam-sgRNACas12ax3 TFSII ATCC, AACA
pKSB-TFIIS 5' Homology Arm AACA, CTAA
pKSB-attP-3xP3-GFP-sv40-attP CTAA, AATT
pKSB-TFIIS 3' Homology Arm AACA, CTAA

Works Cited

Akbari, O. S., Bellen, H. J., Bier, E., Bullock, S. L., Burt, A., Church, G. M., … Knoblich, J. (2015). Safegaurding gene drive experiments in the laboratory. *Science*, *349*(6251), 927–929.

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*(8), 671–677. https://doi.org/10.1016/j.cub.2013.02.059

Aly, A. S. I., Vaughan, A. M., & Kappe, S. H. I. (2009). Malaria parasite development in the mosquito and infection of the mammalian host. *Annual Review of Microbiology*, *63*, 195–221. https://doi.org/10.1146/annurev.micro.091208.073403

Aminetzach, Y. T., Macpherson, J. M., & Petrov, D. A. (2005). Evolution: Pesticide resistance via transposition-mediated adaptive gene truncation in Drosophila. *Science*, *309*(5735), 764–767. https://doi.org/10.1126/science.1112699

Ashley, E. A., Pyae Phyo, A., & Woodrow, C. J. (2018). Malaria. *Lancet (London, England)*, *391*(10130), 1608–1621. https://doi.org/10.1016/S0140-6736(18)30324-6

Asplen, M. K., Anfora, G., Biondi, A., Choi, D. S., Chu, D., Daane, K. M., … Desneux, N. (2015). Invasion biology of spotted wing Drosophila (Drosophila suzukii): a global perspective and future priorities. *Journal of Pest Science*, *88*(3), 469–494. https://doi.org/10.1007/s10340-015-0681-z

Atella, G. C., Silva-Neto, M. A. C., Golodne, D. M., Arefin, S., & Shahabuddin, M. (2006). Anopheles gambiae lipophorin: Characterization and role in lipid transport to developing oocyte. *Insect Biochemistry and Molecular Biology*, *36*(5), 375–386. https://doi.org/10.1016/j.ibmb.2006.01.019

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., … Horvath, P. (2007). CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes. *Science*, *315*(5819), 1709–1712. https://doi.org/10.1126/SCIENCE.1138140

Barreaux, P., Barreaux, A. M. G., Sternberg, E. D., Suh, E., Waite, J. L., Whitehead, S. A., & Thomas, M. B. (2017). Priorities for Broadening the Malaria Vector Control Tool Kit. *Trends in Parasitology*, *33*(10), 763–774. https://doi.org/10.1016/j.pt.2017.06.003

Basgall, E. M., Goetting, S. C., Goeckel, M. E., Giersch, R. M., Roggenkamp, E., Schrock, M. N., … Finnigan, G. C. (2018). Gene drive inhibition by the anti-CRISPR proteins AcrIIA2 and AcrIIA4 in Saccharomyces cerevisiae. *Microbiology (Reading, England)*, *164*(4), 464–474. https://doi.org/10.1099/mic.0.000635

Beaghton, A., Hammond, A., Nolan, T., Crisanti, A., Godfray, H. C. J., & Burt, A. (2017). Requirements for Driving Antipathogen Effector Genes into Populations of Disease Vectors by Homing. *Genetics*, *205*(4), 1587–1596. https://doi.org/10.1534/genetics.116.197632

Beeman, R. W., Friesen, K. S., & Denell, R. E. (1992). Maternal-effect selfish genes in flour beetles. *Science (New York, N.Y.)*, *256*(5053), 89–92. https://doi.org/10.1126/science.1566060

Bhatt, S., Weiss, D. J., Cameron, E., Bisanzio, D., Mappin, B., Dalrymple, U., … Gething, P. W. (2015). The effect of malaria control on Plasmodium falciparum in Africa between

2000 and 2015. *Nature*, *526*(7572), 207–211. https://doi.org/10.1038/nature15535

Blacklock, B. J., & Ryan, R. O. (1994). Hemolymph lipid transport. *Insect Biochemistry and Molecular Biology*. https://doi.org/10.1016/0965-1748(94)90015-9

Blandin, S., Moita, L. F., Köcher, T., Wilm, M., Kafatos, F. C., & Levashina, E. A. (2002). Reverse genetics in the mosquito Anopheles gambiae: targeted disruption of the Defensin gene. *EMBO Reports*, *3*(9), 852–856. https://doi.org/10.1093/embo-reports/kvf180

Blandin, S., Shiao, S. H., Moita, L. F., Janse, C. J., Waters, A. P., Kafatos, F. C., & Levashina, E. a. (2004). Complement-like protein TEP1 is a determinant of vectorial capacity in the malaria vector Anopheles gambiae. *Cell*, *116*(5), 661–670. https://doi.org/10.1016/S0092-8674(04)00173-4

Brogden, K. A. (2005). Antimicrobial peptides: pore formers or metabolic inhibitors in bacteria? *Nature Reviews Microbiology*, *3*(3), 238–250. https://doi.org/10.1038/nrmicro1098

Bruck, D. J., Bolda, M., Tanigoshi, L., Klick, J., Kleiber, J., Defrancesco, J., … Spitler, H. (2011). Laboratory and field comparisons of insecticides to reduce infestation of Drosophila suzukii in berry crops. *Pest Management Science*, *67*(11), 1375–1385. https://doi.org/10.1002/ps.2242

Buchman, A. B., Ivy, T., Marshall, J. M., Akbari, O. S., & Hay, B. A. (2018). Engineered Reciprocal Chromosome Translocations Drive High Threshold, Reversible Population Replacement in Drosophila. *ACS Synthetic Biology*, *7*(5), 1359–1370. https://doi.org/10.1021/acssynbio.7b00451

Buchman, A., Marshall, J. M., Ostrovski, D., Yang, T., & Akbari, O. S. (2018). Synthetically engineered Medea gene drive system in the worldwide crop pest Drosophila suzukii. *Proceedings of the National Academy of Sciences*, *115*(18), 4725–4730. https://doi.org/10.1073/PNAS.1713139115

Bull, J. J., Remien, C. H., & Krone, S. M. (2019). Gene-drive-mediated extinction is thwarted by population structure and evolution of sib mating. *Evolution, Medicine, and Public Health*, *2019*(1), 66–81. https://doi.org/10.1093/emph/eoz014

Burt, A. (2003). Site-specific selfish genes as tools for the control and genetic engineering of natural populations. *Proceedings of the Royal Society B: Biological Sciences*, *270*(1518), 921–928. https://doi.org/10.1098/rspb.2002.2319

Burt, Austin, & Koufopanou, V. (2004). Homing endonuclease genes: the rise and fall and rise again of a selfish element. *Current Opinion in Genetics & Development*, *14*(6), 609–615. https://doi.org/10.1016/J.GDE.2004.09.010

Callaway, E. (2018). UN treaty agrees to limit gene drives but rejects a moratorium. *Nature*. https://doi.org/10.1038/d41586-018-07600-w

Capillon, C., & Atlan, A. (1999). Evolution of Driving X Chromosomes and Resistance Factors in Experimental Populations of Drosophila simulans. *Evolution*, *53*(2), 506–517. https://doi.org/10.1111/j.1558-5646.1999.tb03785.x

Carareto, C. M., Kim, W., Wojciechowski, M. F., O'Grady, P., Prokchorova, A. V, Silva, J. C., & Kidwell, M. G. (1997). Testing transposable elements as genetic drive mechanisms using Drosophila P element constructs as a model system. *Genetica*, *101*(1), 13–33. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9465407

Cha, S. J., Mori, A., Chadee, D. D., & Severson, D. W. (2006). Cage trials using an endogenous meiotic drive gene in the mosquito Aedes aegypti to promote population replacement. *American Journal of Tropical Medicine and Hygiene*, *74*(1), 62–68. https://doi.org/10.4269/ajtmh.2006.74.62

Chabert, S., Allemand, R., Poyet, M., Eslin, P., & Gibert, P. (2012). Ability of European parasitoids (Hymenoptera) to control a new invasive Asiatic pest, Drosophila suzukii. *Biological Control*, *63*(1), 40–47. https://doi.org/10.1016/J.BIOCONTROL.2012.05.005

Champer, J., Kim, I., Champer, S. E., Clark, A. G., & Messer, P. W. (2019). Suppression gene drive in continuous space can result in unstable persistence of both drive and wild-type alleles. *BioRxiv*, 769810. https://doi.org/10.1101/769810

