Genetic approaches to the development of novel designs for population control of two vector mosquitoes: *Aedes aegypti* and *Culex quinquefasciatus*.

Ilona Lidia Flis

Linacre College

University of Oxford

Thesis submitted for the degree of
Master of Science by Research

Hilary Term 2017

Thesis Abstract

The prevention or reduction of infectious pathogen transmission is essential for safeguarding human health and species conservation, and can be achieved through vector control. New genetics-based innovations have proven to be a viable solution for the successful suppression or eradication of insect vectors, whilst also providing opportunities for further improvements and enhancements. In this thesis, I examine a vector control approach incorporating genetically engineered late-acting lethality, and investigate the character and function of the genetic elements that could lead to improvement of this approach.

I start by assessing the functionality of the late-acting, doxycycline-repressible lethal system in transgenic *Aedes aegypti*. I find that the induced lethality is specific to the late developmental stage and occurs in the vast majority of the individuals carrying the lethal transgene. I also show that the induced phenotype is strongly repressible in the presence of the antidote, which is a crucial prerequisite of a practical RIDL (Release of Insect carrying Dominant Lethal gene) system.

Next, I investigate the *Culex quinquefasciatus Actin-4* gene and its potential use to induce a flightless phenotype. I find that the expression of the *Actin-4* is sex-specific and, by generating a novel mutant via gene editing, that the gene is haploinsufficient or dominant negative in inducing the flightless phenotype in females. Additionally, I provide further support for the effectiveness of the recently discovered CRISPR/Cas9 system by showing that it successfully induces targeted editing of the *Actin-4* gene.

My findings provide novel genetic tools for the development of various genetics-based strategies for control of invasive vector mosquitoes.

Acknowledgements

I would like to thank my supervisors, Luke Alphey, Stuart Wigby and Kelly Matzen for their continued help and support over the past two years, and for allowing me to be a part of the exciting on-going research and development into mosquito population control. All three have been patient and encouraging throughout my studies and their assistance has proved invaluable towards the completion of this thesis.

In particular, I would like to thank Kelly for initially introducing me to the practical side of the mosquito world, and also my colleagues Ed Sulston, Pam Grey and Siân Morgan for their advice and technical support whilst I was conducting my experimental work at Oxitec. Additionally, I am very appreciative of the way in which I was made to feel welcome and a part of the Oxitec family.

I am grateful to Luke Alphey, who was always available to provide scientific advice; I appreciated the fruitful discussions with him, which induced critical and creative thinking and helped to further my knowledge and understanding of my chosen field.

My sincere thanks go to Stuart Wigby for his support during my stay at Oxford. He was always considerate of the direction of my work and how it would marry with my future research goals.

I would also like to offer thanks to two researchers working at The Pirbright Institute, who also aided in this study and were always available to answer any

of my questions. Firstly, Graham Freimanis for processing samples for deep sequencing (NGS) and secondly, Paolo Ribeca, for his preliminary analysis of the NGS data, the results of which are included in this thesis.

The Arthropod Genetics Group (AGG), based at The Pirbright Institute also provided research support. The research contribution of Sanjay Basu from AGG led to the characterisation of the 5′UTR region of the *C. quinquefasciatus Actin-4* gene.

Last but not least, my sincere thank you goes to Graham Leech for his assistance in proofreading of this thesis.

3

Table of contents

Chapter 1: General Introduction

The Sterile Insect Technique

The Sterile Insect Technique (SIT) is an environmentally friendly, species-specific approach to population control. SIT based on ionising radiation can make an insect reproductively sterile by causing chromosomal damage in germ cells, consequently leading to the death of fertilized eggs or embryos (Klassen, 2005; Robinson, 2005). Wild insects produce no offspring after successfully mating with their released, sterile counterparts. As a result, there is a decrease in the number of viable progeny in the next generation of the wild insects. Consequently, this will lead to population decline, under the assumption that the release of a sufficient number of sterile insects is maintained over a long enough period of time (usually several generations) (Klassen, 2005).

SIT has shown to be highly effective in the area-wide control or eradication programmes that were targeted against several major pest insects (Dyck et al., 2005), such as the Mediterranean fruit fly (*Ceratitis capitata* Wiedemann) (Hendrichs et al., 1995), the tsetse fly (Glossina *austeni* Newstead) (Msangi et al., 2000) and the New World screw-worm (*Cochliomyia hominivorax* Coquerel) (Wyss, 2000). SIT-related releases specifically targeting mosquitoes, with the aim of suppressing the wild population, have also previously been carried out (Asman et al., 1981; Benedict & Robinson, 2003). Two small field trials based on chemosterilisation of the insects, implemented against the filariasis vector *Culex quinqefasciatus* Say (Patterson et al., 1970)

and the malaria vector *Anopheles albimanus* Wiedemann (Lofgren et al., 1974),

led to successful suppression and eradication of the local population of these

two mosquito species.

Although the large-scale SIT programmes had a suppressing effect on

the target mosquito populations, they also highlighted some problems that

hindered the successful implementation of these operations, such as the

decreased quality and competitiveness of the released sterile males and the

absence of sexing strains (Alphey, 2002; Benedict & Robinson, 2003). The

obstacles faced during implementation of the SIT programmes drew attention to

the need for the development of new control strategies, with several advantages

potentially available through deployment of recombinant DNA techniques

(Alphey & Andreasen, 2002; Benedict & Robinson, 2003).

Importance of population dynamics in the control of mosquitoes

Understanding of the population dynamics for the control of mosquitoes

is essential, and will likely improve the development and deployment of control

methods, for example by indicating that irradiation-induced embryonic lethality

might not always be the optimal phenotype, or even a desirable one, for

effective population control.

A good example of a mosquito species, for which extensive studies on

population dynamics have been conducted, is the dengue and yellow fever

vector *Aedes* (*Stegomyia*) *aegypti* (Linnaeus). As described by Dye's

mathematical model (Dye, 1984b), population dynamics of this species can be

significantly influenced by density-dependent intraspecific competition among larvae for nutritional resources. This indirect competition in the aquatic larval stage is considered to be the key driver of the density-dependent survival rate in *Aedes aegypti* under laboratory conditions (Dye, 1982, 1984a), though this view has recently been challenged (Legros et al., 2009).

To the extent that decreased competition among conspecifics causes an increase in survival rate to adulthood (Dye, 1984b), introduced early death of the offspring from sterile x wild-type crosses of mosquitoes in the wild will lead to a higher survival of the remaining (wild-type) larvae (M. P. Atkinson et al., 2007; Phuc et al., 2007; Yakob et al., 2008). This compensatory effect of density dependence can consequently result in counterproductive efforts of the population control program (Juliano, 2007; Washburn, 1995). Therefore, density-dependent competition has been suggested as one of the important determinants for the successful implementation of genetic control methods in natural populations (McDonald et al., 1977).

To overcome the problem of existing over-compensatory dynamics in a target population, a population control method based on lethality in later developmental stages – preferably in late larval or pupal stage – can be applied (Phuc et al., 2007). Models based on this strategy predict an equivalent level of control that can be achieved through releases of a lower number of mosquitoes carrying a late-acting lethal gene, compared to the number of those carrying an early-acting lethal genetic system (M. P. Atkinson et al., 2007; Phuc et al., 2007; Yakob et al., 2008).

Engineered late-acting lethality in mosquito population control

A variant of SIT known as RIDL (Release of Insect carrying Dominant Lethal gene) is a self-limiting, tetracycline-repressible and environmentally friendly system (Heinrich & Scott, 2000; Thomas et al., 2000). In this system, the self-limiting gene incorporates a suitable promoter (e.g. female-specific, embryo-specific), which drives the expression of an effector molecule; and a minimal promoter fused to multiple tetO (tetracycline Operator) sites that bind tTAV protein (tetracycline repressible Trans-Activating factor Variant) (Fig. 1). A minimal promoter incorporates sequences of nucleotides that are necessary to bind the transcription factors and to accurately initiate the gene transcription (Roeder, 1996). Without tetracycline (Fig. 1A), tTAV protein is produced, which binds to tetO, and drives the expression of an effector molecule. In contrast, the tetracycline-bound form (Fig. 1B) does not bind tetO and so does not activate expression of the effector, therefore disabling the system. The progeny of released RIDL insects will express the effector gene without tetracycline provided in the wild. If the expressed effector gene is lethal, the absence of tetracycline will inevitably lead to death (Gong et al., 2005). The two-component arrangement of the repressible RIDL system enables a tissue-specific and developmental stage-specific expression of the transgene – a particularly desirable feature in the development of novel system designs for a wide species range.

Figure 1. Tetracycline-repressible system. Schematic representation of the two-component RIDL system adapted from Gong et al. (2005). (A) The expression of an effector molecule in the absence of tetracycline and (B) inactivation of the system in the presence of tetracycline. T – tetracycline.
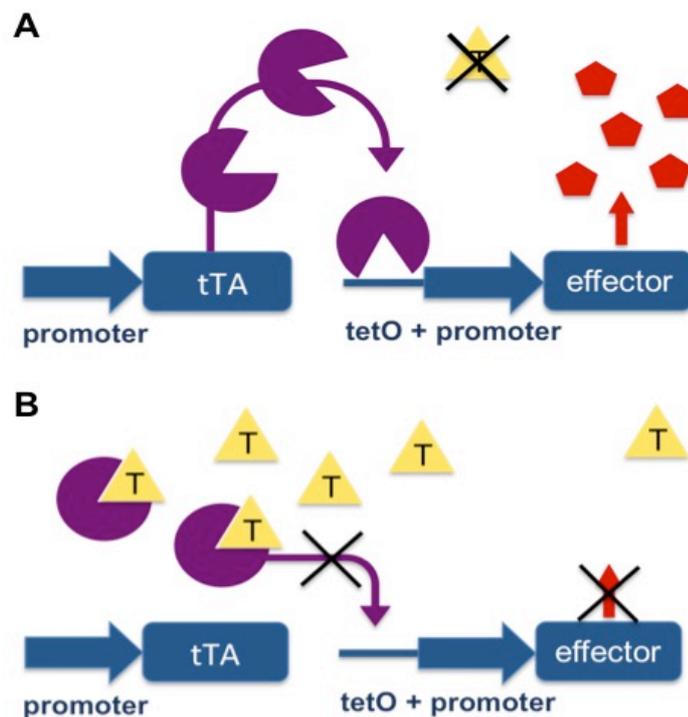
The RIDL system has shown to be a successful approach in driving late-acting lethality in *Aedes aegypti* laboratory strains (Phuc et al., 2007) as well as in the field (Carvalho et al., 2015; Harris et al., 2012; Harris et al., 2011). These open field trials enabled a strong suppression of the target population through controlled releases of the genetically modified OX513A strain. A dominant tetracycline-repressible genetic system of the OX513A line showed near complete penetrance of the lethal gene (Phuc et al., 2007). This is assumed to result from high-level expression and accumulation of the tTAV protein during larval development, as well as a consequence of the so-called "position effect" that is specific to an insert integration event during germ line transformation

(Gong et al., 2005; Phuc et al., 2007).

