Gene Drive in *Drosophila melanogaster* and *Aedes aegypti*

David Navarro Payá

School of Biosciences

Cardiff University

Academic supervisors: Helen White-Cooper & Luke Alphey

Industrial supervisor: Kelly Matzen

This thesis is submitted for the degree of Doctor of Philosophy

September 2017

Table of Contents

Table of Figures

Table Summary

Acknowledgements

I thank my supervisors Helen White-Cooper and Luke Alphey for their invaluable support and for teaching me how to become a better scientist. Also, special thanks to Simona Caporilli for all the help in the laboratory and the fly lab as well as for her priceless company. I would like to thank Kelly Matzen for her guidance during my stay at Oxitec, Joe Turner for teaching me how to inject mosquito eggs, and Pam Gray for showing me how to take care of mosquitoes. Everybody in the Arthropod Genetics lab in The Pirbright Institute would deserve a mention for their warm welcome since my arrival, support and teaching during this last year. Special thanks go to Sanjay Basu for introducing me to the gene I ended up working with for almost a year. Rebekah Ireland also deserves a special thank you for all her support in both cups of tea and food supplies during the write-up of this thesis. Finally, I would also like to extend my gratitude to all the new friends I have made in this journey as well as the old friends, my partner and my family that have always been there.

Abbreviation Index

Ae. aegypti	*Aedes aegypti*
Ae. albopictus	*Aedes albopictus*
AmCyan	*Anemonia majano* cyan fluorescent protein
AWT	*Aedes aegypti* Asian wild type strain
bp	Base pair
Cas9	CRISPR associated protein 9
CDS	Coding DNA sequence
CHIKV	Chikungunya
cDNA	Complementary DNA
CRISPR	Clustered regularly interspersed palindromic repeats
CRP	Cytoplasmic ribosomal proteins
Ct value	Cycle threshold value
DALY	Disability adjusted life year
DBS	Double strand break
DENV	Dengue virus
DFL	Dominant female lethal
DNA	Deoxyribonucleic acid
DSB	Double strand break
dsDNA	Double stranded DNA
DsRed	*Discosoma* species red fluorescent protein
dsRNA	Double stranded RNA
D. melanogaster	*Drosophila melanogaster*
ECSA	East Central South African
fsRIDL	Female specific RIDL
gDNA	Genomic DNA
gRNA	Guide RNA
HDR	Homology directed repair
HR	Homologous Recombination

HEG	Homing endonuclease gene
HRMA	High resolution melt analysis
IIT	Incompatible insect technique
INDEL	Insertion or deletion
LVP	*Aedes aegypti* Liverpool strain
LWT	*Aedes aegypti* Latin wild type strain
mCherry	Monomeric *Discosoma* species red fluorescent protein
MLA	Multi-locus assortment
NEB	New England Biolabs
NHEJ	Non homologous end joining
NIPP1	Nuclear inhibitor of PP1
pBac	*piggyBac*
PCR	Polymerase chain reaction
PP1	Protein phosphatase 1
PSRP	Pool of semi-random primers
RE	Restriction enzyme
RIDL	Release of insects carrying a dominant lethal
RNA	Ribonucleic acid
RNAi	RNA interference
RS	Recognition sequence
RSCU	Relative synonymous codon usage
RT	Room temperature
RNA Pol II	RNA polymerase II
RNA Pol III	RNA polymerase III
sgRNA	Single guide RNA
SIT	Sterile insect technique
SNP	Single nucleotide polymorphism
ssRNA	Single stranded RNA
UD	Underdominance
WT	Wild type

ZFN	Zinc finger nucleases
ZsYellow	Zoanthus sp. yellow fluorescent protein

Abstract

The yellow fever mosquito *Aedes aegypti* is the main vector for several important arboviral diseases such as dengue, yellow fever, chikungunya and zika. With the advent of genetic control strategies, new species-specific tools have emerged for the control of *Aedes aegypti*. This thesis describes attempts at building different gene drive systems aiming for both population suppression or population replacement, as well as exploring the possibility of inserting exogenous sequences in the male locus of *Aedes aegypti*. An underdominance system, consisting of two mutually rescuing killers, was investigated in *Drosophila melanogaster*. It did not work as expected in the configuration tested. The chosen *NIPP1* killer gene could not be upregulated by *tTAV* when under the control of *hsp83, UAS* and *tetO*. *tetO* and *tTAV2* resulted in a lethal positive-feedback loop. *Gal4Groucho* and *LexAGroucho* fusion proteins, previously used as corepressors in the literature, were lethal when under the control of the *tetO-tTAV* system. Males showed the expected feminisation phenotypes, invovling male palp shortening and less feathery antennae, upon *Nix* knock-out using CRISPR-Cas9. However, subsequent homology directed repair into the *Nix* gene in the male locus of Aedes aegypti did not succeed even after reiterated injections. Setting out from the hypothesis of *Act4* haploinsufficiency in *Aedes aegypti*, the building of two different gene drive systems was attempted; female-specific underdominance and RIDL with drive. A CRISPR-Cas9 driven *act4* knock-in unexpectedly confirmed *act4* haplosufficiency in *Aedes aegypti*. Whilst the initially devised gene drive systems could not function as such, *act4* haplosufficiency marked the finding of a new female-specific recessive flightless target (effectively sterile) for use in future population suppression drive systems.

Chapter 1 - **Introduction**

1.1 Mosquito-Borne Diseases

1.1.1 An overview

Mosquito borne-diseases affecting humans such as dengue, malaria and yellow fever represent a major international health concern. More than 50% of the world's population live in areas at risk of these diseases according to the World Health Organisation (September 2017, WHO: http://www.who.int/en/). They are transmitted by mosquitoes from *Aedes*, *Anopheles*, and *Culex* genera (Table 1.1).

Table 1.1: An overview of mosquito-borne diseases:
An overview of infectious diseases transmitted by mosquitoes (WHO: http://www.who.int/mediacentre/factsheets/fs387/en/).

Infectious Disease	Mosquito Vector
Malaria	*Anopheles*
Dengue Fever	*Aedes*
Chikungunya	
Rift Valley Fever	
Yellow Fever	
Zika	
Japanese Encephalitis	*Culex*
West Nile Fever	

1.1.2 Malaria

Malaria is the most deadly mosquito-borne disease causing between 0.8 to 1.2 million deaths and more than 250 million clinical cases per year (Vaughan and Kappe 2012). The disease agents are protozoan parasites of the *Plasmodium* genus of which there are five species known to infect humans (Singh and Daneshvar 2013). It is transmitted by *Anopheles* mosquitoes and more than a

30% of the world's population lives in malaria endemic areas (Figure 1.1). Although a vaccine has been developed, RTS, S/AS01 (RTS 2015), a concern over its relatively limited efficacy has called for further development on other vaccines or alternative strategies (Mahmoudi and Keshavarz 2017).

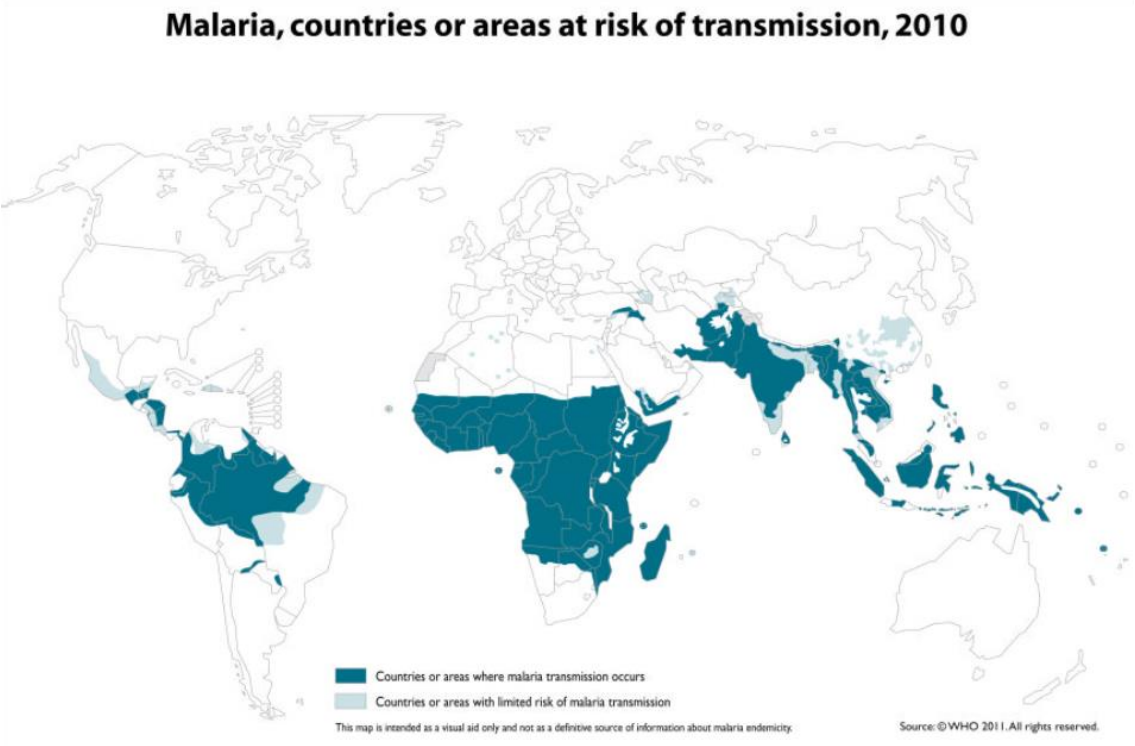


Figure 1.1: Areas at risk of transmission of malaria in 2010:
Taken from Autino *et al.* 2012. The distribution of territories at risk of malaria transmission in 2010 is shown above.

1.1.3 Dengue

Dengue is a viral infection transmitted between humans by mosquitoes (Simmons *et al.* 2012). Dengue viruses (DENVs) are arboviruses of the *Flavivirus* genus (family *Flaviviridae*), and are related to other clinically important arboviruses such as yellow fever and Japanese encephalitis (Messina *et al.* 2014). They consist of positive-sense single-stranded RNA genomes of

approximately 10kb in size. Since the first isolation of DENV in 1943 (Hotta 1951), four different serotypes of the virus have been found to be readily transmitted in humans (DENV1-4) (Gubler 2002). Recently, a new serotype has been found to be able to infect humans, DENV5, even though its prevalence appears to be relatively low (Mustafa *et al.* 2015). A recent study on dengue distribution was given by Bhatt *et al.* 2013, where 390 million dengue infections per year were estimated to occur worldwide, of which 96 million of them manifest clinically.

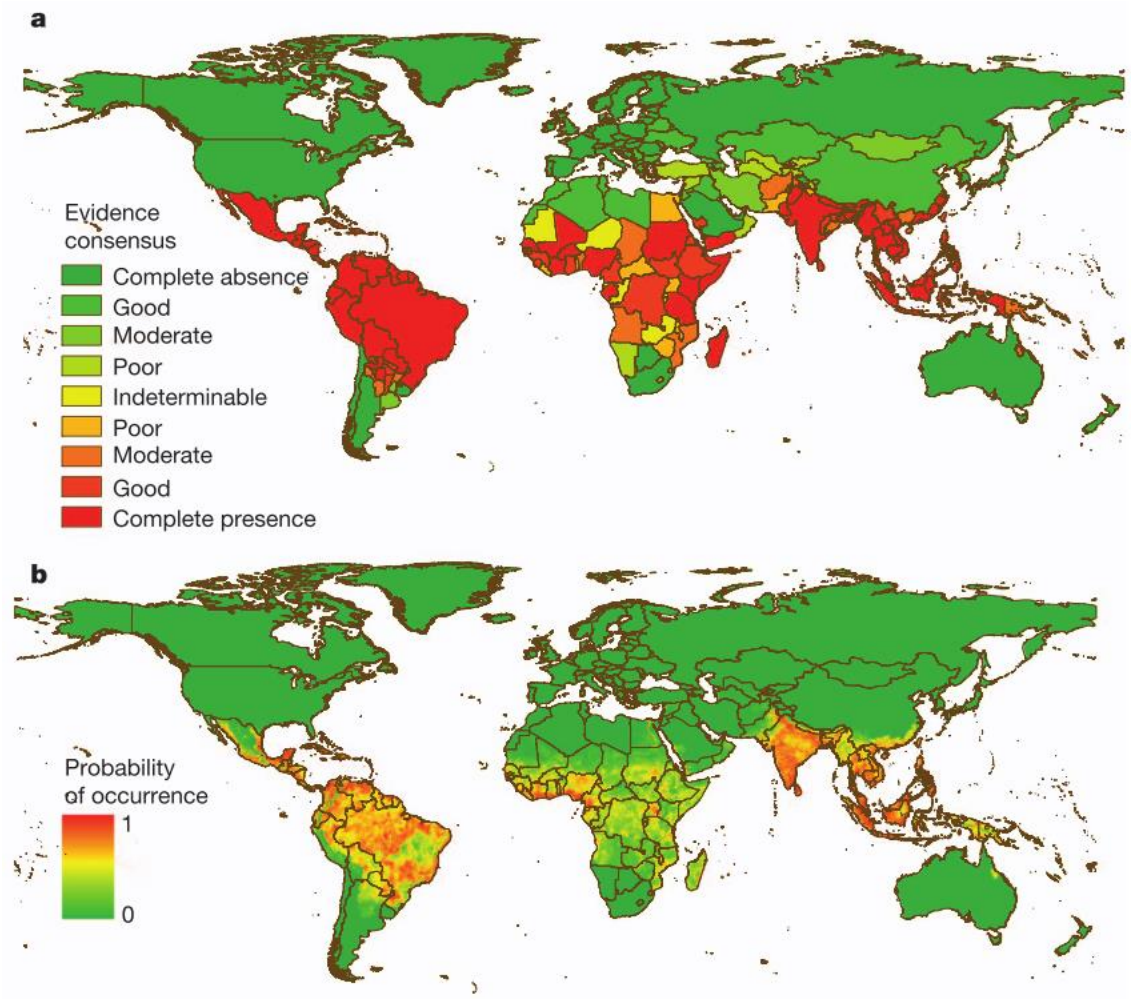


Figure 1.2: Evidence consensus on presence and risk of Dengue in 2010:
Taken from Bhatt *et al.* 2013. (**A**) National and subnational evidence consensus on complete dengue presence (red) through to complete absence (green). (**B**) Probability of dengue occurrence at 5km^2 spatial resolution. A high probability of occurrence is marked with red and a low probability is marked in green.

Dengue is the most prevalent and widespread of mosquito-borne viral diseases, with 2.1-3.7 billion people at risk of infection worldwide (Brady *et al.* 2012). Although rare, dengue can cause haemorrhagic fever, its most life-threatening symptom. Dengue causes fewer deaths than malaria, however, the case fatality rate of dengue can be as high as 15% or as low as 1% depending on the country (Gubler 2002). Its more than 30-fold increase in incidence in the last 30 years makes it a major infectious disease concern. The main vectors of dengue are the

mosquitoes *Aedes aegypti* and *Aedes albopictus* which are also vectors for other arboviruses such as Chikungunya and yellow fever. Infection of the salivary glands in mosquitoes occurs 7 to 10 days after up taking the virus upon a blood meal. This lifelong infection of the mosquito allows further spread to humans during subsequent blood meals (Salazar *et al.* 2007).

Human infection with a particular DENV serotype is thought to provide long-term serotype-specific immunity but only short-lived immunity to the other serotypes (Simmons *et al.* 2012). Development of a dengue vaccine has historically proven difficult due to the need to address all serotypes simultaneously. A recombinant, live, attenuated tetravalent dengue vaccine (CYD-TDV) has been recently developed and tested in two phase 3 trials. Although relatively promising with a 56-61% efficacy against symptomatic dengue, it did significantly increase the risk of hospitalised dengue cases amongst the 2-5 year age group (Hadinegoro *et al.* 2015). Since then, the vaccine has been licensed for use in individuals of 9-45 years of age, however, further evidence of vaccine-enhanced disease across different age groups has been reported (Halstead 2017). The potential to induce a harmful immune response with a dengue vaccine has been stated before due to the nature of dengue immunity itself; protective immunity after dengue infection increases the risk for more severe forms of dengue such as dengue shock or dengue hemorrhagic fever upon a secondary infection (Rothman and Ennis 2016). Two new tetravalent vaccines (NIH/Butantan and DENVax) are currently in Phase III trials (Screaton

and Mongkolsapaya 2017). However, given the challenges for dengue vaccine development, effective vector control still remains an important tool for dengue control.

1.1.4 Yellow fever

Yellow fever is caused by another *Flavivirus* which is also transmitted by *Aedes* mosquitoes, the main vector being *Aedes aegypti*. It is less widespread than dengue or malaria and is mainly endemic in rainforests of Africa and South America (Figure 1.3). Upon human infection with yellow fever virus, after an incubation period of 3 to 6 days, most individuals only suffer mild fever. However, this develops into a serious illness in about 15% of the cases.

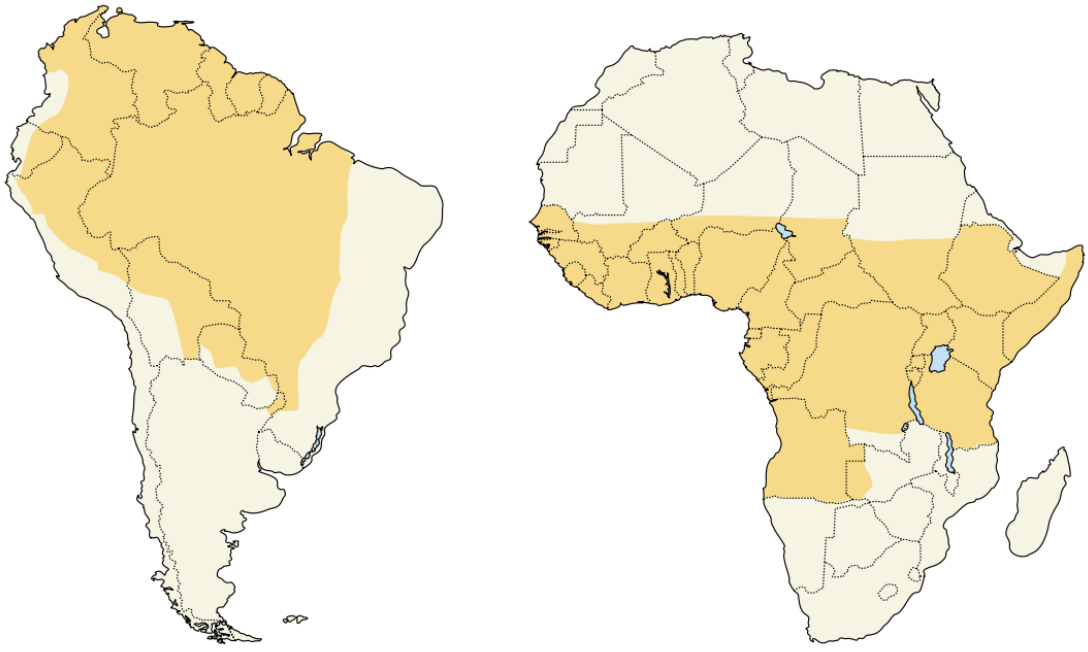


Figure 1.3: Global distribution of yellow fever:
Taken from Barrett and Higgs 2007. Territorial distribution of yellow fever virus in Africa and South America.

The virus has two main transmission cycles; sylvatic and urban. The sylvatic cycle involves transmission of the virus between a range of mosquito species and lower primates. When a mosquito, which normally bites monkeys, incidentally bites a human, potentially transmitting the viral pathogen, the urban cycle can develop. Anthropophilic mosquitoes such as *Aedes aegypti* then start transmission between humans in an urban setting (Barrett and Higgs 2007).

Yellow fever outbreaks were drastically reduced following the introduction of the live-attenuated 17D and FNV vaccines in the late 1930s allowing for a low-incidence period of about 25 years. A relaxation in vaccination and mosquito-control programs have led to a resurgence of yellow fever in recent years (Monath and Vasconcelos 2015). Although the urban yellow fever cycle is currently rare in South American countries there is a considerable risk of urban yellow fever returning to the Americas, given the presently widespread sylvatic cycle and increasing urbanisation (Barrett and Higgs 2007).

1.1.5 Chikungunya

The chikungunya (CHIKV) virus is an arbovirus of the *Alphavirus* genus (family *Togoviridae*). Chikungunya is a positive-sense single-stranded RNA virus with a genome of approximately 12kb (da Cunha and Trinta 2017). Similarly to dengue, chikungunya is rarely fatal except for the more vulnerable in a population (Pialoux *et al.* 2007). The symptoms usually involve acute fever followed by severe and persistent arthralgia (joint paint) during the chronic stage of the

disease (Burt *et al.* 2017). In some patients this debilitating joint pain can last for years. The main vectors for the disease are also *Aedes aegypti* and *Aedes albopictus*. Chikungunya is still a less pressing health concern than dengue or malaria, however, its rapid and recent spread across the world is alarming. Since the year 2000, the incidence of chikungunya outbreaks has increased with recent spread of the virus to previously non-endemic regions. Outbreaks have been detected in both Southern Europe and the Americas (Figure 1.4). With no current treatment or vaccine available chikungunya makes a strong case for mosquito vector control.

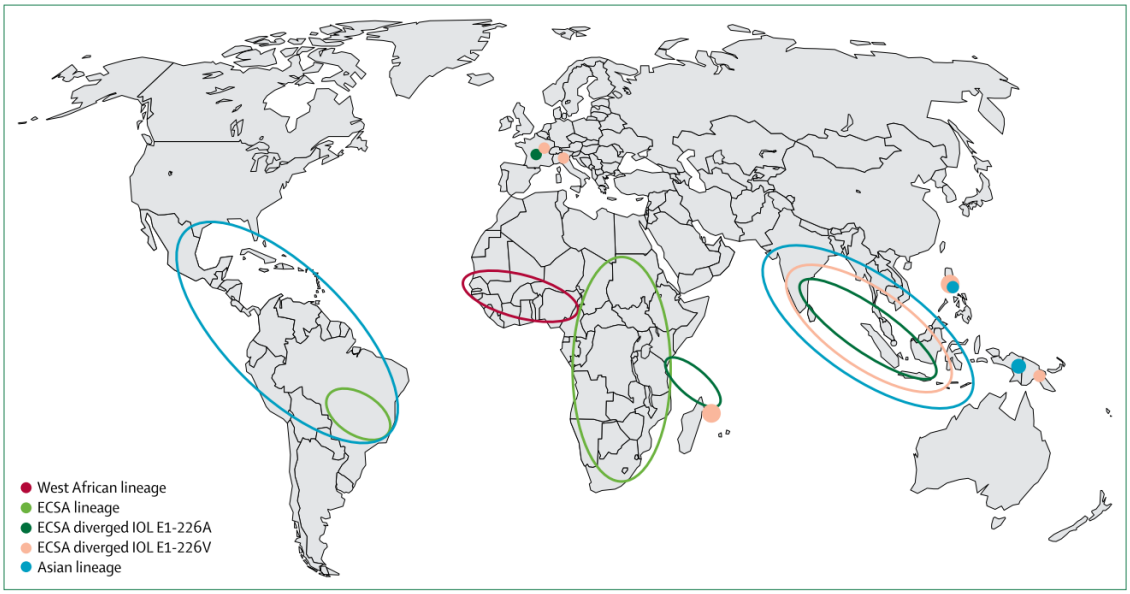


Figure 1.4: The spread of different chikungunya virus lineages across the globe:
Taken from Burt *et al.* 2017. The East Central South African (ECSA) lineage spread and diverged to the Indian Ocean islands (IOLs) and Asia. The Asian lineage in turn spread to the Americas in 2013 whilst the ECSA spread to Brazil in 2014.Outbreaks were also identified in Italy (2007) and France (2014).

1.1.6 Other mosquito-borne diseases

Zika is also an arbovirus from the *Flavivirus* genus (family *Flaviviridae*). The main vectors of Zika are *Aedes aegypti* and *Aedes albopictus* although other *Aedes* species are also thought to be capable of transmission (Song *et al.* 2017). Mosquitoes from other genera, *Culex* and *Anopheles*, have also been found to carry Zika virus (Diallo *et al.* 2014). Presence of the virus does not necessarily mean the species are Zika vectors, and hence more field studies are required to clarify which species can effectively transmit it. Although clinical aspects of Zika are yet to be explored in more detail, the symptoms are thought to range from acute febrile illness, neurological complications and adverse fetal outcomes (Petersen *et al.* 2016). Although known for the last 70 years, the virus has only spread recently. Within the span of a year, Zika virus reached Brazil from the Pacific Islands and thereafter spread rapidly throughout the Americas (Plourde and Bloch 2016).

Rift Valley fever virus (RVFV) is a mosquito-borne virus of the genus *Phlebovirus* (family *bunyaviridae*). The virus causes Rift Valley fever in both ruminants and humans. Current cases in humans are usually asymptomatic, and usually cases which develop clinical symptoms involve a short febrile illness with no chronic effects. However, a small number of cases still develop into severe Rift Valley fever (Mansfield *et al.* 2015). Both *Aedes* and *Culex* mosquitoes are thought to be involved in the disease transmission cycle, and

recent outbreaks outside of Africa are indicative of its potential geographical spread (Al-Afaleq and Hussein 2011).

Japanese encephalitis (JEV) and West Nile virus (WNV), both closely related, are also arboviruses of the *Flavivirus* genus (family *Flaviridae*) and are in this case transmitted by *Culex* mosquito species. Both viruses depend on avian reservoirs and infection in humans tends to be asymptomatic. However, cases of serious viral encephalitis have been reported (Turtle *et al.* 2012).

Due to the limitations of vaccines and preventative treatment in the mosquito-borne diseases outlined above, a case can be made in favour of reinforcing mosquito vector control.

1.2 The Yellow Fever Mosquito, *Aedes aegypti*

1.2.1 Morphology and life-cycle

Aedes aegypti (Linnaeus, 1762) is commonly known as the yellow fever mosquito. Adults are medium sized mosquitoes of approximately 4 to 7 mm in length (Clemons *et al.* 2010). Size is largely dependent on larval diet. It has a distinctive lyre shaped markings of white scales on the dorsal side of the thorax which are also present on abdominal tergites and as bands on their legs.

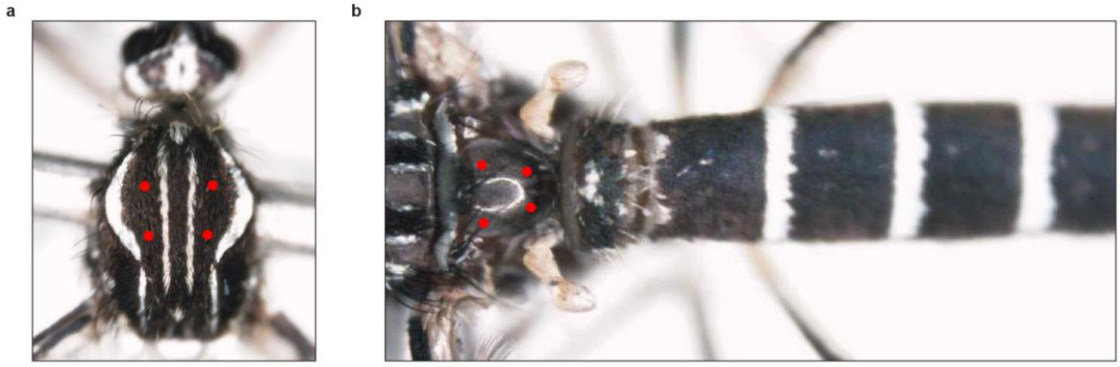


Figure 1.5: Characteristic white scale markings in an *Aedes aegypti* female:
Taken from McBride *et al.* 2014. Red dots do not serve a purpose here, but were used to measure scale colour in the cited study. (**A**) Lyre pattern of white scales on the dorsal part of the thorax. (**B**) White scales can be appreciated on the abdominal tergites.

Like most mosquito species, blood is required for egg development in females. Eggs take 3-5 days to develop after a blood meal (Crampton *et al.* 1997). Adult females can produce around five egg batches in their lifetime, and lay around 70-200 eggs per oviposition (Clemons *et al.* 2010). Eggs are resistant to desiccation and can be stored as long as a year (under optimal low humidity conditions) before hatching. Eggs hatch in response to stimuli, where submersion in water is

a critical stimulus and low oxygen tension is a further stimulus. The four larval stages (L1-L4) and the pupal stage are strictly aquatic. Eggs take around 3-4 days to hatch after oviposition and larvae typically take 6-9 days to pupate, although this is highly dependent on temperature and food availability. *Aedes aegypti* larvae have the distinctive ability to enter a diapause state under starvation conditions allowing them to remain alive for months (Crampton *et al.* 1997). This diapause state of larvae, together with the long term survivability of eggs, make *Aedes aegypti* both a suitable insect in terms of lab maintenance and a difficult pest to control. Pupae take around 48 hours to eclose into adults which become sexually mature in about 1 to 2 days (Lea 1968). Females are capable of taking blood meals 2 to 4 days after eclosion.

1.2.2 Genetics and genomics

A draft genome for *Aedes aegypti* (Liverpool strain) was published in Nene *et al.* 2007. The presented genome was approximately 1.38 Gb in size which makes it about 5 times larger than the published *Anopheles gambiae* genome (Holt *et al.* 2002). However, this increase in size is not due to a greater number of protein coding genes (Table 1.2) but an extensive expansion of transposable elements. Nearly 50% of the *Aedes aegypti* genome consists of transposable elements (Nene *et al.* 2007). Of the 15, 419 protein coding genes in *Aedes aegypti*, 67% of them were found to have orthologues in *Anopheles gambiae*, whilst 58% of them had orthologues in *Drosophila melanogaster*. This is consistent with the fact that estimated divergence from *Drosophila* lineages occurred ~250 million years ago

(Gaunt and Miles 2002) and ~150 million years ago from *Anopheles* lineages (Krzywinski *et al.* 2006).

Both *Aedes aegypti* and *Anopheles gambiae* show a considerable level of synteny between the three pairs of chromosomes in each species. However, the once proposed whole-arm synteny is now thought to not be fully preserved as a number of pericentric inversions have been found (Timoshevskiy *et al.* 2014). Nevertheless, the main difference is that sex chromosomes are homomorphic in *Aedes aegypti* whilst they are heteromorphic in *Anopheles gambiae*. A putative male-determining factor, *Nix*, has been recently found in the M-locus of *Ae. aegypti*'s chromosome 1 (Hall *et al.* 2015).

Table 1.2: Comparison of *Aedes aegypti*, *Anopheles gambiae* and *Drosophila melanogaster* genomes:

Adapted from Nene *et al.* 2007. Statistics were derived from the *Aedes aegypti* genome presented in Nene *et al.* 2007, the *Anopheles gambiae* R-AgamP3 assembly, and the *Drosophila melanogaster* R-4.2 assembly. The data reflects a 5 fold size difference between the *Ae. aegypti* and *Anopheles gambiae* genomes. However, the bigger genome size of *Aedes aegypti* is not reflected in a significant increase in the number of genes or average protein-coding gene length. This is explained in turn by the larger intron and intergenic region size in *Aedes aegypti*.

Feature	Species		
	Ae. aegypti	***An. gambiae***	***D. melanogaster***
Size (Mbp)	1,376	272.9	118
Number of chromosomes	3	3	4
Number of protein-coding genes	15,419	13,111	13,718
Average gene length (bp)	14,587	5,124	3,460
Average protein-coding gene length (bp)	1,397	1,154	1,693
Average intron length (bp)	4,685	808	1,175
Average length of intergenic region (bp)	56,417	17,265	6,043

1.2.3 Systematics and behaviour

Aedes aegypti is a species of African origin where domestication is thought to have occurred. *Aedes aegypti* individuals are divided into two subspecies *Aedes aegypti formosus* (native African) and *Aedes aegypti aegypti* (spread internationally as well as in Africa). *Aedes aegypti formosus* is less anthropophilic and hence is associated with sylvan rather than urban environments. The *Aedes aegypti aegypti* subspecies is thought to have originated from an *Aedes aegypti formosus* population in West African forests, a hypothesis supported by population genetic analyses. Whilst the urban or domestic form usually acquires a browner colour, *Aedes aegypti formosus* remains a darker black colour (Figure 1.6). A recent study found that a change in an odorant receptor gene, *Or4*, is behind the human 'domestication' of the *Aedes aegypti aegypti* subspecies from the *Aedes aegypti formosus* subspecies (McBride *et al.* 2014).

Aedes aegypti are decidedly anthropophilic and this has a strong impact on many aspects of behaviour. For instance, females prefer to take recurrent blood meals instead of relying on sources of sugar for energy reserves (Scott *et al.* 2000). The reason for this adaptation may be behind the fact that sugar sources in an urban environment may be scarcer than blood meals themselves. This adaptation has serious and negative consequences for humans as *Aedes aegypti* females become more efficient vectors of disease. Females of *Aedes aegypti* are monogamous as they are not receptive to sperm uptake after their first mating (Craig 1967).

Although it is not as strict as initially reported, given that polyandry does occur freely 2 hours after the first mating (Degner and Harrington 2016).This behaviour contrasts with that of males which will attempt mating multiple females. Mating of *Aedes aegypti* occurs in flight but it does not involve swarming as it happens mainly around the host. Males, also anthropophilic, will mate with females before and after they take blood meals from their human hosts. Males are drawn to the host by a combination of pheromones (Nijhout and Craig 1971) and the sound frequency of female flight (Arthur *et al.* 2014). Experiments restraining females, and attracting males to the females (with the use of tuning forks at the correct female flight frequency), showed that the males would approach the sound but in no case copulate with the females (Jones and Wheeler 1965). This highlights the critical importance of female flight in successful *Aedes aegypti* mating.

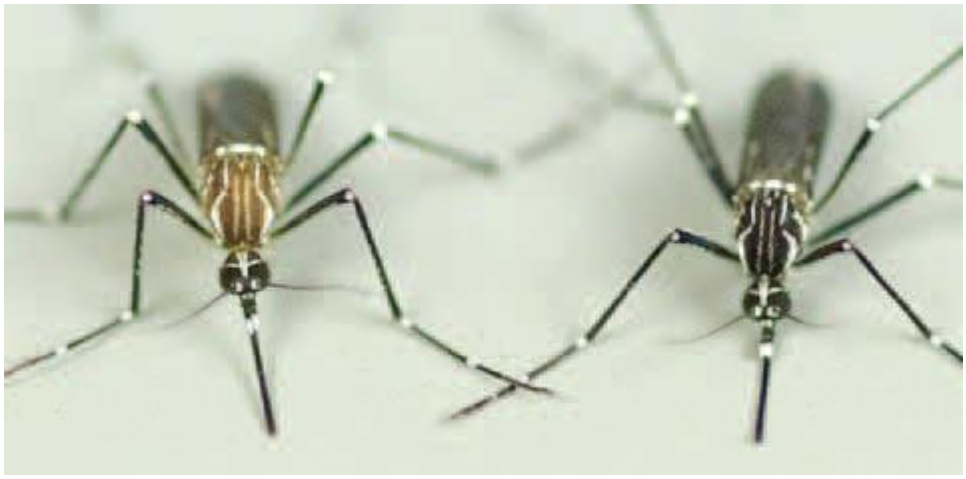


Figure 1.6: Colour variation between urban and sylvan *Aedes aegypti* individuals: Taken from McBride *et al.* 2014. Two *Aedes aegypti* females are compared for their difference in coloration. The domestic or urban form, *Aedes aegypti aegypti*, is shown on the left. The forest or sylvan *Aedes aegypti aegypti* is shown on the right.

1.2.4 An invasive species

Anopheline mosquitoes have been present worldwide before the spread of *Plasmodium* parasites, the disease causing agents of malaria. Malaria parasites originally from Africa have remarkably adapted to new vectors (distinct *Anopheles* species) when introduced into the New World around the 16th century (Molina-Cruz and Barillas-Mury 2014).

The spread of dengue worldwide, however, did not require adapting to a different human feeding vector. The anthropophilic *Aedes aegypti*, originally from Africa, colonised the New World during transatlantic shipping in the 16th to 18th centuries and Asia in the 19th century (Gloria-Soria *et al.* 2016). The pattern of vector spread is now matched by the spread of the dengue serotypes 1 to 4. Moreover, a closely related species; *Aedes albopictus*, originally from Southeast Asia, has spread quickly across the world in the last 40 years now being present in North/South America, parts of Africa, parts of Europe, and Australian overseas territories (Bonizzoni *et al.* 2013).

1.3 Vector Control Strategies

1.3.1 The importance of vector control

Since humans have gradually occupied a larger surface of the earth, habitats have changed for the native species. Native species in human-invaded habitats have, or have had, two main outcomes; i) to face or suffer extinction, ii) to evolve some

sort of commensalism, i.e. ‘domestication’, with humans. When arthropods requiring a vertebrate as a blood source have undergone, or are in the process of, domestication the consequences have been devastating for human health (Powell and Tabachnick 2013). Supporting this concept is the fact that many of the current human vector-borne pathogens infect also animals or have close relatives that do, i.e. malaria or yellow fever. Given the relatively new addition of humans to the earth’s biota, around 2.6 million years ago (White *et al.* 2009), and the much longer presence of blood-feeding insects, more than 100 million years ago (Powell and Tabachnick 2013) this ‘incomplete domestication’ is expected.

This highlights the importance of vector control as well as tackling the diseases themselves as a pressing and constant challenge. Granting human immunity for a particular virus serotype through vaccines is unlikely to eliminate the problem. The closest vaccine to eliminate the problem has been the yellow fever vaccine, which although very effective is not a permanent solution due to yellow fever sylvatic cycles (Frierson 2010). Perhaps, even eliminating a particular mosquito species will not be a permanent solution. Nonetheless, specific diseases in specific vectors may be easier to control, or are more likely to show a tangible benefit on human health from trying. Hence, disease eradication, the ultimate objective, will require awareness of the magnitude of the problem and all efforts possible. Interestingly, climate change poses an extra challenge as it has been linked to recent expansions of mosquito vectors, e.g. the introduction of *Aedes albopictus* in ~25 European countries (Medlock and Leach 2015).

1.3.2 Bed nets and insecticides

Insecticide impregnated bed nets and insecticide spraying of households are one of the most common methods in vector control. Insecticide-treated nets have been largely effective in vector control of Anophelines. They have been used historically as a protection against vector-borne diseases, for instance by the Russian, German and US armies during World War II. A meta-study on the effectiveness of insecticide-treated nets concluded that child mortality can be reduced by one fifth and malaria infections by half (Lengeler 2004), however, large scale coverage is difficult to achieve. Nevertheless, this only works well with Anophelines since females are night-time biters. *Aedes* mosquitoes are generally daytime biters and hence insecticide-impregnated bed nets are not an effective vector control measure against them. The use of bed nets may also favour the behavioural shift of mosquitoes from night to day biting or from indoor to outdoor biting (Russell *et al.* 2011). However, the main disadvantage is that the development of insecticide-resistance in mosquitoes is sped up greatly by their use. Moreover, their usually broad-spectrum action is lethal to a range of species which can pose a serious threat to biodiversity. Indoor residual spreading is thought to be more specific as it targets the resting surfaces of human biting mosquitoes such as *Anopheles gambiae* (Larsen *et al.* 2017).

1.3.3 Environmental management

Domestication of mosquitoes has caused a shift in the oviposition habits of females. In natural environments eggs tend to be oviposited in small pools of

water in tree-holes or leaf axils. Urban environments have provided analogous sites such as pots, drainage ditches or tyres which the domesticated mosquitoes prefer (Crovello and Hacker 1972). Essentially any receptacle containing water may potentially be an artificial breeding site for mosquitoes. Hence, a close control of these sites can reduce the numbers of disease-transmitting adults. Water management is especially effective in big cities, were a majority of breeding sites are artificial. However, implementation is not always simple. In Singapore, the "Control of Vectors and Pesticides Act" (Chapter 59 – Part IV) took effect in 1998 which sets firm law-enforced prohibitions on creating conditions favourable to vectors in private properties, and yet it has not fully tackled the endemic problem of dengue in the area.

1.3.4 Biological vector control

Biological control involves the use of natural predators or pathogens against mosquitoes. Deployment of larvivorous copepods has been successful in eliminating mosquitoes on a local scale in Vietnam (Kay and Nam 2005; Nam et al. 2012), but has not been proven to work in other contexts. However, the risk with biological control strategies is that the control agent can colonise an environment and become the pest themselves (Rupp 1996).

1.4 Mosquito Genetic Engineering

1.4.1 Transposable elements transgenesis

piggyBac is a class II transposable element and belongs to the TTAA-specific family. It was first discovered in a baculovirus genome after passage in *Trichoplusia ni* cells (Fraser *et al.* 1983). These cells are an established insect cell line from the cabbage looper, *Trichoplusia ni*, a moth from the *Noctuidae* family (Hink 1970). The original *piggyBac* element isolated was 2.4kb long, containing a long open reading frame, and flanked by 13bp terminal inverted repeats as well as by two 19bp inverted repeats (Cary *et al.* 1989). The open reading frame of an autonomous *piggyBac* element encodes a transposase which nicks the inverted repeats of the transposon, allowing it to integrate into TTAA sites in the genome (Elick *et al.* 1996, Figure 1.7).

However, for transgenesis applications the *piggyBac* transposase and *piggyBac* are carried in separate plasmids, so as to allow integration but not remobilisation. A wide range of different species belonging to four different orders have been successfully transformed with *piggyBac* including species from Coleoptera, Diptera, Hymenoptera and Lepidoptera (Gregory *et al.* 2016). Crucially, *piggyBac* works both in *Drosophila melanogaster* (Handler and Harrell 1999) and *Aedes aegypti* (Kokoza *et al.* 2001).

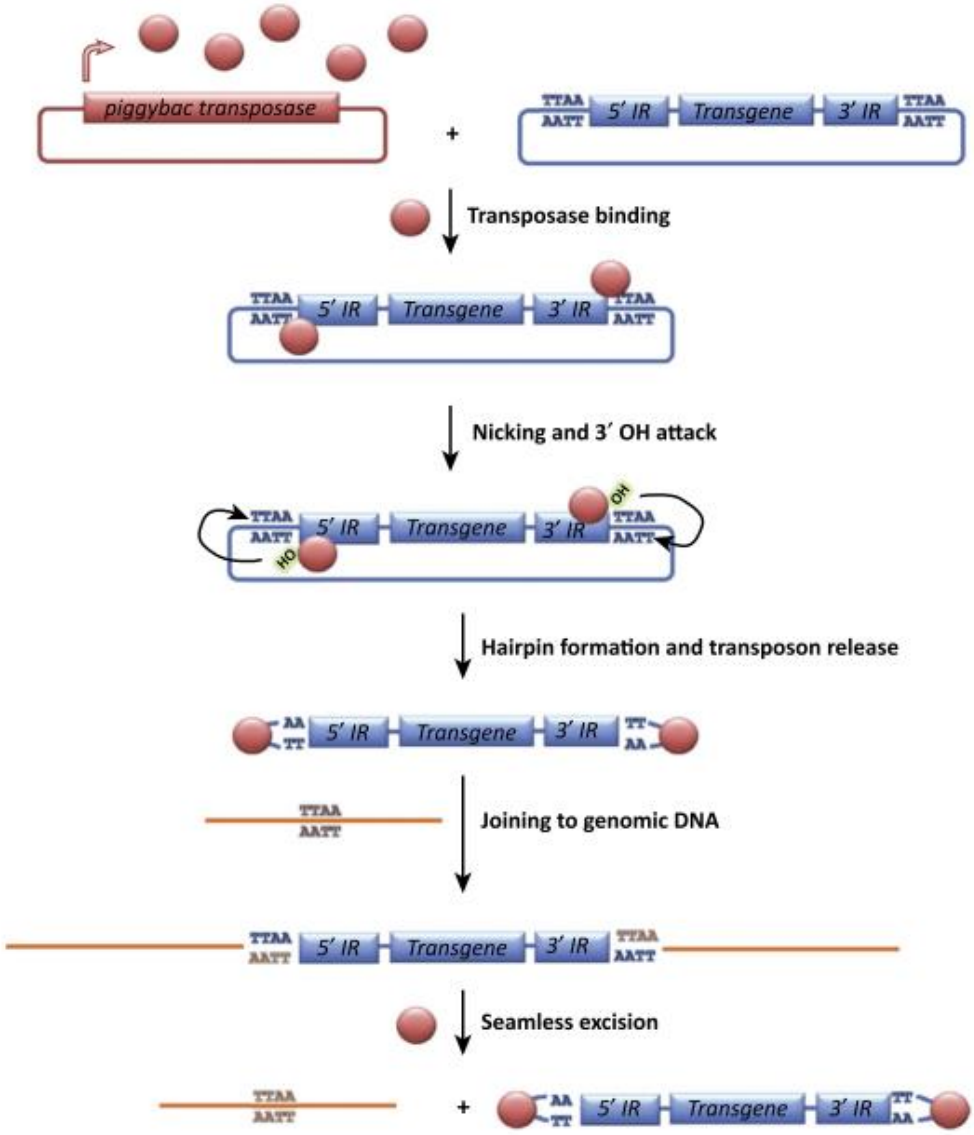


Figure 1.7: *piggyBac* mediated transposition:
Taken from Woodard and Wilson 2015. IRs stand for inverted repeats. Expressed transposase from a plasmid (or from the mobile element in autonomous *piggyBac* sequences) binds the *piggyBac* IRs and induces nicking and 3' hydrophilic attack of the TTAA ends. Hairpin formation then occurs. Joining to the genomic DNA occurs at a TTAA nucleotide sequence resulting in TTAA duplication as well as the integration of the transgene. Should transposase still be expressed the element can be excised thereby recreating the original TTAA target site without any 'scar'.

The *P element* is also a class II transposon, but in this case discovered in *Drosophila melanogaster*. The *P element* works by the same 'cut and paste' mechanism of *piggyBac* and its use in transgenesis is analogous to *piggyBac*,

although it is only used in *Drosophila melanogaster* transgenesis (Majumdar and Rio 2002).

1.4.2 Site-specific recombinases

An issue with transposon mediated transgenesis is that its quasi-random nature results in positional effects in transgene expression. Since enhancers and silencing elements can affect neighbouring sequences, the genomic environment for a transgene will be a key determinant of its expression and thereby function. A recent study comparing 27,000 distinct reporter integrations using *piggyBac* in mouse embryonic stem cells showed more than ~1, 000 fold differences in expression levels (Akhtar *et al.* 2013). Moreover, there is an issue with imposing different fitness costs in the transgenic organism, depending on which elements of the genome are disrupted or not.

Site-specific recombinases provide a solution to the problem as they allow insertion of transgenes into specific genomic sites which normalises the position effect rather than overcoming it. FLP-FRT and Cre-loxP recombinase systems have been shown to work in a range of organisms including *Ae. aegypti* (Haghighat-Khah *et al.* 2015). However, these systems are both reversible which provides certain transgene instability should any residual recombinase be present. An alternative integrase system, ΦC31-*att*, involves recombination between specific phage and bacterial attachment sites (attP and attB respectively), resulting in hybrid sites *attL* and *attR*. Since these new sites are not recognised by

the ΦC31 integrase this makes the integrations irreversible and hence grants more efficient transgenesis. The ΦC31-*att* requires a ‘docking site’ in the genome, preferably *attP* before a transgene of sequence can be inserted into an *attP* target. The transgenesis efficiency of the ΦC31-*att* system in *Aedes aegypti* ranges from 4.7 to 12.5% (Nimmo *et al.* 2006).

1.4.3 HDR gene editing in Aedes aegypti

Homology directed repair in a sequence specific manner using directed DSB nucleases has been successful in *Aedes aegypti*. Both TALENs and CRISPR-Cas9 sequence specific nucleases have been used to catalyse the repair from a co-injected donor plasmid with the appropriate genomic flanks (Basu *et al.* 2015). A comprehensive CRISPR-Cas9 HDR study in *Aedes aegypti* was carried out in Kistler *et al.* 2015. HDR involved supplying a donor plasmid with the relevant genomic flanks to site specific double strand breaks directed by sgRNA CRISPR-Cas9 activity. The paper also suggests that HDR only occurs in females, and hence G0 male adults were discarded instead of crossed to screen their G1 progeny.

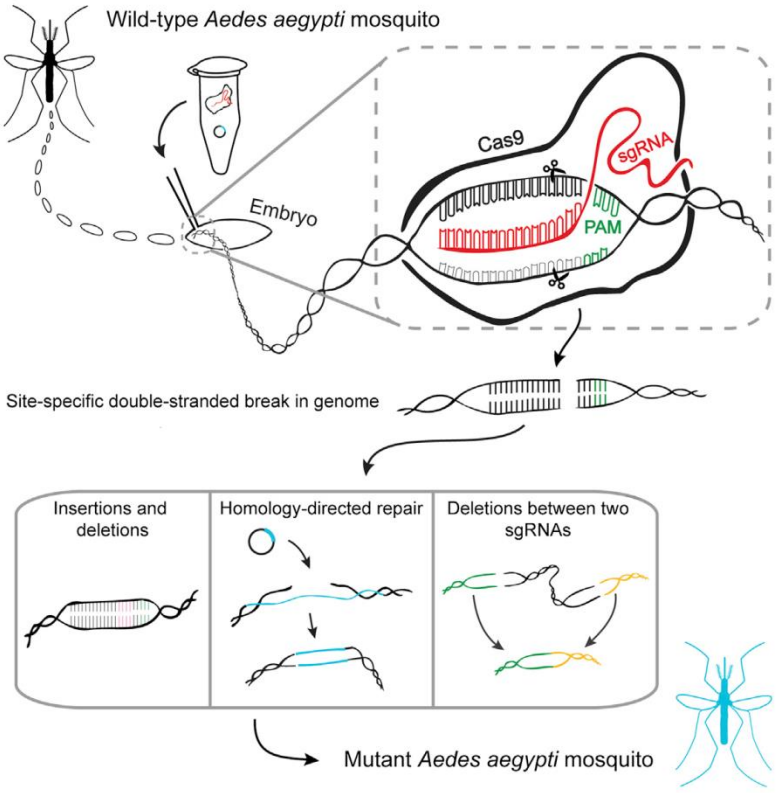


Figure 1.8: Schematic representation of CRISPR-Cas9 driven HDR in *Aedes aegypti*:
Taken from Kistler *et al.* 2015. A WT *Aedes aegypti* lays fertilised eggs, which are microinjected as embryos (Section 3.2.3) with Cas9 protein, sgRNA, and a donor plasmid including a marker flanked by homology arms matching the targeted genomic site. Site-specific double strand breaks in the genome can lead to INDELs, through NHEJ repair, and HDR events through HR repair. NHEJ repair driven deletions are more likely caused by the use of multiple sgRNAs as shown. Screening of mutant G1 mosquitoes will involve checking for a fluorescence marker in case of successful HDR, and checking for mutation phenotypes specific to the gene of interest.

Kistler *et al.* 2015 used flanking homology arms of 766 to 2058kb in length, inserts of 2150 to 2321kb in length (including cassettes for ubiquitous marker expression), a single sgRNA per gene and injected Cas9 protein for 4 different genes. HDR was detected by fluorescence for 3 of the genes, with an efficiency of 16.4, 3.5 and 1.9% respectively. For one of the injections 0% percent of founders resulted in fluorescent offspring, even though a similar amount of founders was obtained for each injection (between 41 and 100). Successful HDR was initially determined by marker expression in G1s, however subsequent PCR

analysis showed that some of the HDR events had been off-target since the targeted genes remained intact and did not have the expected insert. Successfully directed HDR only occurred for two genes and it had an efficiency of 16.4 and 0.9% respectively, given that a proportion of 4 out 11 from screened HDR events were actually not inserted into the target locus.

A more recent study, (Li, Bui, *et al.* 2017), carried out CRISPR-Cas9 HDR in *Aedes aegypti* and compared HDR efficiencies between an integrated Cas9 source (*exu-Cas9*) and protein co-injection. Transformation efficiencies in this paper were atypically calculated as a percentage of marker-positive G1s, which makes comparison with other studies difficult but allows a valid comparison within the study. HDR efficiencies were found to be more than 2 orders of magnitude more efficient when using the integrated Cas9 source with respect to Cas9 protein injection. Hence, this highlights the importance of integrated vs injected Cas9 as well as the value of finding promoters for maternal egg-deposition, such as *exu* (Li, Bui, *et al.* 2017).

1.4.4 The common fruit fly, Drosophila melanogaster as a model for Aedes aegypti

Since the seminal work by Thomas Hunt Morgan in elucidating the role of chromosomes in heredity, *Drosophila melanogaster* has been one of the most studied model organisms (Morgan 1910). Its short generation time, 10 days at 25°C, as well as the simple food and space requirements make it a very easy

insect to work with. Its genetics are very well understood and many genetic tools to facilitate research have been developed in the last century. Balancer chromosomes, with large inversions preventing recombination as well as both recessive and dominant markers exist for the three main *D. melanogaster* chromosomes. Such a tool, absent in mosquitoes, greatly facilitates research involving genetic crosses (Beckingham *et al.* 2005).

Both *Ae. aegypti* and *D. melanogaster* are dipteran insects which makes *D. melanogaster* the most closely related model organism for yellow fever mosquitoes. Although genetic systems designed to function in *Ae. aegypti* will always need to be tested in the mosquito, there are cases were testing a genetic system in the better understood *D. melanogaster* is advisable. Where a complex and relatively unknown genetic system is to be tested, interpretation of results would prove harder in *Ae. aegypti* than in *D. melanogaster*. Therefore, it may be worth carrying out the experiments in the common fruit fly first, where understanding the system would be easier and faster.

1.5 Genetic Vector Control Strategies

There are many naturally occurring selfish genetic elements, able to spread through their hosts irrespective of their fitness costs (Hurst and Werren 2001; Burt and Trivers 2006), which hold unexploited potential in insect vector control. The concept of insect vector replacement with engineered gene drive systems was established long ago (Curtis 1968; Curtis and Graves 1988; Knipling *et al.* 1968); however, it has only recently become feasible to attempt a design of such a system.

For the purpose of disease vector population replacements it should ideally; i) be strong enough to tolerate fitness costs and reach gene fixation on a relatively short time-scale (Sinkins and Gould 2006; Boëte and Koella 2003), ii) be able to spread multiple genes so as to reduce the chances of cargo being mutated or lost (Marshall 2008), iii) be as safe as possible i.e. allowing for population isolation (Hay *et al.* 2010), and finally, iv) be possible to remove it from the wild type population in case of unforeseen negative effects of cargo genes or other components (Moreira *et al.* 2004).

Genetic control designs can be classified within a spectrum of persistent to invasive drives where persistent drives are generally self-limiting and invasive drives self-sustaining (Alphey 2014). Furthermore, the further down the self-limiting side of the spectrum the easier drive reversal is. Several naturally

occurring selfish genetic elements have been proposed to be used as gene drive mechanisms. These encompass transposons (Charlesworth *et al.* 1994), homing endonucleases (Burt 2003; Deredec *et al.* 2008), *Medea* (Chen *et al.* 2007) and *Wolbachia* elements (Stouthamer *et al.* 1999). However, each has their own advantages and disadvantages.

Although different applications would benefit from more or less invasive approaches, homing endonucleases, *Medea* and transposons are relatively invasive examples, meaning that reversal and population isolation would be a challenging issue (Marshall and Hay 2012). If worked as predicted they could become global drives (Noble *et al.* 2016), i.e. capable of worldwide spreading, which is something to be avoided due to safety concerns and the simple fact that regulatory bodies dealing with gene drive do not operate worldwide (Alphey 2014). Self-limiting genetic control designs such as RIDL or simple sterile-male releases provide this safety aspect and ease of control. However, in this case release numbers and frequencies required for population suppression or population replacement (introgression by inundative release) may be too high for many applications. Such large releases may not be currently feasible due to the limitations in mass rearing itself and the prohibitive cost it may suppose. Figure 1.9 shows a spectrum of genetic control mechanisms ranging from self-limiting to self-sustaining strategies.