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

Champer, J., Reeves, R., Oh, S. Y., Liu, C., Liu, J., Clark, A. G., & Messer, P. W. (2017). Novel CRISPR/Cas9 gene drive constructs reveal insights into mechanisms of resistance allele formation and drive efficiency in genetically diverse populations. *PLoS Genetics*, *13*(7), e1006796. https://doi.org/10.1371/journal.pgen.1006796

Champer, S. E., Oh, S. Y., Liu, C., Wen, Z., Clark, A. G., Messer, P. W., & Champer, J. (2019). Computational and experimental performance of CRISPR homing gene drive strategies with multiplexed gRNAs. *BioRxiv*, 679902. https://doi.org/10.1101/679902

Chan, Y.S., Huen, D. S., Glauert, R., Whiteway, E., & Russell, S. (2013a). Optimising Homing Endonuclease Gene Drive Performance in a Semi-Refractory Species: The Drosophila melanogaster Experience. *PLoS ONE*, *8*(1), e54130. https://doi.org/10.1371/journal.pone.0054130

Chan, Y. S., Takeuchi, R., Jarjour, J., Huen, D. S., Stoddard, B. L., & Russell, S. (2013b). The Design and In Vivo Evaluation of Engineered I-OnuI-Based Enzymes for HEG Gene Drive. *PLoS ONE*, *8*(9), e74254. https://doi.org/10.1371/journal.pone.0074254

Chan, Y. S., Naujoks, D. A., Huen, D. S., & Russell, S. (2011). Insect Population Control by Homing Endonuclease-Based Gene Drive: An Evaluation in Drosophila melanogaster. *Genetics*, *188*(1), 33–44. https://doi.org/10.1534/genetics.111.127506

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 (New York, N.Y.)*, *316*(5824), 597–600. https://doi.org/10.1126/science. 1138595

Chiu, J. C., Jiang, X., Zhao, L., Hamm, C. A., Cridland, J. M., Saelao, P., … Begun, D. J. (2013). Genome of Drosophila suzukii, the spotted wing drosophila. *G3 (Bethesda, Md.)*, *3*(12), 2257–2271. https://doi.org/10.1534/g3.113.008185

Chu, F. C., Klobasa, W., Grubbs, N., & Lorenzen, M. D. (2018). Development and use of a piggyBac-based jumpstarter system in Drosophila suzukii. *Archives of Insect Biochemistry and Physiology*, *97*(3), e21439. https://doi.org/10.1002/arch.21439

Chu, V. T., Weber, T., Wefers, B., Wurst, W., Sander, S., Rajewsky, K., & Kühn, R. (2015a). Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. *Nature Biotechnology*, *33*(5), 543–548. https://doi.org/10.1038/nbt.3198

Chu, V. T., Weber, T., Wefers, B., Wurst, W., Sander, S., Rajewsky, K., & Kühn, R. (2015b). Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. *Nature Biotechnology*, *33*(5), 543–548. https://doi.org/10.1038/nbt.3198

Cini, A., Anfora, G., Escudero-Colomar, L. A., Grassi, A., Santosuosso, U., Seljak, G., & Papini, A. (2014). Tracking the invasion of the alien fruit pest Drosophila suzukii in Europe. *Journal of Pest Science*, *87*(4), 559–566. https://doi.org/10.1007/s10340-014-0617-z

Clements, A. N. (1992). *The Biology of Mosquitoes* (1st ed.). London New York Wallingford, Oxfordshire, UK: Chapman & Hall,Cambridge, MA, USA CABI.

Collins, F. H., Paskewitz, S. M., & Finnerty, V. (1989). Ribosomal RNA Genes of the Anopheles gambiae Species Complex (pp. 1–28). Springer, New York, NY. https://doi.org/10.1007/978-1-4612-3292-6_1

Conde, R., Zamudio, F. Z., Rodríguez, M. H., & Possani, L. D. (2000). Scorpine, an anti-malaria and anti-bacterial agent purified from scorpion venom. *FEBS Letters*, *471*(2–3), 165–168. https://doi.org/10.1016/s0014-5793(00)01384-3

Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., … Zhang, F. (2013). Multiplex Genome Engineering Using CRISPR/Cas System. *Science*, *339*(6121), 819–823. https://doi.org/10.1126/science.1231143

Cordaux, R., Pichon, S., Ling, A., Perez, P., Delaunay, C., Vavre, F., … Greve, P. (2008). Intense Transpositional Activity of Insertion Sequences in an Ancient Obligate Endosymbiont. *Molecular Biology and Evolution*, *25*(9), 1889–1896. https://doi.org/10.1093/molbev/msn134

Costa, G., Eldering, M., Lindquist, R. L., Hauser, A. E., Sauerwein, R., Goosmann, C., … Levashina, E. A. (2017). Mosquito lipids regulate Plasmodium sporogony and infectivity to the mammalian host. *BioRxiv*, 149443. https://doi.org/10.1101/149443

Cox, F. E. (2010). History of the discovery of the malaria parasites and their vectors. *Parasites & Vectors*, *3*(1), 5. https://doi.org/10.1186/1756-3305-3-5

Curtis, C. F. (1968). A possible genetic method for the control of insect pests, with special reference to tsetse flies (Glossina spp.). *Bulletin of Entomological Research*, *57*(4), 509–523. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/5659159

Curtis, C. F., & Graves, P. M. (1988). Methods for replacement of malaria vector populations. *The Journal of Tropical Medicine and Hygiene*, *91*(2), 43–48. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/3288764

Darboux, I., Charles, J.-F., Pauchet, Y., Warot, S., & Pauron, D. (2007). Transposon-mediated resistance to Bacillus sphaericus in a field-evolved population of Culex pipiens (Diptera: Culicidae). *Cellular Microbiology*, *9*(8), 2022–2029. https://doi.org/10.1111/j.1462-5822.2007.00934.x

Davis, S., Bax, N., & Grewe, P. (2001). Engineered underdominance allows efficient and economical introgression of traits into pest populations. *Journal of Theoretical Biology*, *212*(1), 83–98. https://doi.org/10.1006/jtbi.2001.2357

Deprá, M., Poppe, J. L., Schmitz, H. J., De Toni, D. C., & Valente, V. L. S. (2014). The first records of the invasive pest Drosophila suzukii in the South American continent. *Journal of Pest Science*, *87*(3), 379–383. https://doi.org/10.1007/s10340-014-0591-5

Deredec, A., Burt, A., & Godfray, H. C. J. (2008). The Population Genetics of Using Homing Endonuclease Genes in Vector and Pest Management. *Genetics*, *179*(4), 2013–2026. https://doi.org/10.1534/genetics.108.089037

Dhole, S., Lloyd, A. L., & Gould, F. (2019). Tethered homing gene drives: A new design for spatially restricted population replacement and suppression. *Evolutionary Applications*, *12*(8), 1688–1702. https://doi.org/10.1111/eva.12827

DiCarlo, J. E., Chavez, A., Dietz, S. L., Esvelt, K. M., & Church, G. M. (2015). Safeguarding CRISPR-Cas9 gene drives in yeast. *Nature Biotechnology*, (November), 1–8. https://doi.org/10.1038/nbt.3412

DiGiacomo, G., Hadrich, J., Hutchison, W. D., Peterson, H., & Rogers, M. (2019). Economic Impact of Spotted Wing Drosophila (Diptera: Drosophilidae) Yield Loss on Minnesota Raspberry Farms: A Grower Survey. *Journal of Integrated Pest Management*, *10*(1). https://doi.org/10.1093/jipm/pmz006

Ding, D., Chen, K., Chen, Y., Li, H., & Xie, K. (2018). Engineering Introns to Express RNA Guides for Cas9- and Cpf1-Mediated Multiplex Genome Editing. *Molecular Plant*, *11*(4), 542–552. https://doi.org/10.1016/J.MOLP.2018.02.005

Dong, Y., Aguilar, R., Xi, Z., Warr, E., Mongin, E., & Dimopoulos, G. (2006). Anopheles gambiae Immune Responses to Human and Rodent Plasmodium Parasite Species. *PLoS Pathogens*, *2*(6), e52. https://doi.org/10.1371/journal.ppat.0020052

Dong, Y., Das, S., Cirimotich, C., Souza-Neto, J. A., McLean, K. J., & Dimopoulos, G. (2011). Engineered Anopheles Immunity to Plasmodium Infection. *PLoS Pathogens*, *7*(12), e1002458. https://doi.org/10.1371/journal.ppat.1002458