Importantly, a construct design that enables transcriptional control of the time of the lethal gene expression at a relatively late developmental stage is also feasible and can be achieved through the use of a stage-specific promoter. Such promoters have been previously characterised and evaluated in *Aedes aegypti*, and have exhibited a similar transgene expression pattern to that which was determined for the respective endogenous gene (Coates et al., 1999; Moreira et al., 2000).

A proof-of-principle for a construct enabling transcriptional control through the use of a late-acting, stage-specific promoter has been already demonstrated in two mosquito species: *Aedes aegypti* and *Aedes albopictus* (Skuse) (G. Fu et al., 2010; Labbé et al., 2012) and resulted in development of a new approach to population suppression. In these studies, *Actin4* promoter enabled the engineering of a conditional flightless phenotype in females through exploiting sex-specific alternative splicing of the native *Actin-4* gene, which is predominantly expressed in the female indirect flight muscles (IFMs) at the pupal stage. Flightless females are likely to be unable to attract a male (Hartberg, 1971) in the field; they are also unable to leave the breeding site in search of a blood meal, or to escape from predators. These artificially induced behaviours appear to be the apparent advantages of this female-specific RIDL system (fsRIDL), leading to female sterility and early adult lethality. Although this system has not yet been studied in the field, it has been shown that the periodic release of males from the *A. aegypti* OX3604C strain – carrying a transgene that imposes a tetracycline-repressible female-flightless phenotype – led to the eventual eradication of the target population in a laboratory cage system (Wise de Valdez et al., 2011) though, for reasons that are not clear, did

not show such an effect in a field cage trail (Facchinelli et al., 2013).

fsRIDL can also provide further – and considerably more effective – control relative to other self-limiting genetic systems, such as SIT or bi-sex lethal RIDL – a variant of RIDL, in which the lethal phenotype is expressed in both sexes of offspring (Alphey, 2014). A key advantage of the fsRIDL approach is the possibility of providing genetic sexing of the release cohort of mosquitoes homozygous for a transgene. Presence of the heterozygous males for the female-flightless transgene, as the result of the wild-type x fsRIDL matings, can additionally strengthen self-limiting properties of this system by inducing a flightless phenotype of the female offspring in the following generations (Alphey, 2014). Furthermore, the fsRIDL is an example of the late-acting lethal system that allows for competition among larvae for nutritional resources and, as a result, can lead to a reduction of the compensatory effect of density dependence in mosquito population control.

Thesis overview

In this thesis, I study the character and function of the candidate genes or genetic elements that can be utilised in future development of new technologies that aim to genetically control populations of vector mosquitoes. My findings contribute to the assessment of functionality of the novel late-acting lethality system design for population control of the *Aedes aegypti* mosquito. Furthermore, I contribute to understanding of the transcriptional properties of the *Culex quinquefasciatus Actin-4* gene sequence, and show – for the first time

15

in mosquitoes – the functional character of this gene based on the relationship between its alleles.

In Chapter 2, I examine the functionality of the *Osiris 2* promoter in transgenic *A. aegypti* lines in driving a late-acting, doxycycline-repressible lethal system *in vivo.* I evaluate the performance of this promoter in the heterozygous and homozygous context by analysing survival data for four transgenic lines. I find that the transgene incorporating the *Osiris 2* promoter has the ability to induce the desirable lethal phenotype manifesting itself predominately at the pupal stage. I also show that the lethality is strongly repressible when the antidote (doxycycline) is applied. I report, however, a strongly reduced survival rate for the lines homozygous for the transgene. This finding demonstrates the need for further evaluation, which is required for any future downstream field use of transgenic mosquitoes of this type.

In Chapter 3, I characterise the 5′UTR region of the *C. quinquefasciatus Actin-4* gene (*Cxq-A4*). I also perform targeted mutagenesis for the coding region of the *Cxq-A4* and investigate the effects of the introduced mutations on the target sequence, as well as on the adult female and male phenotype. Furthermore, I evaluate the editing efficiency of the four sequence-specific components (sgRNAs) of the CRISPR/Cas9 genome editing system, which was used here. I find that the 5′UTR of the *Cxq-A4* exhibit sex-specific splicing and that the gene is either haploinsufficient or a specific mutant is dominant negative in inducing the female flightless phenotype. I also show the all four sgRNAs used during my study are active both *in vitro* and *in vivo*; and their *in vitro* editing efficiency strongly corresponds to that predicted *in silico*. This finding demonstrates the potential for utilising the genetic properties of the *Cxq-A4* in developing population control strategies for this species of mosquito.

Chapter 2: '*Osiris*' promoter: a potential genetic tool for development of a novel dominant, late-acting, bi-sex RIDL system design.

Abstract

The *Aedes aegypti* mosquito is a highly invasive and potent vector of life-threatening infectious diseases such as yellow fever and dengue. To reduce the impact of these diseases, effective population control strategies, which aim to suppress this mosquito species, are required. One such strategy is the RIDL (Release of Insect carrying Dominant Lethal gene) system based on late-acting lethality that can be engineered with the use of specific regulatory elements.

In this study, I investigate the performance of a putative late-acting promoter (*Aedes aegypti Osiris 2*) in four *A. aegypti* RIDL lines. I analyse the stage-specific survival data collected for the lines of mosquitoes that are heterozygous or homozygous for the transgene by constructing generalised linear models (GLM) and by applying repeated *G*–test of goodness-of-fit.

I find that the presence of the transgene in the heterozygous lines induces stage-specific lethality without having a major adverse effect on mosquito survival up until the desired lethal stage. Additionally, I show that the induced lethality is strongly repressible and the presence of the transgene does not affect survival of the insects reared with doxycycline. This stage-specific, doxycycline-repressible lethality exhibits near full penetrance (97.5% - 99.3%) in all of the heterozygous lines. However, the decrease in survival reported for the homozygous transgenic lines reared on doxycycline, relative to wild type or

heterozygotes, highlights the need for further assessment of this genetic system as far as the establishment of a high-quality release-ready product line is concerned. Overall, the results generated in this study could be used to contribute to the development of a new generation of transgenic lines for the future of mosquito population control.

Introduction

Importance of the vector control: Aedes aegypti

Aedes aegypti is responsible for the transmission of several globally important arboviruses (Gratz, 1999), including yellow fever, dengue (DENV) (Simmons et al., 2012), chikungunya (CHIKV) (Vega-Rúa et al., 2014) and Zika virus (ZIKAV) (Diagne et al., 2015). Although not limited to the tropics, this species of mosquito is particularly potent in the urban environments found in those regions (Rogers et al., 2006). The rapid global spread of this species has been recently reported (Kraemer et al., 2015), including its introduction to Europe (Schaffner & Mathis, 2014). Given the impact on public health of the aforementioned diseases and the highly invasive nature of *A. aegypti*, urgent planning and development of effective mosquito population control strategies is strongly required. This is of particular importance because the conventional methods of vector control have proven either ineffective or unsustainable (WHO Technical Report Series, 2009). In addition, there is no commercial vaccine against chikungunya and the efficacy of the existing Sanofi-Pasteur dengue

vaccine (Dengavaxia) – available only in six countries – depends on the recipient's prior vaccination immunity (Ferguson et al., 2016).

The modest number of previous studies that embraced the challenge of *A. aegypti* population control through the implementation of SIT have met with limited success (Benedict & Robinson, 2003). Due to fundamental problems with current methods, such as the reduced competitiveness and life span of irradiated males and lack of an effective mechanism for separation of sexes before the insect release; methods based on the use of transgenics insects for the control of mosquito-borne diseases have been proposed (Alphey, 2002). The effectiveness of the control RIDL method, based on the release of the transgenic strain of *A. aegypti* mosquitos exhibiting late-acting lethality, has previously been shown in open field studies (Carvalho et al., 2015; Harris et al., 2012; Harris et al., 2011). These outcomes indicated the desired improvement in the efficacy and utility of this sterile insect technique.

Osiris promoters: novel regulatory elements in driving late-acting lethality

Regardless of the apparent advantages of the female-specific flightless RIDL system (discussed in in the first chapter), there is a trade-off in the approach to developing a new genetic system suitable for future field releases. The trade-off is particularly evident between a genetic system that can lead to more desirable outcomes in the field and a system that enables both cost-effective and time-effective maintenance under laboratory conditions. Candidate promoters that could enable development of a genetic system with a preference

to the latter approach have already been identified and partially characterized. Exhibiting the required late-acting expression profile, these novel promoters derived from *A. aegypti* putative *Osiris* genes have confirmed the ability to induce tetracycline-repressible stage-specific lethality in an engineered bi-sex dominant lethal system of *A. aegypti* and *A. albopictus* (Conway, 2014)

One of the tested *Osiris* promoters induced near full penetrance (98-100%) of the effector gene – pro-apoptotic michelob_x (mx) with clear pupa-specific expression (Conway, 2014). This regulatory element *Osiris 2* (hereafter, '*Os2*') derives from *A. aegypti* AAEL004298 gene (Altschul et al., 1997). Endogenous orthologs of the *Os2* promoter exhibit a similar expression profile in *A. aegypti*, *A. albopictus* and *Drosophila melanogaster*, with peaks in 2-3-day-old pupae, suggesting its conserved function across dipteran species. These results show promise for the future application of the *Osiris* gene-based system in developing a more effective and versatile RIDL approach for other insect species targeted in population control.