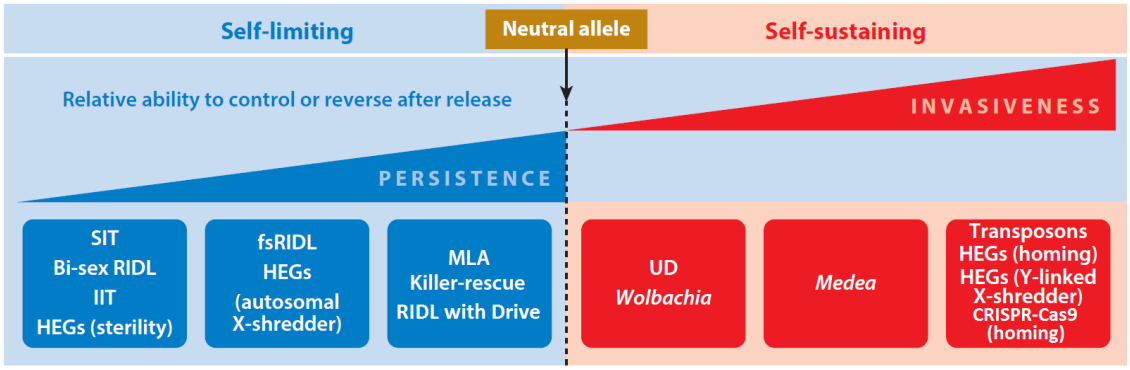


Figure 1.9: A spectrum of invasiveness for different genetic control mechanisms: Taken from Alphey 2014. Genetic systems for vector control can be classed as self-limiting or self-sustaining. Reversibility or control over releases is mostly provided by the fitness cost of the genetic themselves, which would be eliminated from a population by natural selection at a rate dependent on the severity of the fitness cost and the persistence/invasiveness of the genetic system. The more severe the fitness cost, or the more self-limiting the genetic system the more frequent or larger the releases would have to be for effective population suppression or replacement (depending on the mode of action of the genetic system). Generally, self-limiting strategies involve population suppression systems and self-sustaining strategies involve population replacement systems, however this is not always the case. Along this continuum, more invasive approaches would reduce the size and/or frequency of releases required, but would make restriction of a genetic system within a target population challenging, as well as the removal of the genetic system from the targeted population itself. On the other hand more self-limiting approaches would allow for easier control and reversal whilst increasing the size or frequency of releases, and in turn increase the cost. Abbreviations: Sterile insect technique (SIT), release of insects carrying a dominant lethal (RIDL), incompatible insect technique (IIT), homing endonuclease genes (HEGs), female specific RIDL (fs RIDL), multi-locus assortment (MLA), underdominance (UD), CRISPR-Cas9 (Clustered Regularly Interspaced Short Palindromic Repeats-CRISPR associated protein 9).

1.5.1 tetO-tTAV system and the release of insects carrying a dominant lethal

The *tetO-tTAV2* system consists of an enhancer (*tetO*) which is normally bound by a transactivator (*tTAV* or *tTA*) except in the presence of tetracycline (Gossen and Bujard 1992). The system was created, in the cited study, by fusing a tetracycline repressor (*tetR*) domain (which normally blocks expression from the tetracycline operator, *tetO*) to the transcription activator C-terminal domain of VP16 from Herpes simplex virus (HSV). A *tetO* binding transactivator able to respond to tetracycline was synthesised (Gossen and Bujard 1992). Two versions of the transactivator exist, one that becomes inactive in the presence of

tetracycline (Tet-Off system) and one that becomes active instead (Tet-On system) (Lycett *et al.* 2004).

The Tet-Off system has been used to engineered repressible genetic sterility in insects (Thomas *et al.* 2000). This involved a two-component system whereby the *tTAV* is under the control of a tissue-specific promoter, and an exogenous lethal gene under the control of *tetO*. In the absence of tetracycline the lethal gene was upregulated through the interaction of the *tetO* enhancer and the *hsp70* minimal promoter (Figure 1.10). A simplified version of this system, a one-component positive feedback system, was devised; placing *tTAV* under the control of a minimal promoter and an adjacent *tetO* sequence (Gong *et al.* 2005). The minimal promoter, in the absence of tetracycline, would express a basal level of the transactivator which in turn would enhance its own expression through interaction with *tetO*. Since the excessive accumulation of VP16 in a cell is toxic due to transcription squelching (Berger *et al.* 1990), *tTAV* is both the lethal effector and the transactivator (Figure 1.10).

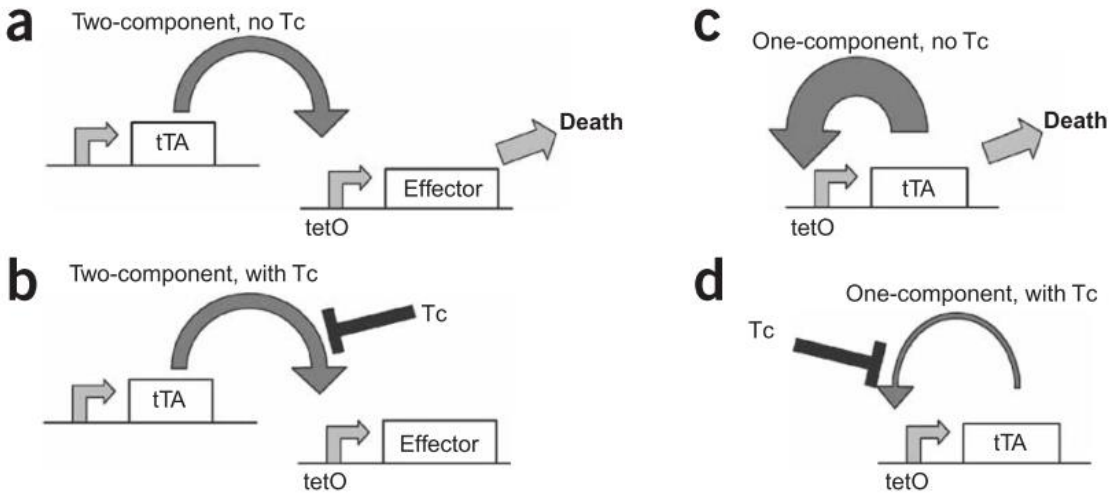


Figure 1.10: Tetracycline-repressible systems for genetic sterility in insects:
The Tet-Off systems are repressible by tetracycline. *tTA* (*tTAV*) is the transactivator open reading frame. Minimal promoters are not shown. Tc stands for tetracycline (**A**-**B**) In the absence of tetracycline the expressed *tTA* leads to the upregulation of a lethal effector and hence killing the organism. In the presence of tetracycline, this upregulation is prevented. (**C**-**D**) In the absence of tetracycline *tTA* is first expressed at a basal level by the action of a minimal promoter which then starts the positive-feedback loop between *tetO-tTAV* which ultimately leads to a lethal accumulation of VP16. In the presence of tetracycline, this positive-feedback loop is prevented.

Insects carrying either of the systems can be reared in the lab under the presence of tetracycline as homozygous and then released into the wild as a refined version of the sterile-insect technique (SIT). Hence the name of this variant of the technique, release of insects carrying a dominant lethal (Thomas *et al.* 2000). Given the higher fitness of RIDL insects (Massonnet-Bruneel *et al.* 2013) compared to traditionally irradiated SIT insects (Helinski *et al.* 2009), a RIDL program should be more effective upon release as well as more economic due to the lower release numbers required. Female-specific RIDL (fsRIDL) systems have been devised in both *Ae. aegypti* and *albopictus* using the *Aedes Act4* promoter to lead to female-flight muscle cell death and hence flightless females (Fu *et al.* 2010; Labbé *et al.* 2012). This not only facilitates genetic sexing upon

release but it creates a stronger version of RIDL in the field. RIDL homozygous carrier males crossed to WT females will result in heterozygous RIDL males which will not die themselves and in turn add to the population suppression effect in subsequent generations.

1.5.2 Sequence-specific drives: HEGs, ZFNs, TALENs, and CRISPR-Cas9

With the advent of the highly specific and efficient CRISPR-Cas9 system (Barrangou 2012), homing drives have gained traction as an effective system in mosquito control (Hammond *et al.* 2016). As further described in Chapter 4, a wide range of gene drive systems, in terms of different persistence/invasiveness, can be derived from these sequence-specific endonucleases.

Homing is defined as the transfer of an intervening sequence to a homologous allele lacking the sequence (Chevalier and Stoddard 2001). Homing is carried out by an endonuclease capable of recognising and cleaving the target site, causing a double strand break (DSB), in the homologous allele. Such endonucleases are encoded within the intervening sequence itself. This ability was first described for mobile introns and hence the endonucleases encoded by these introns are referred to as homing endonucleases (Dujon *et al.* 1989). Homing endonuclease genes (HEGs) have been described in the three biological domains and their proteins share common structural motifs with a higher degree of similarity than amongst type II restriction enzymes. In addition, they bind to DNA recognition sequences of 12-40bp (Belfort and Roberts 1997). Due to this ability, HEGs have

been proposed for engineered gene drive systems (Burt 2003; Burt and Koufopanou 2004; Deredec *et al.* 2008).

Although not HEGs, strictly speaking, other endonucleases have been proposed for engineered gene drives due to their homing ability (Simoni *et al.* 2014; Gantz and Bier 2015). These include zinc finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs), and CRISPR-Cas9 (Gaj *et al.* 2013). All of these endonuclease based systems would share the same homing mechanism if designed appropriately (Figure 1.11). RNA-guided nucleases such as CRISPR-Cas9 have recently shown the most promise out of the available homing mechanisms due to their high effectiveness and specificity, but especially for simplicity of assembly. In contrast to other systems mentioned above, CRISPR-Cas9 does not rely on specific protein domains for sequence recognition but rather relies on RNA guide(s) for sequence specificity. This RNA guided specificity is much easier to change to a new sequence (or reprogram) than protein domain driven specificity. Only 18-21bp sgRNAs (single guide RNAs) are required to guide the Cas9 and hybridise with the genome sequence for *in situ* cleavage (Radzisheuskaya *et al.* 2016). A disadvantage of the Cas9 system is that the sequence 'NGG' (protospacer adjacent motif or PAM site) must follow the sgRNA sequence in the genome. This poses some restriction over which positions can be chosen for targeting in the genome (Gaj *et al.* 2013). However, 'GG' sequences are sufficiently common for it to not cause a major challenge. Another way around this is the use of other class 2 CRISPR effectors (Cas9-like

proteins) such as Cpf1 (CRISPR from *Prevotella* and *Francisella* 1). Cpf1 has a different PAM site; i.e. ‘TTN’ (Zetsche *et al.* 2015).

Although sequence-specific homing systems are very precise, effective and especially suited for gene drives they also have their caveats. Sequence-specificity can in turn become a disadvantage when considering the genomic variation of wild type populations or the potential generation of resistant alleles (Hammond *et al.* 2017). This becomes a special challenge when assembling a CRISPR-Cas9 drive chain, i.e. a daisy chain drive (Noble *et al.* 2016) where several components must undergo sequence-specific homing effectively. Sequence independent drives, such as underdominance, avoid these hurdles whilst facing others.

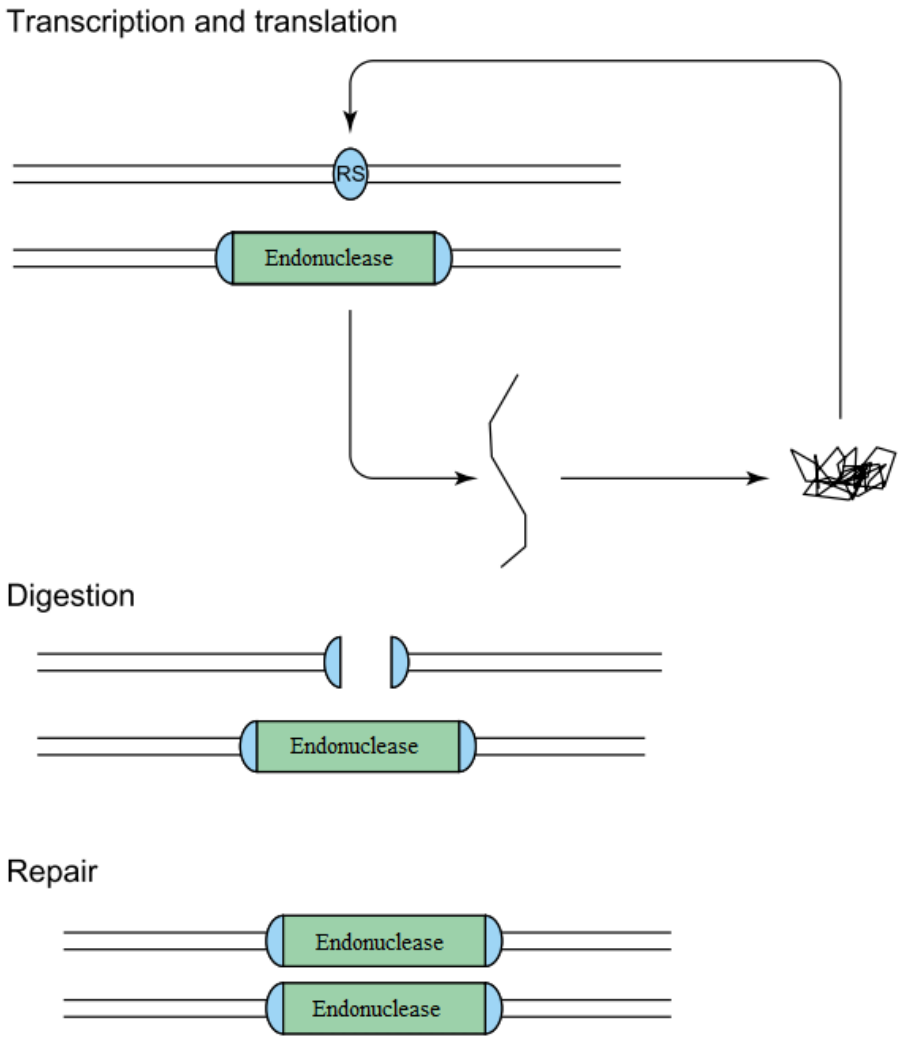


Figure 1.11: Endonuclease driven homing:
Adapted from (Burt and Koufopanou 2004). When an endonuclease is embedded within its recognition sequence (RS) in the genome a homing mechanism is obtained. The translated protein from an integrated endonuclease will recognise the RS of the neighbouring allele and lead to a double strand break. In the event of homology directed repair in a diploid genome the endonuclease bearing chromosome will be used as a repair template and hence carrying out the copying/ homing of the endonuclease.

1.5.3 Suitability of Aedes aegypti for genetic vector control

Genetic control systems work in a species-specific manner, which is an advantage in terms of reducing environmental impact, but also a challenge in terms of tackling several species at once. The fact that there are fewer *Aedes* species transmitting arboviral diseases than species of *Anopheles* transmitting

malaria make *Aedes* species more suitable for genetic control. The high genomic variation amongst anopheline vectors of malaria compromises the implementation of sequence-specific gene drives (Deitz *et al.* 2016; Feng *et al.* 2017). Conversely, the comparatively recent expansion of *Aedes* populations worldwide should be reflected with a lower genetic variability which would in turn facilitate the implementation of sequence-specific gene drives.

1.6 Scope of Work

This thesis describes three main lines of research. The first is the investigation of an underdominance gene drive system in *Drosophila melanogaster* (Chapter 2) which was undertaken at Cardiff University. The second is the exploration of the use of homology directed repair to insert exogenous sequences into the M-locus of *Aedes aegypti* with the outlook of male specific linkage of components for future gene drive applications (Chapter 3) which was undertaken at Oxitec. Finally, this thesis explores the genetics of *Act4*, a putative gene for flight in *Aedes aegypti* females, in the attempt to create distinct gene drive systems (Chapter 4) which took place at the Pirbright Institute.

Chapter 2 - **Underdominance Gene Drive in *Drosophila melanogaster***

2.1 Introduction

2.1.1 Project aim

The aim of the project was to develop an underdominance-based gene drive system in *Drosophila melanogaster*, through a synthetic biology approach, to spread desirable genes through a population of wild type disease-vectors in the face of natural selection. Desirable genes would in this case render the vectors resistant to disease itself. Such genes are often referred to as refractory genes.

2.1.2 Insects as disease vectors

As indicated in Chapter 1, insect-borne diseases have a devastating impact on health around the globe. Mosquito-transmitted diseases cause over 1 million deaths every year. DALY (disability-adjusted life year), a measure of disease impact, is based on the number of years lost due to mortality and/or morbidity caused by a disease (McGraw and O'Neill 2013). DALY more accurately reflects a disease burden than simple mortality counts since many infectious diseases may be severely debilitating but self-limiting, to keep the host alive. It is estimated that around 17% of the world's infection-related DALY is caused by

insect-transmitted diseases, whilst 90% of this fraction is attributed to mosquito-borne diseases alone (WHO 2004) (Figure 2.1).

Indicators such as DALY or simple mortality counts are very useful but they also underestimate the disease burden on communities as they do not account for its social and economic aspects. When considering infection cases per year, malaria is the most widespread, infecting between 200 and 500 million people annually. Dengue is only second to malaria, infecting an estimated 390 million people annually (Bhatt *et al.* 2013), and is a rapidly growing mosquito-transmitted disease. Therefore, the impact is not fully appreciated without considering the great social and economic effects of malaria or dengue on communities where these are endemic (McGraw and O'Neill 2013). Hence, there is a need for effective strategies to reduce mortality in the journey towards disease eradication.

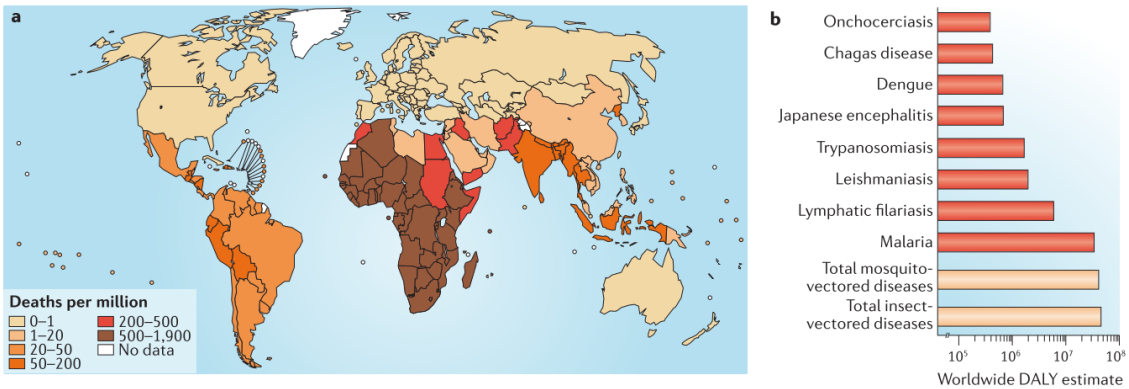


Figure 2.1: Vector-borne disease burden on the world:
Taken from (McGraw and O'Neill 2013). **(A)** The map shows the deaths per million caused by vector-borne diseases, showing much greater incidences amongst developing countries. **(B)** A log scale graph shows DALY estimates for different vector-borne diseases. It is worth noting how mosquito-vectored diseases are responsible for most of insect-borne DALY.

2.1.3 Tackling insect-borne diseases

Due to the severity of the problem, many different attempts to solve it have taken place. There has been a focus of the medical research community to develop vaccines against the malaria parasite or the dengue virus. However, the complex life cycle of the malaria parasite has challenged this, and it has become apparent that multiple vaccines for the different life stages may be required (Vaughan and Kappe 2012), which hinders the approach. Similarly, vaccine development for dengue has been challenging due to the multiple serotypes of the arbovirus which causes the disease (Wan *et al.* 2013). The use of insecticides has shown to be effective in targeting of mosquitoes (Ramirez *et al.* 2009), in a wider range of contexts. Nevertheless, the high economic and ecological costs of insecticide application, together with the development of resistance in their targets make these methods insufficient for disease eradication, although still useful as an effective control method (McGraw and O'Neill 2013). Control of vector species through genetic modification has recently arisen as a promising tool in the efforts to combat insect-borne diseases.

2.1.4 Underdominance gene drive

The available tools described in Chapter 1 seek to achieve population elimination or population replacement of disease-vectors to reduce vector competence. Population elimination involves developing technologies such as RIDL (Release of Insects with a Dominant Lethal), which is an improved version of the sterile insect technique (SIT) that consists of copious releases of sterile males to reduce

successful mating of wild type females (Thomas *et al.* 2000; Alphey *et al.* 2010; Fu *et al.* 2007). RIDL is capable of highly efficient sex separation facilitating (Fu *et al.* 2007) the release of only male populations and hence has been shown to be more cost-effective than previous SIT techniques (Atkinson *et al.* 2007).

Population eradication could, however, open a niche for other species to continue disease transmission and also have detrimental effects for the environment. On the other hand, population replacement would reduce the disease transmission competence of vectors by introducing novel traits into a population without removing it from the environment (Hay *et al.* 2010). Nevertheless, such a novel trait generated through transgenesis is likely to be associated with a fitness cost, at least from the insertion itself (Marrelli *et al.* 2006). In addition, the genetic background of lab-reared insect strains is likely to be sub-optimal in the wild, hindering the spread of such a trait under natural selection (Sinkins and Gould 2006). Therefore, there is a need for molecular mechanisms to force an engineered allele into a population in the face of natural selection (Hay *et al.* 2010)**.**

As mentioned in Chapter 1, whilst the relatively strong drives above offer a low release cost, they would be difficult to control. Hence, it would be useful to have a middle ground system capable of persisting in a target population for an extended time but not so invasive as to be predicted to spread significantly beyond the target population, even when accounting for expected levels of

migration. This would not be ideal for all applications, but it would fill a much needed niche, of less invasive gene drives, in the varying range of genetic control mechanisms available.

An underdominance gene drive could be such a middle ground system, between both ends of the spectrum. This project focused on engineered underdominance as a gene drive mechanism, which draws on a naturally occurring process. Underdominance is the advantage of homozygous states over heterozygotes. In extreme cases, heterozygotes are inviable. The simplest underdominance system would consist of a single locus for which the heterozygote is less fit than either homozygote. Such a proposed system involves the use of a pair of alleles; each one contains a killer factor, and an antidote to the killer factor produced by the other allele. Thus the alleles mutually suppress the deadly effects (Figure 2.2). The presence of both engineered alleles, which would suppress their individual fitness costs, isn't strictly speaking homozygous. However, for arguments sake, this transheterozygote can be regarded as a 'homozygous state' for engineered alleles, whereby the wild type homozygous state and our engineered 'homozygous state' have an extreme advantage over a heterozygous mix of both these states, i.e. the presence of just one of the engineered alleles. More complex systems have been proposed to be engineered at two loci (Davis *et al.* 2001).

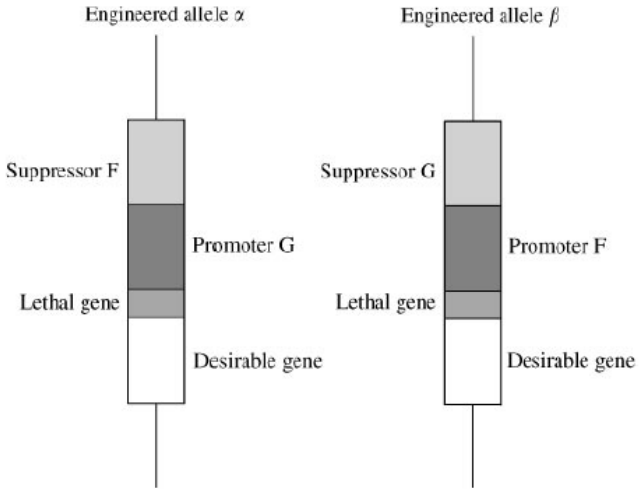


Figure 2.2: Proposed mechanism for engineered underdominance:
Taken from (Davis *et al.* 2001). Alleles α and β are shown in the diagram, where each allele contains a lethal gene under a different promoter. Each allele contains a different suppressor which blocks the expression of the lethal gene in the opposite allele. This results in a form of underdominance, as both alleles are required for survival. A desirable gene can then be carried or spread through a population as long as it is linked with the alleles.

The key aspect of underdominance is that it is a frequency-dependent drive, i.e. releasing above or below a threshold frequency will offer distinct population dynamics. The advantage of underdominance is that to achieve an unstable equilibrium between WT and engineered alleles a relatively high release would be required. In this unstable equilibrium, alleles of higher frequency will tend to increase, whilst lower frequency alleles tend to decrease over time (generations). The release size required, or threshold frequency, will depend on the inherent fitness costs carried by the alleles for engineered underdominance (or the cargo itself). For equal fitness costs between WT and engineered alleles the required release frequency would likely be of 50% of the total population. Hence, depending on the size of the release, this unstable equilibrium can be reached, surpassed, or simply not met (Edgington and Alphey 2017).

Below this threshold release frequency the element will have a higher chance of decreasing in a population over time, whilst above it there will be a higher chance of increasing in frequency over time (Figure 2.3). This unstable equilibrium, or threshold frequency, can be seen as the invasion threshold and will vary with the intrinsic properties of the genetic element as well as with its inherent fitness costs. Although deterministic models appear to convey a certainty in the predicted spread or disappearance of the genetic element when above or below the invasion threshold respectively, they are merely indicating the likely spread or removal of a particular element for a given release frequency and set parameters (Jansen *et al.* 2008).

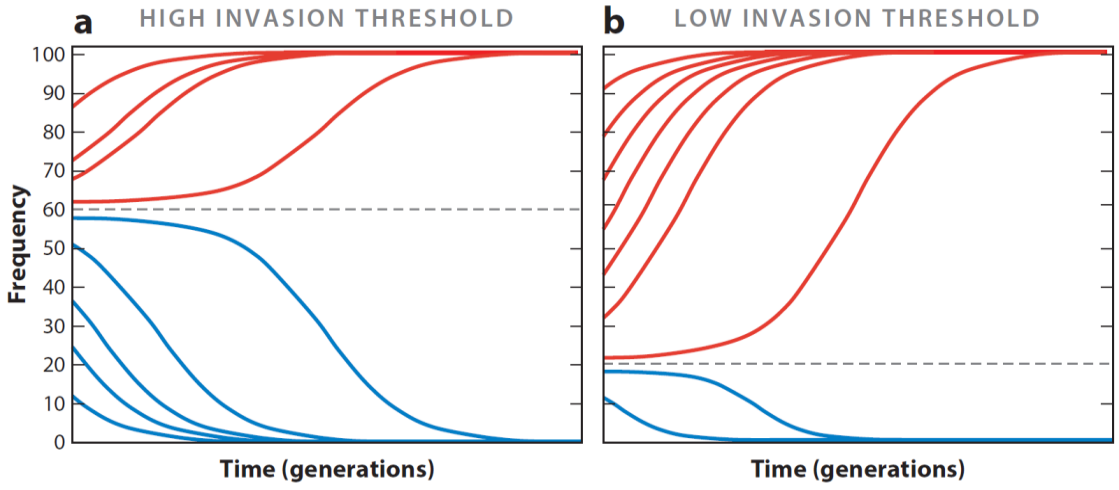


Figure 2.3: Frequency-dependent nature of releases of genetic systems such as underdominance:
Taken from (Alphey 2014). The unstable equilibrium frequency, or invasion threshold, is marked by a horizontal dashed line. Release frequencies of a particular element above or below its invasion threshold will tend to increase or decrease respectively over time, usually expressed as generations. The element can then increase to fixation, i.e. present in 100% of individuals, or attain an alternative equilibrium short of fixation depending on the intrinsic properties of the genetic system and its fitness costs. Graphs show frequency development over time, for two hypothetical systems (with (**a**), a high or (**b**), a low invasion threshold) for a range of different release frequencies on the y-axis.

As a result, underdominance drives can be relatively weak, requiring relatively large releases of insects for fixation, and can therefore be reversed with comparative ease (Marshall and Hay 2012)**.** In addition, the modification of the environment is minimal compared to population eradication techniques. Single-locus underdominance, provides the advantage of population isolation, as any mating with the wild type population will be futile (Altrock *et al.* 2010; Magori and Gould 2006). Population isolation is useful to prevent invasion of WT insects in other areas. This could be a useful tool if only a local population had to be eliminated or to fulfil regulatory body requirements. On the other hand, two-loci underdominance provides a more effective gene drive profile as it allows some gene-flow between the wild-type and engineered population. Figure 2.4 illustrates how this system would result in higher than Mendelian inheritance rates of engineered alleles and hence allow for gene drive. Moreover, it would still allow for a relative confinement of the engineered alleles as migration rates would have to be high for fixation in neighbouring populations to occur (Marshall and Hay 2012). Finally, two-locus underdominance can carry extra copies of refractory genes, maximising the chance of full refractoriness. To minimise the chance of resistance to refractoriness two different refractory alleles could be used, one for each underdominance allele. This could be implemented for both one-locus and two-loci systems making it a general advantage of underdominance gene drives.

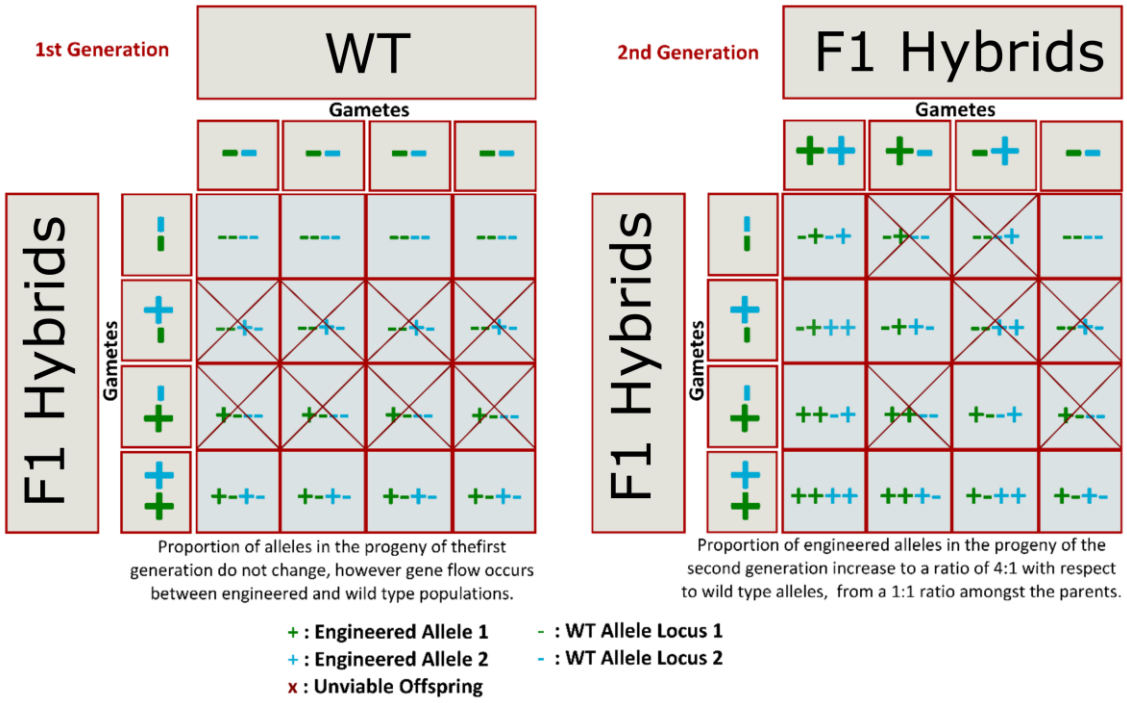


Figure 2.4: Engineered underdominance alleles increase in frequency when introduced into a wild population:
First generation hybrids between wild type and the released drive strain (homozygous for each of the two underdominance drive constructs) are heterozygous for at both loci. Such individuals have two lethals but are viable as they also have a suppressor for each lethal. These may mate wild type, or other such hybrids, or the release strain, for example. Outcomes of the first two of these potential crosses are shown. Parents labelled engineered 'homozygous', are actually transheterozygotes for engineered alleles in two different loci. Progeny which do not inherit at least one copy of each allele will not be viable. Each of the crosses have several non-viable progeny classes. Consequently, these F1 hybrids are less fit than either parental type, other things being equal, which establishes an under-dominant situation and the basis for a gene drive system.

2.1.5 Synthetic Biology

Synthetic biology is a developing field, with the ultimate aim of bringing the 'engineering' into genetic engineering. Synthetic biology is the assembly of novel and modular biological components in a rational attempt to construct a complex system with a desired new function (Kwok 2010). This project will make use of these concepts in the optimisation of underdominance-based constructs. The progression of synthetic biology relies on the characterisation of

simple molecular components with the ultimate goal of being able to combine parts to obtain a predictable complex output (Sprinzak and Elowitz 2005; Nandagopal and Elowitz 2011)**.** Any components characterised in this project may become part of the wider toolset in synthetic biology, available for use beyond intended purposes (Andrianantoandro *et al.* 2006).

The characterisation of novel modular components is likely to be able to expand the available synthetic biology toolset for uses beyond the contrived gene drive system. The gene drive system devised could be adapted for other uses; such as in genetically modified crops, which are in need of population isolation mechanisms to guarantee confinement and improve public acceptance (Kwit *et al.* 2011). The successful underdominance-based gene drive would have many applications beyond the mosquito species *Aedes aegypti*, which is the target of this project. The characteristics of such a drive; relative confinement in a local population, ease of reversal and considerable longevity makes it an ideal system for field trials. Ultimately, this project could allow for a safe control of a wide range of insect vector species, with minimal environmental impacts, that could potentially save many communities from devastating diseases such as malaria and dengue (McGraw and O'Neill 2013).

2.1.6 Prototypes for an underdominance system

The starting underdominance prototypes proposed for this project were constructed by Oxitec (Figure 2.5); tetO-LexA-NIPP1-AeHex1g-tTAV2-tetO-Gal4Groucho and tetO-UAS-NIPP1-AeHex1g-tTAV2-tetO-LexAGroucho.

The validation of the constructs will involve their injection into *D. melanogaster* and the generation of stable transgenic lines, following procedures described in Materials and Methods 2.2. Transposable elements (*piggyBac* and *P element*) will be used for transgenesis instead of site directed systems. This is to provide a range of transgenic line expression strengths with which to optimise the system. *Drosophila* will be fed with tetracycline to inhibit killer elements to allow individual generation of transgenics. Both the prototypes and individual components will be tested for their required output, i.e. correct spatial expression of fluorescence, or expected damage of certain tissues. Finally, validation would ideally continue in an analogous manner in other more relevant vector species such as *Aedes aegypti*.

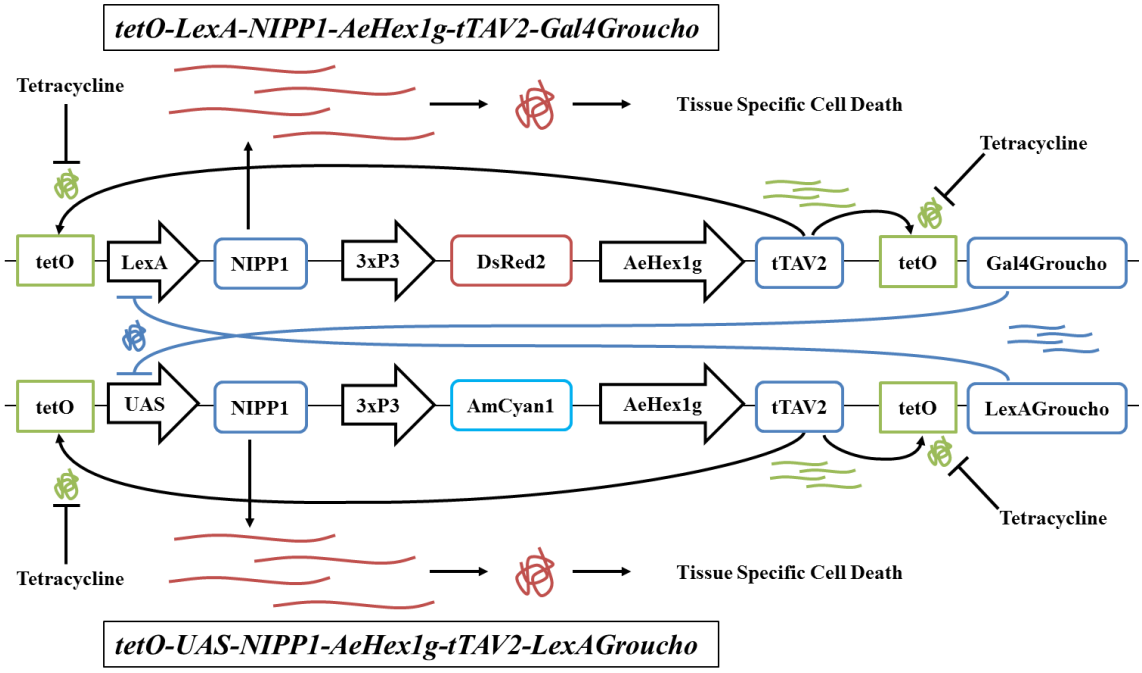


Figure 2.5: Proposed prototype constructs for underdominance:
NIPP1 is the killer protein, as it is a nuclear inhibitor of PP1 (Protein phosphatase 1) and its overexpression has been shown to be cell lethal in a range of tissues and developmental stages in *Drosophila* (Parker *et al.* 2002). DsRed2 and AmCyan1 are transformation markers (eye expression). tTAV2 is the activator which drives expression of components from tetO enhancers in the absence of tetracycline. AeHex1g promoter (Hex) drives the expression of tTAV2; the cell, tissue and developmental activity of this promoter defines the activity of the whole system, since expression of all components apart from the transformation marker depend on tTAV2 presence (ignoring basal expression). *AeHex1g* is derived from the promoter of *Hex1γ* in *Aedes aegypti* (Totten *et al.* 2013); the promoter fragment used is predicted to drive expression in the female fat body in late larval stages (Telfer and Kunkel 1991; Korochkina *et al.* 1997). Finally, the Gal4 Groucho repressor, in the first construct, suppresses expression of NIPP1 by binding at the UAS sequences and the LexA-Groucho repressor, in the second construct, binds the LexA sequences for NIPP1 suppression.

2.2 Materials and Methods

Figure 2.6 shows a flow-chart summary for Materials and Methods showing the different subheadings in this Section in an approximate chronological order.

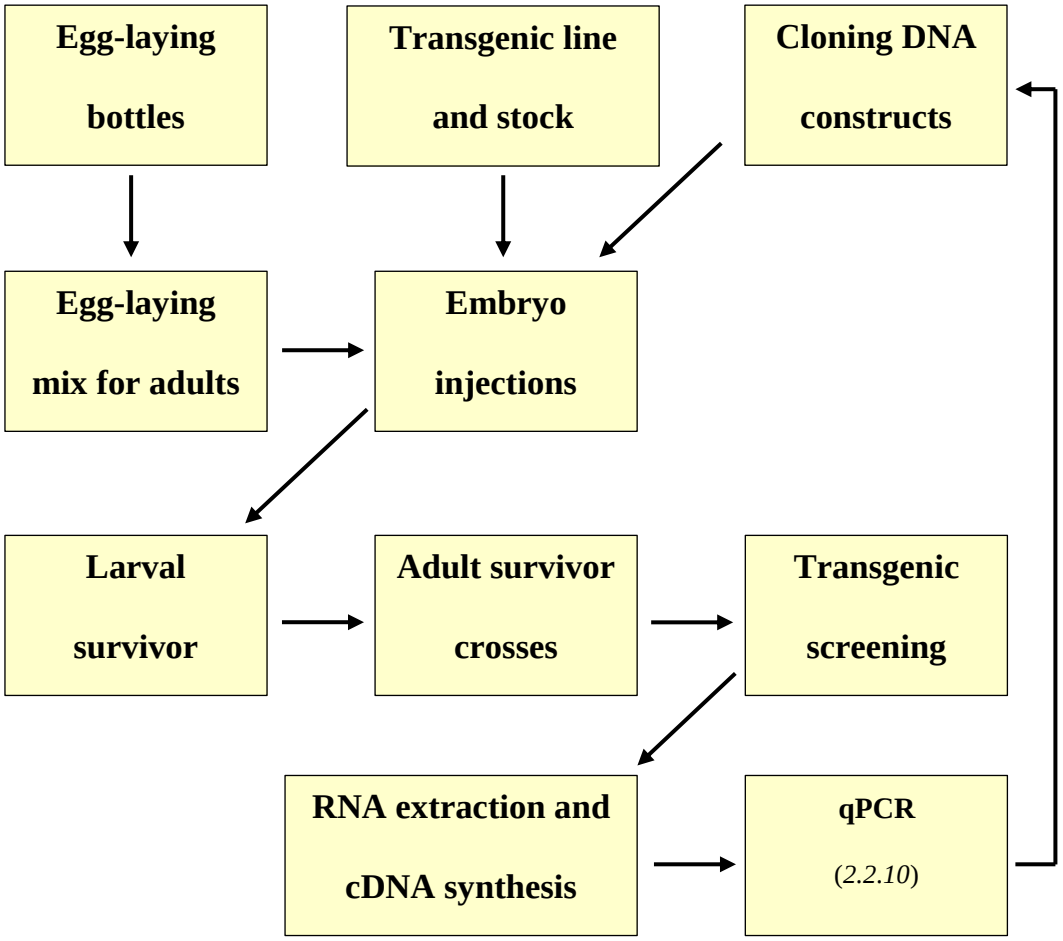


Figure 2.6: Materials and Methods summary:
This flow chart, starting from the top, states the different methods in chronological progression. This chronological order is only approximated since methods such as stock maintenance were carried out throughout the project.

2.2.1 Egg-laying bottles

For the collection of *D. melanogaster* embryos, adult w^{1118} flies were kept in inverted plastic bottles which were capped by laying pots containing an egg-laying mix. The w^{1118} strain carries a null allele of white which includes a

deletion of exon 1 (Kurkulos *et al.* 1991). The bottles were kept in a 25°C incubator in the dark and laying pots were changed every 40 minutes to collect the embryos at the desired stage for injection (Section 2.2.5)

2.2.2 Egg-laying mix for adults

Female *D. melanogaster* are encouraged to lay eggs on a fruit food based surface. An agar-based fruit mix was prepared to fill the laying pots and serve as the egg-laying surface for the flies. The egg laying mix was prepared in a final volume of 260ml. 6g of sucrose (Fisher Scientific) were weighed into 60ml of blackcurrant fruit juice (Ribena) and the mix was warmed until the solute was dissolved. 6g of agarose were weighed into 200ml of distilled water and were boiled until the solute dissolved. The fruit juice mix was then added to the agarose mix and 3ml of 10% Nipagen/90% Ethanol were added to prevent fungal growth. After letting the mix cool 2 to 4 minutes at room temperature, it was poured into laying pots to have around 5ml of mix per pot. In addition, the same final mix would be poured into petri dishes that would then be used to incubate the injected embryos as they hatch into larvae (Section 2.2.5). All of the prepared pots and petri dishes were stored at +4°C and were used within two weeks to avoid potential fungal contamination.

2.2.3 Transgenic line and stock maintenance

All injections were carried out in a w^{1118} *D. melanogaster* background and all adult survivors after injection were crossed to the same stocks. These stocks were

maintained in plastic bottles containing *D. melanogaster* food and were kept in an incubator at 25°C. An approximate total of 10 bottles were kept at any one time, and stock bottles were turned over every two weeks.

Transgenic individuals were named with a number (according to the G0 adult they originated from) and a letter for every independent insertion found from the same G0 adult. Transgenic lines were then reared On- or Off-Tet as required. On-Tet stands for rearing or feeding on a tetracycline containing diet, whilst Off-Tet stands for feeding or rearing on a diet lacking tetracycline.

Tetracycline was used at a final concentration of 30µg/ml. Tetracycline stocks were prepared and kept at a concentration of 5mg/ml in 70% Ethanol at -20°C. Tetracycline food vials were also prepared in the same manner for developing injection survivors (Section 2.2.6) and for subsequent adult crosses (Section 2.2.7).

2.2.4 Cloning DNA constructs

The following DNA constructs used in this project were provided by Oxitec; the industrial partner in this project; OX4784, OX4785, OX4755, OX5126, and OX4772 (Figure 2.7). 20µl of mini-prep DNA was received for each of these constructs and these samples were subsequently transformed into competent bacterial cells (JM109), grown into a 200ml culture volume of LB Broth

(Sigma), and DNA extracted using an Endo-Free Plasmid DNA Maxi-kit (Omega) to have DNA ready for injection.

The following DNA constructs used in this project were not provided by Oxitec; CU1, CU2, CU3, CU4, and CU5 (Table 2.2). These were cloned by excising the functional cassettes from OX4772, OX4784, and OX4785 and subcloning them into the pCaSpeR4 vector.

2.2.5 Embryo injections

Prior to the collection of *D. melanogaster* embryos, the construct DNA mix was prepared for injection. The injecting DNA mix consisted of 700µg/µl of DNA construct, 250µg/µl of Piggy-Bac or *P element* helper construct in an injection buffer. The helper is essential since it encodes for the transposase which will use the *piggyBac* or *P element* sites of the DNA construct to insert it in the *D. melanogaster* genome. 3µl of the injection mix were loaded on the microneedle (Femtotips®II, Eppendorf).

Embryos were collected at pre-cellularisation stages (before 2 hours post-fertilisation) to maximise the chances of the injected DNA reaching germ cell nuclei since cell membranes would not yet have formed a barrier. Embryos were collected from egg-laying pots every 40min and would then be dechorionated. For the dechorionation of the embryos, two pieces of double sided sticky tape (Scotch) were stuck to two different glass slides and the embryos were collected from the laying pots using a fine paintbrush to be transferred onto one of the

slide's sticky tape. A 0.5mm thick copper wire was placed in-between both slides to prevent crushing the embryos and then the sticky slides were pressed together to stick the chorions of the embryos on both slides. The slides were then prised apart to break open the chorions and leave the embryos exposed to the air.

At a dissection microscope (Olympus SZ51), embryos were individually picked using the tip of a pair of tweezers and were transferred onto a glued 22x22mm coverslip (Menzel-Gläser) on a microscope slide (0.8-1.0mm thick) to put them in the right orientation and immobilise them for injection. To prepare the glued coverslip, 10cm of double sided sticky tape was added to a 15ml tube containing 10ml of 100% heptane and the mix was left on a shaker until the glue dissolved. This mix was then pipetted slowly as a strip at the edge of a coverslip, leaving a strip of glue as the heptane quickly evaporated. Embryos were lined up and orientated on a coverslip/slide so as to have the posterior end of the embryo at the microneedle side of the microscope.

Dechorionation was simultaneous with this method for each batch of embryos and ensured that all embryos would desiccate at the same rate. Embryos were allowed to desiccate on the glued coverslip for around 10min to prevent them from bursting upon injection. Subsequently, the embryos were covered in oil (Halocarbon Oil 700) to prevent further dehydration. The slides were then transferred to the injection microscope (Nikon Eclipse Ti-S) and the joystick control (TransferMan NK2, Eppendorf) was used to position the microneedle at

the same plane as the embryos. The embryos were then subsequently injected. The microneedle was always inserted at the posterior end of the embryo since pole cells form at this end and hence injecting at this end maximised the chances of germ cell DNA insertion. The injection air pressure was provided automatically (Femtojet, Eppendorf) and a compensatory pressure of 70hPa was used to get a constant flow of DNA from the microneedle. If the flow was inadequate, the compensatory pressure was adapted. Once a round of embryos were injected, the coverslips would be removed from the slide and transferred onto a petri dish containing egg-laying mix and small yeast balls to attract the hatching larvae. Subsequent rounds of injections were then carried out every 40min.

2.2.6 Larval survivor recovery

The petri-dishes containing the coverslips with injected embryos were then incubated in the 18°C room inside a humid chamber. The humid chamber consisted of a closed plastic box with wet tissue paper around the petri dishes to maintain a hydrating environment for the embryos. Surviving larvae were collected during the following 3 days and transferred into tetracycline food vials (30µg/ml) and grown at 25°C until hatching into adults.

2.2.7 Adult survivor crosses

As soon as pupae from Section 2.2.6 eclosed into adults, these were crossed individually with w^{1118} *D. melanogaster* flies. Female w^{1118} were selected to be virgin. The vials used for the crosses would also contain tetracycline food.

2.2.8 Transgenic screening

The progeny from the adult crosses from Section 2.2.7 was anaesthetised using CO_2 and was screened using a fluorescence microscope (Leica) using DsRed or AmCyan light filters to detect fluorescence markers depending on the construct injected. Tissue-specific transformation marker expression amongst the screened progeny would indicate the presence of the construct in the organism. Markers included cyan fluorescent eyes from *3xP3-AmCyan1*, red fluorescent eyes from *3xP3-DsRed2*, red fluorescent bodies from *3xP3-DsRed2*, and red pigment eyes from w^{+} (Figure 2.7 and Table 2.2).

2.2.9 RNA extraction and cDNA synthesis

Samples for RNA extraction were prepared by homogenising selected tissues (heads or whole bodies) in 100µl of lysis buffer from RNAqueous™-Micro Total RNA Isolation Kit (Invitrogen, Ambion). RNA extraction was then performed following the kit guidelines. cDNA was made from the whole RNA extracts using oligo dT primers and otherwise following the guidelines of the SuperScript™ III Reverse Transcriptase kit (Invitrogen).

2.2.10 qPCR

qPCR was performed using a SYBR Green kit (Applied Biosystems). Reactions were performed in triplicate in 20µl volumes. cDNA was diluted 1 in 10 and 1µl of diluted cDNA template was used for each reaction. Primers used for product amplification are shown in Table 2.1. qPCR was optimised by melt curve analysis, primers were redesigned until a single product peak was observed. Fold-expression values were calculated by using the Ct (cycle threshold value) value differences between tested primer and RNA pol II subunit 1 control primer reactions. The control reference for qPCR was previously validated by the lab, and as relative values, rather than absolute values between the different tested components were to be measured a single control reference seemed appropriate.Values were then normalised with respect to w^{1118} control samples. Each reaction was run in triplicate, representing the technical replicates. The standard deviation was calculated for the technical replicates and error bars were plotted to display it. Each *D. melanogaster* sample had either tissue coming from 1 or 2 individuals, the number of biological replicates per experiment is indicated in the respective figure legends. Statistical analysis with sufficient power could not be carried out due to the low number of biological replicates. Nevertheless, the qPCR data was in agreement with the phenotypes observed.

Table 2.1: qPCR primers used:

RNA polymerase II primer pair was used to control the amount of DNA input. *tTAV2* and *NIPP1* primer pairs were used to assay levels of expression of these components. All primer pairs were designed to have a melting temperature close to 60°C and to generate an amplicon of no more than 200bp.

Primer Name	Sequence
FW RNA pol II	CCACCCGGCCACGTAAG
RV RNA pol II	AAGAGGGAGAAACACTCGGC
FW tTAV2	TCTGCGGATTGGAAAAGCAAC
RV tTAV2	CGGGGCATCATCATCCGG
FW NIPP1	TCCAGCAGAGCCTAGTTGAC
RV NIPP1	TTTCGGCCGGTAGACCTTG
FW LexAGroucho	CTGCCGGAGAATAGCGAGTT
RV LexAGroucho	TCCAGTGTATCGGCGATGGT
FW Gal4Groucho	CCAGCTGACCGTGTCCTATC
RV Gal4Groucho	CGTTCCAGTGTATCGGCGAT

2.3 Results and Discussion

2.3.1 Underdominance transgenic lines were generated

The initial strategy involved injecting the prototype constructs for underdominance, tetO-LexA-NIPP1-AeHex1g-tTAV2-tetO-Gal4Groucho and tetO-UAS-NIPP1-AeHex1g-tTAV2-tetO-LexAGroucho (Figure 2.7). Injection of both constructs was carried out (Section 2.2.5) on eggs from parents fed on tetracycline. Similarly, the surviving larvae from these injections were raised on tetracycline. This was essential to prevent the constitutive expression of the lethal *NIPP1* and hence the loss of any stable insertions of the transgenes. These constructs were provided by Oxitec and genomic integration relied on the *piggyBac* transposition system. Survival rates from embryo to adult were 6.7% and 4.5% respectively. 170 and 10 G0s were individually crossed and no transgenic adults found amongst their progeny (Table 2.2). Survival rates from embryo to adult after microinjection in *D. melanogaster* normally range between 12.5% and 37.5%, and is strongly user-specific (O'Connor and Chia 2002). The low survival rates and the lack of transgenic adults called for a different approach. With the evidence available, it was difficult to determine the source of the low survival rate; whether it was transient expression toxicity from the construct, the lack of injection experience, or a mix of both factors.

To address both potential explanations an alternative strategy involving injecting individual components into *D. melanogaster* was chosen. Initially, the first set of

individual components were provided by Oxitec and relied on the *piggyBac* transposition system for transgenesis. These were *tetO-UAS-NIPP1* (the killer component of prototype *OX4785*), *AeHex1g-tTAV2* (putative abdomen specific expression of *tTAV2* activator), and *GMR-tTAV2* (eye specific expression of *tTAV2* activator) (Figure 2.7 B and C). Upon injection survival rates from embryo to adult were 11.9%, 6.4%, and 8.4% respectively. These were higher survival rates than for the previous injections which suggested better injection technique, or lower transient expression toxicity of the components. Interestingly, *tTAV2*-containing sequences resulted in the low survival rates observed up to this point.

Transgenic lines were established for *AeHex1g-tTAV2* and *GMR-tTAV2* but not for *tetO-UAS-NIPP1* (Table 2.2). Given that both constructs, for which I had not been able to generate transgenics, shared *3xP3-DsRed* as the transgenic marker, subcloning of *tetO-UAS-NIPP1* into a *pCaSpeR4* vector was pursued. *pCaSper4* carries a mini-white gene which produces eye colour in the white mutant background (w^{1118}). Transgenesis efficiency was not compromised by making use of *P element* transposition which is at least as efficient as *piggyBac* transposition in *D. melanogaster* with regards to transformation rate (Venken and Bellen 2005).

The new construct, *PC4-tetO-UAS-NIPP1*, was injected and resulted in a G0 survival rate of 16.8% (Table 2.2). Transgenic adults were found amongst the G1

progeny, selected by w^+ marker expression. A lack of visible *DsRed2* in the w^+ transgenic G1 progeny, and the fact that 21 out 73 of G0 adults gave transgenic progeny clearly indicated that *DsRed2* was not being expressed in the previous *tetO-UAS-NIPP1* injections. Even though *P element* transformation is more efficient than *piggyBac* the results above cannot be explained just by that fact.

The prototype construct *tetO-LexA-NIPP1-AeHex1g-tTAV2-tetO-Gal4Groucho*, potentially having also the same faulty *3xP3-DsRed2* marker as *tetO-UAS-NIPP1* did not result in transgenic adults after injection even when in the *pCaSpeR4* vector. To attempt to circumvent this problem the *DsRed2* marker was replaced by *AmCyan1* as a parallel strategy. The new construct, *AmCyan-tetO-LexA-NIPP1-AeHex1g-tTAV2-tetO-Gal4Groucho*, did not result in transgenic adults after injection either (Table 2.2). This version of the prototype never yielded transgenics in any of its forms. Other transgenics, detailed below, (Figure 2.7 and Table 2.2), were used to analyse the behaviour of the elements of this construct. In light of these results it was clear that generating a transgenic line of *tetO-LexA-NIPP1-AeHex1g-tTAV2-tetO-Gal4Groucho*, in any of its forms, was unnecessary.

A second attempt at injecting the other prototype construct *tetO-UAS-NIPP1-AeHex1g-tTAV2-tetO-LexAGroucho* was carried out given that only 224 embryos were injected in the first round, and presumably *3xP3-AmCyan* expression would

work just as in *AeHex1g-tTAV2* transgenic lines. Transgenic adults were obtained from these injections, and transgenic lines were established (Table 2.2).

Finally, to help analyse the underdominance based system the remaining individual components were subcloned into *pCaSpeR4* vectors and injected. These were *tetO-LexA-NIPP1*, *tetO-LexAGroucho*, and *tetO-Gal4Groucho* (Figure 2.1), all of which resulted in transgenic lines being established (Table 2.2).

When looking at survival rates across the different injections, there is no evidence of a strong improvement over time. However there was a modest improvement of user injection technique as higher survival rates were obtained (16.4%) than at the beginning (6.4% and 4.7%), and confidence in injection proficiency increased. The lack of a stronger improvement in survival rates could be explained by the fact that transient expression in injected embryos may have been causing lethality. Indeed, *NIPP1*, *tTAV2*, and even *Groucho* overexpression, as determined below in Section 2.3.8, were potentially lethal components. Larval to adult survival rates, ranging from 21% to 48%, were low (Figure 2.7), which further supports the idea of injection derived lethality.

Functional analyses of all these constructs are described below (Section 2.3.2 to 2.3.8). The analyses revealed that having only one of the prototype constructs in transgenic lines was sufficient for the conclusions made about this system.