Dong, Y., & Dimopoulos, G. (2009). Anopheles fibrinogen-related proteins provide expanded pattern recognition capacity against bacteria and malaria parasites. *The Journal of Biological Chemistry*, *284*(15), 9835–9844. https://doi.org/10.1074/jbc.M807084200

Dong, Y., Simões, M. L., Marois, E., & Dimopoulos, G. (2018). CRISPR/Cas9 -mediated gene knockout of Anopheles gambiae FREP1 suppresses malaria parasite infection. *PLOS Pathogens*, *14*(3), e1006898. https://doi.org/10.1371/journal.ppat.1006898

Drury, D. W., Dapper, A. L., Siniard, D. J., Zentner, G. E., & Wade, M. J. (2017). CRISPR/Cas9 gene drives in genetically variable and nonrandomly mating wild populations. *Science Advances*, *3*(5), e1601910. https://doi.org/10.1126/sciadv.1601910

Engels, W. R. (1992, October). The origin of P elements in Drosophila melanogaster. *BioEssays*. John Wiley & Sons, Ltd. https://doi.org/10.1002/bies.950141007

Engels, W. R., Johnson-Schlitz, D. M., Eggleston, W. B., & Sved, J. (1990). High-frequency P element loss in Drosophila is homolog dependent. *Cell*, *62*(3), 515–525. https://doi.org/10.1016/0092-8674(90)90016-8

Esvelt, K. M., Smidler, A. L., Catteruccia, F., & Church, G. M. (2014). Concerning RNA-guided gene drives for the alteration of wild populations. *ELife*, *3*, e03401. https://doi.org/10.7554/eLife.03401

Esvelt, Kevin M., Min, J., Noble, C., Najjar, D., & Esvelt, K. M. (2017). Daisyfield gene drive : harnessing repeated genomic elements as a generational clock to limit spread. *BioRxiv*, 104877. https://doi.org/https://doi.org/10.1101/104877

Evans, B. R., Kotsakiozi, P., Costa-da-Silva, A. L., Ioshino, R. S., Garziera, L., Pedrosa, M. C., … Powell, J. R. (2019). Transgenic Aedes aegypti Mosquitoes Transfer Genes into a

Natural Population. *Scientific Reports*, *9*(1), 13047. https://doi.org/10.1038/s41598-019-49660-6

Facchinelli, L., North, A. R., Collins, C. M., Menichelli, M., Persampieri, T., Bucci, A., … Benedict, M. Q. (2019). Large-cage assessment of a transgenic sex-ratio distortion strain on populations of an African malaria vector. *Parasites & Vectors*, *12*(1), 70. https://doi.org/10.1186/s13071-019-3289-y

Fitz-Earle, M., Holm, D. G., & Suzuki, D. T. (1973). Genetic control of insect population. I. Cage studies of chromosome replacement by compound autosomes in Drosophila melanogaster. *Genetics*, *74*(3), 461–475. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/4200686

Frolet, C., Thoma, M., Blandin, S., Hoffmann, J. A., & Levashina, E. A. (2006). Boosting NF-κB-Dependent Basal Immunity of Anopheles gambiae Aborts Development of Plasmodium berghei. *Immunity*, *25*(4), 677–685. https://doi.org/10.1016/j.immuni.2006.08.019

Galizi, R., Doyle, L. A., Menichelli, M., Bernardini, F., Deredec, A., Burt, A., … Crisanti, A. (2014). A synthetic sex ratio distortion system for the control of the human malaria mosquito. *Nature Communications*, *5*, 3977. https://doi.org/10.1038/ncomms4977

Gantz, V. M., & Bier, E. (2015). The mutagenic chain reaction: A method for converting heterozygous to homozygous mutations. *Science*, *348*, 442–444.

Gantz, V. M., & Bier, E. (2016). The dawn of active genetics. *BioEssays*, *38*(1). https://doi.org/10.1002/bies.201500102

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*, 201521077. https://doi.org/10.1073/pnas.1521077112

Gao, Y., & Zhao, Y. (2014). Self-processing of ribozyme-flanked RNAs into guide RNAs in vitro and in vivo for CRISPR-mediated genome editing. *Journal of Integrative Plant Biology*, *56*(4), 343–349. https://doi.org/10.1111/jipb.12152

Garver, L. S., Dong, Y., & Dimopoulos, G. (2009). Caspar Controls Resistance to Plasmodium falciparum in Diverse Anopheline Species. *PLoS Pathogens*, *5*(3), e1000335. https://doi.org/10.1371/journal.ppat.1000335

Gershenson, S. (1928). A New Sex-Ratio Abnormality in Drosophila Obscura. *Genetics*, *13*(6), 488–507. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/17246563

Ghosh, A K, Ribolla, P. E., & Jacobs-Lorena, M. (2001). Targeting Plasmodium ligands on mosquito salivary glands and midgut with a phage display peptide library. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(23), 13278–13281. https://doi.org/10.1073/pnas.241491198

Ghosh, Anil K., Devenport, M., Jethwaney, D., Kalume, D. E., Pandey, A., Anderson, V. E., … Jacobs-Lorena, M. (2009). Malaria Parasite Invasion of the Mosquito Salivary Gland Requires Interaction between the Plasmodium TRAP and the Anopheles Saglin Proteins. *PLoS Pathogens*, *5*(1), e1000265. https://doi.org/10.1371/journal.ppat.1000265

Girardin, L., Calvez, V., & Débarre, F. (2018). Catch me if you can: a spatial model for a brake-driven gene drive reversal. Retrieved from http://arxiv.org/abs/1812.06641

Goeckel, M. E., Basgall, E. M., Lewis, I. C., Goetting, S. C., Yan, Y., Halloran, M., &

Finnigan, G. C. (2019). Modulating CRISPR gene drive activity through nucleocytoplasmic localization of Cas9 in S. cerevisiae. *Fungal Biology and Biotechnology*, *6*(1), 2. https://doi.org/10.1186/s40694-019-0065-x

Gould, F., & Schliekelman, P. (2004). Population Genetics of Autocidal Control and Strain Replacement. *Annual Review of Entomology*, *49*, 193–217. https://doi.org/10.1146/annurev.biochem.72.121801.161720

Gress, B. E., & Zalom, F. G. (2019). Identification and risk assessment of spinosad resistance in a California population of *Drosophila suzukii*. *Pest Management Science*, *75*(5), 1270–1276. https://doi.org/10.1002/ps.5240

Guirouilh-Barbat, J., Lambert, S., Bertrand, P., & Lopez, B. S. (2014). Is homologous recombination really an error-free process? *Frontiers in Genetics*, *5*, 175. https://doi.org/10.3389/fgene.2014.00175

Gupta, L., Molina-Cruz, A., Kumar, S., Rodrigues, J., Dixit, R., Zamora, R. E., & Barillas-Mury, C. (2009). The STAT Pathway Mediates Late-Phase Immunity against Plasmodium in the Mosquito Anopheles gambiae. *Cell Host & Microbe*, *5*(5), 498–507. https://doi.org/10.1016/j.chom.2009.04.003

Gwadz, R. W., Kaslow, D., Lee, J. Y., Maloy, W. L., Zasloff, M., & Miller, L. H. (1989). Effects of magainins and cecropins on the sporogonic development of malaria parasites in mosquitoes. *Infection and Immunity*, *57*(9), 2628–2633. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/2759705

Haeussler, M., Schönig, K., Eckert, H., Eschstruth, A., Mianné, J., Renaud, J.-B., … Concordet, J.-P. (2016). Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. *Genome Biology*, *17*(1), 148. https://doi.org/10.1186/s13059-016-1012-2

Hamilton, W. D. (1967). Extraordinary Sex Ratios. *Science*, *156*(3774), 477–488. https://doi.org/10.1126/science.156.3774.477

Hamm, C. A., Begun, D. J., Vo, A., Smith, C. C. R., Saelao, P., Shaver, A. O., … Turelli, M. (2014). *Wolbachia* do not live by reproductive manipulation alone: infection polymorphism in *Drosophila suzukii* and *D. subpulchrella*. *Molecular Ecology*, *23*(19), 4871–4885. https://doi.org/10.1111/mec.12901

Hammond, A., Galizi, R., Kyrou, K., Simoni, A., Siniscalchi, C., Katsanos, D., … Nolan, T. (2015). A CRISPR-Cas9 gene drive system targeting female reproduction in the malaria mosquito vector *Anopheles gambiae*. *Nature Biotechnology*, (November), 1–8. https://doi.org/10.1038/nbt.3439