First described in *Drosophila melanogaster*, the *Osiris* gene cluster is a well-conserved and remarkably syntenic gene family of 24 insect-specific orthologous gene groups (Shah et al., 2012). Although the function of the *Osiris* genes remains unknown, this multigene family (Dorer et al., 2003) shows strong expression in a variety of insect tissues, including epidermis, hindgut, foregut, and trachea (Shah et al., 2012) that can serve as targets of a lethal effector gene in the development of a new genetic system for insect population control.

Application of *piggyBac*-based germ line transformation (Handler, 2002) led to the generation of independent genomic insert integrations corresponding to separate lines of transgenic *A. aegypti* mosquitoes established by Michael Conway (Oxitec Ltd.). Functionality of the *Os2* promoter in four of those lines:

OX5055D, OX5055F, OX5055G and OX5055I, has not been previously tested and is the scope of the study presented in this chapter.

Objective of the study

To test the functionality of *Osiris 2* promoter in transgenic *A. aegypti* OX5055 lines in driving late-acting, doxycycline-repressible lethal system *in vivo.*

Materials and Methods

1. Strain background and rearing conditions

An *A. aegypti* wild-type strain with Latin American genetic background was used in germ-line transformation with the OX5055 construct. The insectary was kept at 27°C (±1°C) and 70% (±10%) relative humidity with a 12-hour light/dark cycle. Larvae were fed on crushed dry TetraMin fish food (Tetra GmbH, Germany) and adults on 10% glucose supplemented with 14 µg/ml penicillin and 14 U/ml streptomycin. Females were fed on defibrinated horse blood (TCS Biosciences Ltd., UK) using a *Hemotek* Insect Feeding System (Discovery Workshops, Accrington, UK) set at 37°C. The same rearing conditions were applied to mosquito cohorts used in the experimental assays described in this chapter.

2. Genetic construct

The construct OX5055 used for germ-line transformation of *A. aegypti* was designed and built by Sarah Scaife, Oxitec Ltd (Fig.1). It consists of the AAEL004298 (*Os2*) promoter coupled to a tTAV/tetO-effector system (Conway, 2014). In this system, tTAV (TetR-VP16) protein binds to tetracycline operator (tetO) sites in the absence of tetracycline, driving expression of an effector molecule (Gong et al., 2005; Gossen et al., 1995). In the OX5055 construct, that effector molecule was the pro-apoptotic gene *michelob_x* (*mx*), under the control of a minimal promoter (*hsp^{min}*) derived from the *D. melanogaster hsp70* gene (Conway, 2014) (Fig. 2).

Figure 2. Schematic representation of the OX5055 construct. Diagram of the OX5055 construct (AAEL004298 *Os2*-tTAF3-Dmhsp*70*-3'UTR-tetO-*hsp^{min}*-*michelob_x*) adapted from M. Conway (2014). From left to right: *Osiris* promoter (*Os2*), Ubiquitin-tTAF3-Dmhsp*70* 3'UTR gene fusion, tetO-*hspmin-michelob_x* cassette. A fluorescent transformation marker (AmCyan), under *Hr5/IE1* promoter from the baculovirus *Autographa californica* MNPV (Rodems & Friesen, 1993), as well as the ends of *piggyBac* DNA transposon, are not shown.
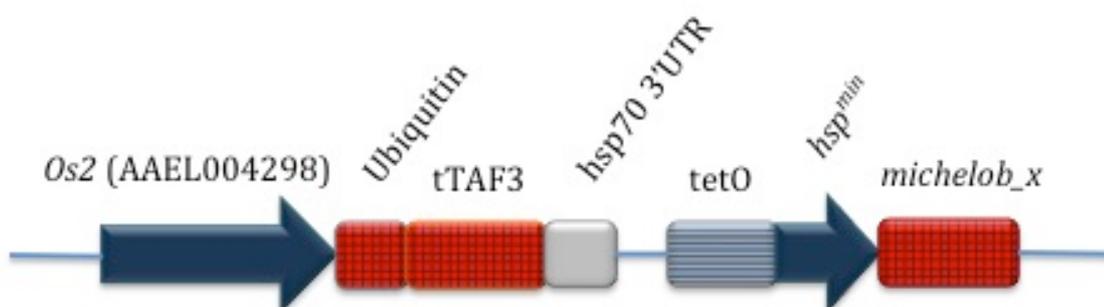
Expression of *mx* gene leads to cell death (Zhou et al., 2005), making it a desirable candidate for a self-limiting gene in the engineered dominant lethal system. tTAF3 used for this construct contains a truncated VP16 activation domain from the *Herpes simplex* virus (HSV) and has previously shown to have lower toxicity in transgenic insects compared to its variant, tTAV protein (Conway, 2014). Use of a 3'UTR sequence derived from the *D. melanogaster hsp70* gene was expected to lead to destabilisation of the tTAF3 transcript by reducing mRNA accumulation and, in consequence, the level of the tTAF3 protein in transgenic insects (Bönisch et al., 2007; Conway, 2014). The ubiquitin fusion technique requires the fusion of ubiquitin to the N-terminus of the protein of choice (Bachmair et al., 1986). It enables the engineering of the *in vivo* half-life of the expressed protein of interest - here the tTAF3 protein.

3. Lethality Assay

The on/off doxycycline survival assay consisted of two experiments carried out for two OX5055 lines at a time. After transformation, the established OX5055 transgenic lines were maintained for 4 generations. Eggs for the 5th generation (F5) of mosquitoes for each line were obtained from a cross of heterozygous for the transgene x wild-type individuals and vacuum-hatched in water for approximately 2.5 hours to ensure synchronous hatching. Hatched first instar larvae (L1) were divided into 3 cohorts of 200 individuals per treatment (off/on doxycycline) per line, placed in 16 oz. pots (Fabri-Kal, USA) with 1 µg/mL of dox solution or deionized water and reared at a density of 1 larva/ml. The larvae were fed the following standard regime of finely ground

TetraMin fish food per larva: day 1 – 0.06 mg, day 2 – no food, day 3 – 0.08 mg, day 4 – 0.16 mg, day 5, 6, 7 and 8 – 0.32 mg, day 9 – 0.16mg, day 10 onwards – as required. On the day of pupation live pupae were removed from pots, screened for fluorescence and their numbers and sex recorded. All fluorescent and wild-type pupae were separated by sex and placed in one insect cage (17.5cm x 17.5cm x 17.5cm, Bugdorm, MegaView Science Co., Ltd., Taiwan) corresponding to the phenotype recorded, and left to eclose. Pupae collection procedure for each line and treatment continued until the remaining 3% or less of the starting cohort of larvae was present. The remaining larvae were counted and discarded. Adults were provided *ad libitum* with 10% sucrose solution supplemented with antibiotics (see the rearing conditions of this section). Cages with adults were assessed four days after the last pupa was added. Assessment included counting the total number of dead pupae, the non-viable adults (dead adults on the water, dead adults on the floor of the cage, flightless adults) and functional (flying) adults.

4. Homozygous Viability Assay

The homozygous viability assay consisted of two experiments carried out for two OX5055 lines at a time. Eggs from the cross of heterozygous (for the transgene) (F4) x wild-type individuals were hatched, reared to adulthood and adults allowed to intercross in cages corresponding to each line. A single cage contained approximately 100 females and 100 males. Eggs collected from these crosses were vacuum-hatched in water for approximately 2.5 hours to ensure synchronous hatching and three pots of 200 L1 larvae (F6) with

doxycycline present were set up for each line. Larval rearing density, doxycycline concentration, feeding regimen, caging and pupae collection procedure, as well as data collection for viable and non-viable individuals, followed the same protocol as for the lethality assay described above.

Mosquitoes from all transgenic lines were reared in the presence of doxycycline (1 µg/mL, 'dox') in both experiments. This analogue of chlortetracycline was shown to be effective at lower concentrations and less toxic to transgenic individuals than tetracycline - routinely used in the rearing process of Oxitec strains (Curtis et al., 2015).

5. Data analysis

Data on stage-specific survival of mosquitoes obtained from both experiments was analysed by constructing generalised linear models (GLM) with Poisson or quasi-Poisson errors for count data (Lethality Assay) and binomial or quasibinomial errors for proportion data (Homozygous Viability Assay), followed by a stepwise removal of factors. The model factors included doxycycline treatment, phenotype, sex and factorial interactions. This approach to data analysis enables the identification of the best-fitting model and to determine the estimates of the model parameters. Differences between saturated and simplified models were analysed by ANOVAs. Additionally, a repeated *G*–test of goodness-of-fit was applied to the survival data obtained from Homozygous Viability Assay. All statistical analyses were performed using RStudio software, version 0.98.1102 (RStudio Team, 2015).

Results and Discussion

1. Lethality Assay

To assay for the presence of a repressible phenotype, I reared L1 larvae from each OX5055 line (D, F, G, I) in presence or absence of doxycycline (as described in Methods). L1 cohorts comprised individuals carrying 0 or 1 copy of (wild-type or heterozygous for) the OX5055 transgene. These were the offspring of heterozygous transgenics x wild-type mosquitoes. Therefore, the expected phenotypic ratio (fluorescent : non-fluorescent) for the screened pupae was 1:1, under the assumption that the transgenic and wild-type mosquitoes were equally viable.

Consistent with the intended function of the construct (and in agreement with previous results obtained by Michael Conway (2014) for established OX5055 lines) the lethal phenotype manifested itself mainly at the pupal stage, with the majority of transgenic cohort reaching pupation regardless of the presence or absence of doxycycline (Fig. 3A).

The expected late-acting, pupal-specific lethal phenotype strongly manifested itself in lines D, G and I. Line F exhibited lethality earlier than expected, with a number of transgenic larvae failing to pupate in the absence of doxycycline. This effect was largely rescued with doxycycline present during larval growth and maturation (Fig. 3A). The ability to repress with doxycycline indicates that the lethality acts via expression of tTAV, even though the Osiris promoter is expected to drive expression of tTAV primarily in pupae, rather than in larvae.