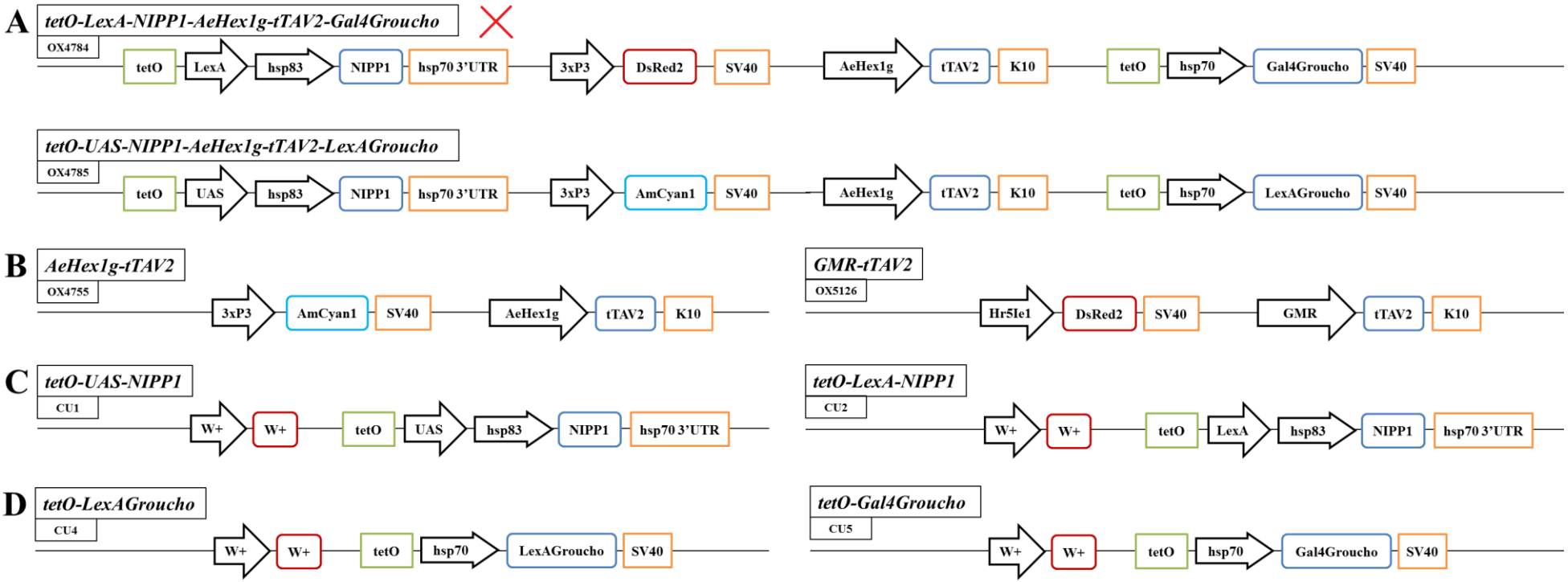


Figure 2.7: An underdominance system was studied by cloning both individual and combined components into injection constructs:
Constructs with the OX code were constructed by Oxitec. Injection data are shown in Table 2.2. Some of the construct variants in Table 2.2 are not shown here. Unsuccessful transgenesis of *OX4784* is marked with a red X. Transgenic visual markers used are either the mini w^+ gene in *D. melanogaster* or tissue specific fluorescent markers. Promoters are shown as black arrows (minimal promoters are smaller). Visible marker proteins are coded with their respective colours. Enhancers are coded in green. Terminators are coded in orange and CDSs are coded in blue. (**A)** Both prototype constructs for underdominance are shown here as described in (Section 2.1). (**B)** Both *tTAV2* components are shown as individual driver lines under the control of tissue specific promoters *GMR* and *AeHex1g*. *tTAV2* should upregulate both killer and rescue components in this system. **(C)** Both NIPP1 components are shown as individual killer lines. Each construct has a different promoter sequence (*UAS/LexA*) which in this case should function as a corepressor binding site. **(D)** Both DNA-binding and corepressor protein fusion constructs are shown. They should work as sequence specific corepressors of *NIPP1* expression, via *LexA/UAS* binding sites respectively.

Table 2.2: Injection of underdominance-based components provided enough transgenic lines for subsequent characterisation of the system:
Constructs with the OX code were cloned and provided by Oxitec. Constructs are listed in a roughly chronological injection order. A summary of the number of injected embryos, surviving adults, embryo to adult survival rates in brackets, and established transgenic stock lines is provided. ≈ Marks approximate values where exact numbers were not recorded. Visible transgenic markers and transposable elements for transgenesis are indicated. In the transgenic lines column the number in brackets represents the number of transgenic-yielding G0 adults. Transgenesis efficiency was calculated as a percentage of transgenic-yielding G0 adults.Although no transgenic lines could be established for prototype *tetO-LexA-NIPP1-AeHex1g-tTAV2-tetO-Gal4Groucho* in any of its variants, established lines for the other prototype and all individual components were sufficient to investigate the function of the whole system.

Construct Code	Construct	Marker	Transposable Element	Embryos	Larvae	Larval to Adult Survival Rate	G0 Adults	Transgenic Lines	Transgenesis Efficiency
OX4784	*tetO-LexA-NIPP1-AeHex1g-tTAV2-tetO-Gal4Groucho*	*3xP3-DsRed2*	*piggyBac*	2525	357 (14.1%)	48%	170 (6.7%)	0	0%
CU3	*PC4-tetO-LexA-NIPP1-AeHex1g-tTAV2-tetO-Gal4Groucho*	w^+ and *(3xP3-DsRed)*	*P element*	≈800	≈70 (≈8.8%)	≈43%	≈30 (≈3.8%)	0	0%
CU6	*AmCyan-tetO-LexA-NIPP1-AeHex1g-tTAV2-tetO-Gal4Groucho*	*3xP3-AmCyan1*	*piggyBac*	≈800	≈60 (≈7.5%)	≈45%	27 (≈3.8%)	0	0%
OX4785	*tetO-UAS-NIPP1-AeHex1g-tTAV2-tetO-LexAGroucho*	*3xP3-AmCyan1*	*piggyBac*	224	26 (11.6%)	38%	10 (4.5%)	0	
				≈500	≈60 (≈12%)	≈48%	29 (≈5.8%)	20 (4)	13.8%
OX5126	*GMR-tTAV2*	*Hr5E1-DsRed2*	*piggyBac*	520	125 (24%)	34%	42 (11.9%)	10 (3)	23.8%
OX4755	*AeHex1g-tTAV2*	*3xP3-AmCyan1*	*piggyBac*	342	106 (31%)	21%	22 (6.4%)	13 (10)	45.5%
OX4772	*tetO-UAS-NIPP1*	*3xP3-DsRed2*	*piggyBac*	1158	317 (27.4%)	44%	138 (11.9%)	0	0%
CU1	*PC4-tetO-UAS-NIPP1*	w^+	*P element*	434	187 (43.1%)	39%	73 (16.8%)	38 (21)	28.8%
CU4	*tetO-LexAGroucho*	w^+	*P element*	≈400	≈30 (≈7.5%)	≈37%	11 (≈2.8%)	12 (2)	18.2%
CU5	*tetO-Gal4Groucho*	w^+	*P element*	≈600	≈90 (≈15%)	≈44%	40 (≈6.7%)	18 (4)	10%
CU2	*tetO-LexA-NIPP1*	*w+*	*P element*	≈400	≈100 (≈25%)	≈40%	40 (≈10%)	10 (2)	5%

2.3.2 tTAV2 expression is detected

To monitor function, mRNA expression levels of components were determined by RT-qPCR as described (Section 2.2.10). One of the first questions was whether *tTAV2* expression was high enough and capable of *NIPP1* upregulation. In turn, would high *NIPP1* levels lead to the expected cell lethality effect (Parker *et al.* 2002).

Since construct integration in the genome was not site-directed, the expression levels of the individual components were checked prior any crosses or functional analysis was done. Transgenic lines, established from G1 marker positive progeny were named with a number (corresponding for the G0 adult number) and a letter (corresponding to the different lines established from the same G0 parent). *tTAV2* expression levels vary amongst the eight transgenic lines picked for RT-qPCR analysis for both *AeHex1g-tTAV2* and *GMR-tTAV2* (Figure 2.8). The *AeHex1g* promoter, which is a promoter derived from and tested in the *Aedes aegypti* mosquito with female fat body specific expression (Zakharkin *et al.* 2001), had not had its spatial expression characterised in *D. melanogaster*. Expression appears to be present in both sexes and present in head, thorax and abdomen fractions (Figure 2.9). It is not restricted to the fat body as in *Aedes aegypti*, nor does it show its expected female specific expression (Totten *et al.* 2013).

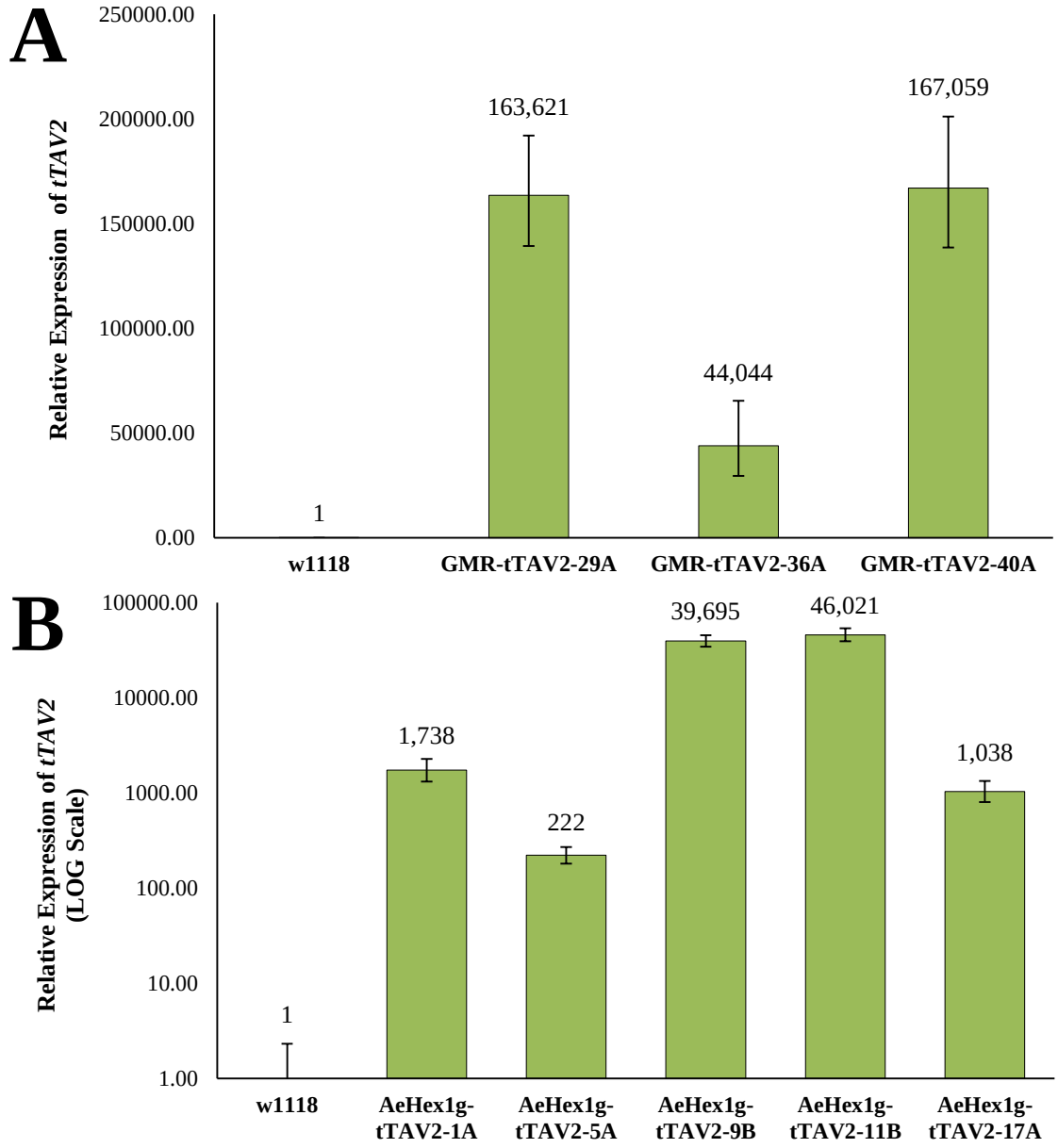


Figure 2.8: *tTAV2* expression varies across insertions for both *GMR-tTAV2* and *AeHex1g-tTAV2* driver lines:

Relative expression of *tTAV2* with respect to w^{1118} is shown on the y-axis for different driver line insertions on the x-axis. w^{1118} is used as a control given that it has no expression of *tTAV2*, the high relative expression values are an artifact as they are fold differences with respect to 0, however it is the relative expression of *tTAV2* between the different lines that is relevant. Relative expression values are normalised for the cDNA input by comparing to RNA polymerase II control amplification. Error bars display the standard deviation for triplicated technical replicates. **(A)** *GMR-tTAV2* lines are shown here. Each cDNA sample was prepared from 2 *D. melanogaster* female adult heads, i.e. 2 biological replicates. *tTAV2* expression seems higher for lines *29A* and *40A*. **(B)** *AeHex1g-tTAV2* lines are shown here. The y-axis is on a logarithmic scale to facilitate the comparison of the strongly expressing lines *9B* and *11B* and the rest of the insertions. Each cDNA sample was prepared from 2 *D. melanogaster* female adults, i.e. 2 biological replicates. *tTAV2* expression seems much higher for lines *9B* and *11B*.

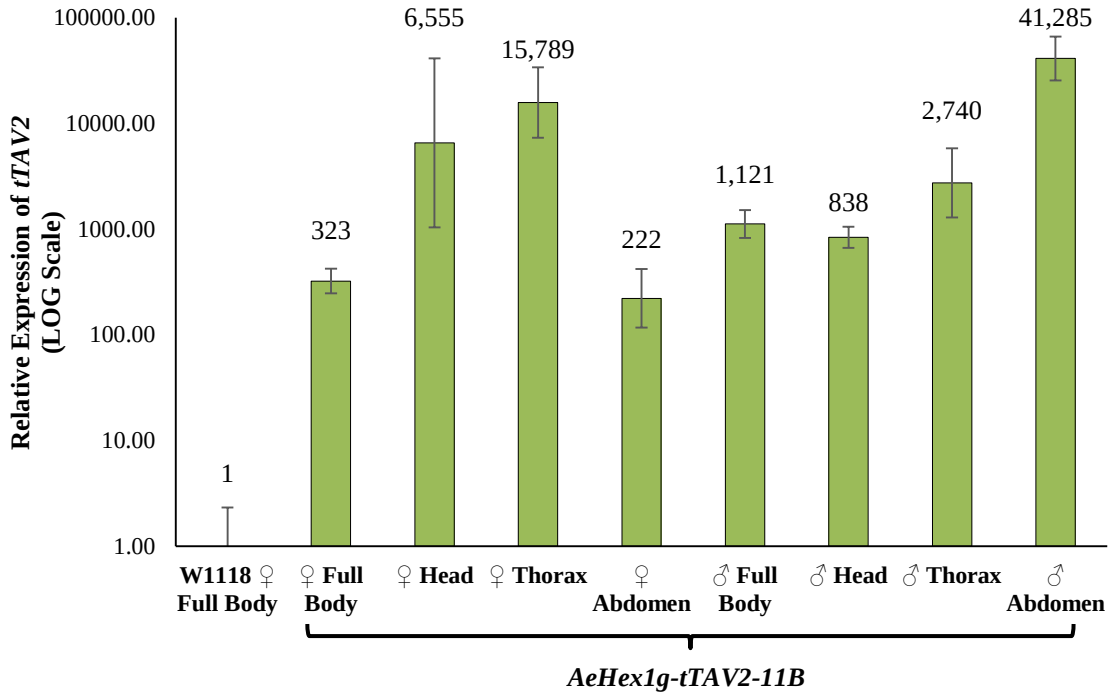


Figure 2.9: *AeHex1g* driven expression of *tTAV2* does not reproduce the expression pattern of this promoter in *Aedes aegypti*:
Relative *tTAV2* expression with respect to w^{1118} on the y axis on a logarithmic scale is shown for different cDNA samples on the x axis. Expression level values are normalised for the amount of cDNA input by comparing to RNA polymerase II control reactions. Error bars display the standard deviation for triplicated technical replicates. Expression fold changes are expressed with respect to the level of *tTAV2* in w^{1118} control which is given a value of 1. Expression values are indicated above the error bars. cDNA samples were prepared from 1 *D. melanogaster* adult per sample (whole or split), i.e. 1 biological replicate. *tTAV2* expression seems highest in the male abdomen of *AeHex1g-tTAV2-11B* and varying across the other body fractions.

2.3.3 NIPP1 expression detected

Having detected tTAV2 expression the next question was whether *NIPP1* was expressed in the *UAS-NIPP1* lines in the absence of GAL4 (basal expression). Unlike the varying *tTAV2* expression amongst the assayed lines, *NIPP1* expression levels were very similar amongst the three stock lines picked for RT-qPCR analysis. The three lines show basal (or leaky) expression of *NIPP1* when compared to w^{1118} endogenous expression, by using primers capable of amplifying both endogenous and transgene *NIPP1* sequences. This could be due to the 5x*UAS* sequences and minimal promoter sequences present (Figure 2.10).

Although *UAS* sequences should need Gal4 for upregulation, *UAS* and/or *hsp83* sequences could be resulting in the observed basal expression. Given the lack of solid evidence for *UAS*-only basal expression as it has been reported to be very low in the absence of Gal4 (Potter *et al.* 2010), and that it is usually present beside a minimal promoter in *Drosophila* (Pfeiffer *et al.* 2010), the basal expression of *NIPP1* is most likely driven by *hsp83* (Xiao and Lis 1989), the minimal promoter in this system.

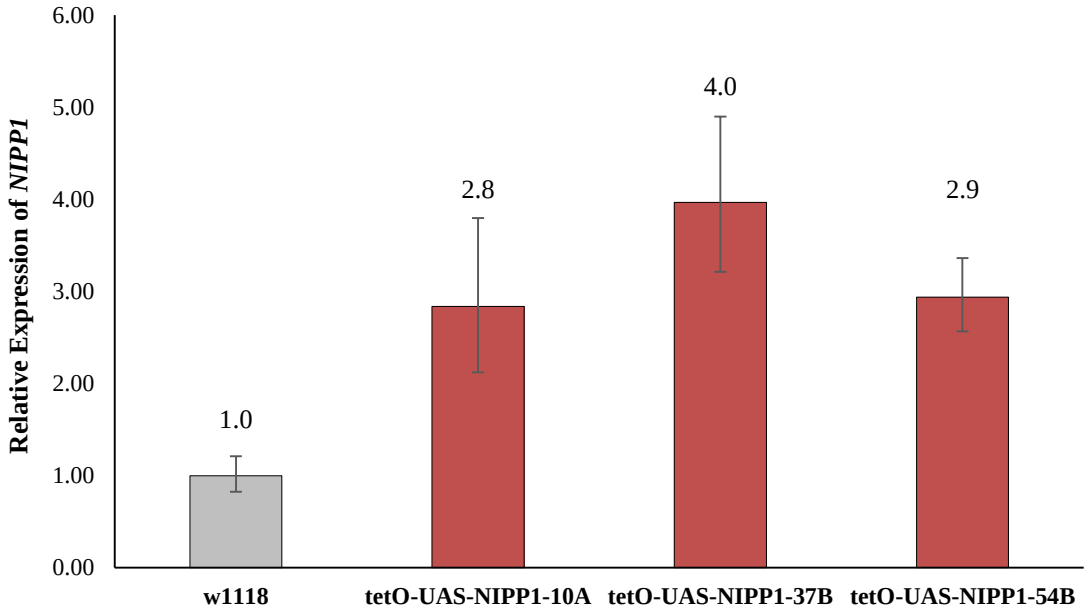


Figure 2.10: Basal *NIPP1* expression is similar across *tetO-UAS-NIPP1* lines:
Relative expression of *NIPP1* with respect to w^{1118} is shown on the y-axis for different killer line insertions on the x-axis. Expression level values are normalised for the amount of cDNA input by comparing to *RNA polymerase II* control reactions. Error bars display the standard deviation for triplicated technical replicates. Expression fold changes are expressed with respect to the level of endogenous *NIPP1* in w^{1118} control which is given a value of 1. Expression values are indicated above the error bars. cDNA samples were prepared from 2 *D. melanogaster*female adult heads per sample, i.e. 2 biological replicates. The reason behind using just heads being that *NIPP1* upregulation was to occur in the eyes upon crossing to *GMR-tTAV2*. *NIPP1* basal expression in *UAS-NIPP1* lines results in an extra 3 to 4 fold *NIPP1* expression when detecting endogenous and transgene *NIPP1* sequences in this RT-qPCR assay. gDNA was digested with TURBO DNase (Thermo Fisher Scientific).

2.3.4 tetO-LexA/UAS-hsp83-NIPP1 do not respond to tTAV2 upregulation as expected

Having established that the driver is expressed, and there is some leaky expression of the responder, the next question was whether tTAV2 was able to upregulate the expression of the responder and result in a phenotype. *tetO-UAS-NIPP1* lines were crossed to *AeHex1g-tTAV2* and *GMR-tTAV2* lines expecting to see cell lethality throughout the body and eye tissues respectively (Figure 2.11). The expected phenotype of pupal or adult lethality (with *AeHex1g-tTAV2*) and loss of eye tissue (with *GMR-tTAV2*) *was* not observed in any of the crosses. Upregulation of *NIPP1* was detected with the presence of *tTAV2* in both crosses but the increase was very low (Figure 2.12). Upregulation of *NIPP1* expression in the presence of *tTAV2* was 12 fold at its highest, but typically 2 fold. The lack of phenotype could be due to insufficient upregulation of *UAS-NIPP1* expression.

Similarly, *tetO-LexA-NIPP1* lines were crossed to *AeHex1g-tTAV2* and *GMR-tTAV2* lines expecting to see cell lethality throughout the body and eye tissues respectively (Figure 2.11). Upregulation of *NIPP1* was detected with the presence of *tTAV2* in both crosses but the increase was again very low (Figure 2.13). The expected phenotype was not observed in any of the crosses. Upregulation of *NIPP1* expression in the presence of *tTAV2* was 16 fold at its highest, but typically between 2 and 5 fold. The lack of phenotype could be due to insufficient upregulation of *LexA-NIPP1* expression.

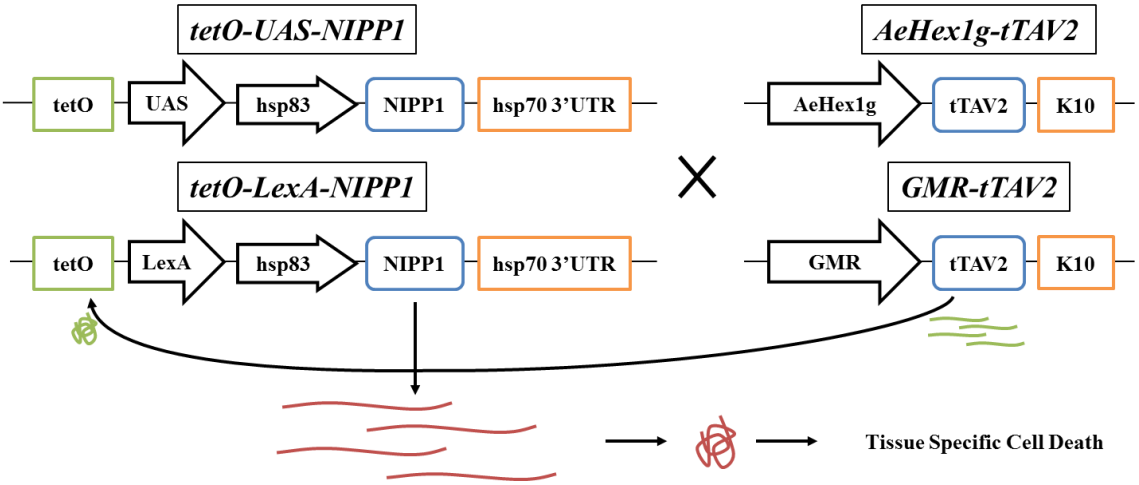


Figure 2.11 *NIPP1* is expected to be upregulated by tTAV2 and cause cell autonomous cell death:
NIPP1 expression in *tetO-UAS-NIPP1*/*tetO-LexA-NIPP1* is designed to be driven by spatially defined *tTAV2* expression from two different constructs resulting in full lethality and eye-specific cell death respectively.

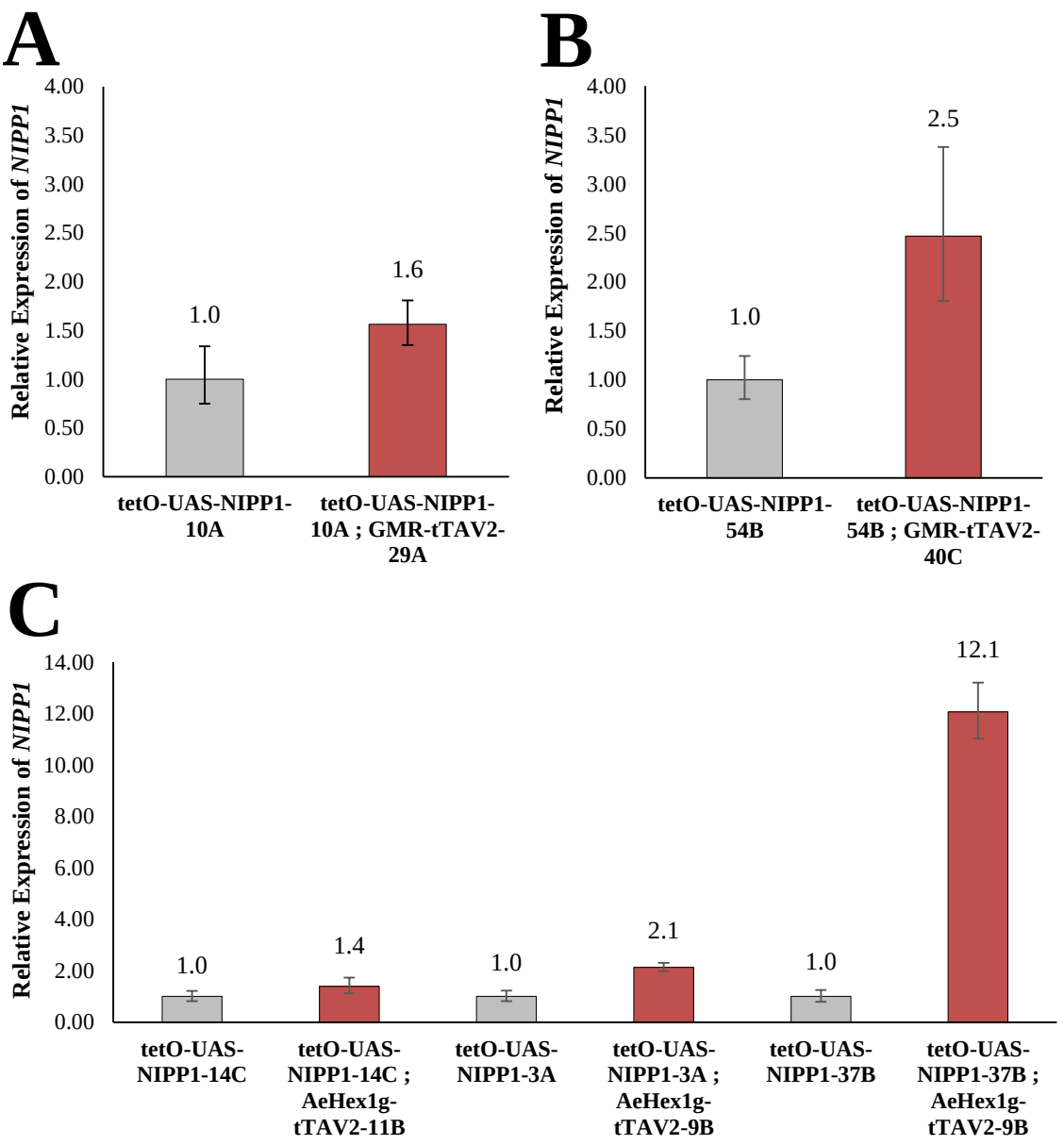


Figure 2.12: *NIPP1* upregulation is no more than 12 fold when *tetO-UAS-NIPP1* and *GMR-tTAV2* or *AeHex1q-tTAV2* lines are crossed Off-Tet:

Relative expression of *NIPP1* is shown on the y-axis for different killer line insertions on the x-axis. Expression level values are normalised for the amount of cDNA input by comparing to *RNA polymerase II* control reactions. Error bars display the standard deviation for triplicated technical replicates. Relative expression is expressed with respect to the level of *NIPP1* in the respective *tetO-UAS-NIPP1* line before the cross, which is given a value of 1. Expression values are indicated above the error bars. **(A)** cDNA samples were prepared from 2 *D. melanogaster* female adult heads per sample, i.e. 2 biological replicates. *NIPP1* expression was only upregulated 1.5 fold in the cross. **(B)** cDNA samples were prepared from 1 *D. melanogaster* female adult head per sample, i.e. 1 biological replicate. *NIPP1* expression was only upregulated 2.5 fold in the cross. **(C)** cDNA samples were prepared from 1 *D. melanogaster* abdomen per sample, i.e. 1 biological replicate. *NIPP1* relative expression was at most 12 fold.

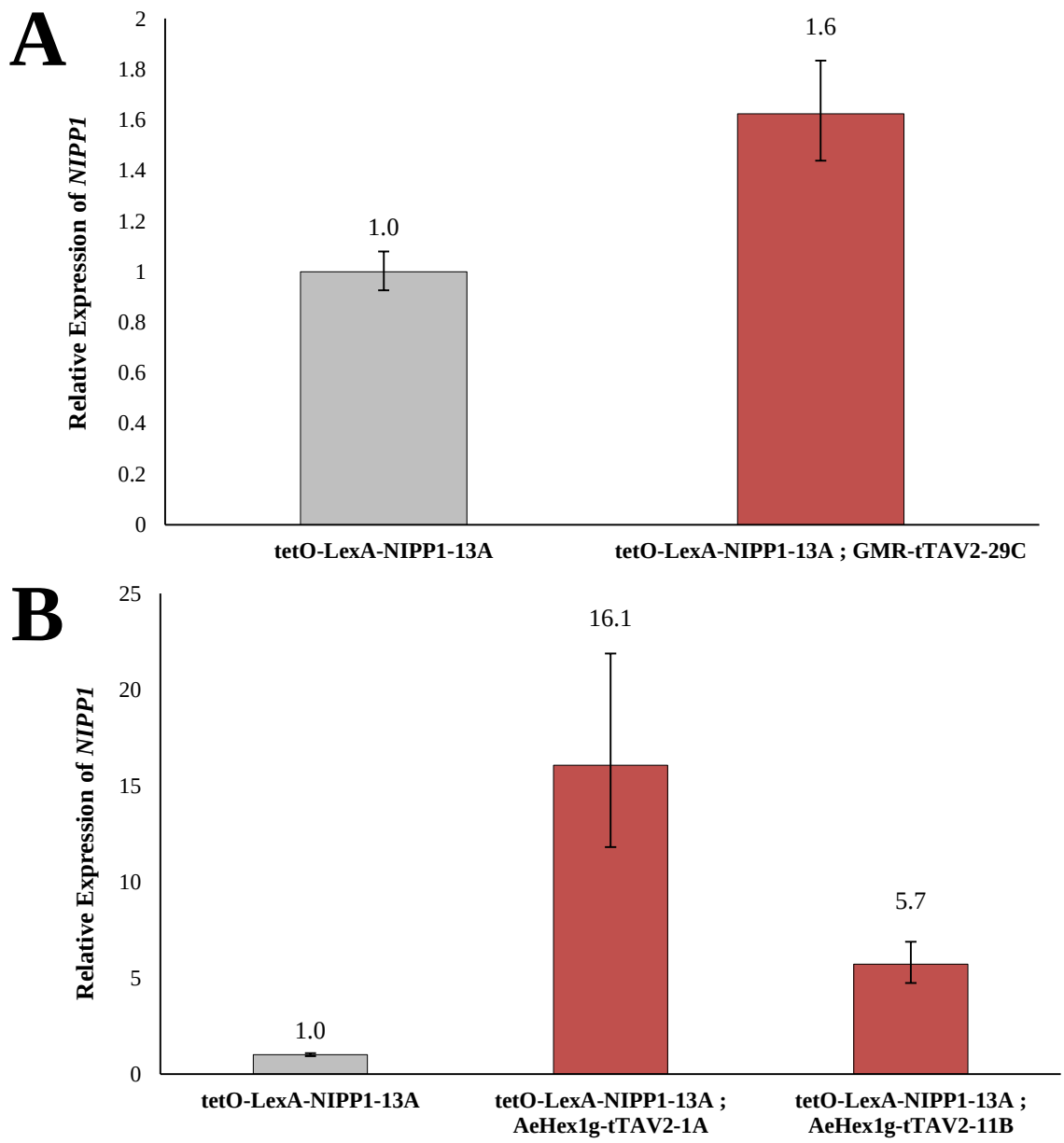


Figure 2.13: *NIPP1* upregulation is no more than 16 fold when *tetO-LexA-NIPP1* and *GMR-tTAV2* or *AeHex1g-tTAV2* lines are crossed Off-Tet:

Relative expression of *NIPP1* is shown on the y-axis for different killer line insertions on the x-axis. Expression level values are normalised for the amount of cDNA input by comparing to *RNA polymerase II* control reactions. Error bars display the standard deviation for triplicated technical replicates. Relative expression is expressed with respect to the level of *NIPP1* in the respective *tetO-UAS-NIPP1* line before the cross, which is given a value of 1. Expression values are indicated above the error bars. cDNA samples were prepared from 1 *D. melanogaster* female adult per sample, i.e. 1 biological replicate. **(A)** *NIPP1* expression was only upregulated 1.6 fold in the cross with *GMR-tTAV2-29C*. **(B)** *NIPP1* expression was only upregulated 16 and 5 fold in the two crosses shown respectively.

In addition, *NIPP1* upregulation was also not seen when *tetO-UAS-NIPP1-AeHex1g-tTAV2-tetO-LexAGroucho* construct lines were removed from a tetracycline diet. As illustrated in Figure 2.14A, the lack of tetracycline should allow for *NIPP1* upregulation. RT-qPCR results show the weak upregulation of *NIPP1* (Figure 2.14B), even in *tetO-UAS-NIPP1-AeHex1g-tTAV2-tetO-LexAGroucho* lines with high *tTAV2* expression due to a positive feedback loop (Section 2.3.7).

Sections 2.3.7 and 2.3.8 show that *tTAV2* is capable of strong upregulation via tetO enhancer sequences, demonstrating that the *tTAV-tetO* system in itself is not faulty. Indeed, a lethal effect was seen on *tetO-UAS-NIPP1-AeHex1g-tTAV2-tetO-LexAGroucho D. melanogaster* pupae reared Off-Tet (Section 2.3.6). However, this effect was not through *NIPP1* upregulation as shown in this Section and Sections 2.3.7 and 2.3.8. An explanation for the lack of upregulation of *NIPP1* could be the fact that the basal levels are already high, but this is unlikely given the results above when comparing expression to WT levels. In addition, *tTAV2* basal levels are likely higher than for *NIPP1*, given its constitutive *AeHex1g* promoter, and this did not seem to dampen the fold upregulation seen when under the effect of the *tetO-tTAV2* feedback loop (Figure 2.16).

The simplest explanation, for the lack of upregulation, is that the *Hsp83* minimal promoter was faulty in some way, independent of context. Although *Hsp70* is

much more commonly used as a minimal promoter, *Hsp83* was used to avoid repetition of minimal promoter sequences within the final underdominance constructs, in order to prevent unwanted recombination events, and subsequent loss of components. A second explanation is that *UAS* and *LexA* sequences are interfering with the minimal promoter itself or are impeding the interaction between the minimal promoter and the *tetO* enhancer. Use of different minimal promoters and a deletion of the *UAS* and *LexA* sequences would provide insights into which of these explanations is correct.

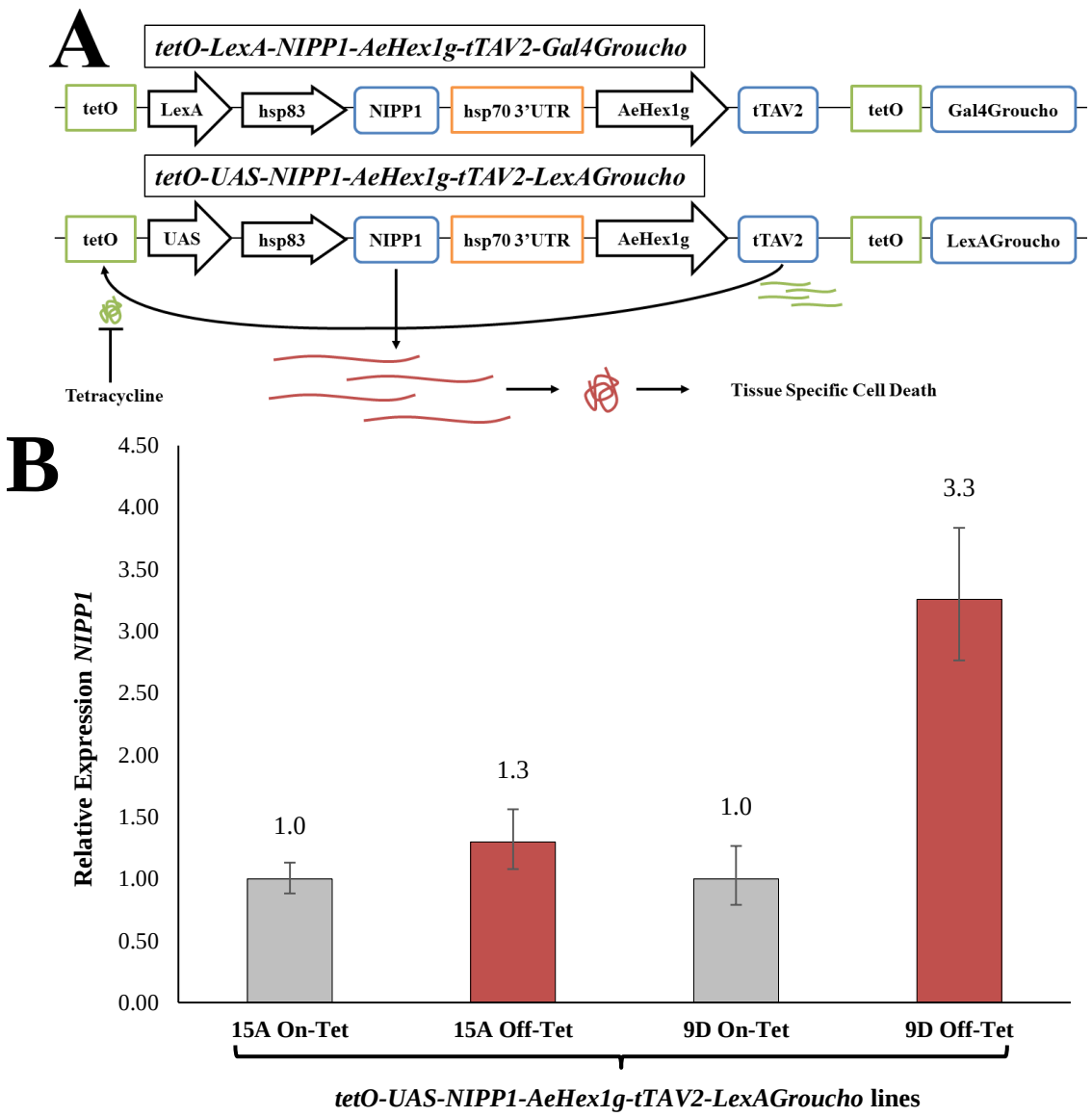


Figure 2.14: *NIPP1* upregulation does not occur as expected in *tetO-UAS-NIPP1-AeHex1g-tTAV2-LexAGroucho* lines Off-Tet:

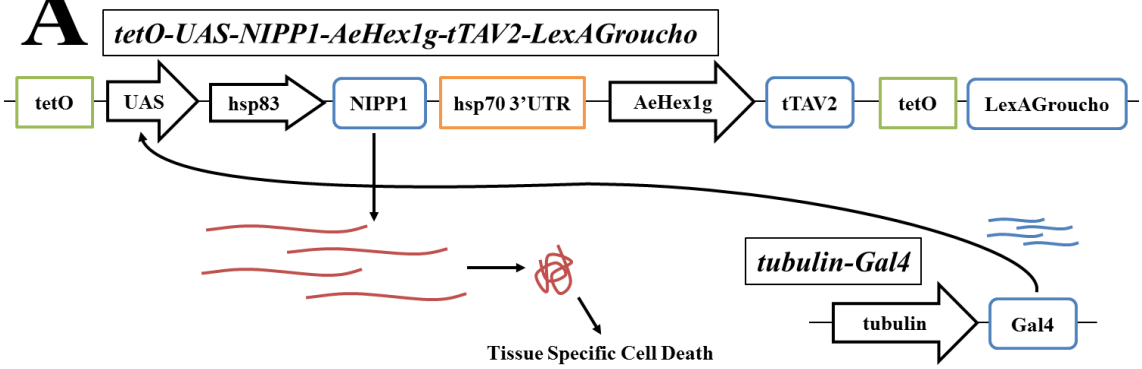
(A) Tetracycline binds to the tTAV2 activator protein and prevents binding to *tetO* sequences and subsequent transcript upregulation. *NIPP1* expression in *tetO-UAS-NIPP1-AeHex1g-tTAV2-LexAGroucho* is designed to be driven by spatially defined *tTAV2* expression resulting in ubiquitous cell death in the absence of tetracycline as it no longer interferes with tTAV2 binding of *tetO* sequences. (**B)** Relative expression of *NIPP1* is shown on the y-axis for different killer line insertions on the x-axis. Expression level values are normalised for the amount of cDNA input by comparing to *RNA polymerase II* control reactions. Error bars display the standard deviation for triplicated technical replicates. Relative expression is expressed with respect to the level of *NIPP1* in the respective *tetO-UAS-NIPP1-AeHex1g-tTAV2-LexAGroucho* line On-Tet. cDNA samples were prepared from 2 *D. melanogaster* early pupae per sample, i.e. 2 biological replicates. *NIPP1* expression was only upregulated 1.3 and 3.3 fold in each line respectively. cDNA samples are the same as those in Figure 2.16 and Figure 2.17.

2.3.5 NIPP1 overexpression phenotype can be achieved in tetO-UAS-NIPP1 lines

Finally, to confirm that *NIPP1* overexpression is possible in this system, and the expected phenotype observed, *tetO-UAS-NIPP1* lines were crossed with a *tubulin-Gal4* line. The known cell lethality phenotype from *NIPP1* expression was driven at 27ºC by a *tubulin-Gal4* construct using the *UAS* promoter sequences (Figure 2.15). Almost no double mutants were recovered. Survivors were analysed for *NIPP1* upregulation and it was only 10 fold.

The results show that *NIPP1* overexpression with GAL4 driven *UAS* sequences is possible in this context. This indicates that *Hsp83* is at least adequately functional in the context of *tubulin-Gal4* driving *UAS* expression. Crossing to a *GMR-Gal4* driver line would have been ideal as the disruption of eye-tissue would be a better visual phenotype to screen for. This cross was indeed attempted but a technical problem with the driver line itself (a lack of *Gal4* upregulation) prompted the use of *tubulin-Gal4* instead.

A



B

Crosses			Progeny Count			
			WT	*tubulin-Gal4*	*tetO-UAS-NIPP1-AeHex1g-tTAV2-tetO-LexAGroucho*	Double Mutant
*If/CyO ; tetO-UAS-NIPP1-AeHex1g-tTAV2-tetO-LexAGroucho-**9D**/tetO-UAS-NIPP1-AeHex1g-tTAV2-tetO-LexAGroucho-**9D*** ♀	X	*tub-Gal4/TM3* ♂	-	-	**14** (7♂ + 7♀)	**2** (1♂ + 1♀)
*If/CyO ; tetO-UAS-NIPP1-AeHex1g-tTAV2-tetO-LexAGroucho-**21B**/tetO-UAS-NIPP1-AeHex1g-tTAV2-tetO-LexAGroucho-**21B*** ♀	X	*tub-Gal4/TM3* ♂	-	-	**20** (10♂ + 10♀)	**1**♀

C

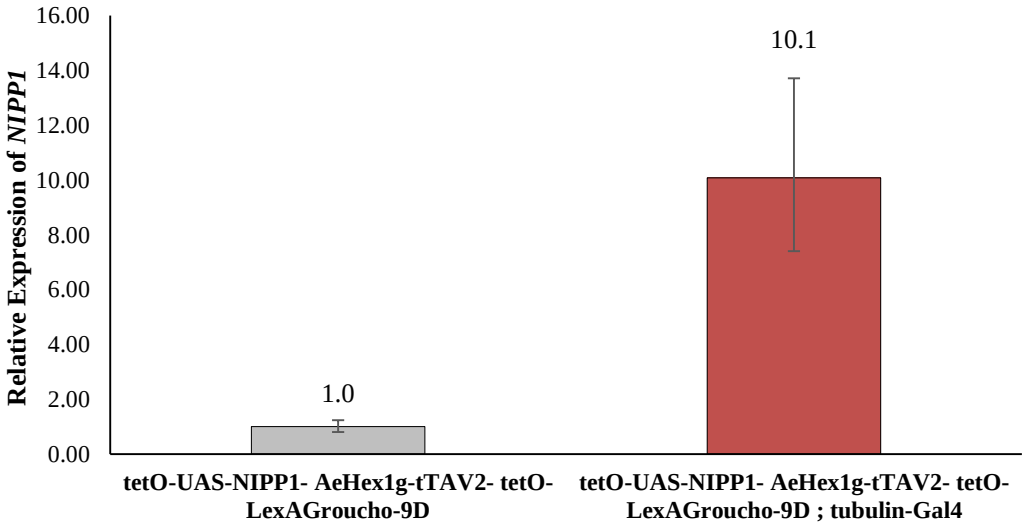


Figure 2.15: *NIPP1* expression was driven effectively at 27ºC by a tubulin-Gal4 construct via *UAS* promoter sequences:

(A) *NIPP1* expression in is to be driven by spatially defined *Gal4* expression resulting in ubiquitous cell death. **(B)** Double mutant individuals have mostly died from *tetO-UAS-NIPP1-AeHex1g-tTAV2-tetO-LexAGroucho* and *tubulin-Gal4* crosses. Survivors are rare, a circle marks the escapers which were analysed for *NIPP1* expression. **(C)** *NIPP1* expression levels in fold changes, on the y axis, are shown for different cDNA samples on the x axis. Expression level values are normalised for the amount of cDNA input by comparing to RNA polymerase II control reactions. Error bars display the standard deviation for triplicated technical replicates. Expression fold changes are expressed with respect to the level of *NIPP1* in the respective *tetO-UAS-NIPP1-AeHex1g-tTAV2-tetO-LexAGroucho* line before the cross, which is given a value of 1. Expression values are indicated above the error bars. cDNA samples were prepared from 1 male plus 1 female *Drosophila* adult per sample, i.e. 2 biological replicates. *NIPP1* expression was only upregulated 10 fold in the double mutants which escaped the lethal phenotype.

2.3.6 tetO-UAS-NIPP1-AeHex1g-tTAV2-tetO-LexAGroucho transgenic lines do not survive in long term Off-Tet rearing

Prototype *tetO-UAS-NIPP1-AeHex1g-tTAV2-tetO-LexAGroucho* adults fail to generate viable adults when kept on a long-term Off-Tet diet. Off-Tet will be used for individuals reared or fed on a diet lacking tetracycline, whilst On-Tet will be used for individuals on a diet containing tetracycline (Section 2.2.3). Transgenic lines, homozygous for independent insertions, were mostly unable to result in transgenic adults when moved to an Off-Tet diet (Table 2.3). As an inbreeding homozygous transgenic line, all of the progeny should be transgenic. Out of the 6 independent insertion lines; 5 of them were unable to produce progeny which would survive past late pupal stages, and one of them was able to produce a subsequent generation of adults which were in turn unable to generate a further generation on the same Off-Tet diet. Some of the lethality seen at pupal stages was associated with various pupal case abnormalities. In addition, a reduced level of third instar larvae was seen which indicated lethality effects were not restricted to pupal stages. The only component of the construct predicted to cause lethality in a tetracycline repressible manner is *NIPP1*. However, I had already determined that there is only limited *NIPP1* upregulation in the individual component analysis. This led me to investigate other factors to explain the lethality in this context.

Table 2.3: *tetO-UAS-NIPP1-AeHex1g-tTAV2-tetO-LexAGroucho adults fail to generate viable progeny for stable stocks when Off-Tet:*

Most insertion lines (5 out of 6) failed to produce viable adults when healthy adults of each transgenic line (reared On-Tet) were moved to an Off-Tet diet. Insertion line *21B* managed to produce a first generation progeny which was able to develop into adults. These adults were kept Off-Tet and crossed with each other, not being able to even result in viable larvae.

***tetO-UAS-NIPP1-AeHex1g-tTAV2-tetO-LexAGroucho* lines**	**Transgenic larvae amongst the first generation Off-Tet**	**Transgenic adults amongst the first generation Off-Tet**	**Transgenic larvae amongst the second generation Off-Tet**
9C	Yes	No	-
9D	Yes	No	-
12A	Yes	No	-
15A	Yes	No	-
15E	Yes	No	-
21B	Yes	Yes	No

2.3.7 tetO sequences can enhance expression at distances of at least 1900bp upstream

A lack of an explanation for the lethality phenotype observed in *tetO-UAS-NIPP1-AeHex1g-tTAV2-tetO-LexAGroucho* transgenic lines after rearing in a diet without tetracycline suggested other components than *NIPP1* might have been the cause for the lethality. Since tTAV2 is known to be toxic to insects at high enough concentrations (Gong *et al.* 2005) its expression levels were assayed in the same early pupae cDNA samples from Section 2.3.2, Figure 2.14. *tTAV2* expression levels are upregulated very strongly in early pupae reared Off-Tet (Figure 2.16). This suggests that *tetO* sequences are enhancing the expression of *tTAV2* in a positive feedback loop. This implies that the enhancement interaction can occur even when the *tetO* operator and responsive promoter are at least 1900bp apart since these are the closest *tetO* sequences in the construct to *tTAV2*. The *tetO* downstream is the closest enhancer to *tTAV2* in the final design construct and is the best candidate *tetO* for the feedback loop (Figure 2.16B). A second possibility would be that the *tetO* enhancer upstream, although 4800bp

apart, would be engaging in this feedback loop as well or instead, with the same consequences. This positive feedback loop could explain the observed lethality in this system. Evidence from this was also observed in *OX513A* in *Aedes aegypti* as marker expression would increase when reared Off-Tet. The *tetO* sequence was in this case almost 2kb away from the *Act5C-DsRed* putative transcription start site.

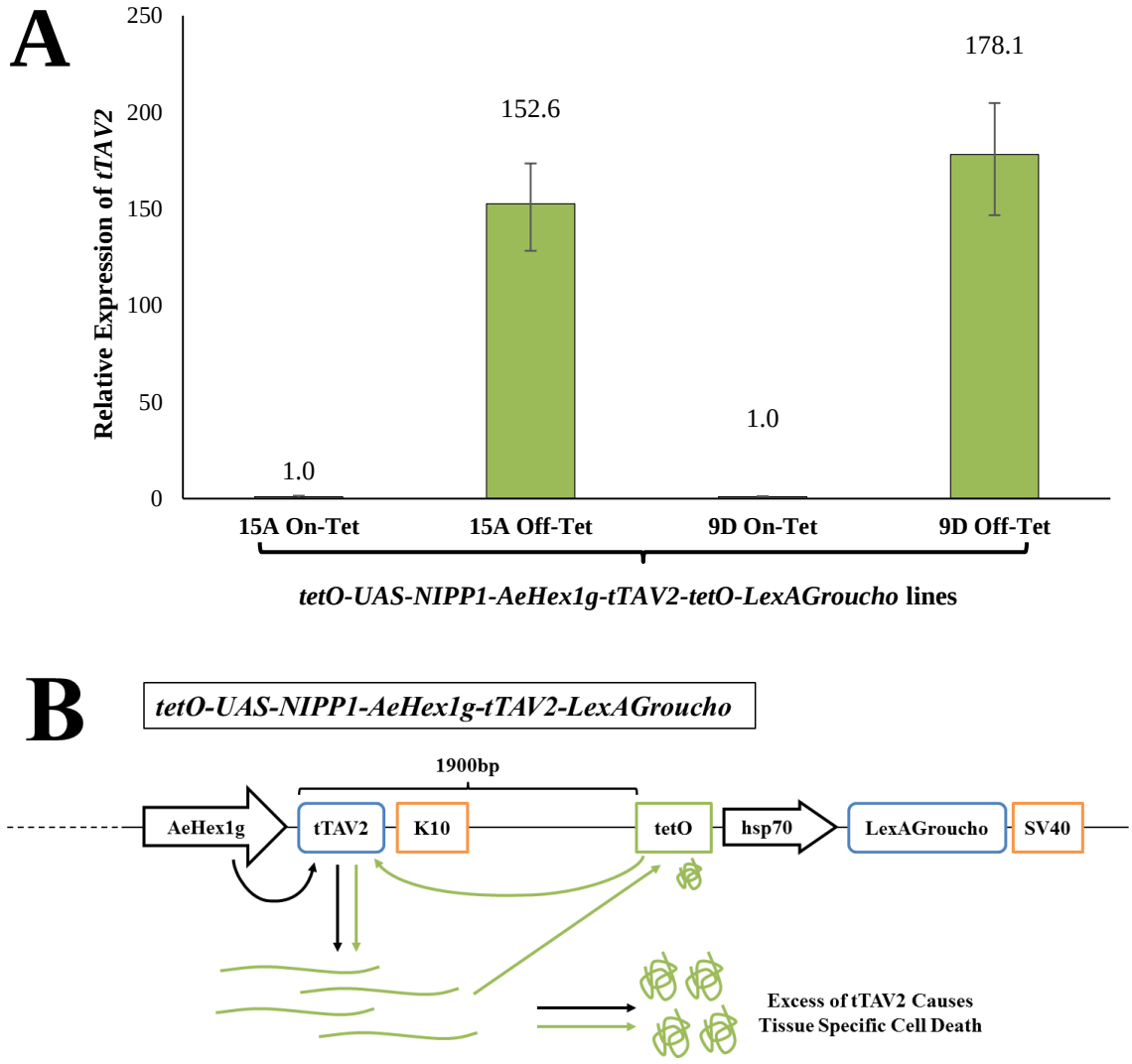


Figure 2.16: *tTAV2* upregulation is more than 150 fold in *tetO-UAS-NIPP1-AeHex1g-tTAV2-tetO-LexAGroucho* lines when Off-Tet:
Relative expression of *tTAV2* is shown on the y-axis for different killer line insertions on the x-axis. Expression level values are normalised for the amount of cDNA input by comparing to *RNA polymerase II* control reactions. Error bars display the standard deviation for triplicated technical replicates. Expression fold changes are expressed with respect to the level of *tTAV2* in the respective *tetO-UAS-NIPP1-AeHex1g-tTAV2-tetO-LexAGroucho* lines fed on tetracycline, which are given a value of 1. Expression values are indicated above the error bars. cDNA samples were prepared from 2 *D. melanogaster* early pupae per sample, i.e. 2 biological replicates. *tTAV2* expression was upregulated 152 and 178 fold in each line respectively. cDNA samples are the same as those analysed in Figure 2.14 and Figure 2.17.

2.3.8 LexAGroucho and Gal4Groucho upregulation is lethal

LexAGroucho repressor expression in *tetO-UAS-NIPP1-AeHex1g-tTAV2-tetO-LexAGroucho* lines was also assayed in the same early pupae cDNA samples from Section 2.3.2, Figure 2.14. *LexAGroucho* expression levels are upregulated

strongly in early pupae (46 to 90 fold) when larvae have fed Off-Tet (Figure 2.17). This together with earlier data suggests that the *tetO-NIPP1* element is defective specifically rather than other elements of the system. This repressor was expected to be harmless, but since it is the only other upregulated component in these lines apart from *tTAV2* in early pupae destined to die before reaching adulthood, one hypothesis was that its overexpression was also lethal, either alone or in combination with *tTAV2*.