Hammond, A., Kyrou, K., Gribble, M., Karlsson, X., Morianou, I., Galizi, R., … Nolan, T. (2018). Improved CRISPR-based suppression gene drives mitigate resistance and impose a large reproductive load on laboratory-contained mosquito populations. *BioRxiv*, 360339. https://doi.org/10.1101/360339

Hammond, A. M., Kyrou, K., Bruttini, M., North, A., Galizi, R., Karlsson, X., … Nolan, T. (2017). The creation and selection of mutations resistant to a gene drive over multiple generations in the malaria mosquito. *PLOS Genetics*, *13*(10), e1007039. Retrieved from https://doi.org/10.1371/journal.pgen.1007039

Helleu, Q., Gérard, P. R., Dubruille, R., Ogereau, D., Prud'homme, B., Loppin, B., & Montchamp-Moreau, C. (2016). Rapid evolution of a Y-chromosome heterochromatin protein underlies sex chromosome meiotic drive. *Proceedings of the National Academy*

of Sciences of the United States of America, *113*(15), 4110–4115. https://doi.org/10.1073/pnas.1519332113

Hemingway, J., Shretta, R., Wells, T. N. C., Bell, D., Djimdé, A. A., Achee, N., & Qi, G. (2016). Tools and Strategies for Malaria Control and Elimination: What Do We Need to Achieve a Grand Convergence in Malaria? *PLoS Biology*, *14*(3), e1002380. https://doi.org/10.1371/journal.pbio.1002380

Hickey, W. A., & Craig, G. B. (1966). Genetic distortion of sex ratio in a mosquito, Aedes aegypti. *Genetics*, *53*(6), 1177–1196. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/5958915

Hinnen, A., Hicks, J. B., & Fink, G. R. (1978). Transformation of yeast. *Proceedings of the National Academy of Sciences of the United States of America*, *75*(4), 1929–1933. https://doi.org/10.1073/pnas.75.4.1929

Hollingdale, M. R., Nardin, E. H., Tharavanij, S., Schwartz, A. L., & Nussenzweig, R. S. (1984). Inhibition of entry of Plasmodium falciparum and P. vivax sporozoites into cultured cells; an in vitro assay of protective antibodies. *Journal of Immunology (Baltimore, Md. : 1950)*, *132*(2), 909–913. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/6317752

Houck, M. A., Clark, J. B., Peterson, K. R., & Kidwell, M. G. (1991). Possible horizontal transfer of Drosophila genes by the mite Proctolaelaps regalis. *Science (New York, N.Y.)*, *253*(5024), 1125–1128. https://doi.org/10.1126/science.1653453

Hu, X., Wang, C., Liu, Q., Fu, Y., & Wang, K. (2017). Targeted mutagenesis in rice using CRISPR-Cpf1 system. *Journal of Genetics and Genomics*, *44*(1), 71–73. https://doi.org/10.1016/j.jgg.2016.12.001

Isaacs, A. T., Jasinskiene, N., Tretiakov, M., Thiery, I., Zettor, A., Bourgouin, C., & James, A. A. (2012). Transgenic Anopheles stephensi coexpressing single-chain antibodies resist Plasmodium falciparum development. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(28), E1922–E1930. https://doi.org/10.1073/pnas.1207738109

Isaacs, A. T., Li, F., Jasinskiene, N., Chen, X., Nirmala, X., Marinotti, O., … James, A. A. (2011). Engineered Resistance to Plasmodium falciparum Development in Transgenic Anopheles stephensi. *PLoS Pathogens*, *7*(4), e1002017. https://doi.org/10.1371/journal.ppat.1002017

Ishino, Y., Shinagawa, H., Makino, K., Amemura, M., & Nakata, A. (1987). Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. *Journal of Bacteriology*, *169*(12), 5429–5433. https://doi.org/10.1128/jb.169.12.5429-5433.1987

Ito, J., Ghosh, A., Moreira, L. A., Wimmer, E. a, & Jacobs-Lorena, M. (2002). Transgenic anopheline mosquitoes impaired in transmission of a malaria parasite. *Nature*, *417*(6887), 452–455. https://doi.org/10.1038/417452a

Jaenike, J. (1996). Sex-Ratio Meiotic Drive in the Drosophila quinaria Group. *The American Naturalist*, *148*(2), 237–254. https://doi.org/10.1086/285923

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 (New York, N.Y.)*, *337*(6096), 816–821. https://doi.org/10.1126/science.1225829

Jones, M. S., Delborne, J. A., Elsensohn, J., Mitchell, P. D., & Brown, Z. S. (2019). Does the U.S. public support using gene drives in agriculture? And what do they want to know? *Science Advances*, *5*(9), eaau8462. https://doi.org/10.1126/sciadv.aau8462

Joung, J. K., & Sander, J. D. (2013). TALENs: a widely applicable technology for targeted genome editing. *Nature Reviews Molecular Cell Biology*, *14*(1), 49–55. https://doi.org/10.1038/nrm3486

Kenis, M., Tonina, L., Eschen, R., van der Sluis, B., Sancassani, M., Mori, N., … Helsen, H. (2016). Non-crop plants used as hosts by Drosophila suzukii in Europe. *Journal of Pest Science*. https://doi.org/10.1007/s10340-016-0755-6

Kidwell, M. G., & Ribeiro, J. M. C. (1992). Can transposable elements be used to drive disease refractoriness genes into vector populations? *Parasitology Today*, *8*(10), 325–329. https://doi.org/10.1016/0169-4758(92)90065-A

Kim, H. K., Song, M., Lee, J., Menon, A. V., Jung, S., Kang, Y.-M., … Kim, H. (2016). In vivo high-throughput profiling of CRISPR–Cpf1 activity. *Nature Methods*, *14*(2). https://doi.org/10.1038/nmeth.4104

Kim, W., Koo, H., Richman, A. M., Seeley, D., Vizioli, J., Klocko, A. D., & O'brochta, D. A. (2004). Ectopic Expression of a Cecropin Transgene in the Human Malaria Vector Mosquito *Anopheles gambiae* (Diptera: Culicidae): Effects on Susceptibility to *Plasmodium*. *Journal of Medical Entomology*, *41*(3), 447–455. https://doi.org/10.1603/0022-2585-41.3.447

Kim, Y. G., Cha, J., & Chandrasegaran, S. (1996). Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proceedings of the National Academy of Sciences of the United States of America*, *93*(3), 1156. https://doi.org/10.1073/PNAS.93.3.1156

Kimble, M., Dettman, R. W., & Raff, E. C. (1990). The beta 3-tubulin gene of Drosophila melanogaster is essential for viability and fertility. *Genetics*, *126*(4), 991–1005. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/2127581

Kleinstiver, B. P., Pattanayak, V., Prew, M. S., Tsai, S. Q., Nguyen, N. T., Zheng, Z., & Keith Joung, J. (2016). High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects. *Nature*. https://doi.org/10.1038/nature16526

Kleinstiver, B. P., Prew, M. S., Tsai, S. Q., Topkar, V. V., Nguyen, N. T., Zheng, Z., … Joung, J. K. (2015). Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature*, *523*(7561), 481–485. https://doi.org/10.1038/nature14592

Kleinstiver, B. P., Tsai, S. Q., Prew, M. S., Nguyen, N. T., Welch, M. M., Lopez, J. M., … Joung, J. K. (2016). Genome-wide specificities of CRISPR-Cas Cpf1 nucleases in human cells. *Nature Biotechnology*, *34*(8), 869–874. https://doi.org/10.1038/nbt.3620

Kokoza, V., Ahmed, A., Woon Shin, S., Okafor, N., Zou, Z., & Raikhel, A. S. (2010). Blocking of Plasmodium transmission by cooperative action of Cecropin A and Defensin A in transgenic Aedes aegypti mosquitoes. *Proceedings of the National Academy of Sciences*, *107*(18), 8111–8116. https://doi.org/10.1073/pnas.1003056107

Kondo, S., & Ueda, R. (2013). Highly Improved gene targeting by germline-specific Cas9 expression in Drosophila. *Genetics*, *195*(3), 715–721. https://doi.org/10.1534/genetics.113.156737

Koufopanou, V., & Burt, A. (2005). Degeneration and Domestication of a Selfish Gene in Yeast: Molecular Evolution Versus Site-Directed Mutagenesis. *Molecular Biology and*