26

In order to compare survival to pupation of transgenic and wild-type individuals, I performed statistical analysis for counts of pupation rates by constructing GLM models for each transgenic line. A significant bias from 1:1 sex ratio in transgenic lines can be an indicator of sex-specific lethality due to differential expression of the transgene in both sexes or sex-linked effects caused by a linkage of the insert to the sex-determining locus (M-locus), therefore the 'sex' response variable was included in further analysis. Counts of the number of pupae of each sex for OX5055 lines are presented in Figure 4.

Figure 3. Pupation and eclosion rate in OX5055 lines. The figure presents barplots of average pupation (A) and eclosion (B) rate for four OX5055 lines (D, F, G and I) reared in the presence ('ON DOX') or absence ('OFF DOX') of doxycycline. All the rates for transgenic lines are normalized by pupation rates for their wild-type counterparts (1:1 pupation rates are indicated here with a dashed line). Error bars show standard error of the mean (SEM) of three replicates, each replicate comprising a cohort of heterozygous for the transgene and wild-type individuals (200 in total). Estimation of eclosion rate includes all individuals (viable and non-viable) that fully emerged from the pupal case as a fraction of those that pupated.
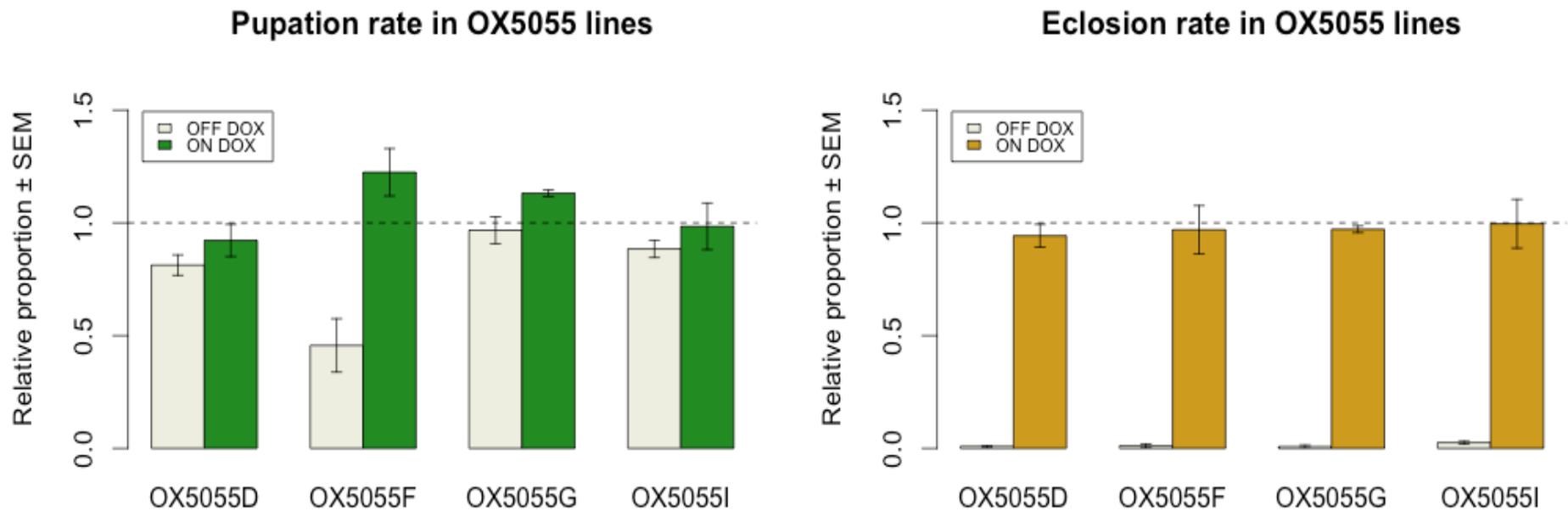
Figure 4. Effect of doxycycline on pupation in both sexes of OX5055 lines. The figure presents the effect of doxycycline ('DOX') on number of female and male larvae reaching pupation in OX5055 lines (D, F, G and I). Each data point represents a sum of pupae of a given sex and phenotype from a single line replicate reared in the presence ('ON DOX') or absence ('OFF DOX) of doxycycline. The horizontal bar represents the mean number of pupae from 3 replicates. A single replicate comprises a cohort of heterozygous for the transgene and wild-type females and males L1 (200 individuals in total).
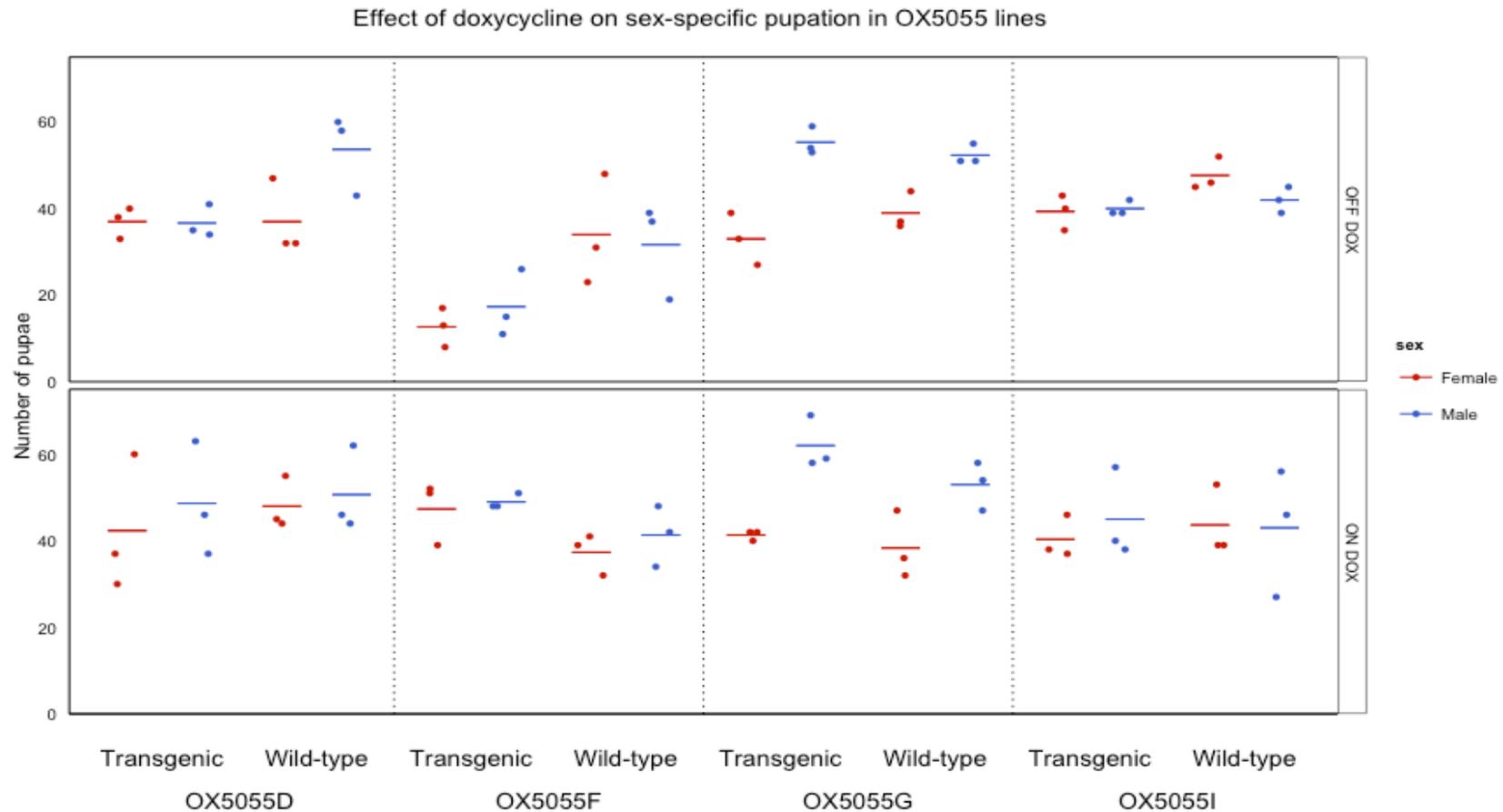
I found that pupation rates for line D and I did not significantly deviate from the null models with quasi-Poisson errors, indicating that the presence of the transgene, doxycycline treatment, or sex, did not significantly influence pupation rates in this assay: df = 1; F = 2.38; p = 0.137 and df = 1; F = 0.008; p = 0.93; for line D and I, respectively. The presence of the transgene and doxycycline treatment in line G were found not to be significant factors for pupation rates. However, both transgenic and wild-type males in line G, had higher survival to pupation regardless of the doxycycline treatment (one factor model with Poisson errors: df = 1; χ^{2} = 40.65, p = 1.82x10^{-10}). This effect is not associated with zygotic presence of the transgene; other potential explanations include environmental differences relative to wild type controls, or the presence of a sex-ratio-distorting element in the line (Hickey & Craig, 1966).

Finally, the interaction term between transgene presence and doxycycline treatment was found to be a significant factor in the models for line F, indicating better survival to pupation of transgenic individuals than that of their wild-type counterparts in the presence of doxycycline, but lower for transgenic individuals in the absence of doxycycline (two factor model quasi-Poisson error, df = 1, F = 22.96; p = 0.0001). This significant interaction might have been the effect of an insert position in the genome and a presumed tTAV-dependent toxicity of the transgene. Sex did not have a significant effect on pupation in this line. **Overall, this data shows that the presence of the transgene does not have an effect on survival to pupation in most of the OX5055 lines.**

Lethality that was exhibited later than expected (when reared 'off dox') was observed in line I, with deaths occurring at the stage of emergence from

the pupal carapace or soon after completion of metamorphosis (Photo 1A, Appx. 1.S1, p. 181).