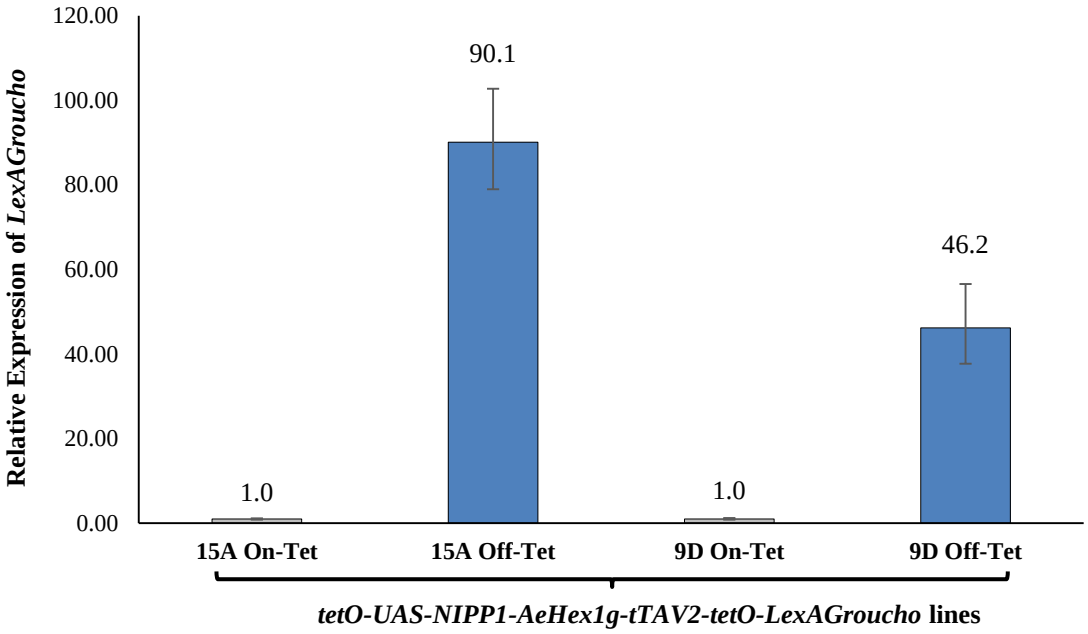


Figure 2.17: *LexAGroucho* upregulation is between 46 and 90 fold in *tetO-UAS-NIPP1-AeHex1g-tTAV2-tetO-LexAGroucho* lines when Off-Tet:
Relative *LexAGroucho* expression on the y axis for different cDNA samples on the x axis. Expression levels are normalised for the amount of cDNA input by comparing to *RNA polymerase II* controls. Error bars display the standard deviation for triplicated technical replicates. Relative expression was compared to respective *tetO-UAS-NIPP1-AeHex1g-tTAV2-tetO-LexAGroucho* lines fed On-Tet, which are given a value of 1. Expression values are indicated above the error bars. cDNA samples were prepared from 2 *D. melanogaster* early pupae per sample, i.e. 2 biological replicates. *LexAGroucho* expression was upregulated 90 and 46 fold for each line respectively. cDNA samples are the same as those analysed in Figure 2.14 and Figure 2.15.

To address whether over expression of groucho-containing fusion proteins contributes to the lethality *tetO-LexAGroucho* and *tetO-Gal4Groucho* individual constructs were subcloned in *pCaSpeR4* vectors and transgenic flies were generated as described (Section 2.2.5). Crossing of these components to individual *GMR-tTAV2* and *AeHex1g-tTAV2* drivers led to pharate adult lethality even when upregulation of *LexAGroucho* and *Gal4Groucho* is restricted to the eye tissue (Table 2.4). Thus the lethality can be caused groucho over-expression as well as, or instead of *tTAV2* positive feedback.

Lethality was observed at pharate adult stages. This was expected for the *GMR-tTAV2* crosses, since severe eye-disruption phenotypes are known to be lethal as they can lead to eclosion failure. The insect's reduced head size is thought to prevent pharate adults from breaching the pupal case (Zhu *et al.* 2017). Indeed, loss of eye tissue was observed amongst dead pharate adults from *GMR-tTAV2* crosses.

Table 2.4: Eye-specific or broad body *LexAGroucho/Gal4Groucho* over-expression is lethal:

Double mutant individuals mostly died from *tetO-LexAGroucho/tetO-Gal4Groucho* and *AeHex1g-tTAV2* and *GMR-tTAV2* crosses. A circle marks the escapers which were analysed for *LexAGroucho* expression. Survivors were found in only one of the crosses, likely due to the fact of weaker *tTAV2* expression of *AeHex1g-tTAV2-1A* line, and sub-lethal induction levels.

Crosses			**Progeny Count**			
			WT	***tTAV2***	***LexAGroucho* or *Gal4Groucho***	**Double Mutant**
*tetO-LexAGroucho-**5A**/+* ♀	X	*GMR-tTAV2-**29C**/+* ♂	**2**♂	**3**♀	**18**♂	**0**
*tetO-LexAGroucho-**5A**/+* ♀	X	*GMR-tTAV2-**40C**/CyO* ♂	**5**♂	**2**♀	**11** (4♂ + 7♀)	**0**
*tetO-LexAGroucho-**5A**/+* ♀	X	*AeHex1g-tTAV2-**9B**/Y* ♂	**15**♂	**12**♀	**12**♂	**0**
*tetO-LexAGroucho-**5A**/ tetO-LexAGroucho-**5A*** ♀	X	*AeHex1g-tTAV2-**1A**/CyO* ♂	**-**	**-**	**4** (3♂ + 1♀)	**5** (2♂ + 3♀)
*GMR-tTAV2-**36A**/CyO* ♀	X	*tetO-Gal4Groucho-**1A**/+* ♂	**2**♂	**2** (2♂ + 2♀)	**5** (3♂ + 2♀)	**0**
*AeHex1g-tTAV2-**9B**/ AeHex1g-tTAV2-**9B*** ♀	X	*tetO-Gal4Groucho-**1A**/+* ♂	**-**	**11** (6♂ + 5♀)	**-**	**0**

Finally, one of the *tetO-LexAGroucho* and *AeHex1g-tTAV2* crosses resulted in double transgene-containing survivors (black circle in Table 2.4). RT-qPCR revealed that an upregulation of 16 fold does not result in a lethal phenotype (Figure 2.18) suggesting that *D. melanogaster* is tolerant to some induction of *LexAGroucho.* This is somewhat expected as *groucho* is an endogenously expressed gene during *D. melanogaster* development, albeit with different tissue

and time specificity. The over-expression level is only a sub-lethal value, as the individuals that died are expected to have had much higher induction levels. Hence, Figure 2.18 underestimates the 'real' induction level from *tetO-tTAV2* seen for most of the crosses. Since different *tetO-LexAGroucho*, and *tTAV2* driver lines were used, it is very possible that the insertion dependent strength of each differs enough that when two weakly expressing lines are crossed some individuals may survive as double mutants. Therefore the measured upregulation in these experiments can only be considered as a sub-lethal value.

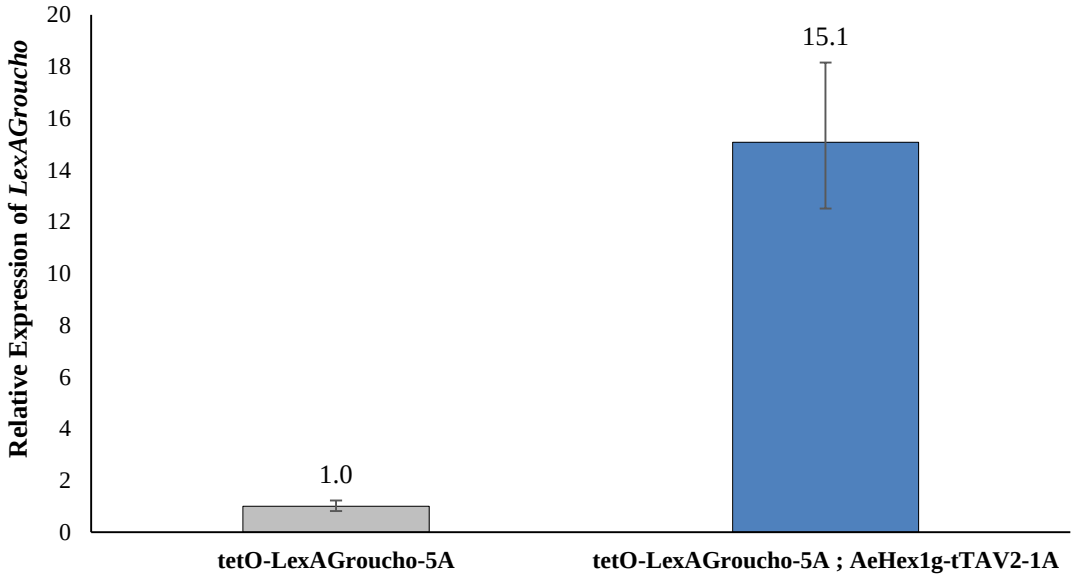


Figure 2.18: *LexAGroucho* 15 fold upregulation when driven by *AeHex1g-tTAV2* is not lethal in *D. melanogaster*:
Table 2.4 double mutant survivors, circled in black, were analysed for *LexAGroucho* expression. Relative *LexAGroucho* expression on the y axis is shown for different cDNA samples on the x axis. Expression levels are normalised for the amount of cDNA input by comparing to *RNA polymerase II* controls. Error bars display the standard deviation for triplicated technical replicates. Expression fold changes are expressed with respect to the level of *LexAGroucho* in the respective *tetO-LexAGroucho* line before the cross, which is given a value of 1. Expression values are indicated above the error bars. cDNA samples were prepared from 1 female *D. melanogaster* adult per sample, i.e. 1 biological replicate. *LexAGroucho* expression was only upregulated 15 fold in the double mutant which escaped the lethal phenotype.

2.3.9 A design for underdominance using the results as feedback

The experiments presented in this Chapter have evaluated the potential for a two construct killer plus mutual repressor system for engineered underdominance in *Drosophila melanogaster*. From a practical standpoint the decision was made not to pursue such a system in pest insects in light of these results. However there is scope for further refinement, which could lead to a functional system. Figure 2.19 illustrates one of the options for a new refined design. The Groucho fusion proteins could be used as two independent killer genes which could be independently silenced through different RNAi sequences targeting *LexA* and *Gal4* sequences respectively. Although, it would be theoretically feasible, many questions are raised.

The levels of RNAi vs Groucho fusion protein expression would require tight tuning to reduce fitness costs. Controlling this when each construct is on a different locus of the genome may be somewhat challenging. Site specific integration in the genome in a similar chromatin environment would allow for better comparison of constructs and for better tuning of components. Perhaps expressing the RNAi in the same tissue and at sufficient levels for repression of the protein could be the biggest challenge. To tackle this issue, the RNAi element could be included as part of the same transcript as the Groucho mRNA. Although an RNAi hairpin (of broader targeting than miRNA) could be included within an intron, longer hairpins may not be correctly processed in this arrangement so miRNAs could be used instead. RNAi could be achieved by including miRNA

sequences within an intron as speculated in Figure 2.19, or by expressing them under the control of a different promoter. This raises a new concern around RNAi dependent degradation. The system relies on RNAi degradation of the transcript suppressing Groucho but not preventing RNAi function, should the miRNAs prevent transcription of the whole transcript then one of the alleles would stochastically end up expressed highly and hence underdominance would not obtained.

Such a system, even when working, would most likely have associated fitness costs due to over usage of the RNAi pathway, incomplete suppression of *Groucho* expression, or other factors. Measuring the fitness costs of the system in comparison to WT and relying on modelling to estimate if the system would drive with such costs would be a good starting point before population cage experiments or release trials.

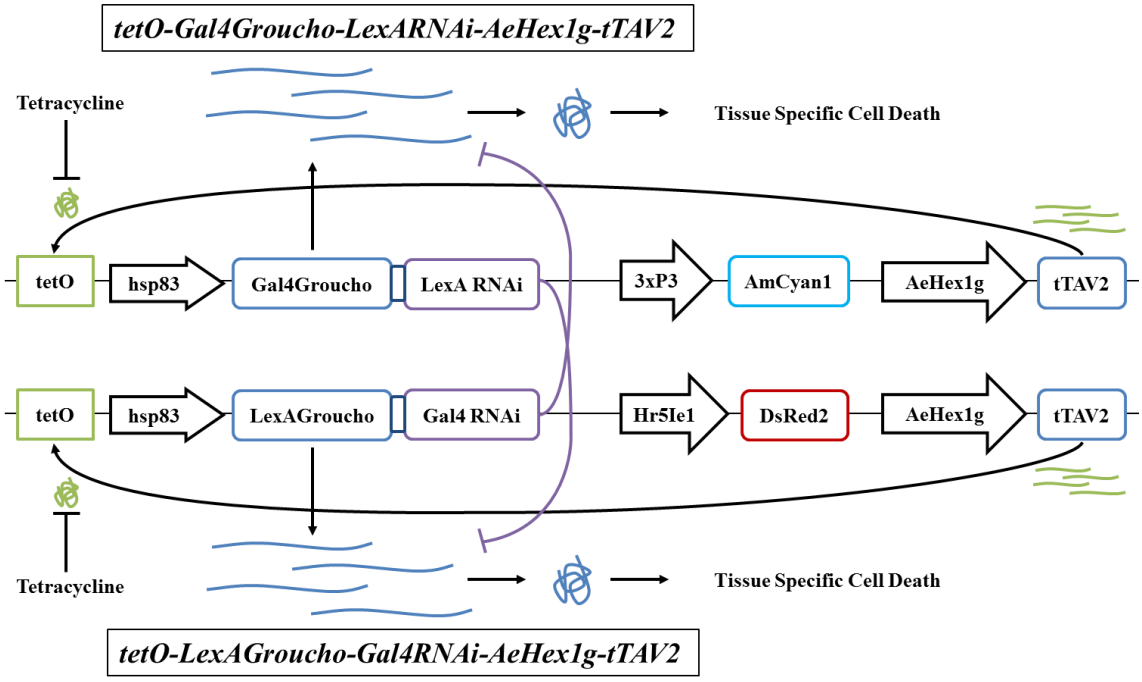


Figure 2.19: A new design for two-loci underdominance:
Groucho fusion proteins are to result in tissue specific cell death when driven by *tTAV2* expression. *LexA* and *Gal4* regions serve as distinct RNAi targets. Groucho fusions and miRNA are kept part of the same transcript but the miRNA is within an intron which will be spliced out and the miRNA processed. miRNAs (RNAi activity) mutually repress each other's killer protein transcript. Finally, *tetO* and *tTAV2* sequences are kept at the maximum distance possible, assuming enough distance will prevent positive *tetO/tTAV2* feedback.

2.4 Conclusions

2.4.1 The relevance of an individual component approach

This study suggests an individual component analysis may be advisable when dealing with highly variable systems. The starting prototypes had a rational design; however, they did not work as expected. In hindsight, this is not that surprising given the complexity and context dependent nature of biological systems. As described in more detail below, previous knowledge obtained from functionality in other systems does not necessarily provide predictability in a complex biological system, and may even be misleading. This example shows how multiple malfunctions within a set of components can provide the expected working phenotype (Section 2.3.6) and yet dissecting the functional contribution, within this complex context, of the different components was not feasible. To be able to explain how the particular malfunctions gave rise to the overall expected phenotype, individual components had to be analysed separately. This supports the need for a careful individual analysis of components in their specific context for a more robust approach.

On the other hand, this does not guarantee success either since individual functions may not ‘add up’ as expected when combined. Hence, making a rationally designed complex configuration could potentially save a lot of time in terms of individual component analysis. In this particular example, making full constructs directly without thorough testing of the individual components was a

considerable risk, but it might have worked. This could have saved time testing individual components.

Ultimately, a balance between these two main concerns will provide the best starting point in designing complex biological systems efficiently.

2.4.2 Perspectives on synthetic biology

The unexpected functions of different components in this project certainly call for reconsideration on how to approach synthetic biology as defined by some key reviews introducing the field (Andrianantoandro *et al.* 2006).

For instance, *hsp83* minimal promoters have been extensively used before and so have *UAS* and *LexA* sequences (Gnerer *et al.* 2015). Nevertheless, *NIPP1* upregulation by tTAV2 was very low when *UAS* or *LexA* were present between the *tetO* and *hsp83*, pointing at faulty interactions between these components or *tetO* itself. In addition, a strong positive feedback loop between *tTAV2* and *tetO* though 1900bp apart, is evidence of the importance of context. To be able to tackle this problem enhancer-blocking insulators could be used such as dCTCF, a distant *D. melanogaster* orthologue of the well-characterised mammalian CTCF insulator (Raab and Kamakaka 2010). Finally, Gal4Groucho and LexAGroucho fusion proteins when driven by *tTAV2* in *D. melanogaster* appear to be lethal when under the control of the *tetO-tTAV2* system unlike previous uses of

Gal4Groucho fusion proteins in the literature (Fisher and Caudy 1998; Fisher *et al.* 1996).

Synthetic biology is typically defined as the assembly of modular components in a rational attempt to construct a complex system with a desired new function (Andrianantoandro *et al.* 2006). Nothing much appears to be strictly modular in biology. Although modularity may be found in some examples, i.e. certain promoter sequences such as 3xP3 reliably drive tissue-specific mRNA expression of any gene situated downstream; a lot of biology is context dependent. Even in the best examples, biological systems are subject to relatively high variability and noise, *3xP3* driven expression, amongst other cases may vary, in level or location, between individuals, different developmental stages or between different genomic insertions. This variability and inherent noise may appear significantly strong when compared to other physics based systems which traditional engineering deals with. Attempting to make biology fit traditional engineering criteria, as encouraged in some of the literature (Wang *et al.* 2016), is a difficult task with the current limited knowledge. This is not to say that the construction of complex biological components of a desired function is not possible. In the future, after an extensive analysis of different biological parts there may well be a more functional library of quasi-modular components. Online databases for these kind of libraries are being proposed and some already exist like the Registry of Standard Biological Parts (Kwok 2010). Although their use may be limited at the moment, they may become crucial in the future.

Furthermore, current synthetic biology mostly focuses on microbial systems which are simpler in many ways, and better understood than insects, with the result that the noise or biological variation observed is lower. There is also a better control of environmental conditions with bacteria than with insects, such as growth in fermenters with tightly controlled parameters (Fang *et al.* 2017) which may also help reduce another source of variation. Given that synthetic biology in insects is less developed and fundamentally harder, the results from this work may provide a critical vision of synthetic biology principles which may not apply as strictly on microbial work.

With the current level of unknown variables, mainly due to the inherent nature of biology, a refined approach to synthetic biology or at least a more accurate definition may be informative and helpful. Context in biology is often very powerful, as exemplified in this work, it can even ‘override’ and ‘reverse’ the functionality of a particular relatively well understood component. This questions the concept of modularity and abstraction in biology. Noise and context are such powerful factors in biology that a user should always bear these in mind when trying to use modular components.

Of course, in every area of engineering the power of rational prediction is limited, or testing experiments would be unnecessary and merely confirmatory. If a scale of predictability in a given science or field could be drawn, biology would be relatively unpredictable if compared to physics-based engineering disciplines.

In other words, for a successful approach given the nature of biology, synthetic biology should be much more readily testing rational predictions through trial and error than traditional engineering disciplines. An argument could be made by suggesting that no achievable critical mass of knowledge in biology will bring these two different disciplines to the same point in the scale as the distinct levels of system complexity may make this impossible. In line with this reasoning, (Vilanova *et al.* 2015) questions the ambitious endeavour of component standardisation suggesting there is a surprisingly limited reuse of biological parts from the Registry of Standard Biological parts (Vilanova and Porcar 2014). However, this should not undermine the usefulness of the endeavour to understand biology through building it. It may just be a question of keeping scientific scepticism high when dealing with modular components whilst still aiming to efficiently construct complex systems through rationally informed approaches.

Chapter 3 - **Homology Directed Repair into the M-locus of *Aedes aegypti* and its Significance for Gene Drives**

3.1 Introduction

3.1.1 Sex determination in mosquitoes

In Culicine mosquitoes, which includes both *Aedes* and *Culex* genera, a dominant male-determining locus (M-locus) found on a homomorphic sex-determining chromosome, is the trigger for male development in mosquitoes (McClelland 1962; Newton *et al.* 1978). Conversely, *Anopheles* mosquitoes have fully differentiated heteromorphic sex chromosomes with the male-determining locus located on the non-recombining Y chromosome (Marín and Baker 1998; Hall *et al.* 2014).

Heteromorphic sex chromosomes have evolved multiple times independently in both plants and animals (Charlesworth 1996; Rice 1996). In principle, these are hypothesised to arise from a pair of autosomes that at some point acquire a sex-determining region or locus. It is thought that selection pressures would favour the sorting of sex-specific beneficial alleles, into linked or unlinked positions from the sex-determining region, depending on which sex it determines. Such

alleles would have antagonistic roles depending on the sex, i.e. provide a benefit in one whilst being detrimental in the other. In theory these selection pressures would also in turn favour the suppression of recombination near or at the sex-determining locus (Charlesworth *et al.* 2005).

Homomorphic chromosomes are those where this recombining region is restricted to a portion of the sex-determining chromosome whilst heteromorphic chromosomes are those were the non-recombining region occupies most of the sex chromosome (Toups and Hahn 2010). The non-recombining region tends to expand from the sex-determining locus to almost the entirety of the chromosomes as is evidenced in many species of distant lineages such as dioecious plants, chickens and humans (Nicolas *et al.* 2005; Handley *et al.* 2004; Lahn and Page 1999). In fact, comparative genetics between *Drosophila melanogaster*, *Aedes aegypti*, and *Anopheles gambiae* genomes suggest sex determining chromosomes were homomorphic in the common ancestor of both mosquito species (Toups and Hahn 2010). *Anopheles gambiae* would hence have evolved hetermorphic sex chromosomes once the lineages had diverged (Figure 3.1).

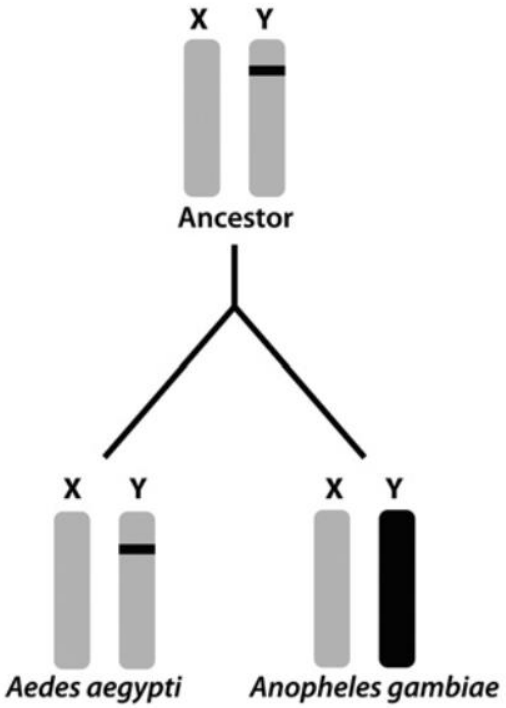


Figure 3.1: Sex-chromosome evolution in *Anopheles gambiae* and *Aedes aegypti*:
Taken from (Toups and Hahn 2010). The ancestor of both *Aedes aegypti* and *Anopheles gambiae* is thought to have had homomorphic chromosomes as *Aedes aegypti* does currently and would suggest that *Anopheles* lineages saw an expansion of the non-recombining region to occupy the whole Y chromosome. This expansion must have been rapid since all extant *Anopheles* have heteromorphic chromosomes (Hall *et al.* 2014).

Although modelling also suggests that homomorphic sex-determining chromosomes will eventually become heteromorphic (Charlesworth *et al.* 2005), there may be occasions were homomorphic chromosomes are preserved by selection pressures. An example of this may be *Aedes aegypti* given its presumed lengthy evolutionary history of having homomorphic chromosomes. It has been recently shown that an antagonistic *Aedes aegypti* gene, *myo-sex* (beneficial in males, detrimental in females), linked to the M-locus however it can recombine and be inherited by females occasionally. This is surprising, since such a male-specific gene should have faced evolutionary pressures to join the M-locus non-recombining region. However, *myo-sex* was found to be expressed much lower in

females than in males even when inheriting the gene through rare recombination events (Hall *et al.* 2014). Hence, it is possible that the homomorphic chromosomes in *Aedes aegypti* have remained as such for so long due to differential sexual expression of antagonistic genes such as *myo-sex*. This could in turn eliminate or reduce selection pressures encouraging the expansion of the non-recombining region. There is however, no evidence for dosage compensation in *Aedes aegypti*, nor is thought to be required given the viability of *nix* knockouts (Hall *et al.* 2015). Another possibility for the observed homomorphic preservation of sex determination is the fact that in *Aedes aegypti*, and also in *Culex pipiens* which shares this homomorphic preservation, the M-locus is located near tandem repeated ribosomal genes which could be promoting recombination (Timoshevskiy *et al.* 2013). Hence, it could be that recombination hotspots serve as an extra evolutionary threshold against the expansion of the non-recombination region (Hall *et al.* 2014).

3.1.2 M-locus in Aedes aegypti

The M-locus in *Aedes aegypti*, as described above, is found in a homomorphic sex chromosome. Out the three pairs of chromosomes in *Aedes aegypti*, chromosome 1 is the shortest, chromosome 3 of intermediate length, and chromosome 2 the longest (Sharakhova *et al.* 2011; Rai 1963; Timoshevskiy *et al.* 2013) The M-locus is found in chromosome 1, specifically in band 1q21 (McClelland 1962). Although chromosome 1 pairs are homomorphic, they are

referred to for simplicity as the M-chromosome (M-locus bearing) and m-chromosome (M-locus lacking) in the literature (Motara and Rai 1978).

M-locus sequences were omitted from the *Aedes aegypti* genome sequencing assembly described (Nene *et al.* 2007). This was not so surprising given the repetitive nature of Y like sequences as evidenced by the difficult sequence assembly of other sequenced heteromorphic chromosomes from *Drosophila* and *Anopheles* (Carvalho 2002; Krzywinski *et al.* 2004). Correspondingly, when a putative male determining gene in *Aedes aegypti* was finally cloned in *Aedes aegypti* its sequence was also found missing in the assembly (Hall *et al.* 2015). Information on the *Nix* mRNA sequence allowed the finding the *Nix*-containing BAC clone in a LVP genome library and sequencing a 200kb region of the M-locus (Turner *et al.* 2017), including the *Nix* gene itself. Apart from providing insights into the structure of the *Nix* gene (described below), the sequencing of the region showed the highly repetitive nature of the non-recombining M-locus in *Aedes aegypti* which is a characteristic feature of a sex chromosome.

3.1.3 The role of Nix in Aedes aegypti sex determination

The first strong candidate for a primary male determining factor (or M factor) in insects was found in *Aedes aegypti* and it was the *Nix* gene (Hall *et al.* 2015). The difficulty in finding it was as mentioned above due to the characteristic repeat-rich regions of Y chromosomes, or Y like regions (Charlesworth and Mank 2010). The *Nix* gene was found from RNA-seq data as a distant orthologue

of *transformer-2* in *D. melanogaster* which has the role of splicing *doublesex* and *fruitless* genes which are two key regulators of sexual differentiation in *D. melanogaster* (Salz and Erickson 2010). However, sexual determination is not driven by the Y chromosome but by X chromosome dosage in *D. melanogaster*.

Nix was found to be specifically present in male DNA and knock-out with CRISPR-Cas9 demonstrated its requirement for male development. Feminisation of G0s just upon somatic knock-out of the gene was observed in a large proportion of injected males (55 out of 79). Although the degrees of feminisation were variable, as expected of mosaic knock-outs, feminisation was seen in all dimorphic phenotypes; feminised external genitalia and antennae (Hall *et al.* 2015). *Nix* was also determined as sufficient for male development as ectopic expression of *Nix* in females resulted in almost fully formed male external genitalia and female specific antennal structure. Although some masculinised genitalia appeared to have defects, they looked male overall (Hall *et al.* 2015). In addition, ectopic expression of *Nix* led to the internal development of male genitalia, including testes, accessory glands, and vas deferens. Hence *Nix* is necessary and sufficient for the initiation of male development in *Aedes aegypti*.

The *Nix* cDNA sequence is 985bp and it encodes a 288 amino acid polypeptide. The *Nix* gene is formed of two exons and has a long intron of 99kb, one of the largest of the *Aedes aegypti* genome (Turner *et al.* 2017). The *Nix* region is full of repetitive sequences, especially retrotransposons, which may be the reason for

the lengthy expansion of the *Nix* region. Genes expressed early in embryonic development such as *Nix* typically have very short introns or no introns which would suggest there is an evolutionary pressure for a smaller intron to do with short cell cycles (Biedler *et al.* 2012). However, the non-recombining region of the M-locus allows the accumulation of repeats as it is known to happen in other sex chromosomes. Hence, the expansion of the *Nix* gene intron will carry on unless it is limited by the need for a shorter intron in an early embryonic development gene (Biedler *et al.* 2012).

3.1.4 Sex-specific gene drives

With the success of sterile insect technique (SIT) (Alphey *et al.* 2010) and release of dominant lethal (RIDL) strategies, the discovery of the male determining factor in *Aedes aegypti* represents a milestone in gene drive research. Ultimately, understanding the genetic programs underlying sex determination in mosquitoes will allow the development of desirable gene drive mechanisms (Adelman and Tu 2016). For instance, CRISPR-Cas9 systems carrying male-determining factors such as *Nix* would drive maleness in a population. Whether fertile or not, this represents a good trait to spread in a population; maleness could be seen as the ultimate disease refractory phenotype for population suppression. Conversely, CRISPR-Cas9 components could be linked to the M-locus for male specific expression, which would be useful in particular drives as discussed in Chapter 4.

3.1.5 Ommochrome synthesis in insects, and *kmo* in *Aedes aegypti*

Much of the knowledge on eye pigment development in insects comes from studies in *Drosophila melanogaster*. The first mutant isolated in *Drosophila melanogaster* was in the *white* gene (Morgan 1910). Mutations in *white* are recessive and result in a white eye phenotype and its sex-linked nature, *white* is on the X chromosome, facilitated its discovery. A single mutant copy would express phenotypically in males but not females so it is not surprising that the isolated mutant was a male (Morgan 1910).

The *white* gene encodes a monomer of a heterodimer ATP-binding cassette (ABC) transporter which localises at the surface of pigment granules in ommatidia (Mackenzie *et al.* 2000). The other monomer is encoded by *scarlet*. The *white* gene's main role is to deliver ommochrome precursors into pigment granules such as tryptophan and kynurenine (Sullivan *et al.* 1974). Ommochromes are compounds derived from the amino acid tryptophan and are widely used as eye pigments among insects (Linzen 1974). Ommochrome pathway terminates with xanthommatin synthesis, the brown pigment in *D. melanogaster* (Phillips *et al.* 1973). A model of ommochrome synthesis in *D. melanogaster* is shown in Figure 3.2 . With this model as a framework, ommochrome biosynthesis has since been studied in mosquitoes to find considerable similarities (Han *et al.* 2007).

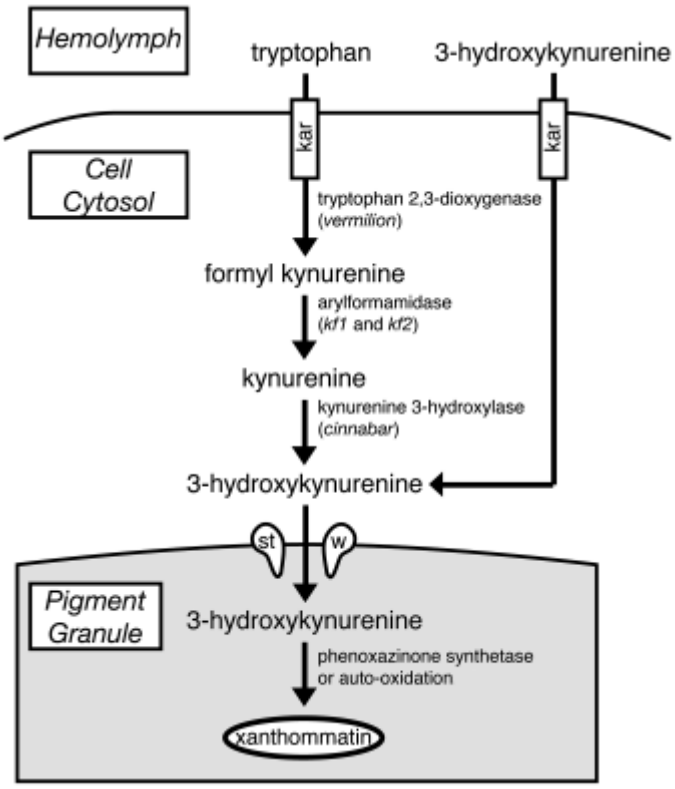


Figure 3.2: Ommochrome synthesis in *Drosophila melanogaster* eyes:
Adapted from Reed and Nagy 2005. In *D. melanogaster* Karmoisin, *kar*, is the cell membrane transporter that transports aromatic amino acids into the cell (Kim *et al.* 2001). A series of enzymes, starting with *vermillion* and ending with *cinnabar*, transform tryptophan into 3-hydroxykynurenine. 'st' and 'w' mark the *scarlet* and *white* monomers forming the heterodimer ABC transporter responsible of uptake of 3-hydroxykyunurenine into the pigment granule. The last pathway step leading to xanthommatin (brown pigment) synthesis is not as well understood.

kmo in *Aedes aegypti* is the orthologue of *Drosophila melanogaster cinnabar* and encodes for kynurenine 3-monooxygenase (KMO) (Sethuraman and O'Brochta 2005; Han *et al.* 2003). It was initially described in a white eyed *Aedes aegypti* mutant and the allele causing the mutation renamed kh^{w}, as the *kmo* gene was then known as *kynurenine hydroxylase* (Cornel *et al.* 1997). *kmo* mutations have also been generated in *Aedes aegypti* using CRISPR-Cas9 (Basu *et al.* 2015). There is also a *white* orthrologue in *Aedes aegypti*. Recent CRISPR-Cas9 studies in *Aedes aegypti* generated mutations in both *kmo* and *white*, both resulting in a recessive white eye phenotype (Li, Bui, *et al.* 2017).

None of the ommochrome genes discovered so far are sex-linked in *Aedes aegypti*, owing to its homomorphic sex chromosomes. Since mutations in this pathway are often recessive, such as with *kmo* and *white*, screening for mutants in *Aedes aegypti* is more challenging than in *D. melanogaster*. The *white* gene has also been successfully mutated (using CRISPR-Cas9) in three different *Anopheles* species, and although the mutation is also recessive, it is found in the X chromosome just as in *D. melanogaster* which means that mutant hemizygous males show the white eye phenotype (Li, Akbari, *et al.* 2017).

3.1.6 Project aim

The aim of the project was to explore the possibility of using homology directed repair to insert exogenous sequences into the M-locus of *Aedes aegypti*. Male specific linkage of genetic components would provide additional options for future gene drive applications. Successful development of such a method would also facilitate basic-science investigations of the nature and mechanism of the primary sex determination signal and of the evolution of sex loci and chromosomes.

3.2 Materials and Methods

3.2.1 Aedes aegypti rearing procedures

The *Aedes aegypti* strain used for transformation and crossing experiments was an Asian wild type (AWT) which had undergone five generations of full-sibling mating (Joe Turner). AWT is a laboratory strain originating from Jinjang, Kuala Lumpur, Malaysia, that has been reared in the Institute of Medical Research since the 1960s (Lacroix *et al.* 2012). Transgenic lines maintained were *OX5226a* and *OX5226b* (*nanos-Cas9* lines generated by *piggyBac* mediated insertion into AWT). Finally, the generated *NixInt1* transgenic line (generated from AWT) was also maintained.

Mosquitoes were reared using previously described standard procedures for *Aedes aegypti* (Crampton *et al.* 1997). *Aedes aegypti* colonies were kept at 27°C ± 2°C and a relative humidity of 75% ± 5% during a 12 hour day-night cycle in temperature and humidity monitored rooms to best simulate natural climatic conditions. Aquatic life stages (hatching eggs, larvae and pupae) were kept in trays containing approximately 1.5-2.5 litres of deionised water. Larvae were fed with finely ground Tetramin fish food (Tetra GmbH, Germany). Pupae were then picked using 3ml pasteur pipettes (Fisherbrand), with their tips cut-off, and sex-sorted by size (females should be distinctly bigger) should this be required. In rare cases where size separation could not be accurate due to unexpectedly dense rearing of larvae/pupae in trays, pupae were transferred to 100ml weigh boats

(VWR, UK) in order to screen under the microscope (Section 3.2.5). Adults were fed on a 10% sucrose solution containing 14U/ml penicillin and 14 µg/ml streptomycin, to prevent bacterial growth. All adults were kept in small cages.

When egg collection was required, mated females were allowed to take a blood meal from defibrinated horse blood (TCS Biosciences). Parafilm M (Bemis) around a metal plate was used to create a pocket that, once filled with blood, was then sealed. Blood plates were placed on top of the cages, and a microwaved bean bag (20 seconds, 700 watts) was used to warm up the plates. Feeding was carried out three days post mating for 1hr, with bean bag reheating every 20 minutes. Eggs were collected (3 to 4 days post blood feed) on 90mm diameter filter papers (Whatman, UK), placed on top of a petri dish (Thermofisher) containing wet cotton (Synergy Health). Eggs were then desiccated and hatched after three days or stored for up to four months. Hatching was carried out in 200ml deli pots filled with deionised water containing the submerged egg paper or egg coverslips from injections. Hatching was induced by deoxygenating the water via means of a vacuum pump. The ‘vacuum’ was maintained for at least 2 hours before allowing air flow back in and a drop of Liquifry No1 (Interpret) was added to further encourage hatching. Backup egg papers were stored away where necessary.

3.2.2 Preparation of solutions for injections

All design and generation of injection components as well as injection mix preparation were carried out by by Joe Turner in Oxitec (Table 3.1). Injection buffer was added for a final concentration of 5mM KCl and 0.1 mM NaH_2PO_4, pH 6.8. Endonuclease-free water (Ambion, USA) was used for dilutions where required. *OX5346* was the *Nix* donor template which had a *Hr5IE1-DsRed* marker flanked by 2kb *Nix* homology arms reaching up to the predicted cleavage site for *Nix5* and *Nix8* (Figure 3.3b) which are within the exon 1 of *Nix* (Figure 3.3a). Although the homology arms sequence was somewhat repetitive, it was thought to be unique to the *Nix* gene, given the genomic information available at the time (Nene *et al.* 2007; Turner *et al.* 2017).

Table 3.1: Components for injection solutions:

sgRNAs were injected at the concentrations recommended in (Kistler et al. 2015). *Nix* sgRNA sequences were selected from (Hall et al. 2015) and kmo sgRNAs were selected from (Basu et al. 2015). ku70 dsRNA was injected to suppress NHEJ repair as recommended in (Basu et al. 2015). OX5346 donor template was made by Joe Turner and injected at concentrations recommended in (Kistler et al. 2015). Cas9 protein (PNA Bio Inc) was injected at concentrations recommended by the provider.

No Cas9 (injected into *OX5226*)	**Cas9 (injected into AWT)**
Nix5 sgRNA (40ng/µl)	*Nix5* sgRNA (40ng/µl)
Nix8 sgRNA (40ng/µl)	*Nix8* sgRNA (40ng/µl)
kmo281 sgRNA (40ng/µl)	*kmo281* sgRNA (40ng/µl)
kmo460 sgRNA (40ng/µl)	*kmo460* sgRNA (40ng/µl)
kmo519 sgRNA (40ng/µl)	*kmo519* sgRNA (40ng/µl)
ku70 dsRNA (100ng/µl)	*ku70* dsRNA (100ng/µl)
OX5346 donor template (700 ng/µl)	*OX5346* donor template (700 ng/µl)
Injection buffer	Cas9 protein (300 ng/µl)
Water	Injection buffer
	Water

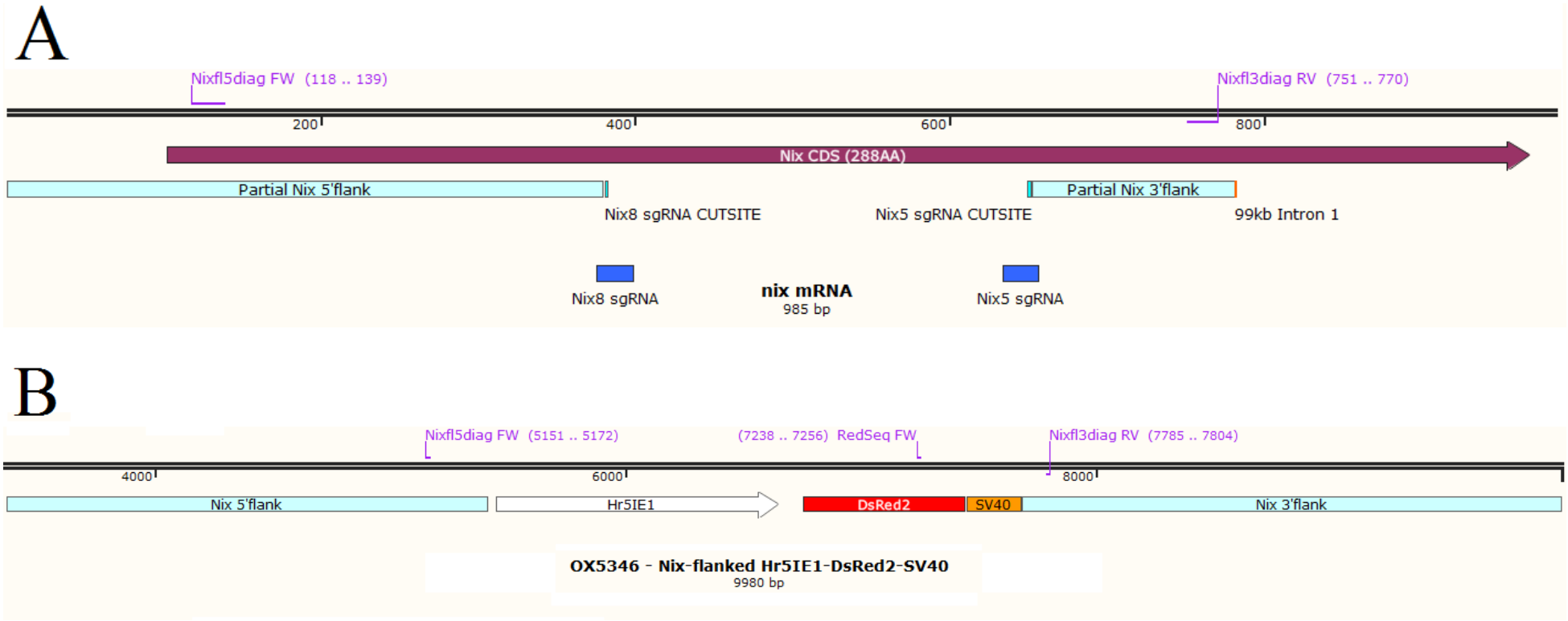


Figure 3.3: cDNA *Nix* gene and *OX5346 Nix* donor sequence maps with their respective diagnostic primers:
Maps were drawn using SnapGene Viewer. Diagnostic primers used for the results below are shown here in purple. (**A**) The 985bp cDNA of the *Nix* gene has a CDS of 288AA. *Nix* sgRNAs are shown, as well as their respective cutsites. The parts of the cDNA which are also present in the *Nix* donor flanks. (**B**) Plasmid map for the *Nix* donor is shown but the Kanamycin resistance cassette for the *Nix* donor is not included. The *Hr5IE1-DsRed-SV40* marker is flanked by 2kb *Nix* homology arms reaching up to the predicted cleavage site for *Nix5* and *Nix8*.

3.2.3 Aedes aegypti microinjections

Microinjection of 1 to 2 hour old embryos was carried out mostly following out established procedures (Morris 1997). Embryos were collected as described above from the respective injection cages (*OX5226* or AWT) and then kept at 4°C to slow development and be able to inject all the eggs before pole cell formation. Injection cages were kept in the dark during embryo collection to encourage laying. Pole cell formation at RT in *Aedes aegypti* occurs around 2 hours post laying, the melanisation of the eggs and hardening of the chorion occurs simultaneously. Colour guided staging as described in (Lobo *et al.* 2006) allowed selecting the correct embryos for injection. Younger (white) embryos are too fragile to handle securely, whilst older ones (dark grey to black) will not be able to be transformed when injected. Injections were carried out using a Femtojet 4x (Eppendorf) setup. Needles were pulled from aluminosilicate glass (1.0mm x 0.64mm x 10cm, Sutter Instrument) by needle puller Model P-2000 (Sutter Instrument). They were then bevelled before the injection mix (kept on ice and centrifuged at 11000 x g) was loaded using microloader tips (Eppendorf). Halocarbon oil 27 (Sigma) used to cover the embryos was rinsed off post injections and embryos were kept in a closed chamber (of 100% relative humidity) for three days prior vacuum hatching.

3.2.4 Aedes aegypti injected G0 crossing

Both male and female G0s were outcrossed to AWT individuals. They were each divided into G0 pools of no more than 20 individuals each. The sex ratios were

kept 1 to 1 when the pool had G0 females, but instead were increased to 1 to 5 (male to female) when the pool had G0 males. Blood meals and general maintenance was carried out as described above. 3 ovipositions were collected for each G0 pool and hatched without storing (Section 3.2.1) for subsequent screening (Section 3.2.5)

3.2.5 Aedes aegypti transgenic screening

The progeny from G0 crosses was screened at L2-L4 larval stages when screening for the transgenic *Hr5IE1-DsRed* marker (*Nix* integration) or *Hr5IE1-AmCyan* (for *OX5226*) which shows whole body expression and is nuclear localised. Screening was performed using the appropriate light filters on a fluorescence stereo microscope (Leica). G0 adults were screened for feminisation on a visible light stereomicroscope (Leica). G2s from G1 sibling to sibling crosses were screened at pupal stages under the visible light stereomicroscope for white eye colour in *kmo-/-* double knock-outs. Photos were taken using an iPhone 4S on a phone to microscope mount.

3.2.6 Genomic, plasmid and gel DNA extractions

gDNA extractions were carried out using the PureLink Genomic DNA kit (Thermo Fisher Scientific) following instructions from the manufacturer. DNA gel extractions were carried out using the GeneJET Gel Extraction Kit (Thermo Fisher Scientific) following instructions from the manufacturer. Plasmid DNA

extractions were carried out using the GeneJET Plasmid Miniprep Kit (Thermo Fisher Scientific) following instructions from the manufacturer.

3.2.7 General PCR

General PCRs were carried out using Taq DNA Polymerase (PCRBIO) using the recommended manufacturer conditions in final volumes of 50µl. PCR programs were defined with the standard manufacturer recommendations except for primer melting temperatures which were calculated using Primer3. A list of primers used in *Nix* diagnostic PCRs is included in Table 3.2.

Table 3.2: Primers used in *Nix* diagnostic PCRs

Primer names, sequences and a brief description are shown below.

Primer Name	Primer Description	Primer Sequence
Nixfl5diag FW	Binds 5' end of *Nix*	AATGCTGAGATCAATGCAGAAT
Nixfl3diag RV	Binds 3' end of *Nix*	TGTGCAATTCGCTCTTCCGT
RedSeq FW	Anneals with *DsRed* sequence	AAGGGCGAGACCCACAAGG
TD1405	Anneals with *DsRed* sequence	GCCACGAGTTCGAGATCGAG
TD2638	Anneals with *DsRed* sequence	CTCGTACTGCTCCACGATGGT

3.2.8 Flanking PCR

Extracted gDNA from *OX5226a* and *OX5226b* individuals (Section 3.2.6) was digested with two different restriction enzyme (RE) 6bp cutters (EcoRI and XhoI). 100ng of gDNA was digested in a total volume of 20µl for 3 hours in a water bath at 37°C. Flanking PCR was performed with 3 sets of specific nested primers for each *piggyBac* flank (5' or 3'). Two different pools of semi random primers (PSRP1 and PSRP2) were used for each gDNA and enzyme combination. The same PSRP was used in the first, second and third round of PCR of each respective sample. Each PCR round used a 1 in 10 dilution of the

previous PCR and the respective nested *piggyBac* primer. First PCR used 2µl of digested gDNA. PCRs were carried out using Taq DNA Polymerase (PCRBIO) at the recommended reaction concentrations in a final volume of 25µl. Specific nested primers were at a final concentration of 100µM whilst PSRPs were at a final concentration of 25µM in each reaction. The PCR program for first, second, and third round PCRs is shown in Table 3.3. Primers used and their sequences are shown in Table 3.4. Clear bands in third round of PCR were excised and gel extracted (Section 3.2.6) for subsequent pJET1.2 blunt cloning (Section 3.2.9), colony PCR to confirm positives (Section 3.2.10), miniprep plasmid DNA extractions of positives (Section 3.2.6), and final Sanger sequencing (Section 3.2.11).

Table 3.3: PCR programs for 1st, 2nd and 3rd round of flanking PCRs:

Programs for flanking PCR rounds 1 to 3 are shown below. m stands for minute(s) and s for second(s).

	First PCR		Second PCR		Third PCR	
1	95°C	4m	95°C	1m	95°C	1m
2	95°C	30s	97°C	7s	97°C	7s
3	55°C	30s	55°C	40s	55°C	40s
4	72°C	3m. Go to 2 x9	72°C	2m. Go to 2 x1	72°C	2m. Go to 2 x1
5	43°C	1m	97°C	7s	97°C	7s
6	72°C	3m	55°C	20s	55°C	20s
7	97°C	7s	72°C	2m. Go to 5 x32	72°C	2m. Go to 5 x32
8	55°C	30s	72°C	3m	72°C	3m
9	72°C	2m	4°C	∞	4°C	∞
10	72°C	2m. Go to 7 x24				
11	4°C	∞				

Table 3.4: Flanking PCR primers used and their sequences:

Flanking PCR primers used are shown below, with their respective sequences. N stands for a random nucleotide (A, T, G or C) whilst S stands for a G or a C (strong bonds).

Primer Description	Primer Name	Primer Sequence
Pool of Semi Random Primers 1	TD2366	acgctccNNNNNNNNNNSCATGS
	TD2367	acgctccNNNNNNNNNNSGATCS
	TD2368	acgctccNNNNNNNNNNSACGTS
	TD2369	cgcctccNNNNNNNNNNSAGCTS
Pool of Semi Random Primers 2	TD2374	acgctccNNNNNNNNNNNNSGCTCS
	TD2375	acgctccNNNNNNNNNNNNSGCAGS
	TD2376	acgctccNNNNNNNNNNNNSGTGTS
	TD2377	acgctccNNNNNNNNNNNNSCACTS
piggyBac 5' 1	TD 695	TCAATTTTACGCAGACTATCTTTCTAGGG
piggyBac 5' 2	PB2	CAGTGACACTTACCGCATTGACAAG
piggyBac 5' 3	TD225	TGACAAGCACGCCTCACGGGAG
piggyBac 3' 1	TD126	TGTCGAGAGCATAATATTGATATGTGCC
piggyBac 3' 2	TD127	CCGGACCGCGGCT
piggyBac 3' 3	PB3	CAGACCGATAAAACACATGCGTCA

3.2.9 Blunt ligation of DNA fragments into pJET1.2 and transformation into compentent cells

Purified flanking PCR fragments from Section 3.2.8 were ligated into *pJET1.2* (blunt ready) plasmids using the CloneJET PCR Cloning Kit (Thermo Fisher Scientific) following instructions from the manufacturer. Transformations were carried out into Library Efficiency DH5α Competent Cells (Thermo Fisher Scientific) following instructions from the manufacturer.

3.2.10 Colony PCR

For each flanking PCR product (6 different clear bands) 16 colonies were selected for colony PCR (on a 96 well plate). Primers used were specific primers for the *pJET1.2* vector bridging over the insert. 5µl of molecular biology water (Sigma) were added to each of the wells. 2 agar plates had grid marks added so that 96 individual squares could be seen. Labelling of both the PCR plate and the agar plates was done so as to be able to trace back each individual colony PCR

reaction to its plate growth. Pipette tips were used to pick up individual colonies which were dipped into the 96 water containing wells and then streaked on their respective agar plate square (which would then be incubated 8 hours at 37°C). A PCR master mix, using Taq DNA Polymerase (PCRBIO) at recommended manufacturer concentrations for a final 25µl volume was prepared for 100 reactions and then appropriately pipetted into each of the 96 wells. A standard PCR program was used, following Taq manufacturer recommendations except for the reduction in the number of cycles to only 25. There is a risk that too many cycles would start detecting free insert DNA that would have been present on the agar plates following transformation of DH5α cells with the blunt ligation mix (Section 3.2.9). A lower number of cycles, hence prevents false positives. The colony PCRs were then run on a gel and positive colonies were then traced back to the freshly streaked agar plates and standard miniprep cultures were prepared for the positive colonies containing the respective inserts. Plasmid extractions were carried out as described (Section 3.2.6).

3.2.11 DNA sequencing

DNA sequencing was outsourced to GATC Biotech. Samples sequenced in this project included successfully cloned flanking PCR bands (Section 3.2.8), which were screened by colony PCR (Section 3.2.10).

3.3 Results and Discussion

3.3.1 OX5226a and OX6226b nanos-Cas9 lines are not separate insertions

Lines *OX5226a* and *OX5226b* were *nanos-Cas9* lines made by Joe Turner in Oxitec by injecting a *nanos-Cas9* construct into *Aedes aegypti* embryos (AWT strain). Both were recovered from the same G0 adult pool and although initially treated as potentially different due to the apparent differences in the marker expression profile, the possibility remained that they were not independent, i.e. comprised the same genomic insertion. *Aedes aegypti* transgenesis using *piggyBac* is less efficient than in *Drosophila melanogaster* and since no more than 20 G0 adults are usually pooled together, there is a high chance that transgenic G1s appearing in the same G0 adult pool originate from the same individual G0 and are in fact a clonal isolation of the same insertion event.

In order to find out the insertion site of the two potentially different *nanos-Cas9* lines flanking PCR was performed (Section 3.2.8) from the known 5' and 3' *piggyBac* ends that are found in *piggyBac* insertions. Two different enzymes were used (XhoI and EcoRI) and two different known pool of semi-random primers (PSRPs) was used (PSRP1 and PSRP2) (Section 3.2.8). Reactions only gave clear bands with the second primer set PSRP2. Whilst the 5' end did not give any clear bands, even after the 2 rounds of nested PCR, the 3' end gave clear bands for both enzymes (Figure 3.4). *OX5226a* and *OX5226b* lines show

the same band profile upon flanking PCR from 3’ *piggyBac*. This suggests they are both the same insertion.

For the purpose of determining the insertion site the clean bands for both *OX5226a* and *OX5226b* (the two bands observed with XhoI and the single band with EcoRI) were gel extracted, purified and cloned into *pJET1.2* (Section 3.2.9). Colony PCR was performed to confirm inserts of the correct size and these were sent for Sanger sequencing (Section 3.2.11). Unfortunately sequence quality was too low to determine the insertion site in the *Ae. aegypti* genome. Given that band profiles were the same between both lines for two different enzymes, and they both originated from the same G0 adult pool, there is a high chance that they were both the same insertion. *OX5226b* became *OX5226* and *OX5226*a was discarded.