Evolution, *22*(7), 1535–1538. https://doi.org/10.1093/molbev/msi149

Kyrou, K., Hammond, A. M., Galizi, R., Kranjc, N., Burt, A., Beaghton, A. K., … Crisanti, A. (2018). A CRISPR–Cas9 gene drive targeting doublesex causes complete population suppression in caged Anopheles gambiae mosquitoes. *Nature Biotechnology*, *36*(11), 1062–1066. https://doi.org/10.1038/nbt.4245

Larracuente, A. M., Presgraves, D. C., Cicchetti, R., & Ganetzky, B. (2012). The selfish Segregation Distorter gene complex of Drosophila melanogaster. *Genetics*, *192*(1), 33–53. https://doi.org/10.1534/genetics.112.141390

Laven, H., Cousserans, J., & Guille, G. (1972). Eradicating mosquitoes using translocations: A first field experiment. *Nature*, *236*(5348), 456–457. https://doi.org/10.1038/236456a0

Le, C. (n.d.). Addgene: Zhang Lab's CRISPR Frequently Asked Questions. Retrieved November 5, 2019, from https://www.addgene.org/crispr/zhang/faq/#HR

Leach, H., Van Timmeren, S., & Isaacs, R. (2016). Exclusion Netting Delays and Reduces *Drosophila suzukii* (Diptera: Drosophilidae) Infestation in Raspberries. *Journal of Economic Entomology*, *109*(5), 2151–2158. https://doi.org/10.1093/jee/tow157

Lengeler, C. (2004). Insecticide-treated bed nets and curtains for preventing malaria. *Cochrane Database of Systematic Reviews*, (2), CD000363. https://doi.org/10.1002/14651858.CD000363.pub2

Li, F., & Scott, M. J. (2016). CRISPR/Cas9-mediated mutagenesis of the white and Sex lethal loci in the invasive pest, Drosophila suzukii. *Biochemical and Biophysical Research Communications*, *469*(4), 911–916. https://doi.org/https://doi.org/10.1016/j.bbrc.2015.12.081

Li, J., & Handler, A. M. (2017). Temperature-dependent sex-reversal by a transformer-2 gene-edited mutation in the spotted wing drosophila, Drosophila suzukii. *Scientific Reports*, *7*(1), 12363. https://doi.org/10.1038/s41598-017-12405-4

Lin, J., Chen, H., Luo, L., Lai, Y., Xie, W., & Kee, K. (2015). Creating a monomeric endonuclease TALE-I-SceI with high specificity and low genotoxicity in human cells. *Nucleic Acids Research*, *43*(2), 1112–1122. https://doi.org/10.1093/nar/gku1339

Lyttle, T W. (1979). Experimental population genetics of meiotic drive systems. II. Accumulation of genetic modifiers of segregation distorter (SD) in laboratory populations. *Genetics*, *91*(2), 339–357. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/407130

Lyttle, Terrence W. (1977). Experimental population genetics of meiotic drive systems. I. Pseudo-Y chromosomal drive as a means of eliminating cage populations of Drosophila melanogaster. *Genetics*, *86*(2), 413–445.

Makarova, K. S., Wolf, Y. I., & Koonin, E. V. (2018). Classification and Nomenclature of CRISPR-Cas Systems: Where from Here? *The CRISPR Journal*, *1*(5), 325–336. https://doi.org/10.1089/crispr.2018.0033

Mancini, M. V., Spaccapelo, R., Damiani, C., Accoti, A., Tallarita, M., Petraglia, E., … Favia, G. (2016). Paratransgenesis to control malaria vectors: a semi-field pilot study. *Parasites & Vectors*, *9*(1), 140. https://doi.org/10.1186/s13071-016-1427-3

Manzoni, G., Briquet, S., Risco-Castillo, V., Gaultier, C., Topçu, S., Ivănescu, M. L., … Silvie, O. (2014). A rapid and robust selection procedure for generating drug-selectable marker-free recombinant malaria parasites. *Scientific Reports*, *4*(1), 4760.

https://doi.org/10.1038/srep04760

Marshall, J. M., & Hay, B. A. (2014). Medusa: A Novel Gene Drive System for Confined Suppression of Insect Populations. *PLoS ONE*, *9*(7), e102694. https://doi.org/10.1371/journal.pone.0102694

Meredith, J. M., Basu, S., Nimmo, D. D., Larget-Thiery, I., Warr, E. L., Underhill, A., … Eggleston, P. (2011). Site-Specific Integration and Expression of an Anti-Malarial Gene in Transgenic Anopheles gambiae Significantly Reduces Plasmodium Infections. *PLoS ONE*, *6*(1), e14587. https://doi.org/10.1371/journal.pone.0014587

Moreira, C. K., Rodrigues, F. G., Ghosh, A., Varotti, F. de P., Miranda, A., Daffre, S., … Moreira, L. A. (2007). Effect of the antimicrobial peptide gomesin against different life stages of Plasmodium spp. *Experimental Parasitology*, *116*(4), 346–353. https://doi.org/10.1016/j.exppara.2007.01.022

Moreira, L. A., Ito, J., Ghosh, A., Devenport, M., Zieler, H., Abraham, E. G., … Jacobs-Lorena, M. (2002). Bee Venom Phospholipase Inhibits Malaria Parasite Development in Transgenic Mosquitoes. *Journal of Biological Chemistry*, *277*(43), 40839–40843. https://doi.org/10.1074/jbc.M206647200

Moreno-Mateos, M. A., Fernandez, J. P., Rouet, R., Vejnar, C. E., Lane, M. A., Mis, E., … Giraldez, A. J. (2017). CRISPR-Cpf1 mediates efficient homology-directed repair and temperature-controlled genome editing. *Nature Communications*, *8*(1), 1–9. https://doi.org/10.1038/s41467-017-01836-2

Mori, A., Chadee, D. D., Graham, D. H., & Severson, D. W. (2004). Reinvestigation of an Endogenous Meiotic Drive System in the Mosquito, Aedes aegypti (Diptera: Culicidae). *Journal of Medical Entomology*, *41*(6), 1027–1033. https://doi.org/10.1603/0022-2585-41.6.1027

Mumford, J. D., Long, C. A., Weaver, S. C., Miura, K., Wang, E., Rotenberry, R., … Benedict, M. Q. (2019). Plasmodium falciparum (Haemosporodia: Plasmodiidae) and O'nyong-nyong Virus Development in a Transgenic Anopheles gambiae (Diptera: Culicidae) Strain. *Journal of Medical Entomology*, *56*(4), 936–941. https://doi.org/10.1093/jme/tjz032

Nájera, J. A., González-Silva, M., & Alonso, P. L. (2011). Some lessons for the future from the Global Malaria Eradication Programme (1955-1969). *PLoS Medicine*, *8*(1), e1000412. https://doi.org/10.1371/journal.pmed.1000412

Nardin, E. H., Nussenzweig, V., Nussenzweig, R. S., Collins, W. E., Harinasuta, K. T., Tapchaisri, P., & Chomcharn, Y. (1982). Circumsporozoite proteins of human malaria parasites Plasmodium falciparum and Plasmodium vivax. *The Journal of Experimental Medicine*, *156*(1), 20–30. https://doi.org/10.1084/jem.156.1.20

NAS. (2016). *Gene Drives on the Horizon*. https://doi.org/10.17226/23405

Ndiaga, T. (2019, September 18). Scientists release sterile mosquitoes in Burkina to fight malaria. *Reuters*.