Photo 1. Post-larval stage mortality in OX5055 lines. Examples of death categories scored for all or some of the OX5055 lines during the assay. (A) Dead pupae and 'partially eclosed' individuals from line OX5055I were recorded when reared without doxycycline. (B) Dead adults on water's surface after being reared in the presence of doxycycline – line OX5055F. Images were taken on the 4th day post the last pupal collection and caging.
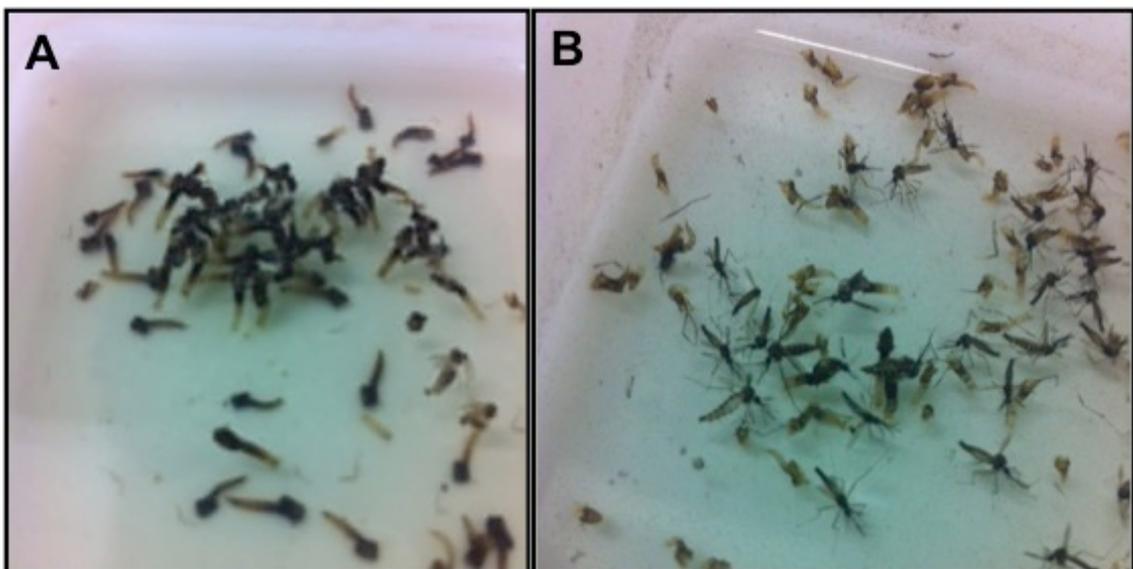
Mortality recorded for 'partially eclosed' individuals in line I comprised about half of all individuals surviving to pupation and was possibly due to the inability of those mosquitoes to emerge from the pupal case. Since *Os2* promoter drives high levels of expression in developing adult appendages, expression of *michelob_x* in the mosquito legs may explain the inability of transgenic individuals to emerge from a pupal carapace (Conway, 2014). Expression of

the *Osiris 2* gene has been also reported in the tracheae of insects (Shah et al., 2012). In *A. aegypti*, some notable changes in re-modelling of the tracheal system take place at the pupal stage, including extension of new tracheal branches in newly forming appendages of the adult (Christophers, 1960). Expression of a pro-apoptotic gene in those structures might therefore lead to a failure in development of the functional respiratory system that is crucial for efficient metabolic energy production.

The highest mortality rate in the other 3 lines: D, F and G, was recorded for the late pupal stage, with almost all of the transgenic individuals unable to survive to eclosion when reared 'off dox' (Appx. 1.S1, p.181). The measure of penetrance in this study was the percentage of individuals that did not succeed in emerging from the pupal carapace in the absence of doxycycline. Based on these assumptions, the observed late-acting lethality was near to full penetrance in all of the lines: D – 99.1%, F – 98.9%, G – 99.3%, I – 97.5% (Tab. 1). Although up to 2.5% of transgenic mosquitoes from each line managed to eclose (Tab. 1), many of them died on the water's surface or on the floor of the cage. Functional adults were recorded only for line F (a single flying adult) and line G (2 flying adults) (Appx. 1.S1, p.181).

In the presence of doxycycline, lethality was highly repressed, resulting in proportions of eclosing individuals and viable adults similar to those observed for the wild-type mosquitoes (Fig. 3B, Appx. 1.S2, p.181 and Tab. 1). **Therefore, the data shows that the presence of the transgene confers the expected repressible lethal phenotype at the pupal stage.** This is a promising outcome in the process for the selection of a successful line that can be used in a future field release study.

Table 1. Doxycycline repressible lethality in OX5055 lines. The table presents the sums of first instar larvae (L1), pupae and emerged adults for three replicates from each transgenic OX5055 line reared in the presence or absence of doxycycline. A single replicate comprised a cohort of heterozygous for the transgene and wild-type L1 (200 in total). The percentage of adults that managed to successfully eclose from the pupal case (comprising dead adults on the water, dead adults on the floor of the cage and flying adults) is presented in the last 'Adults (%)' column.

OX5055 Line	Genotype	DOX	L1	Pupae	Eclosed adults	Adults (%)
D	OX5055D/+	No		221	2	0.9
	Wild-type	No	600	272	268	98.5
	OX5055D/+	Yes		273	254	93.0
	Wild-type	Yes	600	296	292	98.6
F	OX5055F/+	No		90	1	1.1
	Wild-type	No	600	197	197	100
	OX5055F/+	Yes		289	279	96.5

	Wild-type	Yes	600	236	235	99.6
G	OX5055G/+	No		265	2	0.7
	Wild-type	No	600	274	269	98.2
	OX5055G/+	Yes		310	298	96.1
	Wild-type	Yes	600	274	271	98.9
I	OX5055I/+	No		238	6	2.5
	Wild-type	No	600	269	263	97.8
	OX5055I/+	Yes		256	253	98.8
	Wild-type	Yes	600	260	258	99.2

Further analysis of the obtained results for eclosion rate in the presence

of doxycycline was performed by constructing GLM models with binomial or

quasibinomial errors for each transgenic line and its eclosion rate. The eclosion

rate in line D for 'on dox' transgenic individuals compared to their wild-type

counterparts was significantly deviate from the null model with quasibinomial

errors: $df = 1$; $F = 11.09$; $p = 0.029$, indicating a lower survival of transgenic

mosquitoes to adulthood in that line. In contrast, transgenic individuals in line F

had a significantly higher eclosion rate than their wild-type counterparts: null

model with binomial errors, $df = 1$; $\chi^2 = 6.96$; $p = 8.34 \times 10^{-3}$. Regardless of a

better survival rate to pupation (Fig. 3A) and eclosion with doxycycline present

(Fig. 3B), on average about 40% of transgenic adults from line F died on the

surface of the water after eclosion (Photo 1B, Appx.1.S2, p. 181). This

phenomenon resulted in similar average proportion rates of viable flying

transgenic mosquitoes and their wild-type counterparts. Eclosion rates for

doxycycline treated transgenic mosquitoes from line G and I were not

significantly different than those for wild-type mosquitoes: null models with

quasibinomial errors, $df = 1$; $F = 2.44$; $p = 0.193$ and $df = 1$; $F = 0.21$; $p = 0.67$;

for line G and I, respectively. **This data shows that the presence of the**

transgene does not have an effect on survival to eclosion in most of the

doxycycline-rescued OX5055 lines.

Overall, results obtained from the performed analysis showed that

survival to pupation and eclosion for most of the assayed lines, heterozygous

for the transgene and reared in the presence of doxycycline, did not

significantly differ from that observed for their internal wild-type control.

Moreover, the engineered lethal phenotype exhibited near full penetrance in

the absence of doxycycline in all of the OX5055 lines (97.5% - 99.3% considering eclosion, higher in respect of functional adults), delivering a promising approach in developing a new generation of transgenic lines for mosquito population control. Nevertheless, some variation was observed in the stage at which lethality manifested itself during mosquito development, presumably linked to the effect of the insert position in the genome. A variation in the stage at which lethality occurred did not show correlation with the level of penetrance of the lethal gene (Fig. 3, Appx. 1.S1, p.181; Tab.1).

2. *Homozygous Viability Assay*

Since any downstream field use of transgenic mosquitoes of this type would require homozygous lines, each of the OX5055 lines was tested for viability of individuals homozygous for the transgene. To test the viability of the homozygous individuals, fluorescent (heterozygous) males and females from F5 outcross were mated for each OX5055 line. The experimental cohorts (F6) were therefore expected to comprise individuals homozygous for the transgene, heterozygotes, and wild-type individuals in the genotypic ratio of 1:2:1, respectively. The expected phenotypic (fluorescent : non-fluorescent) ratio was 3:1 if transgenic and wild-type mosquitoes are equally viable; 2:1 if the insertion is associated with a recessive lethal phenotype. In some cases, homozygotes show stronger fluorescence than heterozygotes (Photo 2). This cannot always be reliably scored due to subtle differences in the fluorescent protein expression between pupae representing different developmental age at the time of their collection and screening.

Photo 2. Phenotype-based genotype scoring in homozygous OX5055 lines. Examples of homozygous (OX5055I/OX5055I) for the transgene, heterozygous (OX5055I/+) and wild-type (+/+) individuals (L-R) observed for line OX5055I when reared with doxycycline ('on dox') during the homozygous viability assay.
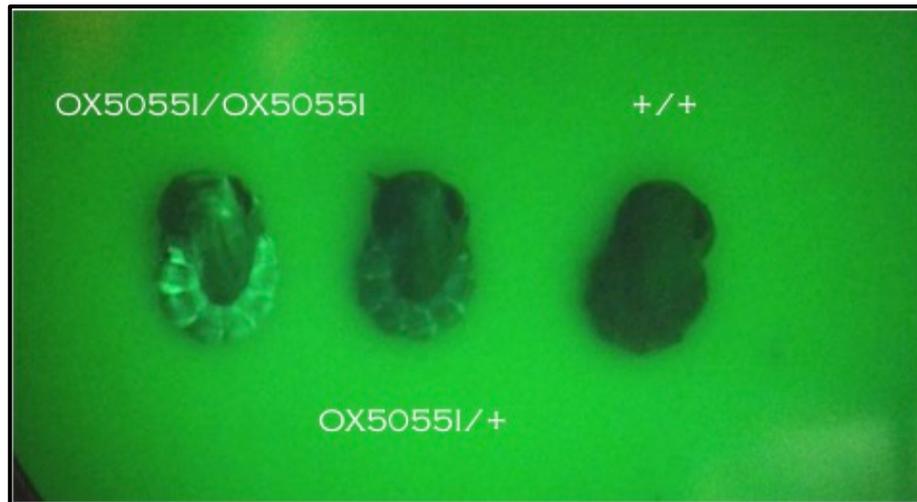
Counts of pupation for all four lines showed considerable variation with a high level of mortality (Fig. 5A). For at least one of 3 replicates within each line, transgenic : non-transgenic ratio below 2:1 was observed. Given that heterozygotes showed reasonable survival relative to wild type (above), this suggests a low survival of homozygotes. Such an effect may be due to an insertional mutagenic effect, whereby the transgene disrupts an essential gene; such effects are typically recessive. It is also possible that the higher transgene dose in homozygotes (two copies) is deleterious, though such effects may be expected to be partially dominant. Under these assumptions, the ratios above 2:1 for line D (2.6:1), F (2.17:1), G (3.6:1 and 3.5:1) and I (2.19:1, 2.34:1) indicated that at least some of the homozygous individuals survived and

successfully reached pupation. The phenotypic ratio above 3:1 in line G might

be an indicator of a double transgene insertion in this line. Furthermore, two

replicates from this line showed better survival to pupation of transgenic males

compared to their transgenic female counterparts (1.40:1 and 1.53:1 ratio),

indicating possible sex-linked insertion of the transgene, which corresponds to

results obtained from the lethality assay for this line.