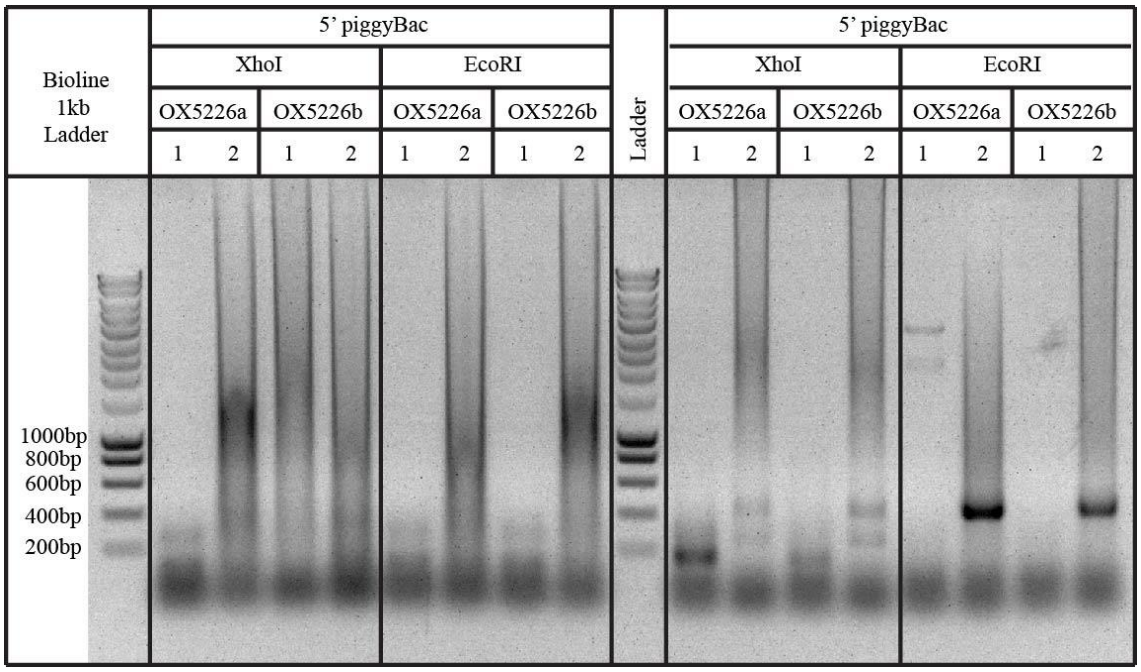


Figure 3.4: *OX5226a* and *OX5226b* yield the same band profile after flanking PCR: 5' end flanking PCRs did not work effectively since no distinct bands are observed, but a characteristic smear of non-specific products when pools of random primers are used. Similarly, no clear bands are observed for PSRP1 containing reactions, 1 and 2 numbers mark PSRP1 and PSRP2 primers respectively. 3' end PCRs using PSRP2s resulted in clear distinct bands for both enzymes. The bands profiles were the same with *OX5226a* or *OX5225b* starting gDNA. In the case of XhoI two somewhat clear bands were seen whilst single, stronger bands were seen for EcoRI.

3.3.2 Nix knock-in and kmo knock-out components were injected into Aedes aegypti

To serve the dual purpose of verifying the *nanos-Cas9 Aedes aegypti* lines (generated by Joe Turner) and obtaining a fluorescent marker *Nix* knock-in the following was injected into *nanos-Cas9* embryos; *Nix* sgRNAs (5 and 8), *Nix* HDR donor plasmid (*OX5346*, see Figure 3.3), *kmo* sgRNAs (281, 460, and 519), and *ku70* dsRNA. *Cas9* protein was also supplied when injections were carried out in AWT *Aedes aegypti* embryos (WT). *kmo* components were coinjected to be able to verify the action of *nanos-Cas9* should *Nix* sgRNAs had not worked efficiently, and although ideally they should have been injected separately, time

constraints indicated otherwise. Concentration of components was kept the same for all injection dates (Section 3.2.2).

The injection mix and its components were designed and prepared by Joe Turner. *Nix* sgRNA sequences were selected from Hall *et al.* 2015, which used them to obtain *Nix* knock-outs. *kmo* sgRNA sequences with the highest cleavage guiding efficiency were selected from Basu *et al.* 2015. *ku70* dsRNA against end-joining repair machinery was also injected following recommendations in Basu *et al.* 2015, the rationale being that inhibiting non-homologous end joining repair would favour homology directed repair of *Nix* double strand breaks from the supplied *Hr5IE1-DsRed Nix-flanked* DNA template. Even if not required for homology directed repair in *Aedes aegypti* (Kistler *et al.* 2015), *ku70* dsRNA injection was been shown to improve HDR rates (Basu *et al.* 2015).

A total of 3135 *OX5226 nanos-Cas9* embryos were injected with the components described above. Survival rates varied between ~2 and ~10% for different injection dates (Table 3.5). Even though Kistler *et al.* 2015 recommended to discard G0 males due to their inability of efficient HDR, a lack of convincing evidence for this meant that G0 males were also outcrossed in this study. No HDR events were detected amongst the G1s as no *Ds-Red* positives were found. Partial feminisation of some G0 adults was observed to varying degrees as described in (Hall *et al.* 2015), whilst no G0 pupae or adults showed mosaic red/white phenotypes as described in the same study. Partial or ‘mosaic’

feminisation would be owed to somatic *Nix* knock-out whilst somatic knock-out of *kmo* was expected to result in mosaic white eyed phenotypes (i.e. patches of red or white in the adult eyes). Section 3.3.3 and 3.3.4 describe these findings in greater detail. Owing to the absence of a white-eyed background *Aedes aegypti* strain to cross injected G0 adults to, the chance of finding *kmo* mutants amongst the G1s was negligible. Any *kmo* G1 mutants would be heterozygous for *kmo-* and hence would not present a phenotype. Instead, G1 adult pools (one for each respective G0 adult cage) were set up for sibling to sibling crosses and the G2 progeny screened for *kmo-/-* white-eyed phenotypes. None of these white-eyed individuals were found (Section 3.3.4).

A total of 2374 AWT embryos were injected with the components described above. Survival rates varied between ~2 and ~8% for different injection dates (Table 3.6). Survival rates were similar to injections into *OX5226* lines, although perhaps lower. An explanation for a lower survival rate when Cas9 protein is injected vs endogenous Cas9 translation in *nanos-Cas9* lines is the fact that there may be more Cas9 in the embryos depending on the amount produced by *nanos-Cas9*. The amount of Cas9 might be variable, depending on the volume of liquid injected, whilst this would be expected to be more constant in *nanos-Cas9* lines. However, this is not supported strongly as the difference between the two injection sets is only slight. Partial feminisation of some G0 adults was observed to varying degrees, whilst no G0 adults showed mosaic red/white phenotypes, as described above. Section 3.3.3 and 3.3.4 describe these findings in detail. G1

adult pools (one for each respective G0 adult cage) were set up for sibling to sibling crosses and the G2 progeny screened for kmo-/- white-eyed phenotypes, as described above. None of these white-eyed individuals were found (Section 3.3.4).

A putative HDR event was found from this set of injections; a total of 38 *Hr5IE1-DsRed* positive individual G1s (19 females and 19 males) were found from a single G0 male adult pool. Transgenesis efficiency (calculated as a percentage of G0 adults bearing transgenic progeny) was of 1.14% (Table 3.6). Given the extremely low occurrence of HDR in the two injection sets (5509 embryos injected, 35 G0 pools were set up, and only 1 in 35 pools gave transgenics), these were assumed to represent a single insertion event. Genetic female transgenics were not expected as a *Nix* knock-in would in theory only be present in genetic males. The hypothesis was that the observed females were actually feminised genetic males. However, in that case it was unclear why *Hr5IE1-DsRed* positive males were found as well. Section 3.3.5 and 3.3.6 describes the downstream analysis for these transgenics.

Table 3.5: *Nix* sgRNAs, *Nix* HDR donor, *ku70* dsRNA, and *kmo* sgRNAs were injected into *nanos-Cas9 Aedes aegypti* embryos:

Aedes aegypti nanos-Cas9 (*OX5226* line) embryos were injected with *Nix* sgRNAs, *Hr5IE1-DsRed Nix* HDR donor plasmid, *kmo* sgRNAs to verify CRISPR-Cas9 activity, and *ku70* dsRNA to reduce NHEJ repair and favour HDR. A summary of the number of injected embryos, larval hatch rates, adult survival rates, and transformation efficiency (percentage of knock-in events per G0 adult) is provided. A total of 18 different G0 adult pools were crossed to AWT. G1 larvae were screened for *Nix* HDR knock-in events and none were found. To be able to screen for *kmo* mutations without a white eyed line to cross to, a pool of G1s from each G0 pool were allowed to interbreed (sibling to sibling crosses) with the hope that two heterozygous *kmo* mutants would cross to phenotypically reveal the mutation amongst G2s. No *kmo-/kmo-* mutants were found amongst the G2 progeny.

Date Injected	Nix sgRNA 5 (ng/µl)	Nix sgRNA 8 (ng/µl)	ku70 dsRNA (ng/µl)	kmo sgRNA 281 (ng/µl)	kmo sgRNA 460 (ng/µl)	kmo sgRNA 519 (ng/µl)	Nix Donor Plasmid (ng/µl)	Embryos	Larvae (Hatch Rate)	Adults (Survival Rate)	Total G0 Pools	G1 HDR +ves	Single HDR +ve Events	kmo G2 KOs
14.7.16	40	40	100	40	40	40	700	1558	186 (11.9%)	151 (9.7%)	13	0	0	0
15.7.16	40	40	100	40	40	40	700							
20.7.16	40	40	100	40	40	40	700	494	14 (2.8%)	8 (1.6%)	2	0	0	0
25.7.16	40	40	100	40	40	40	700							
21.7.16	40	40	100	40	40	40	700	1083	23 (2.1%)	19 (1.8%)	3	0	0	0
22.7.16	40	40	100	40	40	40	700							
							Total	**3135**	**223 (7.1%)**	**178 (5.7%)**	**18**	**0**	**0**	**0**
										Transformation Efficiency			**0.00%**	

Table 3.6: *Nix* sgRNAs, *Nix* HDR donor, *ku70* dsRNA, *kmo* sgRNAs, and Cas9 protein were injected into AWT *Aedes aegypti* embryos:

AWT strain *Aedes aegypti* embryos were injected with *Nix* sgRNAs, *Hr5IE1-DsRed Nix* HDR donor plasmid, *kmo* sgRNAs to verify CRISPR-Cas9 activity, Cas9 protein and *ku70* dsRNA to reduce NHEJ repair and favour HDR. A summary of the number of injected embryos, larval hatch rates, adult survival rates, and transformation efficiency (percentage of knock-in events per G0 adult) is provided. A total of 17 different G0 adult pools were crossed to AWT. G1 larvae were screened for *Nix* HDR knock-in events and 38 individuals (19 females and 19 male) from a single male G0 pool were found to be *Hr5IE1-DsRed* positive. Since they all came from a single G0 pool they originated most likely from a single integration event. Female transgenics were not expected as a *Nix* knock-in would in theory only be present in males. To be able to screen for *kmo* mutations without a white eyed line to cross to, a pool of G1s from each G0 pool were allowed to interbreed (sibling to sibling crosses) with the hope that two heterozygous *kmo* mutants would cross to phenotypically reveal the mutation amongst G2s. No *kmo-/kmo-* mutants were found amongst the G2 progeny.

Date Injected	Nix sgRNA 5 (ng/µl)	Nix sgRNA 8 (ng/µl)	ku70 dsRNA (ng/µl)	kmo sgRNA 281 (ng/µl)	kmo sgRNA 460 (ng/µl)	kmo sgRNA 519 (ng/µl)	Nix Donor Plasmid (ng/µl)	Cas9 (ng/µl)	Embryos	Larvae (Hatch Rate)	Adults (Survival Rate)	Total G0 Pools	G1 HDR +ves	Single HDR +ve Events	kmo G2 KOs
27.7.16	40	40	100	40	40	40	700	300	967	28 (2.9%)	19 (2%)	2	38	1	0
28.7.16	40	40	100	40	40	40	700	300							
1.8.16	40	40	100	40	40	40	700	300	540	22 (4.1%)	16 (3%)	5	0	0	0
2.8.16	40	40	100	40	40	40	700	300							
16.8.16	40	40	100	40	40	40	700	300	540	55 (10.2%)	46 (8.5%)	7	0	0	0
18.8.16	40	40	100	40	40	40	700	300	104						
19.8.16	40	40	100	40	40	40	700	300	223	9 (4%)	8 (3.6%)	3	0	0	0
								Total	**2374**	**9 (4%)**	**89 (3.7%)**	**17**	**38**	**1**	**0**
											Transformation Efficiency			**1.14%**	

3.3.3 G0 males showed signs of feminisation upon Nix sgRNA injection

G0s were screened for NHEJ mutations in *Nix* and *kmo*; through partial feminisation of males as was reported in Hall *et al.* 2015. The expected phenotypes ranged from malformations to feminisation of body parts. For external genitalia these were; gonocoxite rotation with respect to the normal orientation, as well as missing gonocoxites and/or gonostyli. With respect to the characteristic plumose (or feathery) antennae in males, the following was expected; a loss of the feathery appearance into an intermediate feminised form or directly female looking pilose antennae.

The percentage of G0 individuals upon *Nix* injections showing malformed or feminised external genitalia in Hall *et al.* 2015 was 78%. However, the percentage of G0 individuals showing feminised antennae was 50%, a considerably lower value than those showing external genitalia phenotypes. 12 AWT females and 12 males were checked under the microscope in order to gain familiarity on WT sexual dimorphism in *Aedes aegypti*. Gonocoxites were all along the latero-lateral axis and retained the same orientation with respect to the body for the 12 WT males. All WT males had feathery antennae and all gonocoxites as well as gonostyli were all with normal looking morphology. A clear dimorphic feature was the length of the palps, which were considerably longer in males than in females. The length of the palps would be almost as long as the mouthparts in males but in the case of females the palps hardly extend from the head. Interestingly, this dimorphic feature was not discussed as being

also in the control of *Nix* in (Hall *et al.* 2015), even though shorter palps in feminised males can be observed in their supplementary figures.

A number of male G0s were selected, from injected or integrated Cas9 injections, and were examined under the stereomicroscope for signs of feminisation. Injections into *nanos-Cas9* embryos resulted in 4 out of 17 (23.5%) feminised G0 males (Table 3.7). The phenotypes included a male with a shorter palp, two males with gonocoxites at a different angle with respect to the WT orientation, and one male with less feathery antennae. Cas9 protein injections into AWT individuals resulted in 3 out of 19 (15.8%) feminised G0 males (Table 3.7). The phenotypes included 3 males with gonocoxites at different angles with respect to WT orientations, and one of such males also had less feathery antennae. The difference in percentage feminisation is not great, and comparison with such low numbers would not be very meaningful. Instead checking for feminisation in G0s was a confirmation of endogenous Cas9 activity of *OX5226* lines as well as Cas9 activity in Cas9 protein injections into AWT embryos. Figure 3.5 shows images of the G0 individuals screened and the different phenotypes observed when compared to WT.

Table 3.7: Selected G0 males were checked for signs of feminisation:

Selected G0 males from both *nanos-Cas9* and Cas9 protein injections were checked for feminisation, 13 and 16 males respectively. Percentages of 23.5 and 15.8% were observed respectively. Phenotypes seen involved less feathery antennae, orientation changes of gonocoxites, as well as shortening of palps.

Selected G0 ♂s checked for signs of feminisation						
Injected or Integrated Cas9	**Date**	**Normal-looking**	**Feminised**	**Total**	**Percentage Feminised**	**Comments**
nanos-Cas9	14/15 July 2017	13	4	17	23.5%	1 with malformed, shorter palps. 2 with gonocoxites at an angle. 1 less feathery antennae.
Cas9 Protein	27/28 August 2017	16	3	19	15.8%	3 with gonocoxites at an angle as well as 1 of them with less feathery antennae.
	TOTAL	**29**	**7**	**36**	**19.4%**	

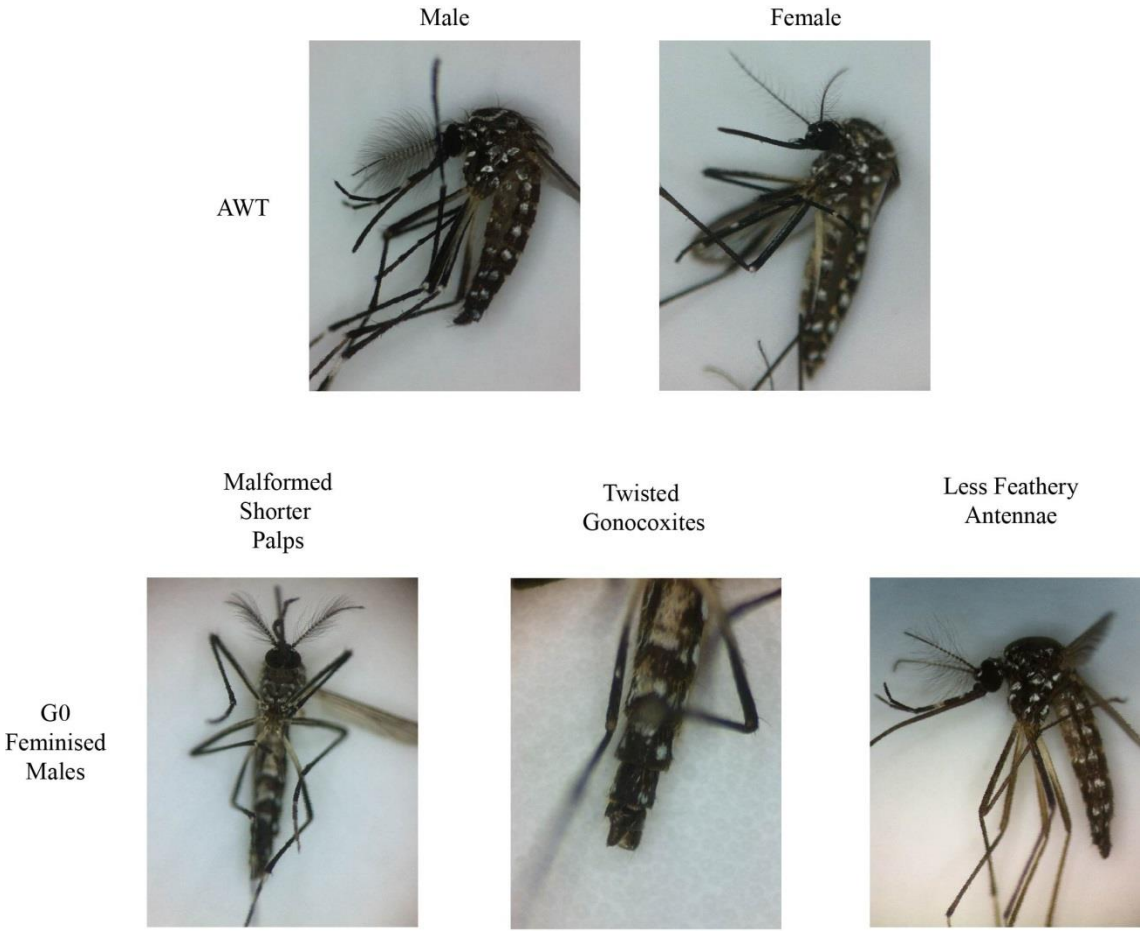


Figure 3.5: Example of feminisation phenotypes amongst the G0s screened:
AWT male and female external genitalia, as well as antennae and palps are shown for comparison with their feminised counterparts. Feminised G0 males are shown above, with the three different phenotypes observed; malformed shorter palps, twisted gonocoxites and less feathery antennae.

3.3.4 No kmo knock-out events were detected

G0 males were screened for mosaic knock-out events in the eye at the same time as they were screened for signs of feminisation (Section 3.3.3). As described in (Hall *et al.* 2015), eye phenotypes were expected to involve red or white mosaic patterns. All eyes screened were the WT black colour, so no somatic activity of Cas9-*kmo*-sgRNAs was detected.

As described above, a white eyed background strain was not available for crossing to at the time of the injections (Table 3.5 and Table 3.6). This meant that G0 adults were crossed to AWT *Ae. aegypti* individuals and any *kmo* knock-out events inherited by the G1s could not show a phenotype as they would always inherit a wild type *kmo* copy from their AWT parent. Given the high efficiency of CRISPR-Cas9 reported in *Ae. aegypti* with the same sgRNAs, 23-90% of G0s resulted in mutant progeny (Basu *et al.* 2015), there was a chance of finding *kmo*-/- individuals amongst G2s upon crossing G1s with themselves. G1s were crossed in sibling to sibling pools and their G2 progeny was screened. A single G1 pool was set up for each initial G0 pool. G1 pools were set up with 50-100 adults. Even after screening 4 oviposition cycles, no white eyed progeny was found.

Obtaining a *kmo* knock-out was not the main objective of these injections. However, there were multiple limitations of such an approach in order to verify CRISPR-Cas9 activity. Firstly, *ku70* dsRNA was injected in order to reduce the rate of NHEJ repair which would, at the same time as increasing the wanted *Nix* HDR, reduce the mutation efficiency of CRISPR-Cas9 as it necessitates the error-prone pathway to generate mutations from the cleaved double strand breaks (DSBs). However, even if *ku70* dsRNA reduces the rate of knock-outs, *Nix* somatic knock-outs were still observed through mosaic feminisation of G0 adults. This appears to be surprising, however whilst the somatic tissues requiring *kmo* function are in the eye, which develops from the anterior end of

the embryo, somatic tissues requiring *Nix* function are probably within the gonads, which develop from the posterior end of the embryo where they were injected. This may in part explain the differential rate at which mosaic mutations are observed for both genes in G0 adults. This disparity is also observed in (Hall *et al.* 2015), as fewer *kmo* than *Nix* mutant mosaic G0s were found. Moreover, co-injecting 5 different sgRNAs simultaneously may not be the best strategy as they may compete for Cas9 loading (Mekler *et al.* 2016), and should Cas9 affinities vary substantially, either the *kmo* or *Nix* mutagenesis strategy might have been severely compromised. Finally, a *kmo* knock-out (white eyed) background line such as "Higgs White Eye" could have been very helpful, as this would have allowed identification of *kmo* mutants in the G1, even if present at rather low frequency.

One way to circumvent these problems would have been to separate HDR from NHEJ orientated injections and maximise the chances of observing NHEJ knock-out events whilst also being able to maximise the chances of successful HDR separately. This would have improved the chances of finding *kmo-/-* individuals amongst G2 individuals.

3.3.5 Hr5IE1-DsRed positive G1 progeny were obtained from one G0 pool

As indicated in Section 3.3.2 a single G0 male adult pool from Cas9 protein injections gave *Hr5IE1-DsRed* positive G1 progeny (Table 3.6). Treated as a single HDR insertion event, the integration individuals were named *NixInt1* (for

Nix integration 1). The observed even split between number of male and female transgenics was unexpected. The expected outcome of the *Nix* integration was to make genetic males into females, and the finding of phenotypic males, as well as females amongst the transgenic progeny was surprising. All G1s from the progeny pool were counted for male and female numbers, whether wild type or *NixInt1* positive (Table 3.8). The hypothesis was that if any genetic males had been feminised, the expected 1:1 sex ratio would have been distorted.

Table 3.8: Sex distribution and ratios for *NixInt1* positive and negative G1s from the same G0 pool:

The normal sex ratio of 1:1 was met almost to perfection. This was unexpected for an integration event into *Nix*. Ratios are given with a precision of 1 decimal place. This suggests a lack of *Nix* knock-out events as well as casting doubt on the integration site of *NixInt1* positives.

	Females	**Males**	**Female:Male Ratio (1dp)**
***NixInt1* positives**	19	19	1.0:1.0
***NixInt1* negatives**	100	101	1.0:1.0
TOTAL	**119**	**120**	**1.0:1.0**

1:1 sex ratios were not even minimally distorted. In addition, any NHEJ *Nix* knock-outs (with incorrectly repaired DSBs) may have also distorted the WT sex ratio. Whilst G0 adult males were screened for mosaic feminisation (Section 3.3.3), G1 pools from the 35 G0 adult pools did not have their sex ratios determined. This meant that there could have been G1 *Nix* knockouts but this was not assayed in this study. The reason behind this was that even if sex ratio skews were found, it would be impossible to determine which individuals had the *Nix* knockout as feminised *Nix*- males and genetic females would in theory be indistinguishable. Nevertheless, sex ratio data was obtained for one of the G1 pools (Table 3.8), even if the purpose was to check for skews driven by *NixInt1*

individuals. In addition, *NixInt1* individuals, male and females, were examined under the stereomicroscope and no signs of feminisation were seen in any individuals. Although the phenotype expected was full feminisation, intermediate forms were not found either, which may have been a sign of some impairment of *Nix* normal function should they have been found.

No stable G1 *Nix* knockouts were obtained in (Hall *et al.* 2015) and hence it would have been interesting to find out if feminised genetic males, due to *Nix* knockout, would be viable, sterile 'pseudo-females' or even fertile 'pseudo-females. *NixInt1* individuals were crossed to each other in sibling to sibling crosses. No obvious reduction in fertility was detected as crosses were successful and resulted in typical numbers of G2 eggs. Hatching and maintaining this line, whilst enriching for *NixInt1* individuals was also straightforward.

HDR was presumed to be highly specific, and misdirected integrations were expected to be rare. However, the simplest explanation for the data above was that the labelled *NixInt1* integration had not occurred in *Nix* but rather in a sex-independent region of the genome. Incidentally, even if dealing with a misdirected integration, the fact that the integration was originated from a male G0 pool is already interesting data in itself. This putative HDR event contradicts the recommendations for HDR given in Kistler *et al.* 2015 where G0 males were just dismissed as incapable of effective HDR.

3.3.6 Integration of Hr5IE1-DsRed2 was not into the expected Nix locus

In order to test the hypothesis, based on the data above, of a misdirected integration of the *Nix Hr5IE1-DsRed* donor template, several PCRs were performed. Three different primer sets were used to be able to confirm the presence of the transgene, rule out the possibility of contamination of the transgenic bearing G0 pool by other *Hr5IE1-DsRed* lines in Oxitec, and assay the integrity of the *Nix* gene in both WT and *NixInt1* individuals (Figure 3.6).

gDNA was collected from 6 individuals (an OX5226 male, an AWT female, an AWT male, a *NixInt1* female, and two *NixInt1* males) and then PCRed with the three different primer sets. The first amplicon, 567bp, involved a primer within the 3' *Nix* flank and a second within *DsRed*; this would only be amplified in individuals containing an insertion from the *Nix Hr5IE1-DsRed2* donor template. The second amplicon, 653bp, has primers which bind at both 5' and 3' *Nix* flanks which are present in the endogenous gene as well as the *Nix Hr5IE1-DsRed2* donor template. However, the amplicon from the donor template would be 2654bp and hence would not be amplified with the choice of PCR conditions (Section 3.2.7). The third amplicon, 575bp, has primers that bind within *DsRed2* which would amplify in potential *Nix* integrations as well as in potential *Hr5IE1-DsRed* contaminants. Amplicon sizes were as expected (Figure 3.6) where a clear band was observed. The *DsRed* to *Nix* amplicon, as well as the *DsRed* only amplicon, showed up for the three *NixInt1* individuals (male or female) and was absent from AWT (male or female) and male *OX5226* individuals. This

suggested *NixInt1* individuals indeed originated from a *Nix Hr5IE1-DsRed2* donor template mediated integration and not from a contamination event. The endogenous *Nix* amplicon showed up for all males and no females, regardless if WT, *OX5226*, or *NixInt1* individuals. This demonstrates that the endogenous *Nix* locus is intact, at the targeted integration site.

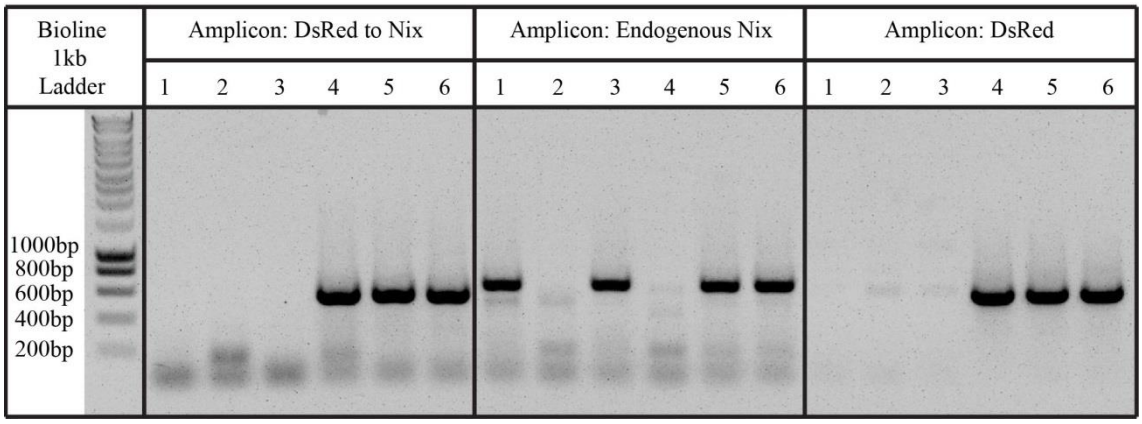


Figure 3.6: The observed integration event was not into *Nix* nor is it sex linked:
gDNA from samples 1 to 6 was PCRed using three different pairs of primers. The three different amplicons for each of the six gDNA samples are shown above. Samples are; 1) OX5226 male, 2) female AWT, 3) male AWT, 4) female *NixInt1*, 5) male *NixInt1*, 6) second male *NixInt1*. *DsRed* to *Nix* amplicon has one primer binding within the DsRed marker and the other binding one of the *Nix* homology arms supplied, 567bp; RedSeqF and Nixfl3diagR. Endogenous *Nix* amplicon has primers bridging the targeted *Nix* region for cleavage, 653bp, (although these primers could also amplify from the *Nix DsRed* donor construct the amplicon would be 2654bp and hence is excluded in this PCR); Nixfl5diagF and Nixfl3diagR. *DsRed* amplicon has internal coding sequence primers; diag6-DsRed FW and diag6-DsRed RV. Primer sequences can be found in the methods Chapter of this Section. DsRed to Nix amplicon appears for both male and female *NixInt1* individuals, suggesting they carry the same integration. DsRed amplicon suggests the same. Endogenous *Nix* is found intact and present in all males regardless if they are WT or *NixInt1* individuals. This confirms HDR into *Nix* did not occur and the integration is elsewhere in the genome where it is not sex linked. Low bands, below 200b, probably represent primer dimers.

This raises several questions, how, where, and how much of the *Nix Hr5IE1-DsRed2* donor template integrated in the *Aedes aegypti* genome. Although this was not explored during this study, it could be discussed as future work to be done. The *Hr5IE1-DsRed* expression observed in *NixInt1* individuals and the

PCR results above, show that at least 2352bp of the *Nix* donor template successfully integrated within the genome, including the full fluorescent marker cassette and part of the 3' *Nix* flank.

It is not possible to determine, from the data of this study, whether the integration observed depended on homology between the integrated site and the *Nix* donor template. Similarly, there is no evidence for the involvement, if any, of Cas9-sgRNA activity in the integration event. A possible explanation for the mechanism of integration would be that off-target Cas9-sgRNA activity resulted in DSBs elsewhere in the genome, and an excess of *Nix* donor template would have favoured it as a template for repair even in the absence of sequence homology. A second option would be that loci of similar sequence to the *Nix* locus would be cleaved, albeit less efficiently, by the Cas9-sgRNAs injected and the partial sequence homology would allow HDR from the *Nix* donor template. This second option appears more likely, given the low coverage of the M-locus and other regions of the *Aedes aegypti* genome in the assembly used in this project to check for potential off-targets (AaegL3, April 2014); similar regions in the genome could easily exist and be currently unaccounted for. In fact, the *Nix* gene itself did not get enough coverage in the AaegL3 assembly and hence was not included as part of the final vector base genome. The *Nix* locus sequence data used in this project came from sequenced M-locus BAC clones later published in Turner *et al.* 2017.

3.3.7 Limitations in the Nix knock-in strategy

The main question arising from this study is why the *Nix Hr5IE1-DsRed2* donor template did not integrate in the M-locus *Nix* gene. A few explanations can be explored; from the repetitive nature of the *Nix* gene and region, to the repressive nature (in terms of chromatin) of y-like chromosomal regions such as the M-locus, to the potential pitfalls of the knock-in strategy itself. Although it should be possible, given the successful genetic modification of the Y chromosome in *Anopheles gambiae* (Bernardini *et al.* 2014).

Although of invaluable prospects, for *Aedes aegypti* genetic control, HDR mediated integrations into the M-locus may not be straightforward. Y-like regions of genomes in different species tend to have particular characteristics. Although not a Y-chromosome, the M-locus appears to be an ancestral version of one, similar to what the common ancestor of *Aedes aegypti* and *Anopheles gambiae* probably had (Toups and Hahn 2010). Y-like regions usually harbour repetitive regions, which is indeed what is found in the M-locus (Hall *et al.* 2015), which may reduce the efficiency and specificity of HDR. In addition, a repressed chromatin environment, due to dosage compensation effects in sex chromosomes (Disteche 2012), may also reduce the efficiency of HDR , or of the expression of the integrated marker gene, or both.

A final concern over the strategy resides in the knock-in objective itself, as it could represent a major pitfall in the strategy. A *Nix* knock-in could in fact be

lethal in itself and this would have prevented any *Nix* specific integrations from being found amongst surviving G1s. A complete lack of *Nix* in early development, where it is usually expressed could be lethal in G1s carrying the knock-in.

3.4 Conclusions

3.4.1 Homology directed repair CRISPR-Cas9 driven strategies can result in misdirected HDR integrations

Even if the mechanism for misdirected integrations during HDR injections is unclear, there is some evidence that misdirected integrations, whether via HDR or not, may not be as uncommon as inititally expected. The misdirected knock-in presented above suggests it is at least possible when trying to get a *Nix* knock-in through HDR.

Moreover, as discussed in this Chapter's introduction, a comprehensive report of HDR in *Aedes aegypti* (Kistler *et al.* 2015) shows that these misguided, or at least unexpected genome integrations are relatively frequent when attempting HDR. Out of 11 independent HDR events (detected by fluorescent marker expression) 4 of them were non-specific. Interestingly some genes seemed specifically refractory to integration. For one of the targeted genes, injections resulted in 3 different G0s bearing fluorescent progeny, but none of the individuals had a directed integration. Conversely, one of the genes showed a total of 6 different G0s bearing fluorescent progeny and all represented directed integrations. The third gene which was successfully knocked-in had 2 different G0s bearing fluorescent progeny but only one of the events was a directed integration. This suggests that some genes are much more amenable to HDR than others. A potential reason for the observations may be that a particular sgRNA is

not efficient at guiding Cas9 to the correct target, or there are other genomic loci which share sequence similarity with the homology arms. However, given the disparity in HDR efficiencies between gene targets and the proportion of misdirected repair this may not be the only reason. There may be some locus specific HDR hurdles such as repressed or repetitive genomic areas which will make integration in certain targets very challenging and hence encourage misdirected integrations if enough embryos are injected.

This chromatin, or otherwise caused, resistance to HDR could prove particularly frustrating for gene drives, especially if the M-locus is one of such regions. In this study a total of 267 G0 adult survivors from *Nix* HDR knock-in injections gave only one marker integration event which was also misdirected. Even when discounting the G0 males, unnecessary as suggested in this study, there was a total of 112 female G0s which is a number greater than for any individual knocked-in gene in Kistler *et al.* 2015. Given that *Nix* sgRNA cutting was confirmed in Hall *et al.* 2015, and there are not any known loci in the genome with similar sequences to the *Nix* homology arms used, the presented body of work may suggest that the *Nix* and perhaps other M-locus knock-ins will be especially challenging in the future.

Chapter 4 - *Act4* Based Gene Drive in *Aedes aegypti*

4.1 Introduction

4.1.1 Mosquito vector-borne and Aedes aegypti control

Mosquito-borne diseases alone are responsible for almost the entirety of the world's insect disease burden (McGraw and O'Neill 2013), it is reasonable to focus efforts on them. Amongst mosquito-borne diseases malaria is the deadliest causing around 600,000 deaths per year (Autino *et al.* 2012). Four species of malaria parasites are traditionally recognised as being responsible for the natural infection in human beings (*Plasmodium falciparum*, *P. malariae*, *P. ovale*, and *P. vivax*) (Mendis *et al.* 2001). The recent upsurge of *P. knowlesi* malaria in South East Asia has led clinicians to consider it as the fifth human malaria parasite as until recently it only infected macaques (Singh and Daneshvar 2013). These different malaria parasites have numerous vector species to transmit it to humans. Over 30 different species of *Anopheles* mosquitoes are capable of malaria transmission, each with different spatial distribution in the world (Kiszewski *et al.* 2004). Even though mosquito transgenesis has been already performed in four species of *Anopheles* mosquitoes; *gambiae*, *stephensi*, *albimanus*, and *arabiensis*, (Catteruccia *et al.* 2005; Perera *et al.* 2002; McGraw and O'Neill 2013) genetic vector control of all the relevant species would be a major challenge.

Dengue is only second to malaria in the number of infections per year and estimates oscillate between 50-100 million cases per year (Mustafa *et al.* 2015). A 2013 study estimates 390 million dengue infections per year with 96 million clinically manifesting (Bhatt *et al.* 2013). The trajectories are different: worldwide malaria cases decreased between 2005 and 2010 from 244 to 216 million despite population increases (Autino *et al.* 2012) whilst dengue cases have increased 30 fold in the last 50 years (Guzman and Harris 2015) (Figure 4.1). Whilst there are several forms of prophylactic treatment for malaria, there are currently no effective treatments against the main four serotypes for human dengue, DENV1-4 (Messina *et al.* 2014). A fifth serotype of dengue virus, DENV-5, has been recently found and, although it mostly infects primates by following the sylvatic cycle (Mustafa *et al.* 2015), this supports the pursuit of genetic vector control solutions.

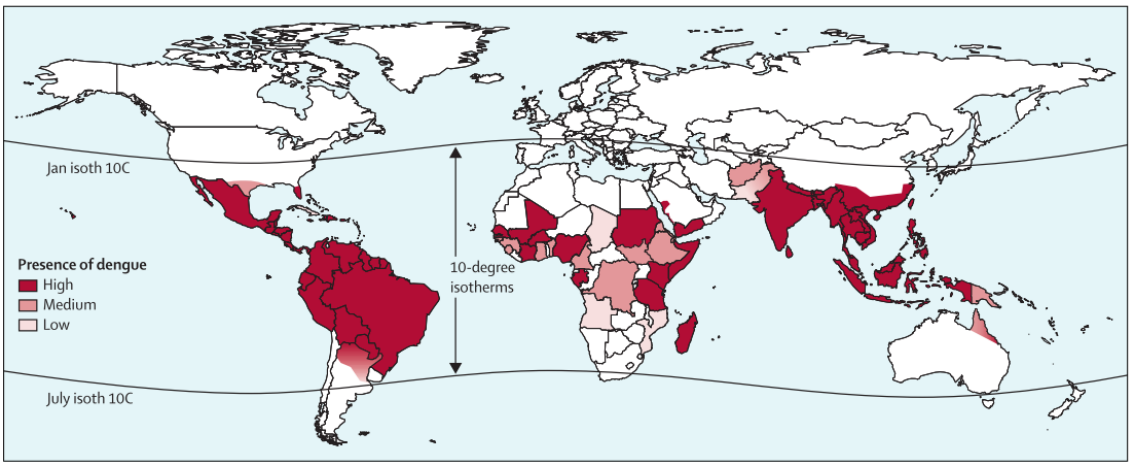


Figure 4.1: Global prevalence of dengue:
Taken from (Guzman and Harris 2015). The map shows the global spread of dengue and its prevalence. This is currently the Americas, parts of Africa, India, and Southeast Asia. Broadly, these areas match areas of *Aedes aegypti* prevalence.

Unlike the breadth of malaria vectors, human dengue is only transmitted by *Aedes aegypti*, to a lesser extent *Aedes albopictus*, and potentially a small number of other *Aedes* species such as *Aedes polynesiensis* (Kraemer *et al.* 2015). This, together with the fact that they occupy distinct enough areas of the world, makes these two species a better choice for genetic control of mosquitoes. Furthermore, population densities of *Anopheles* species in endemic areas may outnumber population densities of *Aedes* species in equivalent areas, making eradication or effective population reduction more feasible. *Aedes aegypti* is also the main mosquito vector for yellow fever and Chikungunya with ~200,000 and ~500,000 cases per year respectively (Barnett 2007; Fredericks and Fernandez-Sesma 2014). It is also responsible for the recent Zika epidemic (Paixão *et al.* 2016). Dealing with a single species, *Aedes aegypti* in this case, would target four different diseases simultaneously. Thus targeting each disease separately, or indeed serotype, is likely to be less efficient.

One of the most commonly used strains of *Aedes aegypti* in research labs is the Liverpool (LVP) strain. Originally from West Africa, it has been maintained at the Liverpool School of Tropical Medicine since 1936 (Vector Base 2017). It was initially selected for greater *Brugia malayi* (filarial worm parasite) susceptibility (Macdonald 1962). The LVP strain has been the one used during this project. A main reason for choosing this strain is that there are good genome assemblies from which to extract quality sequence data from. At the time of this

work the AaegL3, April 2014 assembly was the best up to date LVP genome available.

4.1.2 Haploinsufficiency as a gene drive tool

Haploinsufficiency can be defined as the lack of full function provided by a single copy of a gene. Consequently, mutations in haploinsufficient genes would be dominant mutations in the same way as mutations of haplosufficient genes would be recessive. Haploinsufficiency is rarer than haplosufficiency in diploid organism. One reason for this is the evolutionary advantage of having expression level redundancy from an extra allele and hence a buffer against mutation effects. This begs the question of why haploinsufficient genes exist in the first place. Aside from simple fortuitous occurrence, the reason for haploinsufficiency could be that an excess of a particular gene product incurs a fitness cost. Hence, where this fitness cost outweighs the benefit of allele redundancy haploinsufficiency may be evolutionarily maintained for a particular gene.

Haploinsufficient genes can be used for underdominance based gene drives of surprisingly simple architecture. Only an RNAi and an RNAi insensitive rescue component are required, where both are linked in the same locus, for function. Whilst one copy of the RNAi against the endogenous haploinsufficient gene should be enough to knockdown expression from both alleles, one rescue copy would not be enough to restore function. In the case of lethal or sterile haploinsufficient targets, extreme one locus underdominance can be achieved.

Homozygous RNAi-Rescue individuals would be just as viable as homozygous WT individuals, provided the rescue is complete and no constitutive fitness cost of the transgenes, whilst heterozygotes would be dead or inviable.

This concept has been proposed in insects making use of the well-known haploinsufficient cytoplasmic ribosomal protein (CRP) genes. In *Drosophila melanogaster* mutations in one allele of these genes, of which there are 88, often result in what is called the Minute phenotype. The Minute phenotype confers prolonged development as well as poor fertility and viability (Marygold *et al.* 2007). 64 of the 65 known Minute loci in *D. melanogaster* most likely correspond to CRP genes (Marygold *et al.* 2007). These genes have also been shown to be haploinsufficient in yeast, zebrafish, *Arabidopsis*, humans and mice (Kim *et al.* 2010; Marygold *et al.* 2007). In a particular study a working underdominance system was devised (Figure 4.2) using *RpL14*, a Minute CRP gene, in *D. melanogaster* (Reeves *et al.* 2014). This system was shown to work in *D. melanogaster*, as an effective but reversible underdominance drive.

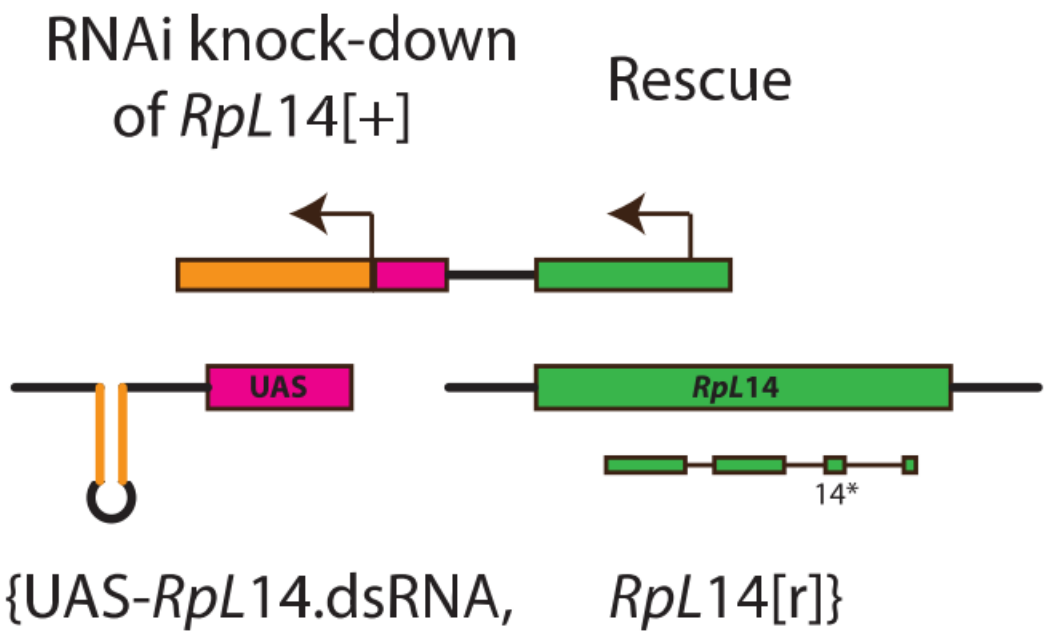


Figure 4.2: Underdominance on *RpL14* haploinsufficiency in *D. melanogaster*:
Taken from (Reeves *et al.* 2014). All transgenic components are inserted at the same locus. The RNAi component is under a UAS promoter making the knockdown inducible and hence facilitating the building of the system for proof of principle. The rescue copy includes the full copy of the endogenous *RpL14* gene with exons and introns as well as regulatory regions. Both population transformation and reversal was demonstrated with this system.

4.1.3 Is Act4 haploinsufficient in Aedes aegypti?

Act88F is one of the few known haploinsufficient genes in insects aside from CRPs. An Actin isoform, *Act88F*, specifically expressed in the indirect flight muscle in *D. melanogaster* was shown to be haploinsufficient for flight in both males and females (An and Mogami 1996). 88F marks the cytological chromosome location of the gene in *D. melanogaster*.

The gene has orthologues in *Aedes aegypti*, *Aedes albopictus*, and *Culex quinquefasciatus*. The orthologue of interest in mosquitoes is called *Act4* and although it is expressed in the indirect flight muscles, just as in *Drosophila*, it does so only in females. There is also a male gene orthologous to *Act88F* in

mosquitoes; however, there is limited use in making males unable to fly in terms of genetic control strategies, since they are not the actual vectors of disease. Evidence from *Act4* Crispr-Cas9 mutagenesis in *Culex quinquefasciatus* suggests *Act4* is haploinsufficient for flight in females from this species of mosquito. A caveat to this is that the mutation obtained in that study was a 6bp in frame deletion which could conceivably be a gain of function dominant negative (antimorphic) rather than a loss of function (amorphic or hypomorphic) allele (Ilona Flis/Luke Alphey, unpublished work).

There is currently no direct evidence of *Act4* haploinsufficiency in *Aedes aegypti* although it is a strong hypothesis given all of the above. Experiments driving a dominant lethal gene under the control of the *Aedes aegypti/albopictus Act4* promoter resulted in flightless females in both *Aedes aegypti* and *albopictus* (Fu *et al.* 2010; Labbé *et al.* 2012). However, this does not necessarily indicate a role for *Act4* in female flight, other than is implied by its expression pattern and putative encoded protein.

4.1.4 Act4 haploinsufficiency would allow for female-specific underdominance

A similar system to the above mentioned CRP underdominance designs can be devised using *Act4* haploinsufficiency instead. The main drive difference would be that whilst CRP based underdominance functions in both sexes, *Act4* based underdominance would be female specific. Such a system could be built using *Act4 RNAi* plus *Act4* rescue components together into a single construct allele.

One copy of the *RNAi* is assumed to be able to practically knock-out endogenous *Act4* function, whilst two copies of the *Act4* RNAi-insensitive rescue would be required for full recovery of function (Figure 4.3). Thus, female-specific underdominance is attained as +/+ and -/- females are fertile and +/- females have flightless derived sterility. Inability to fly is likely also a lethal trait in the field due to inability to avoid predators, though this would not typically apply in a laboratory context.

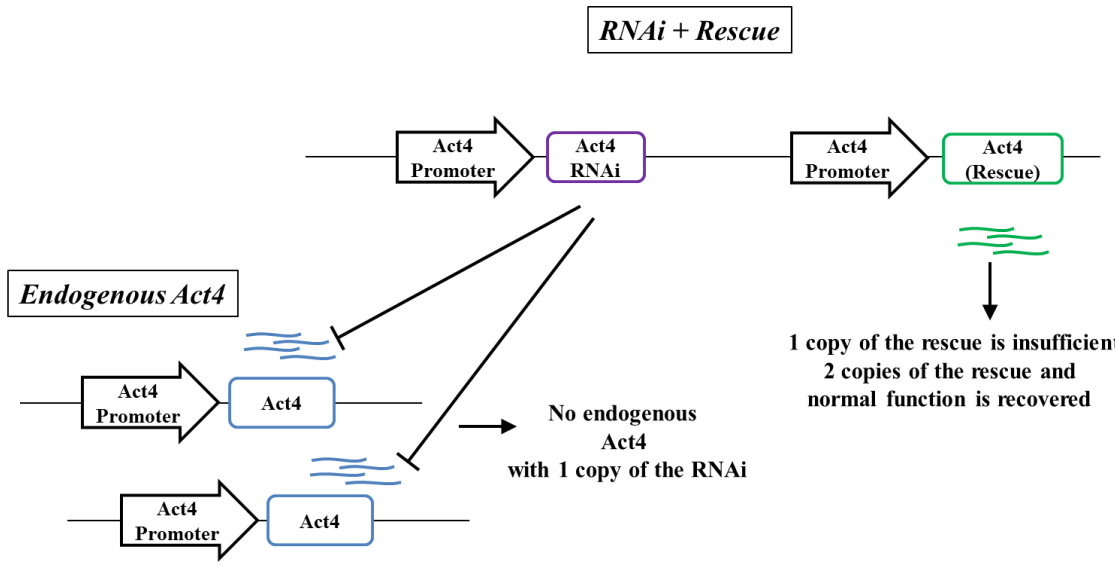


Figure 4.3: RNAi-Rescue system based on *Act4* haploinsufficiency leads to female-specific underdominance:
An *Act4 RNAi* plus *Act4* rescue construct would in theory provide female-specific underdominance, dependent on *Act4* haploinsufficiency. Whilst one copy of the RNAi component should be capable of practically knock-ing out endogenous *Act4* function, a single copy of the RNAi insensitive rescue would not be sufficient for normal *Act4* function. Hence, underdominance is established.

4.1.5 Act4 haploinsufficiency would allow for a RIDL with drive system

As theoretically postulated in (Thomas *et al.* 2000) if a dominant female lethal (DFL) were to be inherited by males at a rate of higher than the normal 50% then a more effective version of RIDL would be created. It was also suggested that

this could be achieved via coupling the DFL to a meiotic drive. Such a system could be called ‘RIDL with drive’ and was modelled to be more effective at population suppression than either SIT or RIDL (Thomas *et al.* 2000).

Implementing this concept together with CRISPR-Cas9 homing into *Act4* would generate such a system provided *Act4* is haploinsufficient and insertions into its coding sequence are dominant sterile. Sterility in this case is akin to lethality as flightless females will not be able to mate or participate in courtship behaviour. Although the females would not be strictly sterile, they could in principle be forced mated in the lab, they would be effectively sterile in the wild.

4.1.6 Act4 haplosufficiency would provide a global or daisy-chain drive target

If *Act4* were haplosufficient the same construct proposed for the RIDL with drive would in turn become a global drive. Female sterility would be recessive and this would allow drive-carrying females to mate successfully and further spread the construct. This raises a concern over the RIDL with drive design as there is a risk of haplosufficient alleles existing in the wild even if *Act4* haploinsufficiency was confirmed in the lab. These alleles would inadvertently change the RIDL with drive system into a global drive. Haplosufficient alleles could exist in the wild, even if the majority are haploinsufficient, with relevant changes in the regulatory region of *Act4*, or the gene itself. This risk can be dealt with by making use of the M-locus in *Aedes aegypti* as described in Section 4.1.7.

A split drive system was proposed in yeast whereby the Cas9 would be separated from the locus containing the relevant sgRNAs for homing (DiCarlo *et al.* 2015). The need for such drives aroused from concerns on the reversibility or control of CRISPR-Cas9 homing drives (Akbari *et al.* 2015). RNA guided homing drives, due to the small release size required for their spread, could be classed as global drives as they could spread worldwide relatively easily with difficult reversibility (Esvelt *et al.* 2014). A split drive, although much less effective for population replacement or suppression, would persist in a population, dependent on its fitness cost, without global spread.

A daisy-chain drive is a development from the split drive concept where sgRNAs and Cas9 components are unlinked. It consists of a linear series of components which drive the next component in the chain and never itself (Noble *et al.* 2016). A chain with three of these elements could have the following configuration; i) The bottom element of the chain, ‘C’ in Figure 4.4, containing Cas9 and sgRNAs targeted at locus ‘B’, ii) ‘B’ in turn consisting of Cas9 only, and iii) the top element ‘A’, called the payload element, could consist of sgRNAs targeting its own homing. Such a configuration could allow for modular construction of daisy chain drives as elements C and B would work for different payload elements. The payload elements would ideally be recessive knock-out insertions of target genes haplosufficient for fertility or life, preferably female-specific. In fact, different ‘A’ elements could be used in the same daisy chain drive if multiple

targets were to be used. Hence, if *Act4* were to be haplosufficient, it would make an ideal 'A' target for daisy-chain drives.

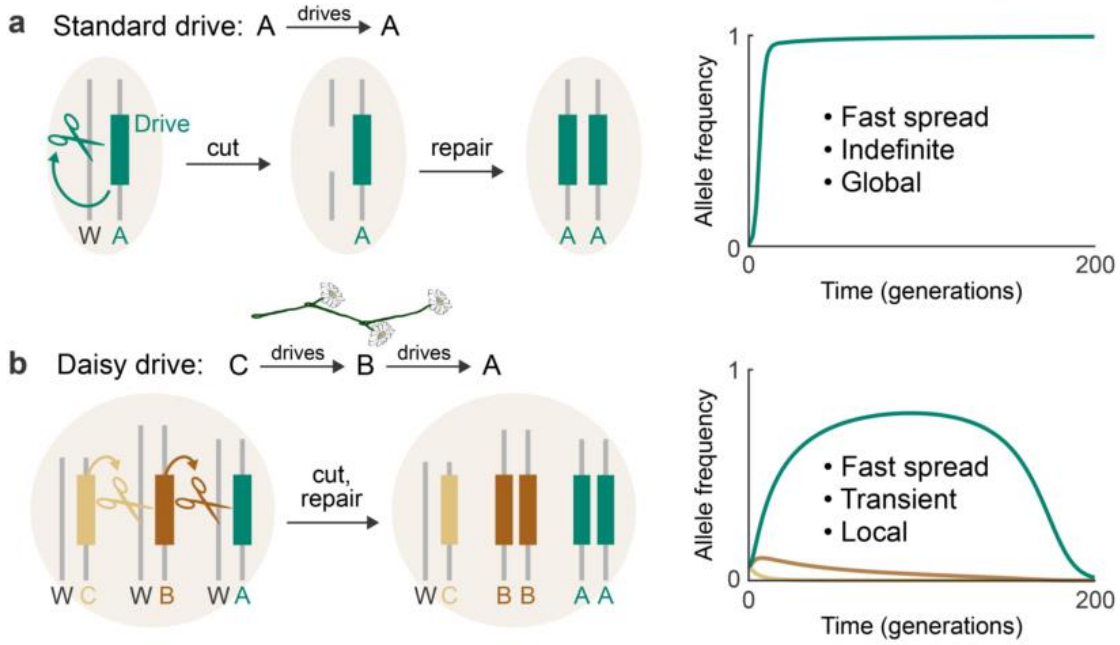


Figure 4.4: Daisy-chain drives, unlike global drives, allow a fast but transient spread of local populations:
Taken from (Noble *et al.* 2016). **(A)** Standard drive versions are modelled to spread quickly and indefinitely and hence potentially becoming global. **(B)** On the other hand a daisy drive has several elements driving each other sequentially. The bottom of the chain 'C' could consist of Cas9 and a sgRNA targeting the locus of 'B'. 'B' in turn would have Cas9 only. And the top element 'A' (the payload or target gene) would have sgRNAs to promote its own homing. This chain would increase the frequencies of 'B' and to a much greater extent 'A' but only transiently. The bottom element would never undergo homing and hence would be naturally lost from a population over time. Gradually, all elements from the chain would also drop out from the population as there is no basal element to drive them anymore.