Noble, C., Adlam, B., Church, G. M., Esvelt, K. M., & Nowak, M. A. (2018). Current CRISPR gene drive systems are likely to be highly invasive in wild populations. *ELife*, *7*. https://doi.org/10.7554/eLife.33423

Noble, C., Olejarz, J., Esvelt, K. M., Church, G. M., & Nowak, M. A. (2017). Evolutionary dynamics of CRISPR gene drives. *Science Advances*, *3*(4), e1601964.

https://doi.org/10.1126/sciadv.1601964

Noedl, H., Se, Y., Schaecher, K., Smith, B. L., Socheat, D., Fukuda, M. M., & Artemisinin Resistance in Cambodia 1 (ARC1) Study Consortium. (2008). Evidence of Artemisinin-Resistant Malaria in Western Cambodia. *New England Journal of Medicine*, *359*(24), 2619–2620. https://doi.org/10.1056/NEJMc0805011

O'Brochta, D. A., Alford, R., Harrell, R., Aluvihare, C., Eappen, A. G., Li, T., … Billingsley, P. F. (2019). Is Saglin a mosquito salivary gland receptor for Plasmodium falciparum? *Malaria Journal*, *18*(1), 2. https://doi.org/10.1186/s12936-018-2634-5

O'Brochta, D. A., Sethuraman, N., Wilson, R., Hice, R. H., Pinkerton, A. C., Levesque, C. S., … Atkinson, P. W. (2003, November 1). Gene vector and transposable element behavior in mosquitoes. *Journal of Experimental Biology*. The Company of Biologists Ltd. https://doi.org/10.1242/jeb.00638

Oberhofer, G., Ivy, T., & Hay, B. A. (2018). Behavior of homing endonuclease gene drives targeting genes required for viability or female fertility with multiplexed guide RNAs. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(40), E9343–E9352. https://doi.org/10.1073/pnas.1805278115

Orhan, A., Aslantas, R., Önder, B. Ş., & Tozlu, G. (2016). First record of the invasive vinegar fly Drosophila suzukii (Matsumura)(Diptera: Drosophilidae) from eastern Turkey. *Turkish Journal of Zoology*, *40*, 290–293. https://doi.org/10.3906/zoo-1412-25

Petersen, M., Armisén, D., Gibbs, R. A., Hering, L., Khila, A., Mayer, G., … Misof, B. (2019). Diversity and evolution of the transposable element repertoire in arthropods with particular reference to insects. *BMC Evolutionary Biology*, *19*(1), 11. https://doi.org/10.1186/s12862-018-1324-9

Pike, A., & Dimopoulos, G. (2018). Genetic modification of Anopheles stephensi for resistance to multiple Plasmodium falciparum strains does not influence susceptibility to o'nyong'nyong virus or insecticides, or Wolbachia-mediated resistance to the malaria parasite. *PLOS ONE*, *13*(4), e0195720. https://doi.org/10.1371/journal.pone.0195720

Pike, A., Dong, Y., Dizaji, N. B., Gacita, A., Mongodin, E. F., & Dimopoulos, G. (2017). Changes in the microbiota cause genetically modified *Anopheles* to spread in a population. *Science*, *357*(6358), 1396–1399. https://doi.org/10.1126/science.aak9691

Port, F., & Bullock, S. L. (2016). Augmenting CRISPR applications in Drosophila with tRNA-flanked sgRNAs. *Nature Methods*, *13*(10), 852–854. https://doi.org/10.1038/nmeth.3972

Port, F., Chen, H.-M., Lee, T., & Bullock, S. L. (2014). Optimized CRISPR/Cas tools for efficient germline and somatic genome engineering in Drosophila. *Proceedings of the National Academy of Sciences*, *111*(29), E2967–E2976. https://doi.org/10.1073/pnas.1405500111

Port, F., Muschalik, N., & Bullock, S. L. (2015). Systematic Evaluation of Drosophila CRISPR Tools Reveals Safe and Robust Alternatives to Autonomous Gene Drives in Basic Research. *G3 (Bethesda, Md.)*, *5*(7), 1493–1502. https://doi.org/10.1534/g3.115.019083

Raag, R., & Whitlow, M. (1995). Single-chain Fvs. *The FASEB Journal*, *9*(1), 73–80. https://doi.org/10.1096/fasebj.9.1.7821762

Rai, K. S., Grover, K. K., & Suguna, S. G. (1973). Genetic manipulation of Aedes aegypti:

incorporation and maintenance of a genetic marker and a chromosomal translocation in natural populations. *Bulletin of the World Health Organization*, *48*(1), 49–56. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/4541148

Ran, F. A., Hsu, P. D., Lin, C. Y., Gootenberg, J. S., Konermann, S., Trevino, A. E., … Zhang, F. (2013). Double nicking by RNA-guided CRISPR cas9 for enhanced genome editing specificity. *Cell*, *154*(6), 1380–1389. https://doi.org/10.1016/j.cell.2013.08.021

Rattray, A., Santoyo, G., Shafer, B., & Strathern, J. N. (2015). Elevated mutation rate during meiosis in Saccharomyces cerevisiae. *PLoS Genetics*, *11*(1), e1004910. https://doi.org/10.1371/journal.pgen.1004910

Richman, A. M., Dimopoulos, G., Seeley, D., & Kafatos, F. C. (1997). Plasmodium activates the innate immune response of Anopheles gambiae mosquitoes. *The EMBO Journal*, *16*(20), 6114–6119. https://doi.org/10.1093/emboj/16.20.6114

Robinson, A. S. (1976). Progress in the use of chromosomal translocations for the control of insect pests. *Biological Reviews*, *51*(1), 1–24. https://doi.org/10.1111/j.1469-185X.1976.tb01118.x

Rodriguez, M. D., Zamudio, F., Torres, J. A., Gonzalezceron, L., Possani, L. D., & Rodriguez, M. H. (1995). Effect of a Cecropin-like Synthetic Peptide (Shiva-3) on the Sporogonic Development of Plasmodium berghei. *Experimental Parasitology*, *80*(4), 596–604. https://doi.org/10.1006/expr.1995.1075

Roggenkamp, E., Giersch, R. M., Schrock, M. N., Turnquist, E., Halloran, M., & Finnigan, G. C. (2018). Tuning CRISPR-Cas9 Gene Drives in Saccharomyces cerevisiae. *G3 (Bethesda, Md.)*, *8*(3), 999–1018. https://doi.org/10.1534/g3.117.300557

Rono, M. K., Whitten, M. M. A., Oulad-Abdelghani, M., Levashina, E. A., & Marois, E. (2010). The Major Yolk Protein Vitellogenin Interferes with the Anti-Plasmodium Response in the Malaria Mosquito Anopheles gambiae. *PLoS Biology*, *8*(7), e1000434. https://doi.org/10.1371/journal.pbio.1000434

Ros, G., Anfora, G., Grassi, A., & Ioriatti, C. (2012). The potential economic impact of Drosophila suzukii on small fruits production in Trentino (Italy).

Rossi Stacconi, M. V., Grassi, A., Ioriatti, C., & Anfora, G. (2019). Augmentative releases of Trichopria drosophilae for the suppression of early season Drosophila suzukii populations. *BioControl*, *64*(1), 9–19. https://doi.org/10.1007/s10526-018-09914-0

Rudin, N., Sugarman, E., & Haber, J. E. (1989). Genetic and physical analysis of double-strand break repair and recombination in Saccharomyces cerevisiae. *Genetics*, *122*(3), 519–534. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/2668114

Ryan, G. D., Emiljanowicz, L., Wilkinson, F., Kornya, M., & Newman, J. A. (2016). Thermal Tolerances of the Spotted-Wing Drosophila *Drosophila suzukii* (Diptera: Drosophilidae). *Journal of Economic Entomology*, *109*(2), 746–752. https://doi.org/10.1093/jee/tow006

Schetelig, M. F., Yan, Y., Zhao, Y., & Handler, A. M. (2019). Genomic targeting by recombinase-mediated cassette exchange in the spotted wing drosophila, Drosophila suzukii. *Insect Molecular Biology*, *28*(2), 187–195. https://doi.org/10.1111/imb.12537

Schetelig, Marc F., & Handler, A. M. (2013). Germline transformation of the spotted wing drosophilid, Drosophila suzukii, with a piggyBac transposon vector. *Genetica*, *141*(4–6), 189–193. https://doi.org/10.1007/s10709-013-9717-6

Schwirz, J., Yan, Y., Franta, Z., & Schetelig, M. F. (2019). Bicistronic expression and

differential localization of proteins in insect cells and Drosophila suzukii using picornaviral 2A peptides. *BioRxiv*, 766188. https://doi.org/10.1101/766188

Scott, J. G., Michel, K., Bartholomay, L. C., Siegfried, B. D., Hunter, W. B., Smagghe, G., … Douglas, A. E. (2013). Towards the elements of successful insect RNAi. *Journal of Insect Physiology*, *59*(12), 1212–1221. https://doi.org/10.1016/j.jinsphys.2013.08.014

Scott, M. J., Morrison, N. I., & Simmons, G. S. (2014). Transgenic approaches for sterile insect control of dipteran livestock pests and lepidopteran crop pests. In *Transgenic insects: techniques and applications* (pp. 152–167). Wallingford: CABI. https://doi.org/10.1079/9781780644516.0152

Serebrovsky, A. (1969). On the possibility of a new method for the control of insect pests. *Sterile-Male Technique for Eradication or Control of Harmful Insects. Proceedings of a Panel on Application of the Sterile-Male Technique for the Eradication or Control of Harmful Species of Insects, Organised by the Joint FAO/IAEA Division of Atomic Ener*. Vienna, Int. Atom. Energy Ag. Retrieved from https://www.cabdirect.org/cabdirect/abstract/19701000720