Analysis of the pupation rates was performed by applying a repeated *G*–

test of goodness-of-fit for 3 replicates of the experimental cohorts of

mosquitoes from each OX5055 line, which were reared in the presence of

doxycycline. This likelihood ratio test was applied in order to estimate an

overall deviation from the expected 3:1 proportions (pooled *G*), as well as

variation among the replicates (*G*-test of independence). Individual *G*–test of

goodness-of-fit for a single replicate was performed on the data for each line,

followed by a test for significant variation between replicates (*G*-test of

independence). Further analysis, including calculation of "pooled" and "total *G*-

value", was only performed for replicates with a non-significant "heterogeneity

G-value". No significant difference between replicates was found for line D, F

and I (heterogeneity *G*-value: 0.240; 0.551 and 0.312, respectively). Estimated

total *G*-values for those lines: D - $5.1x10^{-4}$; F - $9.9x10^{-4}$ and I - $5.0x10^{-4}$,

indicated a significant difference from the expected 3:1 ratio. Moreover,

significant pooled *G*-values indicated a loss in the number of transgenic

individuals in those lines (pooled *G*: line D - $2.2x10^{-4}$, line F - $1.0x10^{-4}$, and line I

- $8.7x10^{-5}$), which is the cause of reported significant difference from 3:1 ratio.

Since heterogeneity *G* in line G showed significant difference (*p* = 0.037) in

proportions between 3 replicates, an estimation of the size of deviation with

confidence was not feasible, therefore pooled and total G for this line were not calculated. Observed variation in survival between cohorts from different replicates within a line could have been influenced by random factors, such as mosquito handling during the experimental procedures. **Overall, this data suggests that for most of the tested lines homozygosity for OX5055 transgene decrease survival to pupation.** Summary of the results based on the repeated *G*–test of goodness-of-fit analysis are presented in Table 2.

Counts of eclosion rates showed that transgenic individuals from all four OX5055 lines experienced further lethality with a severe example of line I, in which a possible complete loss of homozygous individuals, as well as loss of some of the heterozygous mosquitoes, was observed (phenotypic ratio below 2:1 in all 3 replicates).

Presumed survival to eclosion for homozygous individuals was observed for 3 other lines, in which replicates with phenotypic ratio above or equal to 2:1 ratio were reported: line D – 2.3:1, line F – 2:1 and 2.2:1, and line G – 2.2:1 and 2.8:1 (Fig. 5B).

Figure 5. Pupation and eclosion rate in the intercrossed OX5055 lines. The figure presents the barplots of average pupation (A) and eclosion (B) rate in the progeny obtained from intercrosses of heterozygous for the transgene individuals for OX5055 lines (D, F, G and I) reared in the presence ('ON DOX') of doxycycline. The presented rates for transgenic progeny are normalised by pupation rates for their wild-type counterparts (relative pupation rates are indicated here with dashed lines, where the 2:1 and the 3:1 transgenic : wild-type ratio implies the homozygous lethal and homozygous viable line, respectively). Error bars show standard error of the mean (SEM) of three replicates, each replicate comprising a cohort of homozygous and heterozygous for the transgene, as well as wild-type individuals (200 in total). Estimation of eclosion rate includes all individuals (viable and non-viable) that fully emerged from the pupal case.
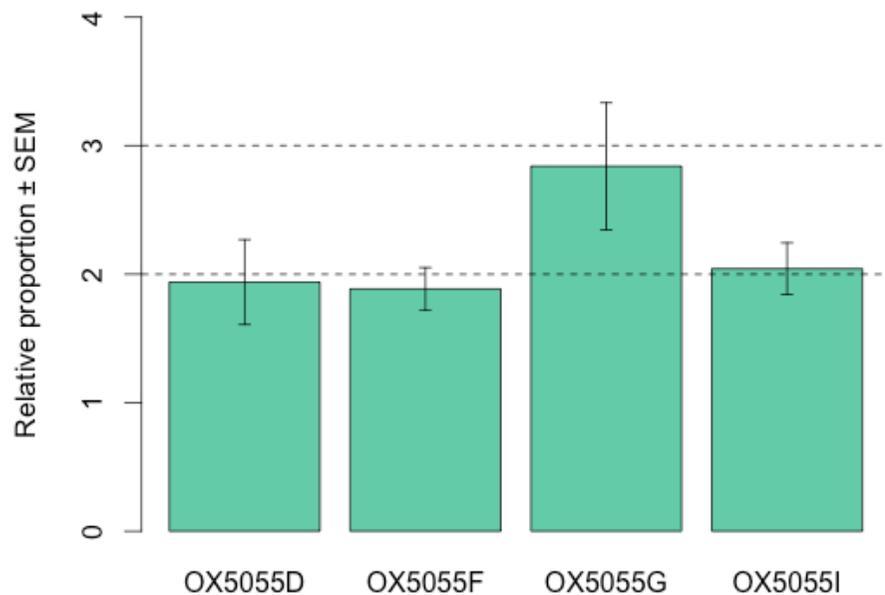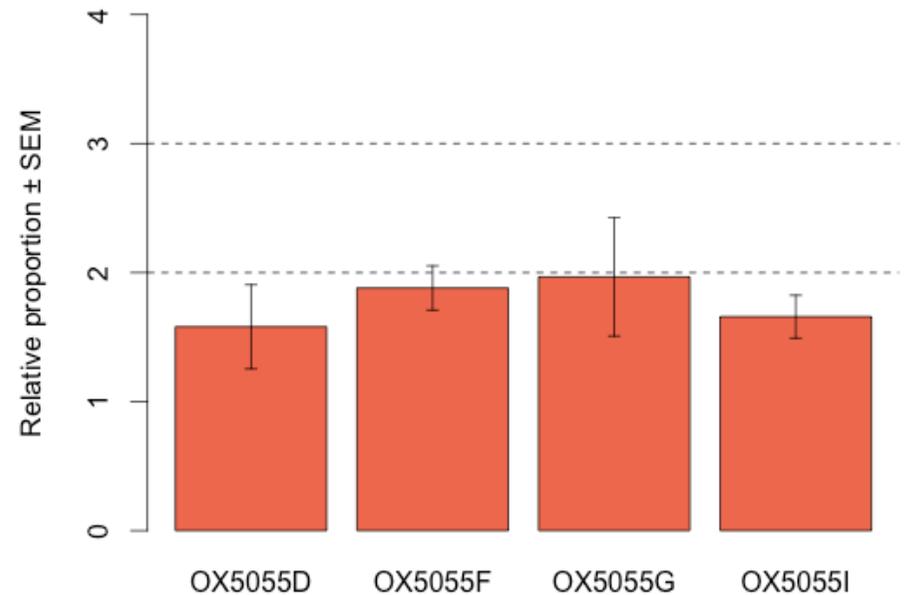
Table 2. Results of the *G*–test of goodness-of-fit analysis for pupation rate in the intercrossed OX5055 lines. The summary of the results for the repeated *G*–test of goodness-of-fit analysis for pupation rates in transgenic and wild-type mosquitoes from the intercrosses of heterozygous for the transgene individuals (from F5 outcross of individual OX5055 lines). G-values for each replicate within a line were obtained by performing a *G*–test of goodness-of-fit. *G*-test of independence allowed for the estimation of *G*-values for heterogeneity. Pooled *G*-values were calculated for data summed across each replicate within a line. An asterisk symbol (*) corresponds to a significant *p*-value for each replicate after applying Bonferroni correction. Each replicate comprised a cohort of 200 individuals (F6) obtained from the intercross (F5) for each OX5055 line.