A further challenge arises in this case where intermediate elements of the chain require conserved recognition sites in the genome as well as 'inert' integration sites, i.e. which impose a low to no fitness cost. These two requirements are, to an extent, incompatible as conserved regions of a genome are usually required for function. To circumvent this, conserved loci could be chosen as 'B' elements

as long as a sgRNA insensitive rescue copy was included within the elements themselves (Noble *et al.* 2016). CRPs could be potential 'B' loci, as well as having a relatively conserved sequence, if a haploinsufficient sequence is used as the insertion site (with suitable restoration of function), some types of cut-resistant mutants would be non-viable. This therefore would also reduce the rate of resistant allele formation.

4.1.7 Act4 drive components in the M-locus of Aedes aegypti

As described in the previous Chapter, *Aedes aegypti* does not have an X or a Y chromosome. Instead, it has a male-specific genomic region called the M-locus in a Y-like portion of the chromosome (Hall *et al.* 2015). The *Nix* gene, within the M-locus, has been shown to be the male determining factor in *Aedes aegypti* as its knock-out results in feminised genetic males (Hall *et al.* 2015). It has been recently shown to be around 100kb in length owing to the large 99kb intron present (Turner *et al.* 2017).

Hence, there is an opportunity to have gene drive components linked to the M-locus. The *Nix* gene was omitted from the AaegL3, April 2014, genome assembly (Nene *et al.* 2007), due to the repetitive nature of the M-locus and the underrepresentation in terms of sequence reads due to its absence from females. However, a new PacBio assembly provides more robust data on the M-locus ((Matthews *et al.* 2017)), and could be a starting point for M-locus integration of

components. Integrating components downstream of *Nix* would be an appropriate initial objective.

As mentioned above, even if *Act4* haploinsufficiency in lab LVP strains was confirmed, haplosufficient alleles could exist in nature. This could turn a RIDL with drive system into a global drive inadvertently, by resulting in sgRNA-Cas9 driven homing in females. Integrating Cas9 into the M-locus instead of *Act4* (with the sgRNA element in *Act4*) would generate a drive with the same dynamics as the RIDL with drive system described above (Section 4.1.5), except it would be much less prone to uncontrolled spread. Cas9 would only ever be present in males and hence homing at the *Act4* locus would be restricted to males, while the Cas9 would not drive and so would limit the geographic spread and (assuming a fitness cost to Cas9) the temporal persistence of the transgene system. The difficulty of homology directed repair into a highly repetitive genome region such as the M-locus could be a challenge. However, since integrating components in the M-locus could have many applications (Adelman and Tu 2016), including as daisy chain elements, it would be a valid pursuit.

4.1.8 sgRNA multiplexing and off-cutting Cas9 as ways to avoid resistant alleles

Resistant allele formation is one of the biggest hurdles for sequence specific gene drives (Champer, Reeves, *et al.* 2017). Upon endonuclease double strand breaks (DSBs) cell repair goes through one of the two main repair pathways; non-

homologous end joining (NHEJ) or homologous directed repair (HDR) (Jasin and Rothstein 2013). Whilst homologous directed repair is a key requirement for homing, the more common NHEJ pathway does not allow homing and is an error-prone repair process (Lieber 2010). This can result in mutations, from SNPs to small insertions and deletions (INDELs), in the endonuclease targeted site. Changes at the target sequence may result in cut resistant alleles. Provided these alleles would generate functional proteins, these would not only prevent homing but they could compete effectively against the drive should it carry a fitness cost. To prevent this drive dampening or crashing effect, resistant allele formation must be accounted for.

With regards to CRISPR homing drives, multiplexing gRNAs is perhaps the most obvious and widely proposed solution (Xie *et al.* 2015), as a homing drive should tolerate a certain rate of resistant allele formation provided there is cleavage redundancy from several gRNAs against the same target. Modelling in Noble *et al.* 2017 suggested 5 multiplexed gRNAs would be required to overcome drive dampening from resistant allele formation. Since one of the assumptions was that all NHEJ mutations would generate resistant alleles, and this is in practice unlikely, the number of multiplexed gRNAs required for ultimate population replacement or elimination could be somewhat smaller. Modelling in Marshall *et al.* 2017, which does try to calculate fully-resistant allele formation rates, suggests multiplexing of at least 2 gRNAs for a 90% probability of eliminating a population of 10,000, 3 gRNAs for a population of 1

million, and 4 gRNAs for a population of 10 billion. Natural SNPs were not considered in this study, although they would also pose a challenge to homing drives, together with the formation of resistant alleles through mutation. Two of the main gRNA multiplexing solutions proposed involve the use of 5' and 3' ribozymes upstream and downstream of each gRNA (Gao and Zhao 2014), or the use of tRNA self-processing abilities (Xie *et al.* 2015), all occurring within the same RNA transcript.

A second alternative would be to seek endonuclease alternatives to Cas9 with staggered or off cleavage capacity with respect to the target site. Should such an RNA guided endonuclease exist, the target site would be more immune to resistant allele formation. Although not as widely discussed in the literature, this possibility has been suggested. Gantz *et al.* 2015 proposed the use of Cpf1 instead of Cas9 owing to its staggered cut (Zetsche *et al.* 2015) which, even if still within the target sequence, it will cleave at a less critical region for recognition. Cpf1 has the additional advantage of being able to process its own gRNAs and hence facilitating their multiplexing (Zetsche *et al.* 2016).

4.1.9 The importance of germline promoters in homing gene drives

Expression of sgRNAs in insects currently relies on RNA pol III promoters *U6 1-3* (Port *et al.* 2014), which which is thought to result in ubiquitous expression in *D. melanogaster* just as their orthologues in insects such as *Aedes aegypti* (Dong *et al.* 2015), *Anopheles gambiae* (Hammond *et al.* 2016) or *Anopheles stephensi*

(Gantz *et al.* 2015). RNA pol III promoters usually ubiquitous and constitutive in their expression (Snyder *et al.* 2009). Hence, when referring to CRISPR-Cas9 drives, the expression domain of *Cas9* is what delineates CRISPR-Cas9 activity as sgRNAs are not the limiting factor.

Homing drives necessitate germline expression for function. Without germline expression of the relevant nuclease, HDR homing events would not be transmitted to the next generation. Since, supra-Mendelian inheritance is crucial, so is germline expression. Conversely, HDR homing events beyond the germline could bring a range of disadvantages for a homing drive. Unnecessary off-site Cas9-sgRNA presence could lead to somatic cutting and result in a general fitness cost. Therefore, an ideal promoter to drive *Cas9* in homing drives would exclusively drive expression in the germline. Sex-specific or stage-specific germline expression requirements may vary with specific applications. With regards to timing, germline promoters with pre-meiotic expression of transcripts are required since post-meiotic transcription undergoes repression (Olivieri and Olivieri 1965; Gould-Somero and Holland 1974), except for marked exceptions (Barreau, Benson, Gudmannsdottir, *et al.* 2008; Barreau, Benson and White-Cooper 2008; Vibranovski *et al.* 2010). Moreover, once meiosis is complete after meiosis II, cells become haploid which eliminates any chance of homologous recombination (Marston and Amon 2004). Late germline expression is also to be avoided as it may increase the chance of transcript and/or protein contribution of

Cas9 and *sgRNAs* into eggs and/or sperm. This may result in unwanted somatic cutting in the embryo as well as contributing to resistant allele formation.

Such homing drives have already been designed as proof of principle in *D. melanogaster* (Gantz and Bier 2015), *Anopheles gambiae* (Hammond *et al.* 2016), and *Anopheles stephensi* (Gantz *et al.* 2015). They each used species specific orthologues of the *vasa* promoter. However, none of these *vasa* promoters fit the ideal germline promoter criteria, for homing drives, as outlined above. Observations in the three studies illustrate how the choice of promoter is important for effective homing drive design. Although how crucial is the choice will depend on the design context of a particular homing drive. For instance, for homing drives targeted at recessive female sterile targets such as in Hammond *et al.* 2016, a suboptimal choice of promoter can dampen the homing drive to the point of collapse, as shown in a subsequent analysis of the same study (Hammond *et al.* 2017). Another promoter expressing in the male germline is *nanos*, which although shows reduced leaky somatic expression with respect to *vasa*, it also readily results in resistant allele formation in *D. melanogaster* CRISPR-Cas9 homing drives (Champer, Reeves, *et al.* 2017; Champer, Liu, *et al.* 2017). The *β2 tubulin* promoter of *Aedes aegypti* may be a better choice as it is male germline specific, and its transcript is only found in the early spermatogenesis fraction of testes by *in situ* hybridisation (Smith *et al.* 2007).

Two recent studies have sought male and female germline specific promoters in *Aedes aegypti* (Sutton *et al.* 2016; Akbari *et al.* 2014). Amongst the best male germline specific promoters in Sutton *et al.* 2016 is the *D. melanogaster mitoshell* orthologue in *Aedes aegypti* (AAEL010268). This gene was selected on the basis of strongly expressing in early and not late spermatogenesis as well as almost absent in gonadectomised males, ovaries and females. *matotopetli* (or *topi*) is another gene which was shown to be expressed exclusively in primary spermatocytes in *D. melanogaster* (Perezgasga *et al.* 2004). Amongst germline promoters expressing in both males and females in *D. melanogaster* are *zpg* and *bgm*. Promoters with such expression would be useful for bi-sex homing drives.

4.1.10 Project aim

The aim of the project is to use the requirement of *Act4* in *Aedes aegypti* females to build two different gene drives of different persistence; female specific underdominance and RIDL with drive.

4.2 Materials and Methods

4.2.1 Aedes aegypti maintenance

The *Aedes aegypti* mosquitoes used for transformation and crossing experiments in this Chapter were either from Latin wild type (LWT) and Liverpool (LVP) strains. LWT is a Mexico-derived strain (Harris *et al.* 2011), whilst the LVP strain originated from West Africa and was maintained in the Liverpool School of Tropical Medicine since 1936 (Vector Base 2017). Transgenic lines maintained were; *exu-Cas9* (*nanos-Cas9*) and *Act4* potentially different HDR knock-in insertions (2B+ to 2P+) lines.

Mosquitoes were reared using previously described standard procedures for *Aedes aegypti* (Crampton *et al.* 1997). Where *Aedes aegypti* rearing procedures in The Pirbright Institute (Chapter 4) were the same as in Oxitec (Chapter 3) repetition will be avoided (Section 3.2.1). Briefly, the main differences are outlined below.

LVP individuals to be used for injections were maintained by the mosquito team, and adults were reared in large cages rather than the standard small ones used for all other transformation and crossing experiments. Similarly, *exu-Cas9* individuals for injections were also reared to fit larger cages. Blood meals for egg collection were also supplied differently; the Hemotek PS5 system (Hemotek Ltd, UK) was used to heat the blood to 37°C during feeding. Hemotek feeders

were prepared by extending sheep intestine cut-outs over them, followed by a Parafilm M layer (Bemis) and securing the makeshift membrane with the black rings provided by the system. Defibrinated horse blood (TCS Biosciences) was then added and the feeders were sealed with plastic plugs. Blood meals were between 30 to 50 minutes long depending on how fast females would feed. Eggs were collected on unbleached coffee filter paper (Edesia Espress) placed in 1oz polystyrene portion cups (Fabri-Kal) filled with water to keep them wet over the egg laying period. Eggs were then desiccated and hatched after three days or stored for up to four months. Vacuum hatching of the eggs was carried out as described above (Section 3.2.1). However, to screen G1 egg papers for potential transgenics, these were hatched directly on water-filled trays with Liquifry No1 (Interpret) to allow for a more 'natural' hatch and reduce the embryo losses thought to occur with vacuum hatching. Deionised water for mosquito rearing was provided by the Milli-Q system (Elix Technology Inside).

4.2.2 Preparation of solutions for injections

Two different types nuclease-free water were used for molecular biology in this Chapter. DNA work used Water for Molecular Biology (Millipore) whilst RNA work used DEPC Treated Water (Ambion). Since injection solutions all contained sgRNAs, the water of choice was DEPC treated water. For the preparation of injection solutions (Table 4.1); injection buffer was added for a final concentration of 5mM KCl and 0.1 mM NaH_2PO_4, pH 6.8. *AGG1070* plasmid was the *Act4* donor template which had a *3xP3-mCherry* marker flanked

by 2kb *Act4* homology arms reaching up to the predicted cleavage site for *Act4* sgRNA 1 and 3. Two main injection solutions were made, one with and one without Cas9 protein, to be injected into *exu-Cas9* and LVP embryos respectively. *Act4* sgRNA combinations and concentrations (40-100 ng/µl) were altered in some injection dates, as well as the presence and concentration of the *Act4* HDR donor template. Details of which are specified in the results below (Table 4.8, Table 4.11, and Table 4.12). Concentration changes respond to the distinct recommendations in the literature (Basu *et al.* 2015; Kistler *et al.* 2015), whilst different combinations of sgRNAs was thought to provide a general idea on whether some of the *Act4* sgRNAs were more efficient at guiding Cas9 cutting than others.

Table 4.1: Components of injection solutions:

Cas9 protein (PNA Bio Inc) was injected at concentrations recommended by the provider. Two types of injection solutions were made, with or without Cas9 protein, to be injected into *exu-Cas9* and LVP embryos respectively. *Act4* sgRNA concentrations were varied between 40ng/µl to 100ng/µl, for different injection dates and not all *Act4* sgRNAs were always included as indicated in results (Table 4.11 and Table 4.12). Similarly some injections did not include the *AGG1070* donor template as indicated in results, and some other days the concentration supplied was changed (Table 4.8).

Cas9 (injected into LVP)	**No Cas9 (injected into *exu-Cas9*)**
Act4 sgRNA 1 (40-100 ng/µl)	*Act4* sgRNA 1 (40-100 ng/µl)
Act4 sgRNA 2 (40-100 ng/µl)	*Act4* sgRNA 2 (40-100 ng/µl)
Act4 sgRNA 3 (40-100 ng/µl)	*Act4* sgRNA 3 (40-100 ng/µl)
Act4 sgRNA 4 (40-100 ng/µl)	*Act4* sgRNA 4 (40-100 ng/µl)
AGG1070 donor template (700 ng/µl)	*AGG1070* donor template (700 ng/µl)
Cas9 protein (300 ng/µl)	Injection buffer
Injection buffer	DEPC treated water
DEPC treated water	

4.2.3 Aedes aegypti injections

Microinjection of 1 to 2 hour old embryos was carried out mostly following established procedures (Morris 1997). Colour guided staging of embryos as

described in (Lobo *et al.* 2006) allowed selection for injection of embryos of a suitable developmental stage. Embryos were kept at RT (~22°C) whilst lining-up embryos for injection. Embryos too dark to be injected were discarded. Embryos were collected as described above from the respective injection cages (LVP, or *exu-Cas9*). Injection cages were kept in the dark during embryo collection to encourage laying. Occasionally, when embryo collection directly from the cage was scarce, around 20 to 30 females were taken out of the cage into a 50ml tube (Fisherbrand), containing an egg-collection coffee filter paper (Edesia Express), kept moist by wet cotton at the bottom of the tube. The crowded setting appeared to encourage egg-laying in *Aedes aegypti* females. Injections were carried out using a Femtojet 4x (Eppendorf) pneumatic back-pressure system. Needles were pulled from aluminosilicate glass (1.0mm x 0.64mm x 10cm, Sutter Instrument) by needle puller Model P-2000 (Sutter Instrument). They were then bevelled before the injection mix (kept on ice and centrifuged at 11000 x g) was loaded using microloader tips (Eppendorf). Microscope slides 76 x 26mm (Academy) were used as support for plastic coverslips (Fisherbrand). Line-up of embryos for injection was carried out on wet 55mm filter paper (qualitative, Fisherbrand) and embryos were transferred to the taped edge (Scotch double-sided sticky-tape) of a plastic coverslip by applying just enough pressure. Embryos were desiccated just enough to allow for effective liquid uptake from the needle. To stop fatal desiccation halocarbon oil 27 (Sigma) was pipetted over the embryos. The halocarbon oil used to cover the embryos was rinsed -off after injections and embryos were removed gently from the coverslips with a thin paintbrush and

transferred onto new coffee filter papers kept in a closed chamber (with 100% relative humidity) for three days prior hatching.

4.2.4 Aedes aegypti injected G0 crossing

Both male and female G0s were outcrossed to LVP individuals. Females were divided into G0 pools of no more than 20 G0 individuals each at the time of crossing. The sex ratios were kept 1 to 1 when the pool had G0 females, but instead were increased to 1 to 5 (male to female) when the pool had G0 males. Males however were not divided into pools before crossing. Crossing of males took place individually in separate pots before G0s were pooled in cages (3 days after crossing) of no more than 20 G0 individuals. Individual male crosses are thought to give weaker males a chance to mate with females rather than be outcompeted should all males be crossed in a shared pool. Blood meals and general maintenance was carried out as described above. 5 ovipositions were collected for each G0 pool and hatched without storing (Section 4.2.1) for subsequent screening (Section 4.2.5).

4.2.5 Aedes aegypti transgenic screening

The progeny from G0 or G1 crosses were screened at L2-L4 larval stages when screening for the transgenic *3xP3-DsRed* marker (*Act4* integration), which shows eye-specific expression, or *Opie1-DsRed* (*exu-Cas9*) which shows irregular whole body expression. Screening was performed using the appropriate light filters on a fluorescence stereo microscope (M165 FC, Leica). G1 or G2 adults

were screened for flight-impaired phenotypes by transferring pupae into cages and monitoring eclosion and flight. Adults were provided with low access sugar feeders for potential flightless females. Photos were taken with a DFC365FX camera (Leica) connected to the stereomicroscope and to a computer under the control of Leica application suite software.

4.2.6 sgRNA synthesis

Act4 sgRNAs 1 to 4 were designed by Sanjay Basu bearing in mind off-target software predictions. Synthesis of *Act4* sgRNAs 1 to 4 involved the generation of a DNA template for RNA transcription by T7 RNA polymerase. The template was generated by a single DNA polymerase extension cycle between two long overlapping DNA oligos. The first oligo included the T7 promoter, the target site (specific for each *Act4* sgRNA), and part of the sgRNA backbone (overlapping with the second oligo). Table 4.2 shows the oligo sequences used. The second oligo encoded the sgRNA backbone common to the four sgRNAs. After transcription by T7 polymerase and RNAse-free RNA purification, the four sgRNAs were ready for injection.

Table 4.2: DNA oligos for *Act4* sgRNA synthesis:
The first 4 oligo DNA sequences are sgRNA target specific oligos (Forward) whilst the 5th oligo in the list is the general backbone oligo (Reverse), for all *Act4* sgRNA template extension reactions.

Oligo Code	Oligo Description	Oligo Sequence
LA138	Act4 sgRNA 1	GAAATTAATACGACTCACTATAGGGGAGCACTAGTCA TTGACAAGTTTTAGAGCTAGAAA
LA139	Act4 sgRNA 2	GAAATTAATACGACTCACTATAGGGGTCAAAAAGATG CCTACGTGTTTTAGAGCTAGAAA
LA140	Act4 sgRNA 3	GAAATTAATACGACTCACTATAGGGTGCTCTATGGGA TATTTCAGTTTTAGAGCTAGAAA
LA141	Act4 sgRNA 4	GAAATTAATACGACTCACTATAGGGTGGCGAGGGCGG CCGACAAGTTTTAGAGCTAGAAA
LA137	sgRNA General Backbone	AAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATA ACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCT AAAAC

4.2.7 In vitro Cas9 cleavage assays

In vitro digestion of *Act4* DNA with *Act4* sgRNAs and Cas9 was used as an assay of sgRNA guiding activity. An *in vitro* digestion kit (New England Biolabs) containing Cas9 Nuclease from *S. pyogenes*, was used by following the procedure recommended by the manufacturer (Table 4.3). Only 1 *Act4* sgRNA was used per reaction (4 different sgRNAs). Two sets of *Act4* sgRNA preparations (Section 4.2.6) were tested independently. The first set lacked *Act4* sgRNA 1 and hence the total number of reactions was 7. A pre-incubation period of 25°C (10 min) was then followed by incubation at 37°C for 15 minutes.

Table 4.3: Components of *in vitro* Cas9 digests:
Only one sgRNA was used per reaction. Components were incubated as described above. DNA fragments were analysed by gel electrophoresis to check for the expected sequences.

Component	Concentration (In a 30µl Reaction)
10X Cas9 Nuclease Reaction Buffer	1X
Act4 sgRNAs (1 to 4)	30nM
Cas9 Nuclease	30nM
Act4 DNA Template	3nM

70ng of the *Act4* template was used per reaction. Primers used to generate the *Act4* amplicon from gDNA extracted from an individual female LVP (Section 4.2.13) are shown in Table 4.4. A visual representation of the genome region amplified together with the *Act4* sgRNA binding sites (Figure 4.5).

Table 4.4: *Act4* amplicon primers:
Primers used to generate the *Act4* 1276bp template for sgRNA-Cas9 cleavage are shown below.

Primer Name	Primer Sequence
LA410	CCCGATTCCCTCAACTTTGATTGTTA
LA130	GTACAGGGACAGAACAGCTTGGAT

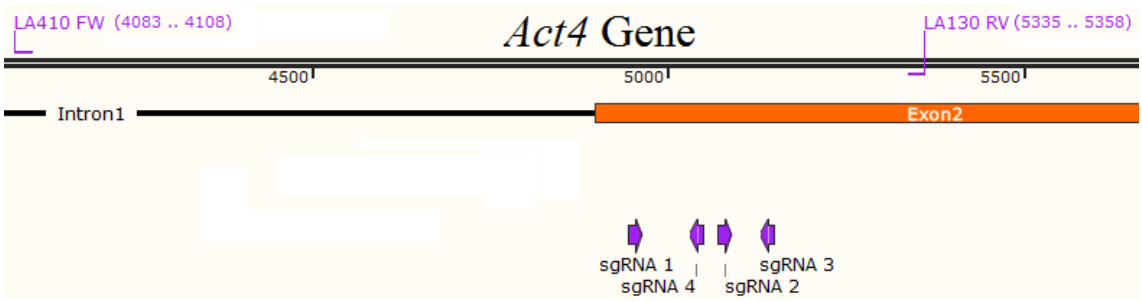


Figure 4.5: *Act4* template for sgRNA guided *in vitro* Cas9 cutting:
The gene map was drawn using SnapGene. A fragment of the *Act4* gene is shown, with the respective primers for *in vitro* template amplification. *Act4* sgRNA 1 to 4 binding sites are shown at the beginning of Exon 2.

4.2.8 sgRNA stability assay

A stability test was performed to assess potential RNAse contamination of sgRNA preparations (Section 4.2.6). 100ng of sgRNAs 1 to 4 at 10ng/µl concentrations were incubated at two different temperatures for 7 hours (24°C and -80°C). Potential degradation at 24°C was checked by analysing the full 100ng of each sample by gel electrophoresis.

4.2.9 Gene recoding, codon optimising, intron detection, and gene synthesis

As a result of the degeneracy of the genetic code (Crick 1966) gene sequences can be altered without a change in the encoded protein sequence. This is especially useful when a genetic rescue of function is required where RNAi against the endogenous gene is being used. The beginning of *Act4* exon 2, 445bp, was recoded to make the new sequence insensitive to the designed *Act4* sgRNAs above (Section 4.2.6). Since codon usage varies between species as well as the relative abundance of different isoacceptor tRNAs, the recoded *Act4* sequence was corrected so as to not include rare codons in *Aedes aegypti*. To be able to determine which codons were rare in *Aedes aegypti* the data provided in (Behura and Severson 2011) was used. In particular relative synonymous codon usage (RSCU) values were used, since absolute codon values in a genome could misrepresent the rarity of a particular codon, especially regarding overused or underused amino acids. In addition, base-pairs were changed to avoid cryptic splice sites. To predict potential cryptic splice sites online software was used (Neural Network 2017), based on pre-existing splice site prediction code (Reese *et al.* 1997). Ultimately, these designed DNA sequences were outsourced for synthesis (Genewiz).

4.2.10 Golden Gate cloning of constructs

Golden Gate cloning involved the use of off-cutting restriction enzymes (REs) BsaI and BsmBI which allow for custom non-palindromic cohesive ends to be designed on any fragment of choice. RE sites were added on oligonucleotide

primers (Sigma) and fragments of constructs shown below (Section 4.3.5) were PCRed or digested from a plasmid as appropriate to then be digested by the relevant RE. Multiple fragments were ligated (up to 3 inserts and one vector at a time) using T4 DNA ligase (NEB). Transformations were carried out into 10µl of XL10-Gold Ultracompetent Cells (Stratagene) following the recommended instructions. Analysis of correct clones was carried out with colony PCR (Section 4.2.14), and subsequently Sanger sequencing (Section 4.2.16).

4.2.11 Genomic, plasmid, and gel DNA extractions

All Macherey-Nagel DNA extraction kits were used by following instructions from the manufacturer. gDNA extractions were carried out using the Nucleospin Tissue kit (Macherey-Nagel). Plasmid DNA extractions for cloning procedures were carried out using the Nucleospin Plasmid kit (Macherey-Nagel). Nucleospin Gel and PCR clean up kit (Macherey-Nagel). Endonuclease free plasmid DNA preparations for injections were carried out using the Nucleobond xtra midi EF kit (Macherey-Nagel).

4.2.12 General PCR

Where sequence fidelity was required, i.e. Golden Gate cloning, LVP and LWT strain sequence analysis, Q5 High-Fidelity DNA Polymerase (NEB) was used due to its proof-reading activity. However, for more routine diagnostic PCRs such as colony PCR the more error-prone DreamTaq Hot Start DNA Polymerase (Thermofisher) was used.

4.2.13 Gene sequence analysis of strains employed

gDNA was extracted for 6 pooled LVP adults (3 males and 3 females). gDNA was extracted for 5 individual LVP flying females and 3 LWT flying females.

Two 2kb *Act4* amplicons were PCRed from the pooled LVP gDNA and cloned into pJET1.2. Two different clones for each were selected for sequencing. Three ~2kb amplicons covering ~6500bp of the *Act4* locus were PCRed from each of the individual samples. Primers used for amplicon PCRs are shown below (Table 4.5).

Amplicons for two individual female LVPs and one individual female LWT were selected for sequencing. A range of primers were used to cover all the samples selected for sequencing (Table 4.6). A small number of clones were sequenced as the main objective of this procedure was to generate homology arms that would be present at least in a proportion of the population.

Table 4.5: Primers used to amplify 2kb *Act4* amplicons in LVP and LWT:
Primers used to amplify the fragments cloned into pJET or sequenced directly are shown below.

Primer Name	Description	Sequence
LA133	pJET cloned 3' Amplicon	GATTTGGCATCACACCTTCTACAAC
LA134		TCAAAATACGACCGTCACCCTATTA
LA135	pJET cloned 5' Amplicon	TCAATGACTAGTGCTCCAGCATCAT
LA136		TCTCAATCGATCCTTCTATGGACTG
LA405	5' *Act4* Amplicon	CTCATGGAGACCTATCTACCCTTCA
LA420		ATCAACATCACATCGGATCGTAATC
LA136	Mid *Act4* Amplicon	TCTCAATCGATCCTTCTATGGACTG
LA130		GTACAGGGACAGAACAGCTTGGAT
LA133	3' *Act4* Amplicon	GATTTGGCATCACACCTTCTACAAC
LA422		AATCAGTTTATCCATGCAACCGTTA

Table 4.6: Primers used for sequencing the *Act4* locus in LVP and LWT:
A list of primers used to cover the ~6500bp region of the *Act4* locus.

Primer Name	Sequence
LA129	TGCCACATGAACAATAACACCAATA
LA133	GATTTGGCATCACACCTTCTACAAC
LA135	TCAATGACTAGTGCTCCAGCATCAT
LA136	TCTCAATCGATCCTTCTATGGACTG
LA159	CGACTCACTATAGGGAGAGCGGC
LA160	AAGAACATCGATTTTCCATGGCAG
LA370	CGTGCATCAGGCTTTGCCAA
LA371	TTACTGCGGACGTAATTCAC
LA372	GGTTGTTCTTATTCCGTAAA
LA373	CCAGCCATCCTTCCTGGGAA
LA405	CTCATGGAGACCTATCTACCCTTCA
LA406	GCCTTAATTGAACAAACTTTTGAAAG
LA408	GTAGTTGAACATCTTCAGACTTCTGG
LA410	CCCGATTCCCTCAACTTTGATTGTTA
LA412	GCCAGGGAAGAAACCATCATTCACC
LA422	AATCAGTTTATCCATGCAACCGTTA
LA423	AAGTAGCATTTTGAACCACTTTTCG

4.2.14 Colony PCR

Colony PCR was extensively used to check for constructs cloned (Section 4.2.10). For each cloned construct in question a number of colonies (≥20) were screened at a time. Primers used were specific for the cloned plasmid and, where possible, pair of primers would bind both within the insert and within the vector to prevent false positives. 5µl of molecular biology water (Millipore) were added to each of the PCR tubes. An agar plate, or more if required, had grid marks added so that individual squares could be seen. Labelling of both the PCR tubes and the agar plates was done so as to be able to trace back each individual colony PCR reaction to its plate growth. Pipette tips were used to pick up individual colonies which were dipped into PCR tubes and then streaked on their respective agar plate square (which would then be incubated 8 hours at 37°C).

A PCR master mix, using DreamTaq DNA polymerase (Thermofisher) at recommended manufacturer concentrations for a final 25µl volume was prepared for the appropriate number of reactions and then pipetted accordingly into each of the PCR tubes. A standard PCR program was used, following Taq manufacturer recommendations except for the reduction in the number of cycles to only 25. There is a risk that too many cycles would start detecting free insert DNA that would have been streaked on the agar plates following transformation (Section 4.2.10). A lower number of cycles, hence prevents false positives. The colony PCRs were then run on a gel and positive colonies were then traced back to the freshly streaked agar plates for the preparation standard overnight bacterial growth cultures in LB. Plasmid extractions were carried out as described, depending if the growth was for a miniprep or a midiprep (Section 4.2.11).

4.2.15 NHEJ and HDR assays

AGG1070 plasmid was the Act4 donor template which had a *3xP3-mCherry* marker flanked by 2kb *Act4* homology arms reaching up to the predicted cleavage site for *Act4* sgRNA 1 and 3 (Figure 4.6B) which are within the beginning of exon 2 of Act4 (Figure 4.6A). Figure 4.6C shows a putative *Act4* marker-only knock-in. Primers designed for NHEJ and HDR assays performed on G0 embryos are shown (Figure 4.6).

Two different batches of 40 LVP embryos were injected with *Act4* sgRNAs, Cas9 protein, and *Act4* marker-only HDR donor (Figure 4.10). A batch of 40

LVP embryos was injected with water as a control. gDNA was extracted from the three batches 1 day post injection for NHEJ and HDR downstream analysis in G0s.

Restriction fragment length polymorphism (RFLP) was used to detect NHEJ mosaic events in the G0 samples. Extracted gDNA was digested overnight with BbsI and BamHI enzymes in order to digest WT DNA *Act4* templates as well as the HDR donor plasmid itself. Should a deletion have occurred between *Act4* sgRNA 1 and 3 a PCR product should be still seen in the digested gDNA samples whilst absent in undigested and water-injected gDNA samples. In addition, even in the undigested gDNA samples, a smaller fragment from a putative deletion should be seen together with the longer WT *Act4* amplicon. For HDR events, the analysis involved running two rounds of PCR trying to amplify two fragments only present in the putative *Act4* knock-in as shown (Figure 4.6). The undigested gDNA samples were used for this purpose. A summary of the diagnostic primers used is shown in Table 4.7.

Table 4.7: NHEJ and HDR assay primers:

Primers used in the diagnosis of NHEJ and HDR mosaic events, upon *Act4* knock-out and knock-in respectively, in G0 embryos are shown below.

Primer Name	Description	Primer Sequence
LA818	1st Round 5' HDR Assay	GAAACCTCACCCGAACGCAC
LA29	1st Round 5' HDR Assay	GGAGCGCGTGATGAACTTCGAGG
LA136	2nd Round 5' HDR Assay	TCTCAATCGATCCTTCTATGGACTG
LA819	2nd Round 5' HDR Assay	ACTCATCAATGTATCTTACTCGACC
LA635	1st Round 3' HDR Assay	GTTTAGCTTGTTCAGCTGCG
LA422	1st Round 3' HDR Assay	AATCAGTTTATCCATGCAACCGTTA
LA816	2nd Round 3' HDR Assay	TAATTGAATTAGATCCCCGGGCG
LA817	2nd Round 3' HDR Assay	ACGACCGTCACCCTATTATTTCTC
LA146	NHEJ Assay	GTGTGACGATGATGCTGGAG
LA147	NHEJ Assay	CCATATCATCCCAGTTGGTG

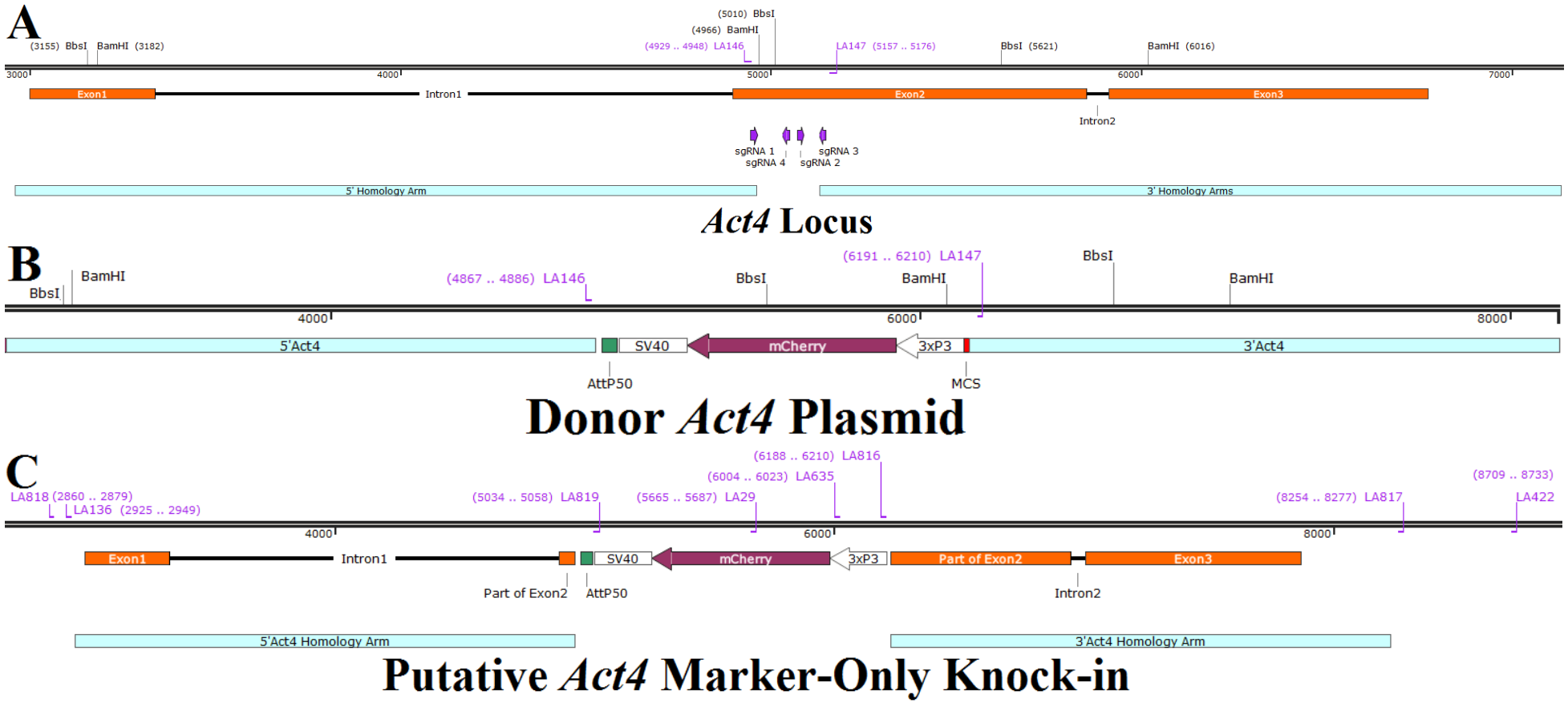


Figure 4.6: Genomic *Act4* and *AGG1070 Act4* donor sequence maps with their respective diagnostic primers:

Maps were drawn using SnapGene Viewer. Diagnostic primers used for the results below are shown here in purple. *Act4* homology arm regions are shown in all maps for orientation. BbsI and BamHI RE sites are shown. (**A**) *Act4* locus with *Act4* sgRNA sites, as well as the NHEJ diagnostic primers. (**B**) *Act4* Marker-only donor plasmid, with the same NHEJ diagnostic primers. (**C**) Putative *Act4* marker-only knock-in as well as HDR primers to detect *Act4* knock-in events either generating an amplicon over the 5' or 3' homology arm regions. Nested primers were also available for the amplicons.

4.2.16 DNA sequencing

DNA sequencing was outsourced to Source Bioscience. Samples sequenced in this project included successfully cloned constructs (Section 4.2.10), PCRs from LVP and LWT strains (Section 4.2.13), and *Act4* PCRs from flight-impaired candidates (Table 4.9).

4.3 Results and Discussion

4.3.1 Act4 genomic sequence is conserved well in two different Aedes aegypti strains

Ae. aegypti LVP strain genome sequence (AaegL3, April 2014 Assembly) was used as a reference for construct design in this Chapter. Even when dealing with the same strain, lab LVP strains could differ in sequence. Unknown sequence variation could have undermined the project as sgRNAs, HDR, and genomic PCRs are sequence dependent. To have an insight into *Act4* sequence variation between lab LVP individuals and different strains the following samples were chosen for Sanger sequencing. Two 2kb *Act4* amplicons were PCRed from pooled adult gDNA and cloned into pJET1.2 (Section 4.2.13). Two different clones for each were selected for sequencing. gDNA from five LVP and three LWT flying females was individually extracted and three amplicons covering ~6500bp of the *Act4* locus were PCRed. Amplicons for two LVP and one LWT were selected for sequencing. A range of primers (Section 4.2.13) were used to cover all the samples selected for sequencing. The fact that sequencing from PCR products usually yields ~600bp of good sequence as opposed to ~900bp from plasmids was taken into account. Whilst the clones provide longer reads, each clone sequence only represents a single allele. However, sequencing PCR products, although giving shorter reads, can reveal allelic variation within the sample as long as gDNA from single adults is used. In total, PCR products from three different single adult gDNA preps and four different PCR product clones

from a pooled adult sample were sent for Sanger sequencing. Only these numbers were sent for sequencing as variation between lab LVPs and the AaegL3 LVP assembly was expected to be small as they are the same strain. However, a greater number of PCR products, from five other single adult gDNA preps, were available should sequence coverage had not been enough.

The sequences were trimmed and then assembled into a contig using the AaegL3 genome as a reference. Sequence coverage of 6325 bp of the *Act4* locus was sufficient to have an insight into the sequence variation in the lab strains available (Figure 4.7). SNPs did not occur in the critical region at the beginning of the CDS where sgRNAs were targeted. Similarly, only four SNPs (all synonymous changes) were found in the CDS. Other SNPs are also present along the introns and the promoter region whilst most of them lie in the 3' UTR and beyond.

Haploinsufficiency depends on the inability of an allele to perform its full function on its own. This may be true in general for genes such like *Act4* in the population, however, there is a possibility for haplosufficient alleles to exist in a wild population. These could either be in the form of a different protein sequence or a regulatory DNA change which would generate sufficient *Act4* from one allele. Hence, the low amino acid variability observed across strains could be a good sign as it may reduce the chances of a haplosufficient allele being present in a wild population. In addition, the few SNPs observed in the upstream promoter

sequence are also indicative of a low variation across LVP and LWT strains and hence a lower risk of potential haplosufficient alleles to have arisen in nature. In any case, this is not a complete survey and only an indication of the conservation of *Act4* sequences between two strains from distant geographical origins. LWT is a Mexico-derived strain (Harris *et al.* 2011), collected near Tapachula, whilst the Liverpool strain was collected in West Africa and maintained in the Liverpool School of Tropical Medicine since 1936 (Vector Base 2017).

Although it is difficult to say what degree of sequence divergence, i.e. SNPs or indels, would pose a problem for homology directed repair it is reassuring that the observed variation is relatively low. The chosen homology arm regions for the HDR donor constructs only have SNPs and no INDELs and hence are expected to be suitable for homology directed repair both in LVP and LWT strains.

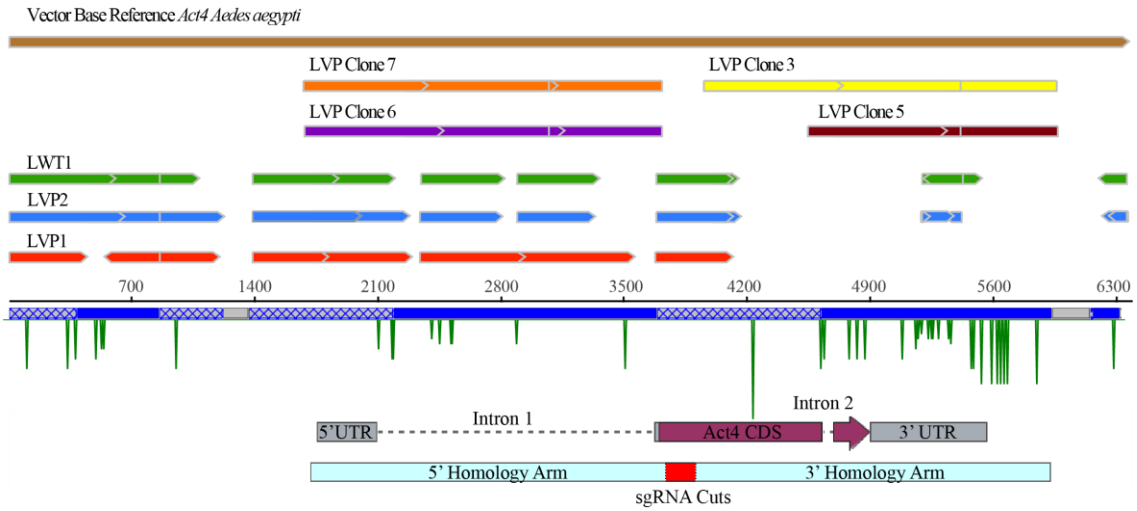


Figure 4.7: *Act4* SNPs when comparing LVP and LWT *Ae. aegypti* strains:
Sanger sequencing of both LVP clones and PCRs from the genome (LVP1, LVP2, and LWT1) provided enough sequence coverage of the *Act4* locus for a Contig to be assembled using the available *Ae. aegypti* genome, from Vector Base, as a reference. LVP1, LVP2, and LWT1 are 3 gDNA samples from single flying *Ae. aegypti* females. Multiple amplicons per sample were amplified and multiple sequencing reactions were carried out. LVP amplicons from pooled adult gDNA samples were cloned into pJET1.2 plasmids and 4 LVP clones were selected for sequencing. Section 4.2.13 shows details on sample preparation and sequencing processing. The Contig, shown below the number line, is of 6325bp. Dark blue regions mark bidirectional sequencing coverage, blue criss-cross regions mark unidirectional sequencing coverage, and grey regions mark an absence of sequencing coverage. Different colours mark sequencing coverage from different sources or the reference genome itself. Green spikes below the Contig mark SNP positions across the *Act4* locus and the length of these indicates the preponderance of the SNP amongst the different sources relative to the reference genome. Maps of the *Act4* gene and potential homology arms for HDR allow more relevant SNP mapping. SNPs present do not compromise the feasibility of HDR into *Act4*. There are 4 SNPs within the CDS of *Act4*, all of which are synonymous. No SNPs were found in the sgRNA cutting region, a few are found within both introns, and most of them are at the 3'UTR and beyond.

4.3.2 G1 flightless screening from Act4 sgRNA injections revealed no knock-outs

Act4 sgRNAs 1 to 4 were designed by Sanjay Basu and synthesised as described (Section 4.2.6). These were co-injected together with Cas9 protein into *Aedes aegypti* embryos. Ideally, sgRNAs would have been injected individually if the efficiency of each was to be compared. However, pairs of sgRNAs were co-injected to increase the chances of seeing any CRISPR-induced effects. If effects

were seen then the two sgRNAs from the pair could be tested independently in follow up studies.

A total of 2538 embryos were injected with an average survival rate to adult of 5.32% Table 4.8). The first round of injections resulted in a combined survival rate of 0.83% out of 1325 embryos injected. This is in stark contrast to the second round of injections which resulted in a combined survival rate of 10.14% out of 1213 embryos injected. The difference between both rounds was better proficiency with embryo line-up and injection, as well as a change in the recovery method of injected embryos which involved transferring the injected from the sticky tape oiled coverslips onto humid filter paper in a closed chamber (Section 4.2.3). It is difficult, however, to pinpoint the cause of such disparity as several factors were changed at once to maximise survival improvement. In addition, injection methods for arthropods are complex themselves and do not control for a lot of variables strictly, such as injection volume or embryo desiccation, which can make different injection rounds of the same construct vary substantially. Indeed, injection efficiency is highly user dependent.

With the improved survival rate 135 adult G0s were obtained. Current reports of CRISPR knock-outs in the literature suggest a knock-out efficiency (percentage of G0s giving mutant G1s) between 23-90% (Basu *et al.* 2015). Hence this number of G0s should have been sufficient to see a result. Seventeen G0 pools were set up and 7135 G1 females were picked as pupae and transferred to cages

to be screened for eclosion and flightless phenotypes. An initial phenotype screen found seven G1 females as potential candidates for some sort of impaired flight phenotype (Table 4.9). These consisted of a bad flier (3A), five flightless females which developed the ability to fly within 1 day post eclosion (9A, 11A, 14A, 15A, and 15B) and two flightless females which never developed the ability to fly (6A and 11A). Forced mating was attempted as described (Wheeler 1962) with no success. 15A was mated normally and all G2s were fliers shortly after eclosion (Table 4.9). The flight delay phenotype in the 15A G1 was not observed in the G2s, suggesting the reason for the phenotype was not necessarily genetic and instead caused by non-heritable factors.

Table 4.8: *Act4* sgRNAs and Cas9 injected into *Aedes aegypti*:

A summary of the number of injected embryos, larval hatch rates, and adult survival rates is provided. A total of 17 different G0 pools were crossed and 7135 G1 ♀ adults were screened for the expected flightless phenotype upon *Act4* knock-out (Table 4.9).

Date Injected	sgRNA 1 (ng/µl)	sgRNA 2 (ng/µl)	sgRNA 3 (ng/µl)	sgRNA 4 (ng/µl)	Cas9 (ng/µl)	Embryos	Larvae (Hatch Rate)	Adults (Survival Rate)	G0 Pools	G1 ♀s	G1 KOs	Single KO Events
11.11.16	40	40	N/A	N/A	300	414	2 (0.49%)	1 (0.25%)	1	57	0	0
15.11.16	40	40	N/A	N/A	300	911	28 (3.08%)	11 (1.21%)	4	248	0	0
16.11.16	N/A	N/A	40	40	300							
23.11.16	N/A	N/A	40	40	300	280	75 (26.79%)	62 (22.15%)	4	3298	0	0
24.11.16	40	40	N/A	N/A	300	385	25 (6.5%)	17 (4.42%)	3	1435	0	0
25.11.16	40	40	N/A	N/A	300	548	68 (12.41%)	44 (8.03%)	5	2097	0	0
					Total	**2538**	**198 (7.81%)**	**135 (5.32%)**	**17**	**7135**	**0**	**0**
									Knock-out Efficiency			**0.00%**

Table 4.9: G1s from *Act4* sgRNA injections selected by flightless phenotype showed no *Act4* knock-outs:

G1s from Table 4.8 injections were screened for flight impaired phenotypes and candidates had their *Act4* gene sequenced. Only 6A and 11A showed a true flightless phenotype. None of the sequenced individuals showed *Act4* putative knock-out sequences. G1 individuals were each given a code consisting of the G0 pool they came from (1-17), letters (A-Z) for different individuals from the same pool, and Ctrl for flying controls. Another 9 flying controls were isolated which are not included in this table as sequencing was not obtained from them. Some individuals were crossed to LVP and flightless individuals were forced mated unsuccessfully. Those successfully mated only gave flying G2s.

♀ G1 Individual	Phenotype	Recovered Flight	Mating (to ♂ LVP)	Mating Comments	Sequencing Results
3A	Flying badly	Always flew badly	N/A	N/A	No mutations.
6A	Not flying	No	Forced Mating	Forced Mating Failed	No mutations.
9A	Not flying	Yes (1 day post-eclosion)	N/A	N/A	No mutations.
11A	Not flying	No	Forced Mating	Forced Mating Failed	No mutations.
14A	Not flying	Yes (1 day post-eclosion)	N/A	N/A	No mutations. 2 synonymous SNPs.
15A	Not flying	Yes (1 day post-eclosion)	Normal Mating	All G2s were fliers	No mutations.
15B	Not flying	Yes (1 day post-eclosion)	N/A	N/A	No mutations.
11Ctrl	Flying well	Always flew well	N/A	N/A	No mutations. 1 synonymous SNP.

Subsequently, the seven candidates plus a flying female control (11Ctrl) had their gDNA extracted, PCRed for *Act4* and submitted for Sanger sequencing (Table 4.9). The flying female control showed normal WT flight and was a G1 sibling to the flightless candidate from the same G0 pool, i.e. 11A. Sanger sequencing results returned only WT sequences, with no changes from the AaegL3 assembly except for two synonymous SNPs. Both were found in 14A and one of them in 11Ctrl. This was specially unexpected for 6A and 11A individuals since they had such a clear flightless phenotype to the point of being unable to mate or be sexually recognised by males.

Nevertheless, given the surprisingly low percentage of G1s showing the expected phenotype convincingly (0.03%), it may not be a surprise that the knock-out of *Act4* is not causing the phenotype. Before the sequencing results, knock-out efficiency as a function of G0s yielding knock-outs, would have been 11.76% which is still low for the standards in the literature. Looking at all the data it is of course 0.00% (Table 4.8). Perhaps, the observed percentage of flightless females (0.03%) is a rough estimate of the natural occurrence of flightless individuals in *Ae. aegypti* LVP strains.

An explanation for the lack of a visible phenotype could be that *Act4* is not haploinsufficient. However, given the fact that evidence in *Drosophila melanogaster* and *Culex quinquefasciatus* strongly suggested otherwise the most likely explanation at this stage was still some sort of experimental pitfall. The following experimental explanations were ruled out first (Section 4.3.3 and 4.3.4).

4.3.3 Act4 sgRNAs successfully guide Cas9 for in vitro cutting

A simple explanation for the lack of knock-outs in Section 4.3.2 would be inefficient cutting of *Act4* by Cas9 guided by the chosen sgRNAs, or inefficient guiding of Cas9 by the chosen sgRNAs. An *in vitro* sgRNA/Cas9 digest of *Act4* DNA template was carried out as described (Section 4.2.7). An *Act4* template of 1276bp was amplified and purified and the expected fragments from Cas9 cutting guided by *Act4* sgRNA 1 to 4 were calculated (Figure 4.8A). All of the fragment

sizes were in theory discernible upon gel electrophoresis of the *in vitro* digests. Two different preparations of the same sgRNAs (Section 4.2.6) were tested simultaneously.

All sgRNAs, except sgRNA 4, guide Cas9 protein to cut the template to some degree (Figure 4.8B). Results between the two different sgRNA preparations were almost identical which suggests *in vitro* results are reproducible and sequence dependent. sgRNAs 2 and 3 were the best at cutting the template whilst sgRNA 1 displayed inefficient cutting. Interestingly, Cas9 appears to be inefficient compared to other double stranded endonucleases. Cas9 was used at a 30nM concentration for only 70ng of template and digestion was incomplete even with the best sgRNAs. This relative inefficiency was expected as the *in vitro* test used recommends a low amount of template per Cas9 (10 moles of Cas9 per 1 mol of target, Section 4.2.7). Given the low amount of starting template visualisation of the cut fragments in the gel was difficult and that is the reason behind showing an additional overexposed image of the same gel (Figure 4.8C).

The results are an indication that guiding/cutting inefficiency was probably not the reason for the lack of observed knock-outs in Section 4.3.2. However, even though there is a general correlation between Cas9 *in vitro* and *in vivo* results (Wu *et al.* 2014), it remains unclear how good this correlation is. Nucleosomes have been shown to be an impediment for Cas9 cleavage of DNA (Horlbeck *et*

al. 2016) and they may explain the higher Cas9 promiscuity and off-target effects observed *in vitro* (Wu *et al.* 2014). sgRNAs 1 to 4 span across 204bp in the genome and therefore nucleosome blocking of all of the sgRNAs seems unlikely as they usually only protect ~147bp of DNA (Cutter and Hayes 2015). In addition, their dynamic behaviour in moving along genomic DNA (Lai and Pugh 2017) may mean that any blocking of sgRNA sites will not necessarily be permanent. Even if these *in vitro* results may not correlate with *in vivo* guiding/cutting efficiencies, of these sgRNAs and Cas9, at least the cleavage assay shows the functional integrity and sequence specificity of the chosen sgRNAs in guiding Cas9 cutting.

A

	Act4 1276bp Template	
sgRNA	**Cut Fragment 1 (bp)**	**Cut Fragment 2 (bp)**
sgRNA 1	879	397
sgRNA 2	1005	271
sgRNA 3	1049	227
sgRNA 4	950	326

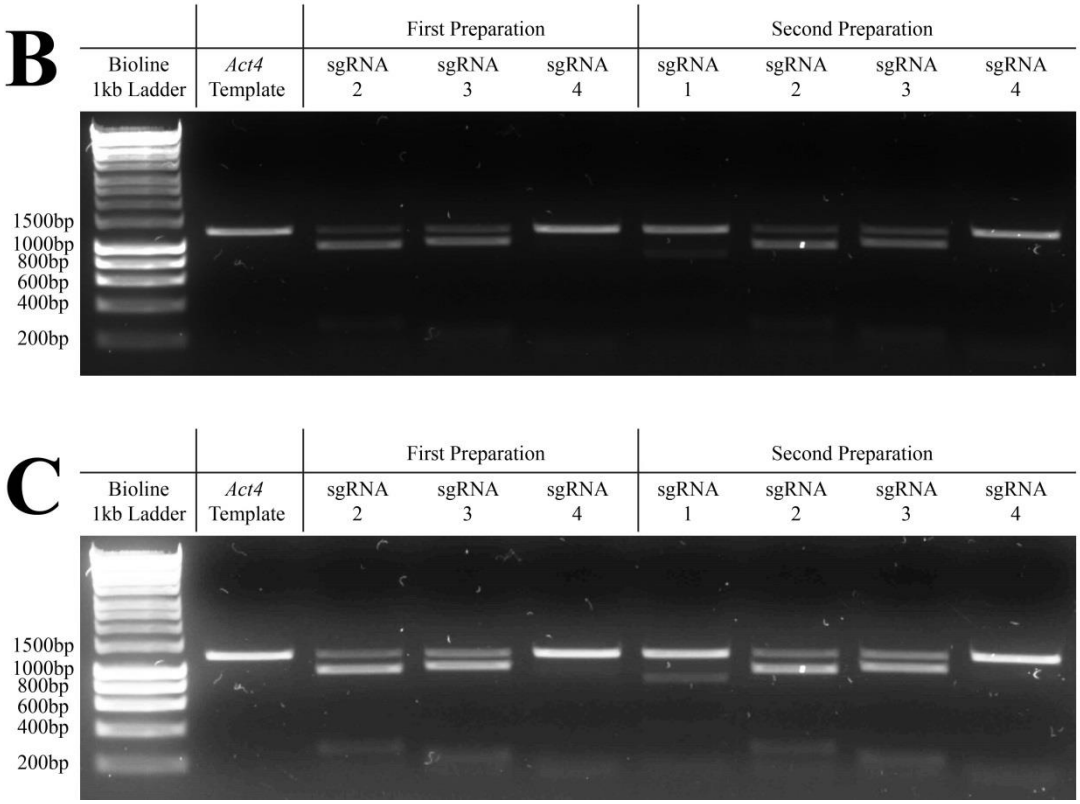


Figure 4.8: *Act4* sgRNAs 1 to 4 guide Cas9 for *in vitro* cutting of an *Act4* DNA template:
sgRNAs 1 to 4 used in injections (Table 4.8) were tested for function in an *in vitro* cleavage assay described in (Section 4.2.8). 70ng of *Act4* template was used per reaction. (**A**) Different sgRNAs guide Cas9 cutting at different positions in the *Act4* template. Predicted fragment sizes are shown. (**B**-**C**) Two different exposures were used to image the same electrophoresis band so both short and long fragments can be visualised clearly. Two different sgRNA batches were prepared and tested with reproducible results. Digestion fragments were as predicted for sgRNA 1, 2, and 3. sgRNA 4 did not lead to any perceivable cutting.