Serrato-capuchina, A., Zhang, S., Martin, W., Peede, D., Earley, E., & Matute, D. R. (2018). Recent invasion of P transposable element into Drosophila yakuba. *BioRxiv*, 453829. https://doi.org/10.1101/453829

Servick, K. (2019). Study on DNA spread by genetically modified mosquitoes prompts backlash. *Science*. https://doi.org/10.1126/science.aaz5392

Shapiro, R. S., Chavez, A., Porter, C. B. M., Hamblin, M., Kaas, C. S., DiCarlo, J. E., … Collins, J. J. (2018). A CRISPR–Cas9-based gene drive platform for genetic interaction analysis in Candida albicans. *Nature Microbiology*, *3*(1), 73–82. https://doi.org/10.1038/s41564-017-0043-0

Shin, S. W., Kokoza, V. A., & Raikhel, A. S. (2003). Transgenesis and reverse genetics of mosquito innate immunity. *The Journal of Experimental Biology*, *206*(Pt 21), 3835–3843. https://doi.org/10.1242/jeb.00640

Simões, M. L., Dong, Y., Hammond, A., Hall, A., Crisanti, A., Nolan, T., & Dimopoulos, G. (2017). The Anopheles FBN9 immune factor mediates Plasmodium species-specific defense through transgenic fat body expression. *Developmental & Comparative Immunology*, *67*, 257–265. https://doi.org/10.1016/j.dci.2016.09.012

Simoni, A., Siniscalchi, C., Chan, Y.-S., Huen, D. S., Russell, S., Windbichler, N., & Crisanti, A. (2014). Development of synthetic selfish elements based on modular nucleases in Drosophila melanogaster. *Nucleic Acids Research*, *42*(11), 7461–7472. https://doi.org/10.1093/nar/gku387

Slaymaker, I. M., Gao, L., Zetsche, B., Scott, D. A., Yan, W. X., & Zhang, F. (2015). Rationally engineered Cas9 nucleases with improved specificity. *Science*, *351*(6268), 84–88. https://doi.org/10.1126/science.aad5227

Smith, R. C., Vega-Rodríguez, J., & Jacobs-Lorena, M. (2014). The Plasmodium bottleneck: Malaria parasite losses in the mosquito vector. *Memorias Do Instituto Oswaldo Cruz*, *109*(5), 644–661. https://doi.org/10.1590/0074-0276130597

Srivastava, M., Nambiar, M., Sharma, S., Karki, S. S., Goldsmith, G., Hegde, M., … Raghavan, S. C. (2012). An inhibitor of nonhomologous end-joining abrogates double-strand break repair and impedes cancer progression. *Cell*, *151*(7), 1474–1487. https://doi.org/10.1016/j.cell.2012.11.054

Sternberg, S. H., Redding, S., Jinek, M., Greene, E. C., & Doudna, J. A. (2014). DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature*, *507*(7490), 62–67. https://doi.org/10.1038/nature13011

Stockton, D. G., Wallingford, A. K., & Loeb, G. M. (2018). Phenotypic Plasticity Promotes Overwintering Survival in A Globally Invasive Crop Pest, Drosophila suzukii. *Insects*, *9*(3). https://doi.org/10.3390/insects9030105

Stork, N. E., McBroom, J., Gely, C., & Hamilton, A. J. (2015). New approaches narrow global species estimates for beetles, insects, and terrestrial arthropods. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(24), 7519–7523. https://doi.org/10.1073/pnas.1502408112

Sumitani, M., Kasashima, K., Yamamoto, D. S., Yagi, K., Yuda, M., Matsuoka, H., & Yoshida, S. (2013). Reduction of malaria transmission by transgenic mosquitoes expressing an antisporozoite antibody in their salivary glands. *Insect Molecular Biology*, *22*(1), 41–51. https://doi.org/10.1111/j.1365-2583.2012.01168.x

Sweeny, T. L., & Barr, A. R. (1978). Sex Ratio Distortion Caused by Meiotic Drive in a Mosquito, Culex pipiens L. *Genetics*, *88*(3), 427–446. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/17248804

Ta, T. H., Hisam, S., Lanza, M., Jiram, A. I., Ismail, N., & Rubio, J. M. (2014). First case of a naturally acquired human infection with Plasmodium cynomolgi. *Malaria Journal*, *13*, 68. https://doi.org/10.1186/1475-2875-13-68

Teem, J. L., Ambali, A., Glover, B., Ouedraogo, J., Makinde, D., & Roberts, A. (2019). Problem formulation for gene drive mosquitoes designed to reduce malaria transmission in Africa: results from four regional consultations 2016–2018. *Malaria Journal*, *18*(1), 347. https://doi.org/10.1186/s12936-019-2978-5

Thomas, D. D. (2000). Insect Population Control Using a Dominant, Repressible, Lethal Genetic System. *Science*, *287*(5462), 2474–2476. https://doi.org/10.1126/science.287.5462.2474

Toxopeus, J., Jakobs, R., Ferguson, L. V, Gariepy, T. D., & Sinclair, B. J. (2016). Reproductive arrest and stress resistance in winter-acclimated Drosophila suzukii. *Journal of Insect Physiology*. https://doi.org/10.1016/j.jinsphys.2016.03.006

Triller, G., Scally, S. W., Costa, G., Pissarev, M., Kreschel, C., Bosch, A., … Wardemann, H. (2017). Natural Parasite Exposure Induces Protective Human Anti-Malarial Antibodies. *Immunity*, *47*(6), 1197-1209.e10. https://doi.org/https://doi.org/10.1016/j.immuni.2017.11.007

Unckless, R. L., Clark, A. G., & Messer, P. W. (2017). Evolution of Resistance Against CRISPR/Cas9 Gene Drive. *Genetics*, *205*(2), 827–841. https://doi.org/10.1534/genetics.116.197285

Unckless, R. L., Messer, P. W., Connallon, T., & Clark, A. G. (2015). Modeling the manipulation of natural populations by the mutagenic Chain reaction. *Genetics*, *201*(2), 425–431. https://doi.org/10.1534/genetics.115.177592

Urnov, F. D., Rebar, E. J., Holmes, M. C., Zhang, H. S., & Gregory, P. D. (2010). Genome editing with engineered zinc finger nucleases. *Nature Reviews Genetics*, *11*(9), 636–646. https://doi.org/10.1038/nrg2842

Van Timmeren, S., & Isaacs, R. (2013). Control of spotted wing drosophila, Drosophila

suzukii, by specific insecticides and by conventional and organic crop protection programs. *Crop Protection*, *54*, 126–133. https://doi.org/10.1016/J.CROPRO.2013.08.003

Vella, M. R., Gunning, C. E., Lloyd, A. L., & Gould, F. (2017). Evaluating strategies for reversing CRISPR-Cas9 gene drives. *Scientific Reports*, *7*(1), 11038. https://doi.org/10.1038/s41598-017-10633-2

Verkuijl, S. A., & Rots, M. G. (2019). The influence of eukaryotic chromatin state on CRISPR–Cas9 editing efficiencies. *Current Opinion in Biotechnology*, *55*, 68–73. https://doi.org/10.1016/J.COPBIO.2018.07.005

Vlachou, D., Schlegelmilch, T., Christophides, G. K., & Kafatos, F. C. (2005). Functional Genomic Analysis of Midgut Epithelial Responses in Anopheles during Plasmodium Invasion. *Current Biology*, *15*(13), 1185–1195. https://doi.org/10.1016/J.CUB.2005.06.044

Volohonsky, G., Hopp, A.-K., Saenger, M., Soichot, J., Scholze, H., Boch, J., … Marois, E. (2017). Transgenic Expression of the Anti-parasitic Factor TEP1 in the Malaria Mosquito Anopheles gambiae. *PLOS Pathogens*, *13*(1), e1006113. https://doi.org/10.1371/journal.ppat.1006113

Wallace, B. (1948). Studies on "Sex-Ratio" in Drosophila pseudoobscura. I. Selection and "Sex- Ratio." *Evolution*, *2*(3), 189. https://doi.org/10.2307/2405380

Walsh, D. B., Bolda, M. P., Goodhue, R. E., Dreves, A. J., Lee, J., Bruck, D. J., … Zalom, F. G. (2011). Drosophila suzukii (Diptera: Drosophilidae): Invasive Pest of Ripening Soft Fruit Expanding its Geographic Range and Damage Potential. *Journal of Integrated Pest Management*, *2*(1), 1–7. https://doi.org/10.1603/IPM10010