Line replicate	Transgenic	Wild-type		*G*-value	*df*	*p*-value
5055D_1	74	37		3.86	1	0.049
5055D_2	71	48		13.44	1	2.4x10-4*
5055D_3	76	29		0.37	1	0.54
			total G	17.67	3	5.1x10-4*
pooled	221	114	pooled G	13.6	1	2.2x10-4*
			heterogeneity G	4.07	2	0.13

5055F_1	70	36		4.24	1	0.039
5055F_2	74	34		2.3	1	0.13
5055F_3	69	43		9.75	1	1.8×10^{-3}*
			total G	16.29	3	9.9×10^{-4}*
pooled	213	113	pooled G	15.1	1	1.0×10^{-4}*
			heterogeneity G	1.19	2	0.551
5055G_1	103	52		5.66	1	0.017
5055G_2	128	36		0.84	1	0.36
5055G_3	124	37		0.36	1	0.55
			total G	NA	NA	NA
pooled	355	125	pooled G	NA	NA	NA
			heterogeneity G	6.58	2	0.037*
5055I_1	112	51		3.27	1	0.07
5055I_2	117	50		2.09	1	0.148
5055I_3	104	62		12.36	1	4.4×10^{-4}*
			total G	17.71	3	5.0×10^{-4}*
pooled	221	114	pooled G	15.39	1	8.7×10^{-5}*
			heterogeneity G	2.32	2	0.312

Further analysis of the obtained results for eclosion rate in the presence of doxycycline was performed by construction of GLM models. Phenotype was the only model factor tested in GLM analysis. Since sample sizes were small, 'sex' as a factor was not included in the model construction. Lower eclosion rate for 'on dox' transgenic individuals compared to their wild-type counterparts was significantly deviate from the null model with binomial errors in line D and I: $df = 1$; $\chi^2 = 27.28$; $p = 1.76 \times 10^{-7}$ and $df = 1$; $\chi^2 = 35.04$; $p = 3.23 \times 10^{-9}$, respectively as well as in line G (null model with quasibinomial errors: $df = 1$; $\chi^2 = 35.17$; $p = 4.05 \times 10^{-3}$). In contrast, eclosion rates for line F indicated that the presence of the transgene did not significantly influence eclosion rates in this assay: null model with quasibinomial errors, $df = 1$; $F = 0.007$; $p = 0.937$. **This data suggests that for most of the tested lines homozygosity for OX5055 transgene decreased survival to eclosion.**

Regardless of doxycycline presence during larval rearing, lethality for transgenic individuals manifested itself early in development (lower than expected survival to pupation for line D, F and I), at the late pupal stage (lower than expected eclosion in line D, G and I) or at the time of emergence resulting in dead pupae and 'partially eclosed' individuals observed for all the assayed lines (Appx. 1.S3, p.181). Such an outcome might indicate an influence of insertion of the transgene, resulting in a recessive or weak dominant lethal condition caused by disruption of an essential gene. Additionally, genomic integration might have led to incorporation of the transgene near genomic enhancer elements causing changes in its expression. Incorporation of the transgene within enhancer region is, however, unlikely to cause the observed non-repressible lethality in four independent insertion lines. Unintended design

of a "leaky" construct may also be a result of cis-acting effects (Wittkopp &
Kalay, 2012) present in the engineered construct itself – here the influence of
the *Os2* promoter on *michelob_x* expression level. Furthermore, toxic over-
expression of tTAF3 may be directly related to the strength of *Os2* promoter.
The incorporation of insulators (Gaszner & Felsenfeld, 2006) into the original
construct as well as the identification and characterisation of a weaker *Osiris*
promoter that could replace *Os2* could be an approach leading to reduction in
toxicity of the transgene.

Non-repressible toxicity of the transgene could be also the result of
using an insufficient concentration of doxycycline in order to repress the lethal
effect caused by the tTAF3-dependent expression of *michelob_x*, given the
presumed higher expression level of tTAF3 in homozygotes than in
heterozygotes. Further study of doxycycline dose response could lead to
successful suppression of the lethal effect under the assumption that this effect
is the result of increased toxicity from expressing two copies of the transgene.
The lack of suppression of the lethal gene with a higher dose of doxycycline
can, on the other hand, indicate the insertion effect of the transgene. Another
approach that would enable distinction between the expression level and
insertion effects would involve production of double heterozygotes carrying a
single copy of the transgene from each insertion line that shows homozygous
lethality and the subsequent evaluation of the transgene effect for the
generated double heterozygous lines. Also, the selection of individuals that are
more resistant to the toxic effects associated with the transgene in the
presence of doxycycline by rearing them through additional generations can
lead to establishing homozygous-viable Osiris lines. Overall, further testing

might help with providing a final answer to whether the OX5055 lines could be made homozygous.

Conclusions

Results presented in this chapter show that a promoter fragment from the putative *Osiris* AAEL004298 gene, coupled to a tTAV/tetO-effector system, has the ability of inducing lethality that is both doxycycline-repressible and late-acting in *A. aegypti* mosquitoes. Further assessment of the genetic system used in this study is important for development of a high-quality product line. The established line should exhibit acceptable mating competitiveness between genetically modified and wild-type insects in order to be considered as suitable for mass rearing. Despite these concerns, the results presented in this chapter indicate potential opportunities for improvement in mosquito control development.

Chapter 3: Targeted mutagenesis in *Culex quinquefasciatus Actin-4* gene using CRISPR/Cas9 system.

Abstract

Culex quinquefasciatus is the primary vector of avian malaria and avian pox. Its unintentional introduction to the Hawaiian Islands, combined with the subsequent introduction of the pathogens that it can vector, was one of the main reasons for the extinction of several bird species native to this part of the world. Bird extinction is an on-going problem and requires an urgent solution, particularly in the face of disease-driven impact from climate change. One of the proposed solutions is the genetic control of the vector population based on field release of genetically engineered mosquitoes carrying a heritable genetic element that causes flightlessness in the wild females.

In this study, I characterise the 5′UTR region of the *C. quinquefasciatus Actin-4* gene (*Cxq-A4*) by applying the RACE (Rapid Amplification of the cDNA Ends) and the sequencing technique. I also use the CRISPR/Cas9 system to induce targeted mutagenesis within *Cxq-A4* both *in vitro* and *in vivo*. I investigate the effects of the introduced mutation by performing behavioural tests, HRMA and sequencing analysis. Additionally, I examine the mutagenesis efficiency of the employed CRISPR/Cas9 system through the *in vitro* digestion and deep sequencing.

I find that the 5′UTR of the *Cxq-A4* is sex-specifically spliced. I identify a loss-of-function mutation driving the flightless phenotype in females and propose that this mutation is either antimorphic (dominant negative) or amorphic (indicative of haploinsufficiency of the *Cxq-A4* gene). Additionally, I show that CRISPR/Cas9 system is active *in vitro* and *in vivo*. I conclude that identified and characterised properties of the *Cxq-A4* show strong potential for their future use in the development of genetic vector control strategies.

Importance of the vector control: Culex quinquefasciatus

The southern house mosquito, *Culex quinquefasciatus* (Say), is a geographically widely dispersed insect and a competent vector of pathogens affecting wildlife and humans (Farajollahi et al., 2011). This species of mosquito is a principal vector of the avian malaria parasite (*Plasmodium relictum*), the avian pox virus (*Avipoxvirus spp.*) (van Riper et al., 1986) and the human parasitic filarial worm (*Wuchereria bancrofti*) (Davis, 1935; Triteeraprapab et al., 2000). In addition, the existing vector competence studies suggest that *C. quinquefasciatus* can potentially play a role in transmitting arboviruses, such as the West Nile Virus (WNV) (Eastwood et al., 2011; Richards et al., 2014; Sardelis et al., 2001) and possibly the Zika virus (ZIKV) (Amraoui et al., 2016; Guo et al., 2016).

Disease-driven extinction of Hawaiian avifauna

Hawaiian ecosystems have been profoundly transformed after Polynesian colonisation in the 13th century, followed by more recent Western contact with the islands. Ecological changes triggered by humans led to the extinction of 74 of 109 known endemic Hawaiian bird species (Reed et al., 2012). Of the remaining 37 species, 33 are federally listed as Endangered Species, although 9 of these may already be extinct (Banko & Banko, 2009; Reed et al., 2012).

One of the major reasons behind the on-going extinction crisis affecting native Hawaiian birds was the accidental introduction of the *C. quinquefasciatus* mosquito to the islands in the early 19th century. The outcome of this was disastrous, as it resulted in the initial transmission of the avian malaria parasite and the avian pox virus between non-indigenous Hawaiian birds (Warner, 1968). Avian populations of the Hawaiian Islands have evolved for millions of years in almost complete isolation (Loope, 1998) and did not have an evolutionary history with the pathogens that had been introduced. This made these immunologically naïve birds highly susceptible to invasion by the newly introduced diseases (Atkinson et al., 2000; Atkinson et al., 1995). After almost a century since the diseases were introduced to the Hawaiian Islands, avian malaria and avian pox have become one of the major culprits responsible for the extensive decline in the abundance, diversity and geographical distribution of Hawaiian native avifauna (Pratt, 2009; Reed et al., 2012; Samuel et al., 2015).

Disease has driven the distribution of Hawaiian forest birds, limiting their presence to higher elevations (above 1,500m above see level) where the cooler temperature hinders development of the vector mosquito and the malaria parasite (van Riper, 1986; Warner, 1968). Because temperature is one of the limiting factors affecting the presence of the vector, its increase - brought about by climate change - will consequently result in the spread of the disease vector mosquitoes to higher elevations, which presently provide refuge for the endangered bird populations. Taking this into account, an adequate long-term conservation strategy for Hawaiian bird species is strongly required (C. T. Atkinson & LaPointe, 2009; Fortini et al., 2015).

Among the several proposed strategies for conservation of the endangered avifauna of Hawaii, are predator removal, feral pig control, or development of malaria resistance or tolerance in the birds themselves. However, the most promising approach to a successful area-wide conservation programme is the genetic control of the invasive vector based on the release of genetically modified mosquitoes (Liao et al., 2017).

Role of actin in inducing the flightless phenotype

Actin, the major cytoskeletal protein, is extremely abundant in all types of eukaryotic cells. It polymerises to form actin filaments which are involved in muscle contraction (Cooper, 2000), such the contraction of the insects' flight muscles.

The understanding of the flight muscle formation and function in adult insects comes from the study of the thorax muscles (Lawrence, 1982; Miller,

1950). The thorax of the fruit fly (*Drosophila melanogaster* Meigen) contains a number of flight muscles, including the indirect flight muscles (IFMs), which are crucial for the production of mechanical power during flight (Miller, 1950). The IFM is the only fibrillar-type of muscle in *Drosophila* and contains proteins that are found exclusively in this muscle (Mogami et al., 1982). The presence of these unique proteins suggested the existence of genes that are expressed only in the IFMs and eventually led to discovery of one such example – the *D. melanogaster act88F*. This gene encodes Actin III, which is the most abundant and major IFM isoform (Hiromi & Hotta, 1985).

Studies based on mutagenesis of the 88F actin (*act88F)* enabled identification and characterisation of numerous dominant negative mutant alleles that are expressed only in the IFM. The induced mutations impaired flight ability without causing any adverse effect on the fly's viability and were further used in an extensive analysis, giving an insight into the function and the structure of actin (An & Mogami, 1996; Hiromi & Hotta, 1985; Mogami & Hotta, 1981). There have been a number of new mutations characterised for *D. melanogaster* that are involved in IFM development (Babu & Ramachandra, 2007; Nongthomba & Ramachandra, 1999) and cause flight disability, flight muscle defects and abnormality in the wing position. Additionally, studies on the *act88F* actin isoforms revealed that their primary sequences confer unique properties that affect the flight ability and the IFM structure (Fyrberg et al., 1998).