4.3.4 sgRNA preparations used for injections were RNAse free

Another potential pitfall to rule out was the potential RNAse contamination of sgRNA preparations and hence the degradation of the sgRNAs before or during injection. To test this, equal amounts at the same concentration of sgRNAs 1 to 4

were subjected to two different temperature conditions for 7hrs (24°C and -80°C). The rationale being any RNase present would degrade the sgRNAs at 24°C but not at -80°C. Both available sgRNA preparations were tested. After the incubation period the samples were analysed by gel electrophoresis. Results show no difference in band intensities between the two temperature treatments, which is indicative of clean preparations (Figure 4.9). sgRNAs ran below 200bp which reflects their expected size. sgRNA band intensity was low for the amount of RNA used per sample (100ng). The low signal intensity of sgRNA bands is probably due to the visualisation dye used (SYBR Safe) as it is designed to visualise dsDNA and not ssRNA. Dye intercalation may only happen along regions of sgRNA secondary structure. These would be smaller than those for an equivalent weight of dsDNA. The fact that both sgRNA preparations remain RNAse free added more confidence to the current sgRNA handling methods.

Given the lack of obvious reasons for which *Act4* sgRNA injections did not produce knock-outs (Table 4.8) the possibility of *Act4* being haplosufficient in *Ae. aegypti* was considered more seriously. A different approach to screen for knock-outs in heterozygotes with no apparent mutant phenotype was necessary. One option would have been to devise some sort of G1 screen of mutant candidates through high resolution melt analysis (HRMA). HRMA uses special intercalating dyes, such as EvaGreen® which saturate dsDNA down to one fluorescent molecule per bp. This improved saturation over other dyes such as SYBR Green, together with thermocyclers capable of accurate temperature

increments as small as 0.15°C per fluorescence intensity read allow for a very high resolution melt curve to be plotted. In a small enough amplicon, 100-150bp, a single bp change and certainly INDELs would set the melt curves apart. Thus, heterozygous mutants could be detected from WT individuals. A second option would be simple restriction fragment length polymorphism analysis (RFLP) exploiting potential RE sites in heterozygous mutants and WT DNA. In the case of bigger deletions these could be detected simply by PCR. However, an *Act4* HDR knock-in approach was chosen (Section 4.3.6) as it was considered the route most likely to yield success, and also more relevant to potential downstream use.

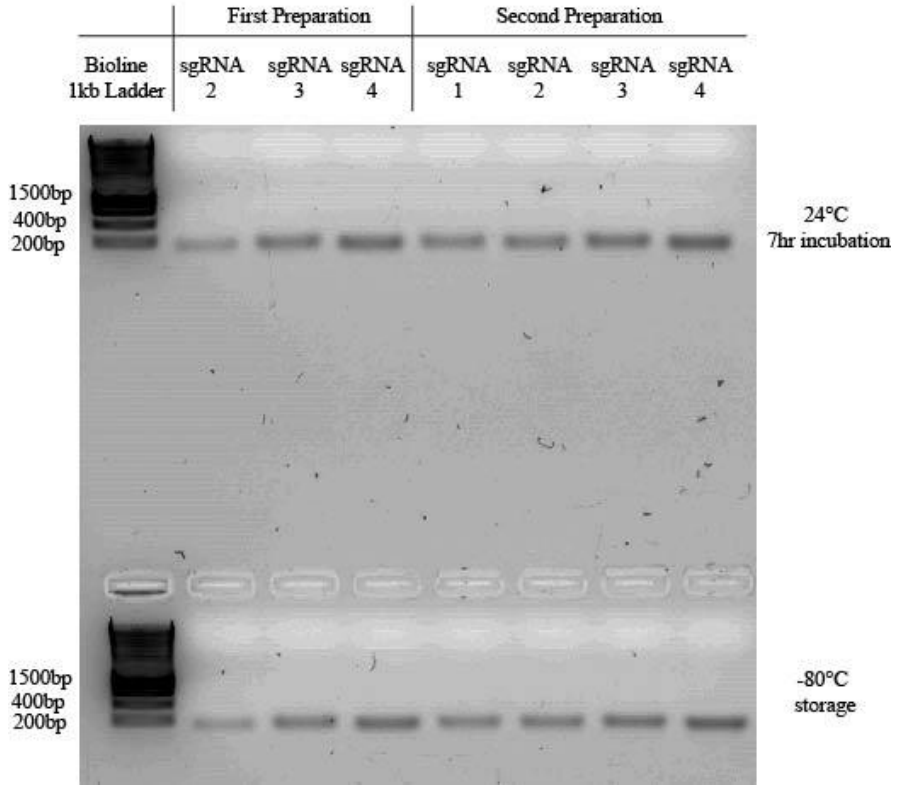


Figure 4.9: sgRNAs synthesised did not degrade after 7 hours at 24°C:
The same sgRNA preps used for injections (Table 4.8) and the cleavage assay (Figure 4.8) were tested for RNAse presence. sgRNAs incubated at 24°C for 7 hours were compared to those stored at -80°C. Two different sgRNA batches were tested. There is no appreciable difference in band intensity between the two conditions which suggests there is not enough RNAse in the preps for significant degradation to occur.

4.3.5 Female-specific underdominance and RIDL with drive constructs were made

As described in this Chapter's introduction (Section **Error! Reference source ot found.**) two different gene drive systems were to be developed. Both systems involve the generation of complex constructs which were built using Golden Gate cloning with the outsourced gene synthesis of key components. Table 4.10 shows a list of constructs made, or in the process, and indicates which constructs were generated by Golden Gate cloning and/or gene synthesis. Whilst details on

the methods are in Section 4.2.10, the rationale behind construct specifics is covered here. Figure 4.10 shows the same constructs as visual maps.

The red eye marker construct, *pBac-3xP3-mCherry-SV40-attP50*, was cloned by Golden Gate cloning from PCRs from existing plasmids. This construct was made first as it was to be used as a precursor for RIDL with drive and female specific underdominance gene drive systems. The *3xP3* promoter was chosen due to its short sequence, low fitness penalty, and strong eye-expression in *Ae. aegypti*. Bearing in mind the eventuality of HDR, shorter sequences are preferred on the assumption that a smaller cargo would integrate at a higher rate as some evidence suggests (Li *et al.* 2014). HDR is thought to have a lower integration rate than *piggyBac* which typically has a transformation rate of 4-11% in *Ae. aegypti* (Labbé *et al.* 2010), and hence the reason to aim for optimal conditions. Daisy-chain drives (Noble *et al.* 2016), which could benefit from components in this work, require several components with different fluorescent markers. Finding different fluorescent proteins in *Ae. aegypti*, even as few as three, with the appropriate excitation and emission properties requires planning. It is vital to readily distinguish them without bleedthrough using standard fluorescence filter sets. Hence, mCherry was chosen for this reason as the fluorescent protein has a red-shifted spectrum relative to DsRed. The emission spectrum of DsRed is more biased towards orange wavelengths, and thus overlaps with ZsYellow. SV40 is also the shortest terminator which has been commonly used in *Ae. aegypti* (Fu *et al.* 2010). An attP site was added to be able to integrate additional components

later in a two-step process. This is crucial for the female specific underdominance design and may also be useful for RIDL with drive if such a two-step process were required.

Even though the final design of the female specific underdominance system consists of a set of components all into a single locus, *Act4* expressed rescue and *Act4 RNAi* (Section 4.1.4), these could not be part of the same initial construct. The assumption is that the system would cause female-specific flightlessness, i.e. effectively female sterility, with one copy of the construct allele, i.e. one copy of the *RNAi* hairpin being enough to knockdown expression of both endogenous copies of *Act4*. *piggyBac* mediated insertion of the final fully designed construct would only be able to generate viable and fertile males. Since homozygosity is required for underdominance, and there would be no viable and fertile females bearing the construct, building a homozygous strain would not be possible. An option chosen to avoid this was to divide the components into two different constructs and transform *Ae. aegypti* in two steps. The first step would have involved transforming the *Act4* rescue components plus the red eye marker, *pBac-3xP3-mCherry-SV40-Act4Rescue-attP50,* as one construct integrated via *piggyBac*. The second step would have involved transforming the *RNAi* components, *pBac-LoxP-attB-Act4RNAi-LoxP-Hr5IE1-AmCyan-K10,* as one construct integrated via attP/attB recombination using φ31 integrase into a homozygous *Act4* rescue background (Figure 4.10).

To make the *Act4* rescue the *Act4* gene was amplified in several fragments from the same LVP2 gDNA from Figure 4.7 as described (Section 4.2.10). The beginning of exon 2 was recoded with synonymous nucleotide changes whilst making sure rare codons were avoided and then outsourced for gene synthesis (Section 4.2.9). This recoding is what makes the *Act4* gene be a rescue copy as it is insensitive to the *RNAi* hairpin targeting the same region. An effort was made to prevent repetition of components between the two constructs, i.e. promoters, fluorescent proteins, and 3'UTRs were all different from each other to avoid the risk of unwanted homology dependent recombination. To be able to express both the rescue and *RNAi* components in the indirect flight muscle of *Ae. aegypti* two different versions of the *Act4* promoter were sourced. The rescue construct, *pB-3xP3-mCherry-SV40-Act4Rescue-attP50,* used the *Ae. aegypti* promoter whilst the *RNAi* construct, *pB-LoxP-attB-Act4RNAi-LoxP-Hr5IE1-AmCyan-K10,* used the orthologous promoter from *Ae. albopictus.* This design has been used before and the *Ae. albopictus* promoter has been shown to work in *Ae. aegypti* with a similar expression pattern and expression levels (Labbé *et al.* 2012).

One disadvantage of *attP/attB* recombination is that the whole plasmid backbone of the second plasmid would integrate itself into the *Act4* rescue locus together with the *RNAi* hairpin. To allow for subsequent excision of this undesired DNA fragment two *loxP* sites were placed to flank the plasmid backbone and the AmCyan marker. The AmCyan marker serves two purposes, first indicating successful *attB/attP* recombination which has a similar efficiency to *piggyBac* in

Ae. aegypti (Nimmo *et al.* 2006), and then to mark successful backbone and marker excision mediated by *Cre-lox*.

At the time of writing neither of these constructs was fully assembled. The construction was halted after the initial findings that suggested *Act4* haplosufficiency (Table 4.8). The *Act4* rescue construct had, at the time of writing, all components incorporated except for the synthesised recoded fragment. These two constructs would not provide female specific underdominance or indeed any interesting drive feature should *Act4* be haplosufficient (Section 4.1.4). One copy of the rescue would be enough to compensate for the knockdown of endogenous *Act4* by the RNAi component. This in turn would mean that being homozygous for the rescue would grant no better fitness than being heterozygous for the rescue. In addition, transgenes usually impose a fitness cost, in this case through RNAi machinery overloading for example (Grimm *et al.* 2010). Hence, the most likely scenario would be that rescue heterozygotes would be fitter than rescue homozygotes, resulting in the opposite concept to underdominance. Subsequently, RIDL with drive constructs gained priority and were continued to be made.

RIDL with drive constructs as described in the introduction do not necessarily depend on *Act4* haploinsufficiency to be useful. If *Act4* were haplosufficient it would still be useful as a recessive female sterile target in future daisy-chain drives (Noble *et al.* 2016). The first construct made for RIDL with drive was

pJet-5'Act4flank-attP50-3xP3-mCherry-SV40-3'Act4flank, which had the marker and attP site cloned within 2kb *Act4* homology arms. Ideally, versions including multiplexed *Act4* sgRNAs and germline Cas9 would have been made first but timing and availability of components were limiting. The *Act4* HDR marker-only construct was a step closer to the RIDL with drive design and a way of finding out if *Act4* is haploinsufficient (Section 4.3.6.).

The next construct to be available was the multiplexed version of *Act4* sgRNAs. As described in the introduction (Section 4.1.8) multiplexing is achieved via the self-processing capacity of tRNA sequences (Xie *et al.* 2015). The long RNA polymerase III transcript is cleaved by excision of the tRNAs to generate the wanted set of sgRNAs. Luke Alphey and Phil Leftwich designed the cassette; selecting tRNAs and sgRNA backbones (Noble *et al.* 2016). The use of the *U63* promoter from *Ae. aegypti* over *U61* and *U62* was predicated on the finding that *U63* is more efficient in *D. melanogaster* (Port *et al.* 2014). In addition, *U61* and *U62* are adjacent and divergent in the *Ae. aegypti* genome and so it is difficult to determine reliably where one promoter starts and the other finishes. The construct, *pUC57-U63-Act4sgRNAs-attP50*, was then outsourced for synthesis (Genewiz).

The other two *Act4* HDR constructs are still under construction, at the time of writing, and will be useful in the future. Incorporating the multiplexed sgRNAs into the *Act4* Marker-only construct will generate *pJet-5'Act4flank-attP50- pUB-*

mCherry-SV40-U63-sgRNAs-3'Act4flank. An additional change planned was to substitute *3xP3* for the ubiquitous *pUB* promoter due to unexpected problems encountered with the marker (Figure 4.11). Primers and a finalised Golden Gate cloning strategy are already available. The third HDR construct, *pJet-5'Act4flank-attP50-pUB-mCherry-SV40-U63-sgRNAs-germlineCas9-3'Act4flank*, is also under construction. Tim Harvey-Samuel designed an *Ae. aegypti* codon optimised Cas9 sequence which was outsourced and synthesised.

To test the RIDL with drive function of the *Act4* HDR plus *sgRNAs* construct had it been completed and integrated in time during the one year placement in The Pirbright Institute, a Cas9 line obtained from Omar Akbari (University of California, Riverside) was obtained. Testing under a split drive setting would have still been very insightful. Such *exu-Cas9* line, which was otherwise used as a source of Cas9 for injections (Table 4.12) expresses Cas9 in the male germline as well as being maternally deposited into the embryo.

A factor to consider is that recessive female sterile targets, often required in somatic cells near the germline, can hinder homing drives with non-specific expression of germline Cas9. Cas9 cutting of the recessive female sterile target in somatic cells surrounding the germ cells has been shown to counter the drive with a reduction of fertility (Hammond *et al.* 2016). However, this factor is much less likely to be detrimental in a system with *Act4* as a target given that the somatic cells requiring functional *Act4* lie in the indirect flight muscles. The

distance between the two tissues, indirect flight muscles and germline, will perhaps reduce the risk of non-specific germline *Cas9* expression reaching the functional *Act4* target gene copy. At least *Act4* is quite convincingly not required in the germline. Whilst this separation would be an advantage if *Act4* were haplosufficient, the risk of non-specific germline *Cas9* expression hindering the drive by reduced fertility would be avoided completely provided *Act4* haploinsufficiency. Since the *Act4* RIDL with drive system only requires germline homing in males, and *Act4* has no function in males, non-specific germline *Cas9* expression will not cause an *Act4* mediated fertility reduction.

Since less strict germline expression of Cas9 may be tolerated this makes *exu-Cas9* a valid source for a split drive test. Moreover, it makes *Act4* an advantageous target for a RIDL with drive system (provided *Act4* is haploinsufficient) or future daisy-chain drives (provided *Act4* is haplosufficient). Nonetheless, finding the ideal germline promoter, with strong and tightly specific gene expression remains a challenge worth solving. Non-specific *Cas9* expression could still have other off-target related fitness costs which would be preferably avoided.

Table 4.10: List of female-specific underdominance and RIDL with drive constructs and intermediates:

The list includes both finalised constructs (those with a construct code) and those still in progress. The constructs share some common components which was very useful in the cloning process as can be seen in the component origin column. Figure 4.10 shows these constructs as visual maps.

Gene Drive	Construct Code	Construct Name	Description	Marker	Method	Component Origin
Female-specific Underdominance and RIDL with Drive	AGG1069	*pBac-3xP3-mCherry-SV40-attP50*	Red Eye Marker and attP Landing Site	3xP3-mCherry	Golden Gate Cloning	Other constructs
RIDL with Drive	AGG1070	*pJet-5'Act4flank-attP50-3xP3-mCherry-SV40-3'Act4flank*	*Act4* HDR: Marker-only	3xP3-mCherry	Golden Gate Cloning	Red eye marker construct, LVP gDNA
RIDL with Drive		*pJet-5'Act4flank-attP50- pUB-mCherry-SV40-U63-sgRNAs-3'Act4flank*	*Act4* HDR: *Act4* sgRNAs	pUB-mCherry	Golden Gate Cloning	Multiplexed sgRNA and Act4 HDR: Marker only constructs
RIDL with Drive		*pJet-5'Act4flank-attP50-pUB-mCherry-SV40-U63-sgRNAs-germlineCas9-3'Act4flank*	*Act4* HDR: *Act4* sgRNAs + Cas9	pUB-mCherry	Golden Gate Cloning	Act4 HDR: Act4 sgRNAs construct, Cas9 construct
RIDL with Drive	AGG1058	*pUC57-U63-Act4sgRNAs-attP50*	Multiplexed *Act4* sgRNAs	N/A	Gene Synthesis	Gene synthesis
Female-specific Underdominance		*pBac-3xP3-mCherry-SV40-Act4Rescue-attP50*	*Act4* Rescue	3xP3-mCherry	Golden Gate Cloning/ Gene Synthesis	LVP gDNA, red eye marker construct, synthesised recoded Act4
Female-specific Underdominance		*pBac-LoxP-attB-Act4RNAi-LoxP-Hr5IE1-AmCyan-K10*	*Act4* RNAi	Hr5-IE1-AmCyan	Golden Gate Cloning	LVP gDNA, other constructs

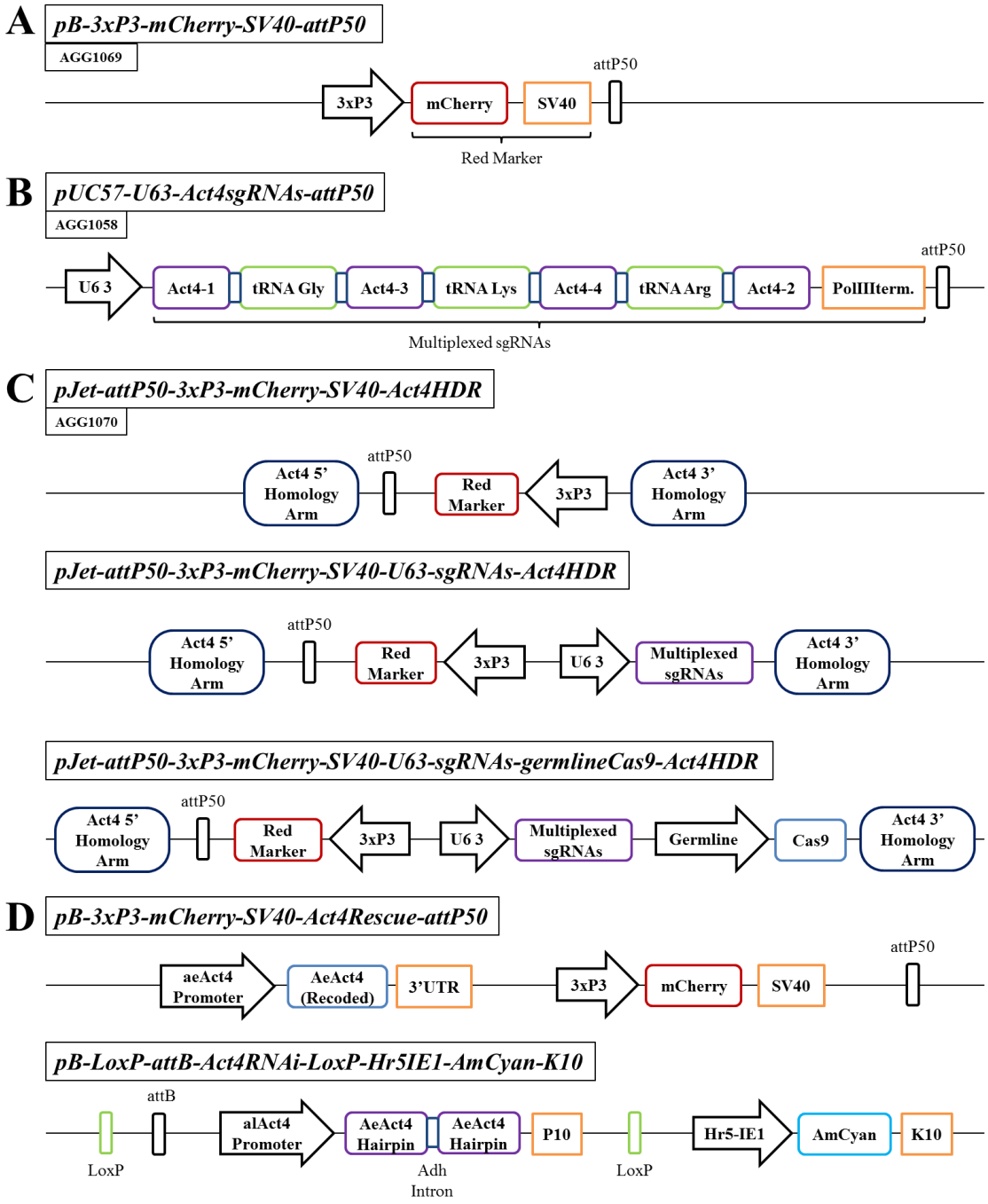


Figure 4.10: Female-specific underdominance and RIDL with drive constructs:
Constructs in Table 4.10 are shown here as maps. **(A)** mCherry is under an eye-specific promoter. **(B)** Act4 sgRNAs were multiplexed into a single polIII transcript with tRNA post transcriptional cleavage for sgRNA individualisation. The U6-3 promoter gives ubiquitous expression. **(C)** HDR constructs have several configurations; marker only, marker plus sgRNAs, and marker plus sgRNAs and Cas9 under a germline specific promoter. The two latter configurations could be used for a RIDL with drive system. **(D)** Both female-specific underdominance constructs are designed to work as a single insertion *in vivo* however; to generate the transgenic line a two-step process is required. The first construct to be integrated via *piggyBac* is the *Act4* rescue itself which is the same as the endogenous gene except for it being RNAi insensitive. The second construct has the RNAi component and a second marker to indicate attP/attB recombination. LoxP sites are positioned to remove the second marker and unwanted plasmid backbone from the insertion site. Repetition of components was minimal.

4.3.6 HDR and NHEJ assays on G0 embryos fail to be conclusive

RFLP assays were the assay of choice to screen *Act4* CRISPR injected G0s for knock-out events (Section 4.2.15). A PCR assay was also used to screen *Act4* HDR knock-in events in G0s (Section 4.2.15).

Two different batches of batches of 40 LVP embryos were injected with Act4 sgRNAs, Cas9 protein, and *Act4* marker-only HDR donor. A batch of 40 LVP embryos was injected with water as a control. gDNA was extracted, digested when required and PCRed with the assay primers as described (Section 4.2.15).

Results from the RFLP assay checking for NHEJ events proved inconclusive as digestion of WT DNA was very inefficient, PCR samples from undigested looked the same. This resulted in only WT *Act4* bands being seen in all samples injected or not, digested or not. Further optimisation of the digestion by reducing the DNA input could improve the assay.

Results from the nested HDR PCRs resulted in a clear strong 2kb band of the expected size in the 2nd round PCR of the 5' HDR assay. However, when extracted and sent for Sanger sequencing, the results did not show a clear sequence. Successful HDR events found in G1s below meant that this assay was not further required. A repeat of the assay with the added step of cloning the positive bands into pJET1.2 would facilitate high-quality Sanger sequencing.

4.3.7 Homology directed repair into Act4 appeared successful in Aedes aegypti

A total of 600 LVP embryos were injected with *Act4* sgRNAs (1 to 4 in different combinations), Cas9 protein, and an *Act4 3xP3-mCherry* HDR donor plasmid (AGG1070) as indicated in Table 4.11. An average survival rate to adult of 16.2% was observed (Table 4.11), resulting in 97 G0 adults. A total of 10 adult G0 pools were crossed to LVP adults and G1 eggs were collected. G1s were hatched into larvae and these screened for *3xP3-mCherry* fluorescence.

A total number of 7198 G1 larvae were screened and 20 *3xP3-mCherry* positives were found from 4 different G0 pools. Not all had the full *3xP3-mCherry* profile expected; i.e. fluorescence in both eyes along the full length of the optical nerve (Table 4.14). As evidenced below (Table 4.15), individuals showing fluorescence in a single eye showed no germline transmission of the fluorescence. Hence, there was probably only a single HDR integration event, amongst all the G1 *3xP3-mCherry* positive individuals. The 14 individuals showing fluorescence expression in both eyes all originated from the same G0 adult male pool, which is consistent with a minimum of a single HDR integration event. Bearing this in mind, the transformation efficiency (as a percentage of G0s bearing transgenic progeny) was of ≥1.0%.

In line with the hypothesis of Act4 haploinsufficiency, 3557 G1 female adults were screened for expected flight-impaired phenotypes upon Act4 knock-out. 18 different flight-impaired candidates were selected, from 5 different pools (Table

4.13). Should all individuals represented *Act4* knock-out events, the minimum number of independent events would be 5. Therefore the knock-out efficiency would be of ≥5.2% (Table 4.11). However, to be able to verify this, further analysis of the knock-out events, i.e. sanger sequencing of G1 *Act4* sites, would be required.

Table 4.11: *Act4* sgRNAs, Cas9 and HDR donor injected into *Aedes aegypti*:

Act4 sgRNAs 1 to 4 in different combinations, Cas9 protein and *Act4* flanked *3xP3-mCherry* as an HDR plasmid donor (AGG1070) were injected in LVP strain *Aedes aegypti* embryos. A summary of the number of injected embryos, larval hatch rates, and adult survival rates is provided. A total of 10 different G0 pools were crossed and 7198 G1 larvae were screened for *3xP3-mCherry* expression marking potential *Act4* knock-ins. 20 *3xP3-mCherry* positive individual G1s were found, however analyses below determined there was only a single *3xP3-mCherry* HDR event amongst them. 3557 G1 ♀ adults were screened for the expected flightless phenotype upon *Act4* knock-out. A total of 18 *Act4* knock-out flightless female adult candidates were found, of these at least 5 were independent events should *Act4* knock-outs be confirmed. Transformation, or knock-in efficiency, as well as knock-out efficiency is calculated as the percentage of transgenic- or mutant-bearing G0s. Single +ve events stand for independent and heritable events. G1 KOs stand for *Act4* G1 knock-out candidates selected by flight impaired phenotypes. Single KO events refer to the minimum number of independent KO events observed which are calculated as one per G0 pool of origin.

Date Injected	sgRNA 1 (ng/µl)	sgRNA 2 (ng/µl)	sgRNA 3 (ng/µl)	sgRNA 4 (ng/µl)	Donor Plasmid (ng/µl)	Embryos	Larvae (Hatch Rate)	Adults (Survival Rate)	G0 Pools	G1 ♀s	G1 ♂s	G1 +ves	Single +ve Events	G1 KOs	Single KO Events
9.6.17	40	40	40	40	443.15	148	38 (25.7%)	29 (19.6%)	4	1693	1752	17	1	16	4
16.6.17	100	N/A	100	N/A	500	175	69 (39.4%)	60 (34.3%)	4	1602	1660	3	0	2	1
20.6.17	150	N/A	N/A	150	600	277	20 (7.2%)	8 (2.8%)	2	262	229	0	0	0	0
					Total	**600**	**127 (21.2%)**	**97 (16.2%)**	**10**	**3557**	**3641**	**20**	**1**	**18**	**5**
										Transformation Efficiency			**≥1.0%**		
													Knock-out Efficiency		**≥5.2%**

A total of 124 *exu-Cas9* line embryos were injected with *Act4* sgRNA 1 and 3, and an *Act4 3xP3-mCherry* HDR donor plasmid as indicated in Table 4.12. An average survival rate to adult of 22.6% was observed (Table 4.12), which resulted in 28 G0 adults. A total of 3 adult G0 pools were crossed to LVP adults and G1 eggs were collected. G1s were hatched into larvae and these screened for *3xP3-mCherry* fluorescence. The average survival rate using the *exu-Cas9* line was slightly higher 22.6% with respect to 16.2% when co-injecting Cas9 protein. Although a small improvement, a similar improvement was observed when switching to *nanos-Cas9* lines for injection of sgRNAs in Chapter 3, also into *Aedes aegypti*.

A total of 2541 G1 larvae were screened and no *3xP3-mCherry* positives of any kind were found. The transformation efficiency was in this case 0.0% (Table 4.12). Nonetheless, the lack of *3xP3-mCherry* positives is puzzling. This line was reported, in Li *et al.* 2017, to increase knock-out and HDR knock-in efficiency by more than two orders of magnitude with respect to co-injecting Cas9 protein. The HDR efficiency in this study for Cas9 protein co-injection was of 1.0%. This number was calculated by omitting the found individuals with single-eye *3xP3-mCherry* expression, and assuming all transgenic progeny from a single G0 pool originate from the same insertion. Nevertheless, since only 1 positive integration event was found amongst the Cas9 protein injections, any comparison is difficult. The number of *exu-Cas9* embryos injected, as well as the survivors (28 G0

adults) were much lower in number than for Cas9 protein injections (97 G0 adults). Hence, the numbers may be too small for a valid comparison.

However, the lack of single-eye fluorescent G1 individuals found amongst *exu-Cas9* injected individuals remains unexplained. Up to 4 out of 10 adult G0 pools for the Cas9 protein injections resulted in at least 1 single-eye fluorescent G1 individual. Should single-eye fluorescence be due to maternal or paternal plasmid deposition, there is no reason why not even this class of fluorescence was seen amongst G1 individuals from injections into *exu-Cas9* individuals. One possible explanation is that *exu-Cas9* injections were carried out with a lower injection volume unintentionally, due to a physical variation in the injection needle, embryo desiccation, or some other factor. The lower injected volume could not only have reduced the rate of plasmid carryover into G1s by eggs or sperm but reduce HDR rate itself. There is still a possibility that single-eye fluorescent G1s do not arise from plasmid carryover from G0s but from something else, although this has not been sufficiently explored in this study. In order to further explore this gDNA from such individuals (which was extracted and purified) could be PCRed in order to figure out if the *3xP3-mCherry* cassette was still within the donor plasmid or somewhere else in the genome.

A total number of 1213 G1 female adults were screened for expected flight-impaired phenotypes upon *Act4* knock-out. A single flight-impaired candidate was selected, from a single pool (Table 4.13). In this case, the minimum number

of independent knock-out events was 1. Therefore the knock-out efficiency was of 3.6% (Table 4.12). Although this value was comparable to the injections with Cas9 protein (5.2%), only 1 flightless candidate was found for *exu-Cas9* injections compared to 18 flightless candidates found for Cas9 protein injections. This difference is considerable, but given the fact that none of them are confirmed knock-outs, the reason for it is yet to be determined.

Table 4.12: *Act4* sgRNAs, and HDR donor injected into *exu-Cas9* lines in *Aedes aegypti*:

Act4 sgRNAs 1 to 3 and *Act4* flanked *3xP3-mCherry* as an HDR plasmid donor (AGG1070) were injected in *exu-Cas9 Aedes aegypti* embryos. A summary of the number of injected embryos, larval hatch rates, and adult survival rates is provided. A total of 4 different G0 pools were crossed and 2541 G1 larvae were screened for *3xP3-mCherry* expression, marking potential *Act4* knock-ins. No *3xP3-mCherry* positive G1 larvae were found. 1213 G1 ♀ adults were screened for the expected flightless phenotype upon *Act4* knock-out. A singe *Act4* flightless female adult candidate G1 adult was found. Transformation, or knock-in efficiency, as well as knock-out efficiency is calculated as the percentage of transgenic or mutant bearing G0s. Single +ve events stand for independent and heritable events. G1 KOs stand for *Act4* G1 knock-out candidates selected by flight impaired phenotypes. Single KO events refer to the minimum number of independent KO events observed which are calculated as one per G0 pool of origin.

Date Injected	sgRNA 1 (ng/µl)	sgRNA 2 (ng/µl)	sgRNA 3 (ng/µl)	sgRNA 4 (ng/µl)	Donor Plasmid (ng/µl)	Embryos	Larvae (Hatch Rate)	Adults (Survival Rate)	G0 Pools	G1 ♀s	G1 ♂s	G1 +ves	Single +ve Events	G1 KOs	Single KO Events
4.7.17	200	N/A	200	N/A	600	124	36 (29.0%)	28 (22.6%)	4	1213	1328	0	0	1	1
										Transformation Efficiency			**0.0%**		
													Knock-out Efficiency		**≥3.6%**

4.3.8 G1 Act4 knock-out flight-impaired candidates were selected

19 flightless female candidates were found amongst the G1s from Table 4.11 and Table 4.12 injections. 6 different controls, one from each of the flightless-bearing G0 adult pools, were taken Table 4.13. These were observed during their lifetime and the recovered ability to fly was assayed. All but 1 individual, which was flying badly in the first place, did not recover the ability to fly and remained flightless during the observation period of two weeks.

Table 4.13: G1s from HDR *Act4* sgRNA injections selected by flight-impaired phenotype:

G1s from Table 4.11 and Table 4.12 injections were screened for flight impaired phenotypes once they were negative for fluorescent marker expression. G1 individuals were each given a code consisting of the G0 pool they came from (numbers 1-10 for Table 4.11 pools and 11-14 for Table 4.12 pools), letters (A-Z) for different individuals from the same pool, and Ctrl for flying controls. 19 flightless female candidate G1s were found from 6 different pools. The candidates, and the relevant controls, had their gDNA extracted for future analysis. Forced mating was not attempted on this occasion as results below meant that it would no longer be necessary. None of the flightless individuals recovered flight in this occasion; future Sanger sequencing on the *Act4* PCR products from their gDNA would be insightful. Individuals were alive as adults for a 10 days post eclosion, living individuals were then processed for gDNA extraction.

G1 ♀ Individual	Phenotype	Recovered Flight
1A	Not Flying	No
2A	Not Flying	No
2B	Not Flying	No
2C	Not Flying	No
2D	Not Flying	No
2E	Not Flying	No
2F	Not Flying	No
2G	Not Flying	No
2H	Not Flying	No
3A	Not Flying	No
4A	Not Flying	No
4B	Not Flying	No
4C	Not Flying	No
4D	Not Flying	No
4E	Not Flying	No
4F	Flying Badly	Yes (2 days post-eclosion)
7A	Not Flying	No
7B	Not Flying	No
13A	Not Flying	No
1Ctrl	Flying	N/A
2Ctrl	Flying	N/A
3Ctrl	Flying	N/A
4Ctrl	Flying	N/A
7Ctrl	Flying	N/A
13Ctrl	Flying	N/A

Given the lack of success of forced mating in Table 4.9, and the observed *Act4* knock-in event, described below, this was not pursued for these individuals as an *Act4* marker knock-in line could be maintained instead and mating flightless

females was no longer necessary. Individuals had their gDNA extracted and Sanger sequencing was postponed for the same reasons. Similarly, flightless candidates in the previous round of injections did not show any knock-out mutations in *Act4*.

On the other hand, unlike the first *Act4* sgRNA and Cas9 protein injections in Table 4.8, this time the number of flight-impaired candidates was greater (19 with respect to 7). Moreover, only one of them recovered the ability to fly in this round of injections, with respect to 5 in the previous round. Although not appropriately measured, several of the candidates were small, ill-looking, and some even appeared to have wing damage. It is still very likely that these individuals were flightless due to non-heritable causes just as the ones in Table 4.9. Further analysis is required to determine whether these individuals have *Act4* mutations or not.

4.3.9 G1 heterozygous Act4 knock-in candidate females fly normally

20 *3xP3-mCherry* positive G1s were found from *Act4* sgRNA, Cas9 protein, and HDR plasmid donor injections into LVP (Table 4.11). These were treated as *Act4* knock-in heterozygous candidates (Table 4.14). All surviving *Act4* knock-in candidates were able to fly upon eclosion, or 1 day post eclosion in the case of 2H+. This strengthened the hypothesis of *Act4* haplosufficiency, unless these individuals were not real *Act4* knock-ins. They were mostly treated as independent knock-in events, and most of them were crossed individually with

LVPs (Table 4.15). Some, suspected to originate from the same knock-in event, were crossed with each other. The purpose of this was to find out if *Act4* knock-in homozygote females would be flightless (Table 4.16).

Table 4.14: G1s from HDR *Act4* sgRNA injections selected by fluorescent marker expression showed varied phenotypes:

G1s from Table 4.11 and Table 4.12 injections were screened for fluorescent marker expression. The positive candidates had their *Act4* gene sequenced. Some of the sequenced individuals showed *Act4* putative knock-in sequences, all of which are likely the same insertion event. G1 individuals were each given a code consisting of the G0 pool they came from (1-10 for Table 4.11 pools and 11-14 for Table 4.12 pools), letters (A-Z) for different individuals from the same pool, and a + sign to differentiate them from flightless candidates (Table 4.13). The same individuals were mated to observe germline transmission in G2s (Table 4.15).

G1 Individual	♀/♂	G1 Fluorescence Phenotype	G1 Flying Phenotype
1A+	♀	1 fluorescent eye	Flying upon eclosion
2A+	♀	1 fluorescent eye	Flying upon eclosion
2B+	♂	2 fluorescent eyes	Flying upon eclosion
2C+	♂	2 fluorescent eyes	Flying upon eclosion
2D+	N/A	2 fluorescent eyes	Died as prepupa
2E+	N/A	2 fluorescent eyes (1 with weaker expression)	Died as prepupa
2F+	N/A	2 fluorescent eyes	Died as L3 Larva
2G+	N/A	1 fluorescent eye	Died as L2 Larva
2H+	♀	2 fluorescent eyes	Flying 1 day post eclosion
2I+	♂	2 fluorescent eyes	Flying upon eclosion
2J+	♀	2 fluorescent eyes (1 with weaker expression)	Flying upon eclosion
2K+	♀	2 fluorescent eyes (1 with weaker expression)	Flying upon eclosion
2L+	♂	2 fluorescent eyes	Flying upon eclosion
2M+	♂	2 fluorescent eyes	Flying upon eclosion
2N+	N/A	2 fluorescent eyes	Died as prepupa
2O+	♂	2 fluorescent eyes	Flying upon eclosion
2P+	♀	2 fluorescent eyes	Flying upon eclosion
5A+	♀	1 fluorescent eye	Flying upon eclosion
5B+	♂	1 fluorescent eye	Flying upon eclosion
7A+	♂	1 fluorescent eye	Flying upon eclosion

The observed eye expression profile of *3xP3-mCherry* was very variable. It ranged from individuals expressing the marker in a single eye (either the right one or the left), to individuals expressing the marker at a different intensity between the eyes, to individuals expressing the marker in both eyes with equal intensity. Marker expression intensity varied in general. This could be due to the

fact that larval ecdysis (Farnesi *et al.* 2012), with the cyclic build-up of chitin, interferes with the signal strength of the fluorescent marker. However, the variation observed was not entirely dependent on the darkness of the larval cuticle, as Figure 4.11 shows. Individuals with a darker cuticle did not show a weaker fluorescent signal. Another option for the observed variation could have been developmental stage variation of *3xP3* promoter activity. However, one eyed individuals such as 1A+ and 2A+ were looked at from L1 to L4 stages and the fluorescence profile remained unchanged.

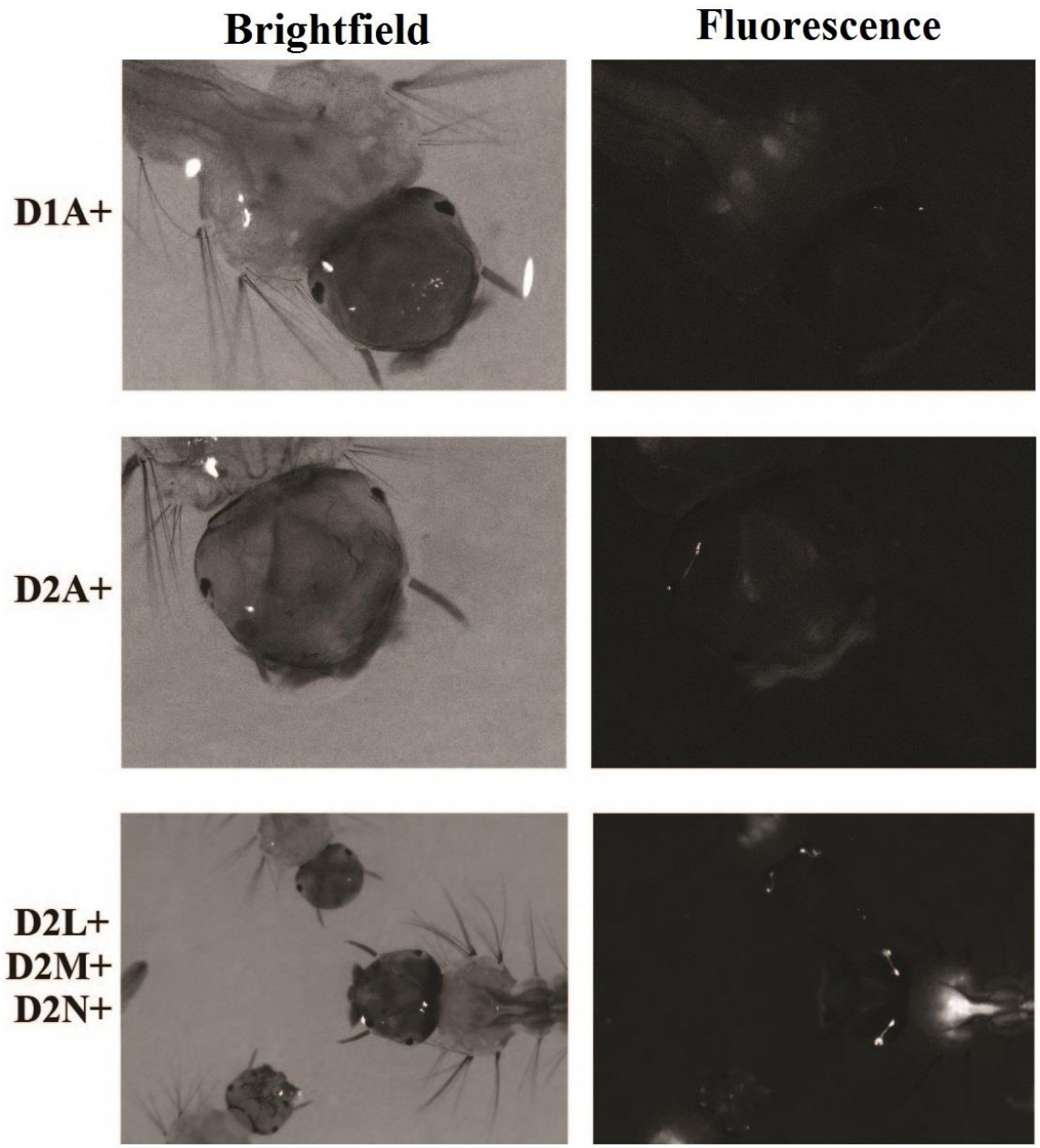


Figure 4.11: G1 fluorescent phenotypes at larval stages for *3xP3-mCherry-SV40* vary for HDR *Act4* injected embryos:

Selected individuals from Table 4.14 evidence the variability of expression at G1 stages. 1A+ and 2A+ only show single eye fluorescence, and even then it seems somewhat incomplete and weak. Strength of expression also varies between individuals as evidenced by 2L-N+ individuals which are most likely the same knock-in event as they come from the same pool. As seen in Table 4.15 germline transmission is unlikely for G1 larvae showing single eye fluorescence. An explanation for this could be transient expression at the G1 stage with donor plasmid being transmitted through the germline of the G0 parents. This is probably more feasible maternally than paternally. However, 1A+ comes from a female G0 pool and 2A+, with stronger fluorescence, comes from a male G0 pool.

4.3.10 Not all Act4 knock-in candidate G1s carried a stable integration of 3xP3-mCherry

15 *Act4* knock-in candidate G1s survived to adulthood and were crossed to LVP or between them. There were 5 single-eye fluorescent individuals originating from 4 different adult G0 pools. The five were treated as individual insertions and were crossed individually to LVP. The other 10 individuals showed fluorescence in both eyes and were all derived from the same G0 pool, 4 of these were crossed individually to LVP whilst the other 6 were crossed amongst each other in three pairs. Table 4.15 shows that only individuals showing fluorescence in both eyes were capable of germline transmission of the marker.

The simplest explanation could be that single-eye fluorescent individuals do not actually have a marker integrated, but have carryover donor plasmid from maternal or paternal G0 contribution. This is especially difficult to picture in terms of paternal contribution given the small volume in sperm compared to eggs. However, two of the single-eye fluorescent-bearing G0 pools consisted of male G0s and two consisted of female G0s. Indeed, if it was due to transient donor plasmid expression, it is difficult to explain why two G2 individuals, amongst the progeny from G1 crosses, displayed single-eye fluorescence. In line with this is the fact that Katharina Von Wyschetzki, another member in the lab, managed to obtain a stable line from a single-eye fluorescent individual with the same marker (Unpublished data), although part of a different construct. If single eye fluorescence is somewhat common amongst G2s of stable integration lines, it

is difficult to understand why out of 5 single-eye fluorescent individuals none carried a stable integration.

The expected Mendelian ratio of 50% fluorescence was approximately met (values from 33% to 61% were observed) amongst the G2 progeny from G1 to LVP crosses. However the expected Mendelian ratio of 75% fluorescence amongst G2s from G1 to G1 crosses was not met for two out of three crosses as the observed percentages were 86 and 93% respectively. All G2s from G1 to LVP crosses flew upon eclosion as expected. When G1s were crossed to other G1s, exhibiting fluorescence in both eyes, a percentage of the female progeny had the flightless phenotype. This is discussed in further detail below.

Table 4.15: G2s from HDR *Act4* sgRNA injections showed germline transmission only when marker expressed in both eyes of G1s:

The same individuals from Table 4.14, i.e. showing *3xP3-mCherry* fluorescence, were mated to LVP or to each other to characterise germline transmission in G2s. Germline transmission only occurred when G1 parents showed 2 fluorescent eyes, it was never observed when G1 parents showed 1 fluorescent eye only. Expression was variable in strength between eyes in several G2 individuals, including 2 with only 1 eyed expression of the marker. This is surprising given that one eyed fluorescence in this case cannot be explained by transient carry-over of the donor plasmid. When G1s were crossed to LVP all G2s were flying upon eclosion. The expected Mendelian ratio of 50% fluorescence was only roughly met (Values from 33% to 61% were observed). When G1s were crossed to other G1s, exhibiting fluorescence in both eyes, a percentage of the female progeny had the flightless phenotype. Suggesting *Act4* is haplosufficient in *Aedes aegypti* (Table 4.16). The expected Mendelian ratio of 75% fluorescence, with these G1 to G1 crosses was not met. Cross 2O+ with 2P+, both heterozygotes for the *3xP3-mCherry* insertion, gave up to 93% of fluorescent G2s.

G1 Individual	Fluorescent Eyes	Mating (to LVP or G1+ve)	Fluorescent G2 Progeny	G2 Fluorescence Comments	G2 Flying Phenotypes
1A+	1	Crossed to LVP	0	N/A	N/A
2A+	1	Crossed to LVP	0	N/A	N/A
2B+	2	Crossed to LVP	31 out of 54 (57%)	1 out of 31 positives showed 1 eyed expression. Expression is variable in strength between both eyes in several individuals	All flying upon eclosion
2C+	2	Crossed to LVP	15 out of 45 (33%)	1 out of 15 positives showed 1 eyed expression. Expression is variable in strength between both eyes in several individuals	All flying upon eclosion
2H+	2	Crossed to LVP	27 out of 44 (61%)	Expression is variable in strength between both eyes in several individuals	All flying upon eclosion
2I+	2	Crossed to LVP	29 out of 63 (46%)	Expression is variable in strength between both eyes in several individuals	All flying upon eclosion
2J+	2	Crossed to 2L+	13 out of 19 (68%)	Expression is variable in strength between both eyes in several individuals	All flying upon eclosion except homozygous females (Table 4.16)
2L+	2	Crossed to 2J+			
2K+	2	Crossed to 2M+	24 out of 28 (86%)	Expression is variable in strength between both eyes in several individuals	All flying upon eclosion except homozygous females (Table 4.16)
2M+	2	Crossed to 2K+			
2O+	2	Crossed to 2P+	41 out of 44 (93%)	Expression is variable in strength between both eyes in several individuals	All flying upon eclosion except homozygous females (Table 4.16)
2P+	2	Crossed to 2O+			
5A+	1	Crossed to LVP	0	N/A	N/A
5B+	1	Crossed to LVP	0	N/A	N/A
7A+	1	Crossed to LVP	0	N/A	N/A

4.3.11 Act4 knock-in was successful and homozygous G2 females are flightless

G1 to G1 crosses described above (Table 4.15) resulted in a significant proportion of flightless females upon eclosion. This is consistent with the hypothesis of *Act4* haplosufficiency and its requirement for flight in females, as well as an indication of a successful integration into *Act4* of the selected individuals (2J+ to 2O+). All of which probably originate from the same *Act4* successful integration event given that they originated from the same G0 pool. The expected percentage of flightless females was 25% given that two G1 heterozygotes had been crossed together. However, the percentages observed were considerably higher at 50, 33, and 67% respectively (Table 4.16). Furthermore, the observed fluorescence in larvae, expected to be approximately 75%, was considerably higher for two of the crosses at 86 and 93% respectively. However, the first cross, with even lower than expected fluorescent larvae still resulted in 50% of G2 females flightless. Given the small size of the G2 eggs collected and hatched, 19, 28 and 44 respectively, not meeting Mendelian inheritance patterns accurately is not entirely surprising. Nevertheless, the current evidence strongly indicates *Act4* haploinsufficiency in *Aedes aegypti*. Interestingly, the fact of a second HDR integration, the first putative HDR insertion described in Chapter 3, originated from a male G0 pool clearly indicates that the procedure of discarding G0 pools from HDR injections is not advisory as previously suggested (Kistler *et al.* 2015).

Table 4.16: *Act4* knock-in heterozygous females are flightless:

When G1s were crossed to other G1s, exhibiting fluorescence in both eyes, a percentage of the female progeny had the flightless phenotype. Whilst the expected G2 larvae fluorescence was 75%; observed percentages were 68, 86, and 93% for each respective cross. The expected percentage of flightless females, should *Act4* be haplosufficient and required for flight in females, was 25% of females being unable to fly. However this percentage was much greater than expected with a 50, 33, and 67% of flightless females respectively. Given the high percentage of G2 fluorescent larvae in the first place, this may not be surprising. However, the first cross, with even lower than expected fluorescent larvae still resulted in 50% of G2 females flightless.

			G2 Adult Progeny				
			Flightless		Flying		Percentage of Flightless Females
G1 ♀	G1 ♂	Fluorescent G2 Larvae	♀	♂	♀	♂	
2J+	2L+	13 out of 19 (68%)	4	0	4	5	50%
2K+	2M+	24 out of 28 (86%)	3	0	6	14	33%
2P+	2O+	41 out of 44 (93%)	8	0	12	20	67%

4.4 Conclusions

4.4.1 Future molecular analysis of Act4 knock-in

Act4 haplosufficiency comes as a substantial surprise given the preliminary evidence for its haploinsufficiency described in the introduction. Yet, the new evidence obtained supports this strongly.

Further molecular analysis on *3xP3-mCherry* germline transmitting individuals (2B+ to 2O+) could not be carried out during the scope of this project. Nevertheless, this additional analysis would be interesting and, should more time be available, the aspects to be explored would be the following. 2B+ to 2O+ individuals could be analysed by gDNA extractions, followed by *Act4* PCRs, and Sanger sequencing. This would confirm whether all had *Act4* integrations. Moreover, with amplicons bridging from *mCherry* to *Act4* sequences at either side of the flanks used, an insight into how the homologous recombination occurred would be attained. Thanks to SNP variations, and the known sequence of the donor plasmid's flanks, it may be possible to find out the approximate points where homologous recombination took place. This would be interesting in terms of finding out how much of the *Act4* flanks were used, and to be able to determine if individuals 2B+ to 2O+ have the same insertion event, the current hypothesis, or there are indeed other independent insertions among them. However, it may be difficult to determine either of these aspects depending on the presence and position of SNPs.

A key aspect to confirm would be to find out if flightless G2 females in Table 4.16 are indeed all *Act4* knock-in homozygotes and whether flying females are *Act4* knock-in heterozygous or WT. There is a possibility that the phenotype does not correlate perfectly with the genotype and that some flightless females are only heterozygotes. *Act4* could be haploinsufficient depending on which WT *Act4* allele complements the *Act4* knock-in allele. The evidence from the injections above would suggest otherwise, however PCRs from these G2 flightless individuals would give a clearer answer to this question. In addition, *Act4* PCRs and Sanger sequencing for gDNA samples from Table 4.13 would help answer this question. These are G1s and must be heterozygotes for *Act4* knock-out, since G0s were outcrossed to LVPs, should they have any *Act4* mutations at all. The current hypothesis is that they are flightless for non-heritable causes, just as individuals in Table 4.9, but this is yet to be confirmed. Moreover, there is a possibility that these individuals are homozygous *Act4* knock-outs if both *Act4* sgRNAs and Cas9 protein were maternally or paternally deposited in G1 zygotes. This possibility remains remote, but some evidence in the literature suggests it may be somewhat plausible (Lin and Potter 2016). However, the evidence is for maternal deposition of endogenously encoded Cas9 and sgRNAs. In the case of this study the deposition would be of Cas9 protein and synthesised sgRNAs injected at the G0 stage, which seems much less plausible if not difficult to picture.

Chapter 5 - **Summary and Concluding Remarks**

Mosquito-borne diseases are a major problem for human health and there will probably be a continued need for vector control. With human population growth and increased urbanisation of once sylvatic habitats, anthropophilic mosquito numbers could increase and spread accordingly. Moreover, the increased abundance of humans may pose selection pressures on sylvatic mosquito species to switch blood meal sources and become anthropophilic. This in turn will pose selection pressures on viruses, and/or parasites to change their host specificity. Examples of this can be seen with the malaria parasite, *Plasmodium knowlesi*, newly identified in humans (Singh and Daneshvar 2013), and dengue serotype (DENV-5), mostly undergoing the sylvatic cycle but already showing an adaptation shift to humans (Mustafa *et al.* 2015). Climate change is also ventured to increase the potential territorial expansion of mosquito vectors (Medlock and Leach 2015). In addition, increased worldwide trade and travel can result in a rapid spread of both mosquitoes and the diseases they transmit. As an example is the rapid expansion of *Aedes albopictus* in southern and central Europe (Bonizzoni *et al.* 2013) as well as the rapid spread of the Zika virus worldwide (Petersen *et al.* 2016). In summary, given these favourable prospects for anthropophilic mosquitoes and the diseases they vector, mosquito vectors cannot be left unchallenged in the interest of world human health.

Targeting diseases themselves is likely to help reduce the disease burden, but is unlikely to eradicate the problem. Research for vaccines against vector-borne diseases has been extensive and although highly effective in some cases such as with yellow fever, dengue and malaria still require the development of effective vaccines. Vector control hence remains a useful option to combat these diseases. Traditional control methods have proven effective in the past, and are still useful, however insecticide resistance and changes in mosquito biting habits threaten their long term viability. In fact, due to *Aedes aegypti's* daytime biting, insecticide-covered bednets have never been particularly effective for the control of this species.

Therefore, increasing the range of available vector control tools is a priority. Genetic engineering of mosquitoes has opened a wide range of new possibilities in vector control. From self-limiting to self-sustaining genetic systems, consolidating a set of different population suppression or replacement tools is likely to be invaluable in the future. It is likely that particular situations, whether regarding the mosquito populations or regulatory bodies, will call for different genetic solutions. This thesis describes efforts, although of modest impact, in increasing the available genetic toolset for mosquito control, in particular against *Aedes aegypti*.