Wang, J., Zhang, Y., Zhao, Y. O., Li, M. W. M., Zhang, L., Dragovic, S., … Fikrig, E. (2013). Anopheles gambiae Circumsporozoite Protein–Binding Protein Facilitates Plasmodium Infection of Mosquito Salivary Glands. *The Journal of Infectious Diseases*, *208*(7), 1161–1169. https://doi.org/10.1093/infdis/jit284

Wang, M., Mao, Y., Lu, Y., Tao, X., & Zhu, J. kang. (2017). Multiplex Gene Editing in Rice Using the CRISPR-Cpf1 System. *Molecular Plant*, *10*(7), 1011–1013. https://doi.org/10.1016/j.molp.2017.03.001

Weers, P. M., Van Marrewijk, W. J., Beenakkers, A. M., & Van der Horst, D. J. (1993). Biosynthesis of locust lipophorin. Apolipophorins I and II originate from a common precursor. *The Journal of Biological Chemistry*, *268*(6), 4300–4303. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8440714

Werren, J. H., Nur, U., & Wu, C. I. (1988, November 1). Selfish genetic elements. *Trends in Ecology and Evolution*. Elsevier Current Trends. https://doi.org/10.1016/0169-5347(88)90105-X

Whitten, M., Shiao, S. H., & Levashina, E. A. (2006). Mosquito midguts and malaria: cell biology, compartmentalization and immunology. *Parasite Immunology*, *28*(4), 121–130. https://doi.org/10.1111/j.1365-3024.2006.00804.x

Wilson, L. O. W., O'Brien, A. R., & Bauer, D. C. (2018). The Current State and Future of CRISPR-Cas9 gRNA Design Tools. *Frontiers in Pharmacology*, *9*, 749. https://doi.org/10.3389/fphar.2018.00749

Windbichler, N, Menichelli, M., Papathanos, P. A., Thyme, S. B., Li, H., Ulge, U. Y., …

Crisanti, A. (2011). A synthetic homing endonuclease-based gene drive system in the human malaria mosquito. *Nature*, *473*(7346), 212–215. https://doi.org/10.1038/nature09937

Windbichler, Nikolai, Papathanos, P. A., & Crisanti, A. (2008). Targeting the X chromosome during spermatogenesis induces Y chromosome transmission ratio distortion and early dominant embryo lethality in Anopheles gambiae. *PLoS Genetics*, *4*(12), e1000291. https://doi.org/10.1371/journal.pgen.1000291

Wirth, D. F. (2002). Biological revelations. *Nature*, *419*(6906), 495–496. https://doi.org/10.1038/419495a

World Health Organization. Global Malaria Programme. (n.d.). *Guidelines for malaria vector control*.

World Health Organization. (2017a). *A Framework for Malaria Elimination*. *WHO Press, World Health Organization*. Geneva.

World Health Organization. (2017b). Vector-borne diseases. Retrieved September 25, 2019, from https://www.who.int/news-room/fact-sheets/detail/vector-borne-diseases

World Health Organization. (2018). *World malaria report 2018*. Geneva.

Wu, B., Luo, L., & Gao, X. J. (2016). Cas9-triggered chain ablation of cas9 as a gene drive brake. *Nature Biotechnology*, *34*(2), 137–138. https://doi.org/10.1038/nbt.3444

Xie, K., Minkenberg, B., & Yang, Y. (2015). Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-processing system. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(11), 3570–3575. https://doi.org/10.1073/pnas.1420294112

Yan, Y., & Finnigan, G. C. (2019). Analysis of CRISPR gene drive design in budding yeast. *Access Microbiology*, acmi000059. https://doi.org/10.1099/acmi.0.000059

Yoshida, S., Matsuoka, H., Luo, E., Iwai, K., Arai, M., Sinden, R. E., & Ishii, A. (1999). A single-chain antibody fragment specific for the Plasmodium berghei ookinete protein Pbs21 confers transmission blockade in the mosquito midgut. *Molecular and Biochemical Parasitology*, *104*(2), 195–204. https://doi.org/10.1016/S0166-6851(99)00158-9

Yoshioka, S., Fujii, W., Ogawa, T., Sugiura, K., & Naito, K. (2015). Development of a mono-promoter-driven CRISPR/Cas9 system in mammalian cells. *Scientific Reports*, *5*, 18341. https://doi.org/10.1038/srep18341

Yoshitake, Y., Inomata, N., Sano, M., Kato, Y., & Itoh, M. (2018). The P element invaded rapidly and caused hybrid dysgenesis in natural populations of Drosophila simulans in Japan. *Ecology and Evolution*, *8*(19), 9590–9599. https://doi.org/10.1002/ece3.4239

Zetsche, B., Gootenberg, J. S., Abudayyeh, O. O., Slaymaker, I. M., Makarova, K. S., Essletzbichler, P., … Zhang, F. (2015). Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell*, *163*(3), 759–771. https://doi.org/10.1016/j.cell.2015.09.038

Zhang, X., Koolhaas, W. H., & Schnorrer, F. (2014). A Versatile Two-Step CRISPR- and RMCE-Based Strategy for Efficient Genome Engineering in Drosophila. *G3: Genes|Genomes|Genetics*, *4*(12), 007864. https://doi.org/10.1534/g3.114.013979

Université de Strasbourg

Emily GREEN

École Doctorale des Sciences de la Vie et de la Santé STRASBOURG

Development of genetic control strategies for insect pests using CRISPR/Cas9

La lutte contre les insectes nuisibles et vecteurs reste un défi économique, environnemental et de santé publique important. Les stratégies de lutte génétique représentent une solution attrayante par rapport aux stratégies traditionnelles comme les pesticides, qui ont des effets négatifs importants. Une nouvelle stratégie génétique est l'utilisation forçage génétique (FG). Les constructions à FG sont des systèmes génétiques capables d'augmenter leur héritabilité et ainsi d'envahir rapidement une population. Le développement du système de modification génomique CRISPR/Cas9 a facilité la création des FGs synthétiques. Ce travail de thèse présente mes efforts pour développer des FGs chez deux espèces nuisibles importantes, *Anopheles gambiae*, un vecteur majeur du paludisme, et *Drosophila suzukii*, un ravageur agricole d'importance mondiale. Les objectifs de ce projet étaient de développer un FG à but d'élimination chez *D. suzukii*, visant à réduire la taille de la population, et un FG à but de modification chez *An. gambiae*, visant à rendre les moustiques incapables de transmettre le paludisme. Mes efforts pour développer un FG chez *D. suzukii* n'ont pas abouti, mais les techniques et protocoles présentés ici peuvent servir de base pour de futurs travaux sur cet insecte nuisible important. Chez *An. gambiae*, j'ai caractérisé avec succès deux lignées transgéniques, dont l'une capable de bloquer significativement la transmission chez un modèle rongeur. J'ai également généré des insertions de sites attP dans deux gènes essentiels, qui pourront servir de sites d'insertion de futures FGs. Enfin, je présente mes avancées dans la mise au point d'une nouvelle stratégie de FG à but de modification de population, le forçage génétique indirect.

Mots-clés : Forçage génétique, CRISPR/Cas9, *Anopheles gambiae*, *Drosophila suzukii*, Transgénèse

Insect pest control remains an important economic, environmental, and public health challenge. Genetic-based strategies are an attractive alternative to traditional control techniques like pesticides, which can have significant off-target effects on other species. One new genetic-based strategy is the use of gene drives. These are genetic systems that can bias inheritance in their favor and rapidly invade a population. The development of the genome editing system CRISPR/Cas9 has facilitated the creation and design of artificial drives. This manuscript presents my efforts to develop gene drives in two important pest species, *Anopheles gambiae*, a major vector of malaria, and *Drosophila suzukii*, a global crop pest. The goals of this project were to develop a suppression gene drive in *D. suzukii,* to reduce population size, and a modification drive in *An. gambiae*, to reduce malaria transmission. While I was unable to produce a functional gene drive in *D. suzukii*, the efforts and protocols presented here can serve as a baseline for future work in this economically important crop pest. In *An. gambiae*, I successfully characterized two transgenic lines, one of which significantly blocks malaria transmission to a rodent model. I also generated dual-*att*P knock-ins into two genes, which could serve as insertion sites for future modification gene drives. Finally, I present my efforts to engineer a new modification gene drive strategy, indirect gene drive.

Keywords: Gene Drive, CRISPR/Cas9, *Anopheles gambiae*, *Drosophila suzukii*, Transgenesis