Chapter 2 describes the attempt of building an underdominance gene drive system based on mutually suppressing killing genetic elements. An

underdominance gene drive system would have the advantage of being frequency-dependent and requiring relatively large releases which provides a release based control over the potential persistence, fixation, or even removal of the genetic system from the wild type population. A much needed system that unfortunately was not successfully constructed for this thesis. A complex genetic system such as an insect, even when relatively well-understood as in the case of *D. melanogaster*, poses an important challenge to successful building of orthologous genetic systems. Unpredictable interactions within a complex genetic system can overturn well-intended rational designs. An example of this was the presented work on underdominance; the selected *NIPP1* killer gene could not be upregulated by *tTAV* when under the control of *hsp83*, *UAS* and *tetO*. *tetO* and *tTAV2* resulted in a lethal positive-feedback loop. *Gal4Groucho* and *LexAGroucho* fusion proteins, previously used as corepressors in the literature, were lethal when under the control of the *tetO-tTAV* system. Although, overall the system failed to produce the expected interactions, lessons learnt on how these elements work would allow for the refinement of the system. Developing a robust underdominance system will always be a worthy effort given its potential applicability to the control of mosquito vectors. Interestingly, with the development of gene drive systems based on CRISPR-Cas9, an advantage of underdominance becomes apparent. CRISPR-Cas9 drives face challenges regarding sequence polymorphisms granting sgRNA resistance in a proportion of a wild population. Underdominance gene drives would not face this issue since they are not sequence specific which may be a more or less critical

advantage to CRISPR-Cas9 drives depending on the degree of sequence variability in wild *Aedes aegypti* populations. A question which is yet to be adequately surveyed.

Chapter 3 describes a persistent attempt at introducing an exogenous DNA sequence into the *Nix* (putative male-determining factor) within the M-locus of *Aedes aegypti*. A knock-in was not successful. After injecting more than 5000 embryos and obtaining more than 250 G0 adult survivors an off-target integration of the fluorescent marker was detected amongst the screened G1 progeny. It is difficult to say whether integration into *Nix* was not seen due to the resistance of the M-locus to HDR or simply the fact that a *Nix* knock-in could be potentially lethal in developing G1s. A more repressed chromatin environment, or repetitive DNA region may prevent effective HDR from occurring. Having genetic components integrated in the M locus of *Aedes aegypti* would allow for very interesting gene drive applications such as restricting Cas9 homing reactions in the male germline should this be required. This restriction would be useful should a female-specific haploinsufficient fertility target be found and a RIDL with drive system built around this target. Fresh attempts at HDR in the M locus of *Aedes aegypti* should look at non-functional sites as a better knock-in target.

Chapter 4 describes the attempted building of two different gene drive systems based on presumed *Act4* haploinsufficiency for female flight; one to result in female-specific underdominance and another for RIDL with drive. However,

act4 was unexpectedly found to be haplosufficient for flight in females. This means that the two initial drive designs will not work as expected; female-specific underdominance will simply not be established, whilst the same RIDL with drive becomes a global drive design with *act4* haplosufficiency. *act4* hence can become a recessive female sterile target for a range of gene drive applications such as the payload element in a daisy-chain drive. It is different to other female sterility targets since *act4* is required in the indirect flight muscles, not cells surrounding the germline, and hence less precise germline-Cas9 promoters may be used with a lower risk of a drop in heterozygous female fertility. The surprise of *act4* haplosufficiency raises the question of whether the mutant *Act4* sequence yielding flightless heterozygous females in *Culex quinquefasciatus* was an anti-morph mutant or indeed *Act4* is in that case haploinsufficient

In the relatively short time span of this work, gene drive designs have undergone a rapid and significant renovation with the onset of CRIPSR-Cas9. This is especially the case with homing drive designs. The ease of CRISPR-Cas9 design in comparison to that of HEG, ZFN, or TALEN facilitated the rapid reproduction of homing drive systems in *D. melanogaster* (Gantz and Bier 2015) as well as different species of *Anopheles* (Gantz *et al.* 2015; Hammond *et al.* 2016). Homing rates reported were about ~98% however results are less clear in *D. melanogaster* since a marker was not incorporated with the Cas9 knock-in and

hence NHEJ repaired knock-out events were undistinguishable from true homing events.

Interestingly, advances in the gene drive field whilst work for Chapter 2 was carried out were key in facilitating the body of work presented in Chapter 4. Although perhaps subjective and anecdotal, this seems to suggest how fast this field is advancing and how research in different areas can ultimately aid applied genetics in mosquito control. This provides a positive note for which to remain expectant of the future, and ready for perhaps small but exciting paradigm shifts.

Bibliography

Adelman, Z.N. and Tu, Z. (2016). Control of Mosquito-Borne Infectious Diseases: Sex and Gene Drive. *Trends in Parasitology* **32:**219–229.

Akbari, O.S., Bellen, H.J., Bier, E., Bullock, S.L., Burt, A., Church, G.M., Cook, K.R., *et al.* (2015). Safeguarding gene drive experiments in the laboratory. *Science* **349:**927–929.

Akbari, O.S., Papathanos, P.A., Sandler, J.E., Kennedy, K. and Hay, B.A. (2014). Identification of germline transcriptional regulatory elements in Aedes aegypti. *Scientific Reports* **4:**3954.

Akhtar, W., De Jong, J., Pindyurin, A. V, Pagie, L., Meuleman, W., De Ridder, J., Berns, A., *et al.* (2013). Chromatin position effects assayed by thousands of reporters integrated in parallel. *Cell* **154:**914–927.

Al-Afaleq, A.I. and Hussein, M.F. (2011). The status of Rift Valley fever in animals in Saudi Arabia: a mini review. *Vector-Borne and Zoonotic Diseases* **11:**1513–20.

Alphey, L. (2014). Genetic control of mosquitoes. *Annual Review of Entomology* **59:**205–224.

Alphey, L., Benedict, M., Bellini, R., Clark, G.G., Dame, D.A., Service, M.W. and Dobson, S.L. (2010). Sterile-Insect methods for control of mosquito-borne diseases: an analysis. *Vector-Borne and Zoonotic Diseases* **10:**295–311.

Altrock, P.M., Traulsen, A., Reeves, R.G. and Reed, F.A. (2010). Using underdominance to bi-stably transform local populations. *Journal of Theoretical Biology* **267:**62–75.

An, H.S. and Mogami, K. (1996). Isolation of 88F actin mutants of Drosophila melanogaster and possible alterations in the mutant actin structures. *Journal of Molecular Biology* **260:**492–505.

Andrianantoandro, E., Basu, S., Karig, D.K. and Weiss, R. (2006). Synthetic biology: new engineering rules for an emerging discipline. *Molecular Systems Biology* **2:**2006.0028.

Arthur, B.J., Emr, K.S., Wyttenbach, R.A. and Hoy, R.R. (2014). Mosquito (Aedes aegypti) flight tones: frequency, harmonicity, spherical spreading, and phase relationships. *The Journal of the Acoustical Society of America* **135:**933–941.

Atkinson, M.P., Su, Z., Alphey, N., Alphey, L., Coleman, P.G. and Wein, L.M. (2007). Analyzing the control of mosquito-borne diseases by a dominant lethal genetic system. *Proceedings of the National Academy of Sciences* **104:**9540–9545.

Autino, B., Noris, A., Russo, R. and Castelli, F. (2012). Epidemiology of malaria in endemic areas. *Mediterranean Journal of Hematology and Infectious Diseases* **4:**1–11.

Barnett, E.D. (2007). Yellow fever: epidemiology and prevention. *Clinical Infectious Diseases* **44:**850–856.

Barrangou, R. (2012). RNA-mediated programmable DNA cleavage. *Nature Biotechnology* **30:**836–838.

Barreau, C., Benson, E., Gudmannsdottir, E., Newton, F. and White-Cooper, H. (2008). Post-meiotic transcription in Drosophila testes. *Development* **135:**1897–1902.

Barreau, C., Benson, E. and White-Cooper, H. (2008). Comet and cup genes in Drosophila spermatogenesis: the first demonstration of post-meiotic transcription. *Biochemical Society Transactions* **36:**540–542.

Barrett, A.D.T. and Higgs, S. (2007). Yellow fever: a disease that has yet to be conquered. *Annual Review of Entomology* **52:**209–229.

Basu, S., Aryan, A., Overcash, J.M., Samuel, G.H., Anderson, M.A.E., Dahlem, T.J., Myles, K.M., *et al.* (2015). Silencing of end-joining repair for efficient site-specific gene insertion after TALEN/CRISPR mutagenesis in Aedes aegypti. *Proceedings of the National Academy of Sciences* **112:**4038–4043.

Beckingham, K.M., Armstrong, J.D., Texada, M.J., Munjaal, R. and Baker, D.A. (2005). Drosophila melanogaster - the model organism of choice for the complex biology of multi-cellular organisms. *Gravitational and Space Biology* **18:**17–29.

Behura, S.K. and Severson, D.W. (2011). Coadaptation of isoacceptor tRNA genes and codon usage bias for translation efficiency in Aedes aegypti and Anopheles gambiae. *Insect Molecular Biology* **20:**177–187.

Belfort, M. and Roberts, R.J. (1997). Homing endonucleases: Keeping the house in order. *Nucleic Acids Research* **25:**3379–3388.

Berger, S.L., Cress, W.D., Cress, A., Triezenberg, S.J. and Guarente, L. (1990). Selective inhibition of activated but not basal transcription by the acidic activation domain of VP16: evidence for transcriptional adaptors. *Cell* **61:**1199–1208.

Bernardini, F., Galizi, R., Menichelli, M., Papathanos, P.-A., Dritsou, V., Marois, E., Crisanti, A., *et al.* (2014). Site-specific genetic engineering of the Anopheles gambiae Y chromosome. *Proceedings of the National Academy of Sciences of the United States of America* **111:**7600–7605.
Bhatt, S., Gething, P.W., Brady, O.J., Messina, J.P., Farlow, A.W., Moyes, C.L., Drake, J.M., *et al.* (2013). The global distribution and burden of dengue. *Nature* **496:**504–507.
Biedler, J.K., Hu, W., Tae, H. and Tu, Z. (2012). Identification of early zygotic genes in the yellow fever mosquito Aedes aegypti and discovery of a motif involved in early zygotic genome activation. *PLoS ONE* **7:**6–14.
Boëte, C. and Koella, J.C. (2003). Evolutionary ideas about genetically manipulated mosquitoes and malaria control. *Trends in Parasitology* **19:**32–38.
Bonizzoni, M., Gasperi, G., Chen, X. and James, A.A. (2013). The invasive mosquito species Aedes albopictus: Current knowledge and future perspectives. *Trends in Parasitology* **29:**460–468.
Brady, O.J., Gething, P.W., Bhatt, S., Messina, J.P., Brownstein, J.S., Hoen, A.G., Moyes, C.L., *et al.* (2012). Refining the Global Spatial Limits of Dengue Virus Transmission by Evidence-Based Consensus. *PLoS Neglected Tropical Diseases* **6:**e1760.
Burt, A. (2003). Site-specific selfish genes as tools for the control and genetic engineering of natural populations. *Proceedings of the Royal Society of London, Series B* **270:**921–928.
Burt, A. and Koufopanou, V. (2004). Homing endonuclease genes: The rise and fall and rise again of a selfish element. *Current Opinion in Genetics and Development* **14:**609–615.
Burt, A. and Trivers, R. (2006). *Genes in Conflict: The Biology of Selfish Genetic Elements*. Cambridge,MA: Belknap Press/Harvard University Press.
Burt, F.J., Chen, W., Miner, J.J., Lenschow, D.J., Merits, A., Schnettler, E., Kohl, A., *et al.* (2017). Chikungunya virus: an update on the biology and pathogenesis of this emerging pathogen. *The Lancet Infectious Diseases* **17:**e107–e117.
Carvalho, A.B. (2002). Origin and evolution of the Drosophila Y chromosome. *Current Opinion in Genetics & Development* **12:**664–668.
Cary, L.C., Goebel, M., Corsaro, B.G., Wang, H.G., Rosen, E. and Fraser, M.J. (1989). Transposon mutagenesis of baculoviruses: analysis of Trichoplusia ni transposon IFP2 insertions within the FP-locus of nuclear polyhedrosis viruses. *Virology* **172:**156–169.
Catteruccia, F., Benton, J.P. and Crisanti, A. (2005). An Anopheles transgenic sexing strain for vector control. *Nature Biotechnology* **23:**1414–1417.
Champer, J., Liu, J., Oh, S.Y., Reeves, R., Luthra, A., Oakes, N., Clark, A.G., *et al.* (2017). Reducing resistance allele formation in CRISPR gene drives. *bioRxiv*1–18.
Champer, J., Reeves, R., Oh, S.Y., Liu, C., Liu, J., Clark, A.G. and Messer, P.W. (2017). Novel CRISPR/Cas9 gene drive constructs in Drosophila reveal insights into mechanisms of resistance allele formation and drive efficiency in genetically diverse populations. *PLoS Genetics* **13:**1–18.
Charlesworth, B. (1996). The evolution of chromosomal sex determination and dosage compensation. *Current Biology* **6:**149–162.
Charlesworth, B., Sniegowski, P. and Stephan, W. (1994). The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature* **371:**215–220.
Charlesworth, D., Charlesworth, B. and Marais, G. (2005). Steps in the evolution of heteromorphic sex chromosomes. *Heredity* **95:**118–128.
Charlesworth, D. and Mank, J.E. (2010). The birds and the bees and the flowers and the trees: Lessons from genetic mapping of sex determination in plants and animals. *Genetics* **186:**9–31.
Chen, C., Schaeffer, L. V, Guo, M. and Hay, B.A. (2007). A synthetic maternal-effect selfish genetic element drives population replacement in drosophila. *Science* **316:**597–600.
Chevalier, B.S. and Stoddard, B.L. (2001). Homing endonucleases: structural and functional insight into the catalysts of intron/intein mobility. *Nucleic Acids Research* **29:**3757–3774.
Clemons, A., Haugen, M., Flannery, E., Tomchaney, M., Kast, K., Jacowski, C., Le, C., *et al.* (2010). Aedes aegypti: An emerging model for vector mosquito development. *Cold Spring Harbor Protocols* **5:**1–9.
Cornel, A.J., Benedict, M.Q., Salazar Rafferty, C., Howells, A.J. and Collins, F.H. (1997). Transient expression of the Drosophila melanogaster cinnabar gene rescues eye color in the white eye (WE) strain of Aedes aegypti. *Insect Biochemistry and Molecular Biology* **27:**993–997.
Craig, G.B. (1967). Mosquitoes: female monogamy induced by male accessory gland substance. *Science* **156:**1499–1501.
Crampton, J.M., Beard, C. Ben and Louis, C. (1997). *The molecular biology of insect disease vectors a methods manual*. Crampton, J. M., Beard, C. Ben, and Louis, C. eds. Chapman and Hall Ltd.
Crick, F.H.C. (1966). Codon-anticodon pairing: the wobble hypothesis. *Journal of Molecular Biology* **19:**548–555.

Crovello, T.J. and Hacker, C.S. (1972). Evolutionary strategies in life table characteristics among feral and urban strains of Aedes aegypti (L.). *Evolution* **26:**185–196.
da Cunha, R. V and Trinta, K.S. (2017). Chikungunya virus: clinical aspects and treatment - a Review. *Memórias do Instituto Oswaldo Cruz* **112:**523–531.
Curtis, C.F. (1968). Possible use of translocations to fix desirable genes in insect pest populations. *Nature* **218:**368–369.
Curtis, C.F. and Graves, P.M. (1988). Methods for replacement of malaria vector populations. *Journal of Tropical Medicine and Hygiene* **91:**43–48.
Cutter, A.R. and Hayes, J.J. (2015). A brief review of nucleosome structure. *FEBS Letters* **589:**2914–2922.
Davis, S., Bax, N. and Grewe, P. (2001). Engineered underdominance allows efficient and economical introgression of traits into pest populations. *Journal of Theoretical Biology* **212:**83–98.
Degner, E.C. and Harrington, L.C. (2016). Polyandry depends on postmating time interval in the dengue vector Aedes aegypti. *American Journal of Tropical Medicine and Hygiene* **94:**780–785.
Deitz, K.C., Athrey, G.A., Jawara, M., Overgaard, H.J., Matias, A. and Slotman, M.A. (2016). Genome-Wide Divergence in the West-African Malaria Vector Anopheles melas. *G3: Genes, Genomes, Genetics* **6:**2867–2879.
Deredec, A., Burt, A. and Godfray, H.C.J. (2008). The population genetics of using homing endonuclease genes in vector and pest management. *Genetics* **179:**2013–2026.
Diallo, D., Sall, A.A., Diagne, C.T., Faye, O., Faye, O., Ba, Y., Hanley, K.A., *et al.* (2014). Zika virus emergence in mosquitoes in Southeastern Senegal, 2011. *PLoS ONE* **9:**4–11.
DiCarlo, J.E., Chavez, A., Dietz, S.L., Esvelt, K.M. and Church, G.M. (2015). Safeguarding CRISPR-Cas9 gene drives in yeast. *Nature Biotechnology* **33:**1250–1255.
Disteche, C.M. (2012). Dosage compensation of the sex chromosomes. *Annual Review of Genetics* **46:**537–560.
Dong, S., Lin, J., Held, N.L., Clem, R.J., Passarelli, A.L. and Franz, A.W.E. (2015). Heritable CRISPR/Cas9-mediated genome editing in the yellow fever mosquito, Aedes aegypti. *PLoS ONE* **10:**1–13.
Dujon, B., Belfort, M., Butow, R.A., Jacq, C., Lemieux, C., Perlman, P.S. and Vogt, V.M. (1989). Mobile introns: definition of terms and recommended nomenclature. *Gene* **82:**115–118.
Edgington, M.P. and Alphey, L.S. (2017). Conditions for success of engineered underdominance gene drive systems. *Journal of Theoretical Biology* **430:**128–140.
Elick, T. a, Bauser, C.A. and Fraser, M.J. (1996). Excision of the piggyBac transposable element in vitro is a precise event that is enhanced by the expression of its encoded transposase. *Genetica* **98:**33–41.
Esvelt, K.M., Smidler, A.L., Catteruccia, F. and Church, G.M. (2014). Concerning RNA-guided gene drives for the alteration of wild populations. *eLife* **3:**e03401.
Fang, H., Kang, J. and Zhang, D. (2017). Microbial production of vitamin B12: a review and future perspectives. *Microbial Cell Factories* **16:**15.
Farnesi, L.C., Brito, J.M., Linss, J.G., Pelajo-Machado, M., Valle, D. and Rezende, G.L. (2012). Physiological and morphological aspects of Aedes aegypti developing larvae: effects of the chitin synthesis inhibitor novaluron. *PLoS ONE* **7:**e30363.
Feng, X., Huang, L., Lin, L., Yang, M. and Ma, Y. (2017). Genetic diversity and population structure of the primary malaria vector Anopheles sinensis (Diptera: Culicidae) in China inferred by cox1 gene. *Parasites & Vectors* **10:**75.
Fisher, A.L. and Caudy, M. (1998). Groucho proteins: Transcriptional corepressors for specific subsets of DNA-binding transcription factors in vertebrates and invertebrates. *Genes and Development* **12:**1931–1940.
Fisher, A.L., Ohsako, S. and Caudy, M. (1996). The WRPW motif of the hairy-related basic helix-loop-helix repressor proteins acts as a 4-amino-acid transcription repression and protein-protein interaction domain. *Molecular and Cellular Biology* **16:**2670–7.
Fraser, M.J., Smith, G.E. and Summers, M.D. (1983). Acquisition of host cell DNA sequences by Baculoviruses: relationship between host DNA insertions and FP mutants of autographa californica and galleria mellonella nuclear polyhedrosis viruses. *Journal of Virology* **47:**287–300.
Fredericks, A.C. and Fernandez-Sesma, A. (2014). The burden of dengue and Chikungunya worldwide: implications for the Southern United States and California. *Annals of Global Health* **80:**466–475.
Frierson, J.G. (2010). The yellow fever vaccine: a history. *Yale Journal of Biology and Medicine* **83:**77–85.
Fu, G., Condon, K.C., Epton, M.J., Gong, P., Jin, L., Condon, G.C., Morrison, N.I., *et al.* (2007). Female-specific insect lethality engineered using alternative splicing. *Nature Biotechnology* **25:**353–357.
Fu, G., Lees, R.S., Aw, D., Jin, L., Gray, P., Berendonk, T.U., White-Cooper, H., *et al.* (2010). Female-

specific flightless phenotype for mosquito control. *Proceedings of the National Academy of Sciences* **107:**18741–18741.
Gaj, T., Gersbach, C.A. and Barbas, C.F. (2013). ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends in Biotechnology* **31:**397–405.
Gantz, V.M. and Bier, E. (2015). The mutagenic chain reaction: A method for converting heterozygous to homozygous mutations. *Science* **3042:**1–7.
Gantz, V.M., Jasinskiene, N., Tatarenkova, O., Fazekas, A., Macias, V.M., Bier, E. and James, A.A. (2015). Highly efficient Cas9-mediated gene drive for population modification of the malaria vector mosquito Anopheles stephensi. *Proceedings of the National Academy of Sciences* **112:**E6736–E6743.
Gao, Y. and 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:**343–349.
Gaunt, M.W. and Miles, M.A. (2002). An insect molecular clock dates the origin of the insects and accords with palaeontological and biogeographic landmarks. *Molecular Bology and Evolution* **19:**748–61.
Gloria-Soria, A., Ayala, D., Bheecarry, A., Calderon-Arguedas, O., Chadee, D.D., Chiappero, M., Coetzee, M., *et al.* (2016). Global genetic diversity of Aedes aegypti. *Molecular Ecology* **25:**5377–5395.
Gnerer, J.P., Venken, K.J.T. and Dierick, H.A. (2015). Gene-specific cell labeling using MiMIC transposons. *Nucleic Acids Research* **43:**e56.
Gong, P., Epton, M.J., Fu, G., Scaife, S., Hiscox, A., Condon, K.C., Condon, G.C., *et al.* (2005). A dominant lethal genetic system for autocidal control of the Mediterranean fruitfly. *Nature Biotechnology* **23:**453–456.
Gossen, M. and Bujard, H. (1992). Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. *Proceedings of the National Academy of Sciences* **89:**5547–51.
Gould-Somero, M. and Holland, L. (1974). The timing of RNA synthesis for spermiogenesis in organ cultures of Drosophila melanogaster testes. *Wilhelm Roux' Archiv für Entwicklungsmechanik der Organismen* **174:**133–148.
Gregory, M., Alphey, L., Morrison, N.I. and Shimeld, S.M. (2016). Insect transformation with piggyBac: Getting the number of injections just right. *Insect Molecular Biology* **25:**259–271.
Grimm, D., Wang, L., Lee, J.S., Schürmann, N., Gu, S., Börner, K., Storm, T.A., *et al.* (2010). Argonaute proteins are key determinants of RNAi efficacy, toxicity, and persistence in the adult mouse liver. *Journal of Clinical Investigation* **120:**3106–3119.
Gubler, D.J. (2002). Epidemic dengue/dengue hemorrhagic fever as a public health, social and economic problem in the 21st century. *Trends in Microbiology* **10:**100–103.
Guzman, M.G. and Harris, E. (2015). Dengue. *The Lancet* **385:**453–465.
Hadinegoro, S.R., Arredondo-García, J.L., Capeding, M.R., Deseda, C., Chotpitayasunondh, T., Dietze, R., Ismail, H.I.H.M., *et al.* (2015). Efficacy and Long-Term Safety of a Dengue Vaccine in Regions of Endemic Disease. *New England Journal of Medicine* **373:**1195–1206.
Haghighat-Khah, R.E., Scaife, S., Martins, S., St. John, O., Matzen, K.J., Morrison, N. and Alphey, L. (2015). Site-specific cassette exchange systems in the aedes aegypti mosquito and the plutella xylostella moth. *PLoS ONE* **10:**1–16.
Hall, A.B., Basu, S., Jiang, X., Qi, Y., Timoshevskiy, V.A., Biedler, J.K., Sharakhova, M. V, *et al.* (2015). A male-determining factor in the mosquito Aedes aegypti. *Science* **348:**1268–1270.
Hall, A.B., Timoshevskiy, V.A., Sharakhova, M. V, Jiang, X., Basu, S., Anderson, M.A.E., Hu, W., *et al.* (2014). Insights into the preservation of the homomorphic sex-determining chromosome of Aedes aegypti from the discovery of a male-biased gene tightly linked to the M-locus. *Genome Biology and Evolution* **6:**179–191.
Halstead, S.B. (2017). Which dengue vaccine approach is the most promising, and should we be concerned about enhanced disease after vaccination? there is only one true winner. *Cold Spring Harbor Perspectives in Biology*a030700.
Hammond, A., Galizi, R., Kyrou, K., Simoni, A., Siniscalchi, C., Katsanos, D., Gribble, M., *et al.* (2016). A CRISPR-Cas9 gene drive system targeting female reproduction in the malaria mosquito vector Anopheles gambiae. *Nature Biotechnology* **34:**78–83.
Hammond, A., Kyrou, K., Bruttini, M., North, A., Galizi, R., Karlsson, X., Carpi, F.M., *et al.* (2017). The creation and selection of mutations resistant to a gene drive over multiple generations in the malaria mosquito. *bioRxiv*1–17.
Han, Q., Beerntsen, B.T. and Li, J. (2007). The tryptophan oxidation pathway in mosquitoes with emphasis on xanthurenic acid biosynthesis. *Journal of Insect Physiology* **53:**254–263.
Han, Q., Calvo, E., Marinotti, O., Fang, J., Rizzi, M., James, A.A. and Li, J. (2003). Analysis of the wild-type and mutant genes encoding the enzyme kynurenine monooxygenase of the yellow fever mosquito, Aedes aegypti. *Insect Molecular Biology* **12:**483–490.
Handler, A.M. and Harrell, R.A. (1999). Germline transformation of Drosophila melanogaster with the

piggyBac transposon vector. *Insect Molecular Biology* **8:**449–457.
Handley, L.-J.L., Ceplitis, H. and Ellegren, H. (2004). Evolutionary strata on the chicken Z chromosome: implications for sex chromosome evolution. *Genetics* **167:**367–376.
Harris, A.F., Nimmo, D., McKemey, A.R., Kelly, N., Scaife, S., Donnelly, C.A., Beech, C., *et al.* (2011). Field performance of engineered male mosquitoes. *Nature Biotechnology* **29:**1034–1037.
Hay, B.A., Chen, C.-H., Ward, C.M., Huang, H., Su, J.T. and Guo, M. (2010). Engineering the genomes of wild insect populations: challenges, and opportunities provided by synthetic Medea selfish genetic elements. *Journal of Insect Physiology* **56:**1402–1413.
Helinski, M.E.H., Parker, A.G. and Knols, B.G.J. (2009). Radiation biology of mosquitoes. *Malaria Journal* **8:**1–13.
Hink, W.F. (1970). Established insect cell line from the cabbage looper, Trichoplusia ni. *Nature* **226:**466–467.
Holt, R.A., Subramanian, G.M., Halpern, A., Sutton, G.G., Charlab, R., Nusskern, D.R., Wincker, P., *et al.* (2002). The genome sequence of the malaria mosquito Anopheles gambiae. *Science* **298:**129–149.
Horlbeck, M.A., Witkowsky, L.B., Guglielmi, B., Replogle, J.M., Gilbert, L.A., Villalta, J.E., Torigoe, S.E., *et al.* (2016). Nucleosomes impede cas9 access to DNA in vivo and in vitro. *eLife* **5:**1–21.
Hotta, S. (1951). Experimental studies on dengue. *Journal of Infectious Diseases* **90:**1–9.
Hurst, G.D. and Werren, J.H. (2001). The role of selfish genetic elements in eukaryotic evolution. *Nature Reviews Genetics* **2:**597–606.
Jansen, V.A.A., Turelli, M. and Godfray, H.C.J. (2008). Stochastic spread of Wolbachia. *Proceedings of the Royal Society B: Biological Sciences* **275:**2769–2776.
Jasin, M. and Rothstein, R. (2013). Repair of strand breaks by homologous recombination. *Cold Spring Harbor Perspectives in Biology* **5:**1–18.
Jones, J.C. and Wheeler, R.E. (1965). An analytical study of coitus in Aedea aegypti (Linnaeus). *Journal of Morphology* **117:**401–423.
Kim, D.-U., Hayles, J., Kim, D., Wood, V., Park, H.-O., Won, M., Yoo, H.-S., *et al.* (2010). Analysis of a genome-wide set of gene deletions in the fission yeast Schizosaccharomyces pombe. *Nature Biotechnology* **28:**617–623.
Kim, D.K., Kanai, Y., Chairoungdua, A., Matsuo, H., Cha, S.H. and Endou, H. (2001). Expression cloning of a Na+-independent aromatic amino acid transporter with structural similarity to H+/monocarboxylate transporters. *Journal of Biological Chemistry* **276:**17221–17228.
Kistler, K.E., Vosshall, L.B. and Matthews, B.J. (2015). Genome engineering with CRISPR-Cas9 in the mosquito aedes aegypti. *Cell Reports* **11:**51–60.
Kiszewski, A., Mellinger, A., Spielman, A., Malaney, P., Sachs, S.E. and Sachs, J. (2004). A global index representing the stability of malaria transmission. *American Journal of Tropical Medicine and Hygiene* **70:**486–498.
Knipling, E.F., Laven, H., Craig, G.B., Pal, R., Kitzmiller, J.B., Smith, C.N. and Brown, A.W.A. (1968). Genetic control of insects of public health importance. *Bulletin of the World Health Organization* **38:**421–438.
Kokoza, V., Ahmed, A., Wimmer, E.A. and Raikhel, A.S. (2001). Efficient transformation of the yellow fever mosquito Aedes aegypti using the piggyBac transposable element vector pBac[3xP3-EGFP afm]. *Insect Biochemistry and Molecular Biology* **31:**1137–1143.
Korochkina, S.E., Gordadze, A. V, York, J.L. and Benes, H. (1997). Mosquito hexamerins: characterization during larval development. *Insect Molecular Biology* **6:**11–21.
Kraemer, M.U.G., Sinka, M.E., Duda, K.A., Mylne, A., Shearer, F.M., Brady, O.J., Messina, J.P., *et al.* (2015). The global compendium of Aedes aegypti and Ae. albopictus occurrence. *Scientific Data* **2:**150035.
Krzywinski, J., Grushko, O.G. and Besansky, N.J. (2006). Analysis of the complete mitochondrial DNA from Anopheles funestus: An improved dipteran mitochondrial genome annotation and a temporal dimension of mosquito evolution. *Molecular Phylogenetics and Evolution* **39:**417–423.
Krzywinski, J., Nusskern, D.R., Kern, M.K. and Besansky, N.J. (2004). Isolation and characterization of Y chromosome sequences from the african malaria mosquito Anopheles gambiae. *Genetics* **166:**1291–1302.
Kurkulos, M., Weinberg, J.M., Pepling, M.E. and Mount, S.M. (1991). Polyadenylylation in copia requires unusually distant upstream sequences. *Proceedings of the National Academy of Sciences of the United States of America* **88:**3038–42.
Kwit, C., Moon, H.S., Warwick, S.I. and Stewart, C.N. (2011). Transgene introgression in crop relatives: molecular evidence and mitigation strategies. *Trends in Biotechnology* **29:**284–293.
Kwok, R. (2010). Five hard truths for synthetic Biology. *Nature* **463:**288–290.
Labbé, G.M.C., Nimmo, D.D. and Alphey, L. (2010). Piggybac- and PhiC31-mediated genetic

transformation of the Asian tiger mosquito, Aedes albopictus (Skuse). *PLoS Neglected Tropical Diseases* **4:**e788.
Labbé, G.M.C., Scaife, S., Morgan, S.A., Curtis, Z.H. and Alphey, L. (2012). Female-specific flightless (fsRIDL) phenotype for control of Aedes albopictus. *PLoS Neglected Tropical Diseases* **6:**e1724.
Lacroix, R., McKemey, A.R., Raduan, N., Kwee Wee, L., Hong Ming, W., Guat Ney, T., Siti Rahidah, A.A., *et al.* (2012). Open field release of genetically engineered sterile male Aedes aegypti in Malaysia. *PLoS ONE* **7:**1–9.
Lahn, B.T. and Page, D.C. (1999). Four evolutionary strata on the human X chromosome. *Science* **286:**964–967.
Lai, W.K.M. and Pugh, B.F. (2017). Understanding nucleosome dynamics and their links to gene expression and DNA replication. *Nature Reviews Molecular Cell Biology* **18:**548–562.
Larsen, D.A., Borrill, L., Patel, R. and Fregosi, L. (2017). Reported community-level indoor residual spray coverage from two-stage cluster surveys in sub-Saharan Africa. *Malaria Journal* **16:**1–9.
Lea, A.O. (1968). Mating without insemination in virgin Aedes aegypti. *Journal of Insect Physiology* **14:**305–308.
Lengeler, C. (2004). Insecticide-treated bed nets and curtains for preventing malaria. *Cochrane Database of Systematic Reviews*.
Li, K., Wang, G., Andersen, T., Zhou, P. and Pu, W.T. (2014). Optimization of genome engineering approaches with the CRISPR/Cas9 system. *PLoS ONE* **9:**e105779.
Li, M., Akbari, O.S. and White, B.J. (2017). Highly efficient site-specific mutagenesis in Malaria mosquitoes using CRISPR. *bioRxiv*1–28.
Li, M., Bui, M., Yang, T., White, B.J. and Akbari, O.S. (2017). Germline Cas9 Expression Yields Highly Efficient Genome Engineering in a Major Worldwide Disease Vector. *bioRxiv*0–47.
Lieber, M.R. (2010). The Mechanism of Double-Strand DNA Break Repair by the Nonhomologous DNA End-Joining Pathway. *Annual Review of Biochemistry* **79:**181–211.
Lin, C.-C. and Potter, C.J. (2016). Non-Mendelian Dominant Maternal Effects Caused by CRISPR/Cas9 Transgenic Components in Drosophila melanogaster. *G3: Genes, Genomes, Genetics* **6:**3685–3691.
Linzen, B. (1974). The tryptophan-ommochrome pathway in insects. *Advances in Insect Physiology* **10:**117–246.
Lobo, N.F., Clayton, J.R., Fraser, M.J., Kafatos, F.C. and Collins, F.H. (2006). High efficiency germ-line transformation of mosquitoes. *Nature Protocols* **1:**1312–1317.
Lycett, G.J., Kafatos, F.C. and Loukeris, T.G. (2004). Conditional expression in the malaria mosquito Anopheles stephensi with Tet-on and Tet-off systems. *Genetics* **167:**1781–1790.
Macdonald, W.W. (1962). The Selection of a Strain of aëdes Aegypti Susceptible to Infection with Semi-Periodic Brugia Malayi. *Annals of Tropical Medicine & Parasitology* **56:**368–372.
Mackenzie, S.M., Howells, A.J., Cox, G.B. and Ewart, G.D. (2000). Sub-cellular localisation of the White/Scarlet ABC transporter to pigment granule membranes within the compound eye of Drosophila melanogaster. *Genetica* **108:**239–252.
Magori, K. and Gould, F. (2006). Genetically engineered underdominance for manipulation of pest populations: a deterministic model. *Genetics* **172:**2613–2620.
Mahmoudi, S. and Keshavarz, H. (2017). Efficacy of phase 3 trial of RTS, S/AS01 malaria vaccine: The need for an alternative development plan. *Human Vaccines and Immunotherapeutics* **5515:**1–4.
Majumdar, S. and Rio, D.C. (2002). P transposable elements in Drosophila melanogaster and other eukaryotic organisms. *Microbiology Spectrum*484–518.
Mansfield, K.L., Banyard, A.C., McElhinney, L., Johnson, N., Horton, D.L., Hernández-Triana, L.M. and Fooks, A.R. (2015). Rift Valley fever virus: A review of diagnosis and vaccination, and implications for emergence in Europe. *Vaccine* **33:**5520–5531.
Marín, I. and Baker, B.S. (1998). The evolutionary dynamics of sex determination. *Science* **281:**1990.
Marrelli, M.T., Moreira, C.K., Kelly, D., Alphey, L. and Jacobs-Lorena, M. (2006). Mosquito transgenesis: what is the fitness cost? *Trends in Parasitology* **22:**197–202.
Marshall, J.M. (2008). The impact of dissociation on transposon-mediated disease control strategies. *Genetics* **178:**1673–1682.
Marshall, J.M., Buchman, A., Sánchez C, H.M. and Akbari, O.S. (2017). Overcoming evolved resistance to population-suppressing homing-based gene drives. *Scientific Reports* **7:**3776.
Marshall, J.M. and Hay, B.A. (2012). Confinement of gene drive systems to local populations: a comparative analysis. *Journal of Theoretical Biology* **294:**153–171.
Marston, A.L. and Amon, A. (2004). Meiosis: cell-cycle controls shuffle and deal. *Nature Reviews Molecular Cell Biology* **5:**983–997.
Marygold, S.J., Roote, J., Reuter, G., Lambertsson, A., Ashburner, M., Millburn, G.H., Harrison, P.M., *et al.* (2007). The ribosomal protein genes and Minute loci of Drosophila melanogaster. *Genome Biology*

8:R216.
Massonnet-Bruneel, B., Corre-Catelin, N., Lacroix, R., Lees, R.S., Hoang, K.P., Nimmo, D., Alphey, L., *et al.* (2013). Fitness of Transgenic Mosquito Aedes aegypti Males Carrying a Dominant Lethal Genetic System. *PLoS ONE* **8:**1–7.
Matthews, B.J., Dudchenko, O., Kingan, S., Koren, S., Antoshechkin, I., Crawford, J.E., Glassford, W.J., *et al.* (2017). Improved Aedes aegypti mosquito reference genome assembly enables biological discovery and vector control. *bioRxiv*240747.
McBride, C.S., Baier, F., Omondi, A.B., Spitzer, S.A., Lutomiah, J., Sang, R., Ignell, R., *et al.* (2014). Evolution of mosquito preference for humans linked to an odorant receptor. *Nature* **515:**222–227.
McClelland, G.A.H. (1962). Sex-linkage in Aedes aegypti. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **56:**4.
McGraw, E.A. and O'Neill, S.L. (2013). Beyond insecticides: new thinking on an ancient problem. *Nature Reviews Microbiology* **11:**181–193.
Medlock, J.M. and Leach, S.A. (2015). Effect of climate change on vector-borne disease risk in the UK. *The Lancet Infectious Diseases* **15:**721–730.
Mekler, V., Minakhin, L., Semenova, E., Kuznedelov, K. and Severinov, K. (2016). Kinetics of the CRISPR-Cas9 effector complex assembly and the role of 3′-terminal segment of guide RNA. *Nucleic Acids Research* **44:**2837–2845.
Mendis, K., Sina, B.J., Marchesini, P. and Carter, R. (2001). The neglected burden of Plasmodium vivax malaria. *American Journal of Tropical Medicine and Hygiene* **64:**97–106.
Messina, J.P., Brady, O.J., Scott, T.W., Zou, C., Pigott, D.M., Duda, K.A., Bhatt, S., *et al.* (2014). Global spread of dengue virus types: Mapping the 70 year history. *Trends in Microbiology* **22:**138–146.
Molina-Cruz, A. and Barillas-Mury, C. (2014). The remarkable journey of adaptation of the Plasmodium falciparum malaria parasite to New World anopheline mosquitoes. *Memorias do Instituto Oswaldo Cruz* **109:**662–667.
Monath, T.P. and Vasconcelos, P.F.C. (2015). Yellow fever. *Journal of Clinical Virology* **64:**160–173.
Moreira, L. a, Wang, J., Collins, F.H. and Jacobs-Lorena, M. (2004). Fitness of anopheline mosquitoes expressing transgenes that inhibit Plasmodium development. *Genetics* **166:**1337–1341.
Morgan, T.H. (1910). Sex limited inheritance in drosophila. *Science* **32:**120–122.
Morris, A.C. (1997). Microinjection of mosquito embryos. In: Crampton, J. M., Beard, C. Ben, and Louis, C. eds. *The Molecular Biology of Insect Disease Vectors*. Dordrecht: Springer Netherlands, 423–429.
Motara, M.A. and Rai, K.S. (1978). Giemsa C-banding patterns in Aedes (Stegomyia) mosquitoes. *Chromosoma* **70:**51–58.
Mustafa, M.S., Rasotgi, V., Jain, S. and Gupta, V. (2015). Discovery of fifth serotype of dengue virus (denv-5): A new public health dilemma in dengue control. *Medical Journal Armed Forces India* **71:**67–70.
Nandagopal, N. and Elowitz, M.B. (2011). Synthetic biology: integrated gene circuits. *Science* **333:**1244–1248.
Nene, V., Wortman, J.R., Lawson, D., Haas, B., Kodira, C., Tu, Z.J., Loftus, B., *et al.* (2007). Genome sequence of Aedes aegypti, a major arbovirus vector. *Science* **316:**1718–1723.
Neural Network (2017). BDGP: Splice Site Prediction [Online].
Newton, M.E., Wood, R.J. and Southern, D.I. (1978). Cytological mapping of the M and D loci in the mosquito, Aedes aegypti (L.). *Genetica* **48:**137–143.
Nicolas, M., Marais, G., Hykelova, V., Janousek, B., Laporte, V., Vyskot, B., Mouchiroud, D., *et al.* (2005). A gradual process of recombination restriction in the evolutionary history of the sex chromosomes in dioecious plants. *PLoS Biology* **3:**.
Nijhout, H.F. and Craig, G.B. (1971). Reproductive isolation in stegomyia mosquitoes. iii evidence for a sexual pheromone. *Entomologia Experimentalis et Applicata* **14:**399–412.
Nimmo, D.D., Alphey, L., Meredith, J.M. and Eggleston, P. (2006). High efficiency site-specific genetic engineering of the mosquito genome. *Insect Molecular Biology* **15:**129–136.
Noble, C., Min, J., Olejarz, J., Buchthal, J., Chavez, A., Smidler, A.L., DeBenedictis, E.A., *et al.* (2016). Daisy-chain gene drives for the alteration of local populations. *bioRxiv*57307.
Noble, C., Olejarz, J., Esvelt, K.M., Church, G.M. and Nowak, M.A. (2017). Evolutionary dynamics of CRISPR gene drives. *Science Advances* **3:**e1601964.
O'Connor, M.J. and Chia, W. (2002). Gene transfer in Drosophila. In: Clarke, A. R. ed. *Methods in molecular biology: transgenesis techniques: principles and protocols*. 2nd ed. Totowa, NJ: Humana Press Inc., 27–36.
Olivieri, G. and Olivieri, A. (1965). Autoradiographic study of nucleic acid synthesis during spermatogenesis in Drosophila melanogaster. *Mutation Research* **2:**366–380.
Paixão, E.S., Barreto, F., Da Glória Teixeira, M., Da Conceição N Costa, M. and Rodrigues, L.C. (2016).

History, epidemiology, and clinical manifestations of Zika: A systematic review. *American Journal of Public Health* **106:**606–612.
Parker, L., Gross, S., Beullens, M., Bollen, M., Bennett, D. and Alphey, L. (2002). Functional interaction between nuclear inhibitor of protein phosphatase type 1 (NIPP1) and protein phosphatase type 1 (PP1) in Drosophila: consequences of over-expression of NIPP1 in flies and suppression by co-expression of PP1. *Biochemical Journal* **368:**789–797.
Perera, O.P., Harrell, R.A. and Handler, A.M. (2002). Germ-line transformation of the South American malaria vector, Anopheles albimanus, with a piggyBac/EGFP transposon vector is routine and highly efficient. *Insect Molecular Biology* **11:**291–297.
Perezgasga, L., Jiang, J., Bolival, B., Hiller, M., Benson, E., Fuller, M.T. and White-Cooper, H. (2004). Regulation of transcription of meiotic cell cycle and terminal differentiation genes by the testis-specific Zn-finger protein matotopetli. *Development* **131:**1691–1702.
Petersen, L.R., Jamieson, D.J., Powers, A.M. and Honein, M.A. (2016). Zika virus. *New England Journal of Medicine* **374:**1552–1563.
Pfeiffer, B.D., Ngo, T.-T.B., Hibbard, K.L., Murphy, C., Jenett, A., Truman, J.W. and Rubin, G.M. (2010). Refinement of tools for targeted gene expression in Drosophila. *Genetics* **186:**735–755.
Phillips, J.P., Forrest, H.S. and Kulkarni, A.D. (1973). Terminal synthesis of xanthommatin in Drosophila melanogaster. III. mutational pleiotropy and pigment granule association of phenoxazinone synthetase. *Genetics* **73:**45–56.
Pialoux, G., Gaüzère, B.A., Jauréguiberry, S. and Strobel, M. (2007). Chikungunya, an epidemic arbovirosis. *Lancet Infectious Diseases* **7:**319–327.
Plourde, A.R. and Bloch, E.M. (2016). A literature review of Zika virus. *Emerging Infectious Diseases* **22:**1185–1192.
Port, F., Chen, H.-M., Lee, T. and 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:**E2967-2976.
Potter, C.J., Tasic, B., Russler, E. V, Liang, L. and Luo, L. (2010). The Q system: a repressible binary system for transgene expression, lineage tracing, and mosaic analysis. *Cell* **141:**536–548.
Powell, J.R. and Tabachnick, W.J. (2013). History of domestication and spread of Aedes aegypti--a review. *Memórias do Instituto Oswaldo Cruz* **108:**11–17.
Raab, J.R. and Kamakaka, R.T. (2010). Insulators and promoters: closer than we think. *Nature Reviews Genetics* **11:**439–446.
Radzisheuskaya, A., Shlyueva, D., Müller, I. and Helin, K. (2016). Optimizing sgRNA position markedly improves the efficiency of CRISPR/dCas9-mediated transcriptional repression. *Nucleic Acids Research* **44:**e141.
Rai, K.S. (1963). A comparative study of the mosquito karyotypes. *Annals of the Entomological Society of America* **56:**160–170.
Ramirez, J.L., Garver, L.S. and Dimopoulos, G. (2009). Challenges and approaches for mosquito targeted malaria control. *Current Molecular Medicine* **9:**116–130.
Reed, R.D. and Nagy, L.M. (2005). Evolutionary redeployment of a biosynthetic module: Expression of eye pigment genes vermilion, cinnabar, and white in butterfly wing development. *Evolution and Development* **7:**301–311.
Reese, M.G., Eeckman, F.H., Kulp, D. and Haussler, D. (1997). Improved Splice. *Journal of Computational Biology* **4:**311–323.
Reeves, R.G., Bryk, J., Altrock, P.M., Denton, J.A. and Reed, F.A. (2014). First steps towards underdominant genetic transformation of insect populations. *PLoS ONE* **9:**e97557.
Rice, W.R. (1996). Evolution of the Y sex in animals: Y chromosomes evolve through the degeneration of autosomes. *BioScience* **46:**331–343.
Rothman, A.L. and Ennis, F.A. (2016). Dengue vaccine: the need, the challenges, and progress. *Journal of Infectious Diseases* **214:**825–827.
RTS, S.C.T.P. (2015). Efficacy and safety of RTS,S/AS01 malaria vaccine with or without a booster dose in infants and children in Africa: final results of a phase 3, individually randomised, controlled trial. *The Lancet* **6736:**1–15.
Rupp, H.R. (1996). Adverse assessments of Gambusia affinis: an alternate view for mosquito control practitioners. *Journal of the American Mosquito Control Association* **12:**155–166.
Russell, T.L., Govella, N.J., Azizi, S., Drakeley, C.J., Kachur, S.P. and Killeen, G.F. (2011). Increased proportions of outdoor feeding among residual malaria vector populations following increased use of insecticide-treated nets in rural Tanzania. *Malaria Journal* **10:**80.
Salazar, M.I., Richardson, J.H., Sánchez-Vargas, I., Olson, K.E. and Beaty, B.J. (2007). Dengue virus type 2: replication and tropisms in orally infected Aedes aegypti mosquitoes. *BMC Microbiology* **7:**9.

Salz, H.K. and Erickson, J.W. (2010). Sex determination in Drosophila: the view from the top. *Fly* **4:**60–70.
Scott, T.W., Morrison, A.C., Lorenz, L.H., Clark, G.G., Strickman, D., Kittayapong, P., Zhou, H., *et al.* (2000). Longitudinal studies of Aedes aegypti (Diptera : Culicidae) in Thailand and Puerto Rico : population dynamics longitudinal studies of Aedes aegypti (Diptera : Culicidae) in Thailand and Puerto Rico : population dynamics. *Journal of Medical Entomology* **37:**77–88.
Screaton, G. and Mongkolsapaya, J. (2017). Which dengue vaccine approach is the most promising, and should we be concerned about enhanced disease after vaccination? the challenges of a dengue vaccine. *Cold Spring Harbor Perspectives in Biology*a029520.
Sethuraman, N. and O'Brochta, D.A. (2005). The Drosophila melanogaster cinnabar gene is a cell autonomous genetic marker in Aedes aegypti (Diptera: Culicidae). *Journal of Medical Entomology* **42:**716–718.
Sharakhova, M. V, Timoshevskiy, V.A., Yang, F., Demin, S.I., Severson, D.W. and Sharakhov, I. V (2011). Imaginal discs - A new source of chromosomes for genome mapping of the yellow fever mosquito aedes aegypti. *PLoS Neglected Tropical Diseases* **5:**e1335.
Simmons, C.P., Farrar, J.J., Chau, N. van V. and Wills, B. (2012). Dengue. *The New England Journal of Medicine* **1:**210–216.
Simoni, A., Siniscalchi, C., Chan, Y.S., Huen, D.S., Russell, S., Windbichler, N. and Crisanti, A. (2014). Development of synthetic selfish elements based on modular nucleases in Drosophila melanogaster. *Nucleic Acids Research* **42:**7461–7472.
Singh, B. and Daneshvar, C. (2013). Human infections and detection of plasmodium knowlesi. *Clinical Microbiology Reviews* **26:**165–184.
Sinkins, S.P. and Gould, F. (2006). Gene drive systems for insect disease vectors. *Nature Reviews Genetics* **7:**427–435.
Smith, R.C., Walter, M.F., Hice, R.H., O'Brochta, D.A. and Atkinson, P.W. (2007). Testis-specific expression of the β2 tubulin promoter of Aedes aegypti and its application as a genetic sex-separation marker. *Insect Molecular Biology* **16:**61–71.
Snyder, L.L., Ahmed, I. and Steel, L.F. (2009). RNA polymerase III can drive polycistronic expression of functional interfering RNAs designed to resemble microRNAs. *Nucleic Acids Research* **37:**1–10.
Song, B.H., Yun, S.I., Woolley, M. and Lee, Y.M. (2017). Zika virus: History, epidemiology, transmission, and clinical presentation. *Journal of Neuroimmunology* **308:**50–64.
Sprinzak, D. and Elowitz, M.B. (2005). Reconstruction of genetic circuits. *Nature* **438:**443–448.
Stouthamer, R., Breeuwer, J.A.J. and Hurst, G.D.D. (1999). Wolbachia Pipientis: Microbial Manipulator of Arthropod Reproduction. *Annual Review of Microbiology* **53:**71–102.
Sullivan, D.T., Grillo, S.L. and Kitos, R.J. (1974). Subcellular localization of the first three enzymes of the ommochrome synthetic pathway in Drosophila melanogaster. *Journal of Experimental Zoology* **188:**225–233.
Sutton, E.R., Yu, Y., Shimeld, S.M., White-Cooper, H. and Alphey, L. (2016). Identification of genes for engineering the male germline of Aedes aegypti and Ceratitis capitata. *BMC Genomics* **17:**1–16.
Telfer, W.H. and Kunkel, J.G. (1991). The function and evolution of insect storage hexamers. *Annual Review of Entomology* **36:**205–228.
Thomas, D.D., Donnelly, C.A., Wood, R.J. and Alphey, L. (2000). Insect Population Control Using a Dominant, Repressible, Lethal Genetic System. *Science* **287:**2474–2476.
Timoshevskiy, V.A., Kinney, N.A., DeBruyn, B.S., Mao, C., Tu, Z., Severson, D.W., Sharakhov, I. V, *et al.* (2014). Genomic composition and evolution of Aedes aegypti chromosomes revealed by the analysis of physically mapped supercontigs. *BMC Biology* **12:**27.
Timoshevskiy, V.A., Severson, D.W., DeBruyn, B.S., Black, W.C., Sharakhov, I. V and Sharakhova, M. V (2013). An integrated linkage, chromosome, and genome map for the yellow fever mosquito Aedes aegypti. *PLoS Neglected Tropical Diseases* **7:**e2052.
Totten, D.C., Vuong, M., Litvinova, O. V, Jinwal, U.K., Gulia-Nuss, M., Harrell, R.A. and Beneš, H. (2013). Targeting gene expression to the female larval fat body of transgenic Aedes aegypti mosquitoes. *Insect Molecular Biology* **22:**18–30.
Toups, M.A. and Hahn, M.W. (2010). Retrogenes reveal the direction of sex-chromosome evolution in mosquitoes. *Genetics* **186:**763–766.
Turner, J., Krishna, R., Van 't Hof, A.E., Sutton, E.R., Matzen, K. and Darby, A.C. (2017). The sequence of a male-specific genome region containing the sex determination switch in Aedes aegypti. *bioRxiv*.
Turtle, L., Griffiths, M.J. and Solomon, T. (2012). Encephalitis caused by flaviviruses. *QJM* **105:**219–223.
Vaughan, A.M. and Kappe, S.H.I. (2012). Malaria vaccine development: persistent challenges. *Current Opinion in Immunology* **24:**324–331.

Vector Base (2017). Liverpool | VectorBase [Online].
Venken, K.J.T. and Bellen, H.J. (2005). Emerging technologies for gene manipulation in Drosophila melanogaster. *Nature Reviews Genetics* **6:**340–340.
Vibranovski, M.D., Chalopin, D.S., Lopes, H.F., Long, M. and Karr, T.L. (2010). Direct evidence for postmeiotic transcription during Drosophila melanogaster spermatogenesis. *Genetics* **186:**431–433.
Vilanova, C. and Porcar, M. (2014). iGEM 2.0—refoundations for engineering biology. *Nature Biotechnology* **32:**420–424.
Vilanova, C., Tanner, K., Dorado-Morales, P., Villaescusa, P., Chugani, D., Frías, A., Segredo, E., *et al.* (2015). Standards not that standard. *Journal of Biological Engineering* **9:**17.
Wan, S.-W., Lin, C.-F., Wang, S., Chen, Y.-H., Yeh, T.-M., Liu, H.-S., Anderson, R., *et al.* (2013). Current progress in dengue vaccines. *Journal of Biomedical Science* **20:**37.
Wang, L.-Z., Wu, F., Flores, K., Lai, Y.-C. and Wang, X. (2016). Build to understand: synthetic approaches to biology. *Integrative Biology* **8:**394–408.
Wheeler, R.E. (1962). A simple apparatus for forced copulation of mosquitoes. *Mosquito News* **22:**402–403.
White, T.D., Asfaw, B., Beyene, Y., Haile-selassie, Y., Lovejoy, O., Suwa, G., Woldegabriel, G., *et al.* (2009). Ardipithecus ramidus and the Paleobiology of Early Hominids. *Science* **326:**64–86.
WHO (2004). *The World Health Report 2004: Changing History*.
Woodard, L.E. and Wilson, M.H. (2015). PiggyBac-ing models and new therapeutic strategies. *Trends in Biotechnology* **33:**1–9.
Wu, X., Kriz, A.J. and Sharp, P.A. (2014). Target specificity of the CRISPR-Cas9 system. *Quantitative Biology* **2:**59–70.
Xiao, H. and Lis, J.T. (1989). Heat shock and developmental regulation of the Drosophila melanogaster hsp83 gene. *Molecular and Cellular Biology* **9:**1746–1753.
Xie, K., Minkenberg, B. and Yang, Y. (2015). Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-processing system. *Proceedings of the National Academy of Sciences* **112:**3570–3575.
Zakharkin, S.O., Headley, V. V, Kumar, N.K., Buck, N.A., Wheeler, D.E. and Beneš, H. (2001). Female-specific expression of a hexamerin gene in larvae of an autogenous mosquito. *European Journal of Biochemistry* **268:**5713–5722.
Zetsche, B., Gootenberg, J.S., Abudayyeh, O.O., Slaymaker, I.M., Makarova, K.S., Essletzbichler, P., Volz, S.E., *et al.* (2015). Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* **163:**759–771.
Zetsche, B., Heidenreich, M., Mohanraju, P., Fedorova, I., Kneppers, J., DeGennaro, E.M., Winblad, N., *et al.* (2016). Multiplex gene editing by CRISPR–Cpf1 using a single crRNA array. *Nature Biotechnology* **Advance On:**1–6.
Zhu, J., Palliyil, S., Ran, C. and Kumar, J.P. (2017). Drosophila Pax6 promotes development of the entire eye-antennal disc, thereby ensuring proper adult head formation. *Proceedings of the National Academy of Sciences* **114:**5846–5853.