In all of the aforementioned studies that led to identification and characterisation of the IFM, a method of chemically induced mutation was applied. Interestingly, the flightless phenotype in *A. aegypti* – reported for the

first time in mosquitoes – was induced using the RIDL technique based on the

piggyBac germ line transformation method (Handler, 2002); subsequently

followed by the successful induction of the same phenotype in its sister

species, *A. albopictus* (Labbé et al., 2012). In both studies, the transgene

employed in generating the flightless phenotype incorporated the *Actin-4*

promoter. This promoter, in combination with sex-specific alternative splicing,

controlled the tissue- and sex-specific expression of tTAV mRNA that was

sufficient for expression of the lethal effector gene (VP16 transactivator) in the

female IFM.

The idea of using the regulatory component of the mosquito Actin-4

(AeAct-4) in developing conditional lethal strains was proposed by Muñoz et al.

(2004), who first identified and characterised the female-specific expression of

this IFM actin gene in *A. aegypti*. Interestingly, the reported predominant

expression of the *AeAct-4* in female pupae corresponds with the findings from

the study on *D. melanogaster*, which demonstrated that the process of adult

IFM formation in the fruit fly starts at initiation of the metamorphosis at the

pupal stage (Fernandes et al., 1991). Furthermore, the gene organisation, high

amino acid identity and expression profiles of *AeAct-4* are similar to its putative

orthologue – *D. melanogaster Act88F* (Muñoz et al., 2004). These findings also

support the hypothesis that the IFM actins have a highly conserved nature.

Application of the CRISPR/Cas9 system in vector mosquitoes research

CRISPR/Cas9 system

Since its recent discovery and development, the CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) and CRISPR-associated (Cas) system (Jinek et al., 2012) has become a remarkably powerful tool that enables the targeting of virtually any genomic sequence. This system has been adapted to modify the genomes of a range of species, including yeast (DiCarlo et al., 2015), zebrafish (Hwang et al., 2013) and monkeys (Niu et al.).

The type II CRISPR/Cas9 system was originally isolated from the bacteria adaptive RNA-guided immune system. It provides its host with the means to defend against invading genetic elements by creating targeted double-strand breaks (DSBs) in the invading DNA (Barrangou et al., 2007; Ishino et al., 1987). In order to be guided to the target site, the *Streptococcus pyogenes* Cas9 endonuclease requires the CRISPR RNAs (crRNAs), with the sequence complementary to the target; and the trans-activating CRISPR RNA (tracrRNA), which recruits crRNA into the Cas9 complex (Deltcheva et al., 2011; Jinek et al., 2012). This two component system has been recently modified into a single synthetic guide RNA (sgRNAs) and successfully used to efficiently create targeted mutations in *D. melanogaster*, enabling the target gene knock-outs (Bassett et al., 2014). In order to produce site-specific DSBs in this context, the Cas9 endonuclease requires an engineered sgRNA, which contains a region of complementarity with a target genomic DNA sequence of

interest (Jinek et al., 2012). The target genomic sequence lies next to a protospacer adjacent motif (PAM) which matches the genetic sequence NGG (where N is any of the four nucleotides) (Jinek et al., 2012). Upon recognition of a PAM sequence – and sufficient complementarity of the sgRNA and the target DNA strand – the Cas9 induces cleavage (3-4 nucleotides upstream of the PAM sequence) of the target DNA strand (Jinek et al., 2012). The generated DSBs can lead to mutations (SNPs, indels, inversions) as a result of imprecise repair by nonhomologous end joining (NHEJ) (Gu & Lieber, 2008). It has been shown that the size of the DNA target sequence (17-20 nucleotides) does not seem to affect the cleavage efficiency, however, the use of truncated sgRNA (17 nucleotide long) substantially decreases off-target cleavage (Y. Fu et al., 2014). Additionally, the design of sgRNAs, based on selection of the most specific seed sequence (7 to 12 nucleotides upstream of the PAM sequence) may be crucial in improving on-target efficiency, especially as mismatches at the PAM distal sequence are more likely to be tolerated (Cong et al., 2013; Jinek et al., 2012).

Targeted mutagenesis

The importance of vector control is the major reason why mosquitoes became one of the first non-model organisms targeted in mutagenesis studies, in which the CRISPR/Cas9 system was employed. Several studies presenting successful genome editing in these insects have already been published and are briefly summarised here. The first report of successful targeted mutagenesis in vector mosquitoes comes from Dong et al. (2015), who

targeted different regions of the *ECFP* (enhanced cyan fluorescent protein) gene in the transgenic *A. aegypti* strain. The outcome of this study was the generation of 2 to 27-nucleotide long indels within the *ECFP* target sequence, as well as the knock-out of *ECFP,* with the reported efficiency of 5.5%. Interestingly, the successful application of the CRISPR-Cas9 editing system in the same species enabled a knock-out of the male determining *Nix* gene and resulted in feminised genetic males (Hall et al., 2015). The *Nix* gene was further confirmed in the same study as being sufficient to initiate male development. The stable and precise loss-of-function mutations were generated for several targeted *A. aegypti* genes, as shown by Kistler et al. (2015), followed by successful integration of an exogenous sequence into the mosquito genome via homology-directed repair, with the reported donor sequence incorporation rate of 0.71%. The editing potential evaluated for 6 different *A. aegypti* genes varied across sgRNAs and a target site as reported by Basu et al. (2015). In this study, the editing rate for identified highly effective RNAs was shown as being between 24% and 90%. A CRISPR/Cas9-mediated genome editing study was also published for another vector mosquito – *C. quinquefasciatus*. In this study, Itokawa et al. (2016) targeted the *CYP9M10* gene that is linked to the pyrethroid resistance phenotype observed in the studied strain of *C. quinquefasciatus*. Disruption of all the copies of this gene led to a significant reduction in mosquito resistance to this insecticide, indicating the key role of *CYP9M10* in conferring resistance. The number of studies based on targeted mutagenesis mediated by CRISPR/Cas9 is not overwhelming at present. However, taking into account the growing interest from the scientific community in this highly adaptable system, more studies

utilising the CRISPR/Cas9-based approach for the purpose of mosquito genome editing research will certainly be conducted in the future.

CRISPR/Cas9-based gene drives

The prospect of vector population control through the release of genetically modified insects led to the development of technologies, such as SIT and RIDL, that have proven to be successful in the suppression or eradication of many pest insect species in the wild (Hendrichs et al., 1995; Msangi et al., 2000; Wyss, 2000), including mosquitoes (Carvalho et al., 2015; Harris et al., 2012; Harris et al., 2011). The use of the aforementioned technologies was, however, not free from its drawbacks. Therefore, both the scientific community and potential investors now seek to develop new technologies that will be cost-effective and relatively quick and easy to establish prior to their successful implementation in the field.

One potential solution has emerged with the discovery of the RNA-guided Cas9 nuclease (Jinek et al., 2012), which is based upon the principles of the gene drive strategy. Such a strategy depends on the properties of the 'selfish' genetic elements, e.g. homing endonuclease genes (HEGs), which can spread through the population in a super-Mendelian inheritance pattern by exploiting host molecular machinery in order to copy themselves into a particular target sequence in the genome (Chevalier & Stoddard, 2001). HEGs can be engineered to alter a target population by transforming it genetically – the property that can be used in control or eradication of a vector population (Burt, 2003). However, retargeting of HEGs to a specific DNA sequence is

more challenging than retargeting the RNA-guided Cas9, which makes the Cas9 nuclease an attractive component of a gene drive-based strategy (Guha et al., 2017).

The CRISPR/Cas9-based gene drive strategy has already been shown to be functional in *Drosophila* (Gantz & Bier, 2015) and two malaria vector mosquitoes - *Anopheles gambiae* Giles (Hammond et al., 2016) and *Anopheles stephensi* Liston (Gantz et al., 2015), and has led to reports of a remarkable homing efficiency (copying rate of a gene drive from heterozygous to homozygous state) of over 90%. The researchers, who based their study on the *A. gambiae* mosquito, managed to create a gene drive incurring a fitness-loss by targeting female fertility genes. The outcome of the study in *A. stephensi* was the production of a gene drive system that enables engineering of malaria resistance. Despite these promising outcomes, which could potentially lead to the development of a successful strategy that can result in suppression or replacement of a vector population with mosquitoes carrying resistance, some problems in the use of these highly efficient RNA-guided homing drives have also become apparent. These disadvantages were the reported reduction in the fertility of females heterozygous for the homing construct (Hammond et al., 2016) and the gene drive instability in the progeny of the drive carriers (Gantz et al., 2015). The example properties of the RNA-guided homing drives, such as the instability and the inability of its control upon release, are the potential source of the context-dependent risks related to the use of this system in the wild. The biggest risk of using such gene drives is the introduction of irreversible changes to natural ecosystems (Alphey, 2016; Champer et al., 2016; Esvelt et al., 2014). Undoubtedly, overcoming the

technical constraints that present potential risks in the use of the gene drives for wild populations control needs to be addressed before their implementation in the field can be considered viable.

Objective of the study

To study effects of targeted mutagenesis in *Culex quinquefasciatus Actin-4* gene by using the CRISPR/Cas9 system.

Materials and Methods

The generation of mutations in the targeted *Culex quinquefasciatus Actin-4* gene through the use of the CRISPR/Cas9 system and detection of the mutations was predominantly based on methodology presented in previously published research studies (Bassett et al., 2014; Kistler et al., 2015). The CRISPR/Cas9 experimental strategy described in this chapter comprised several steps, summarised in Fig. 6.

Figure 6. Schematic workflow for targeted mutagenesis in the *Culex quinquefasciatus Actin-4* gene. The figure presents the consecutive steps of the procedure, starting with the *in silico* experimental design, followed by the experiments conducted *in vitro* and *in vivo*, with the final process of data collection and *in silico* data analysis.
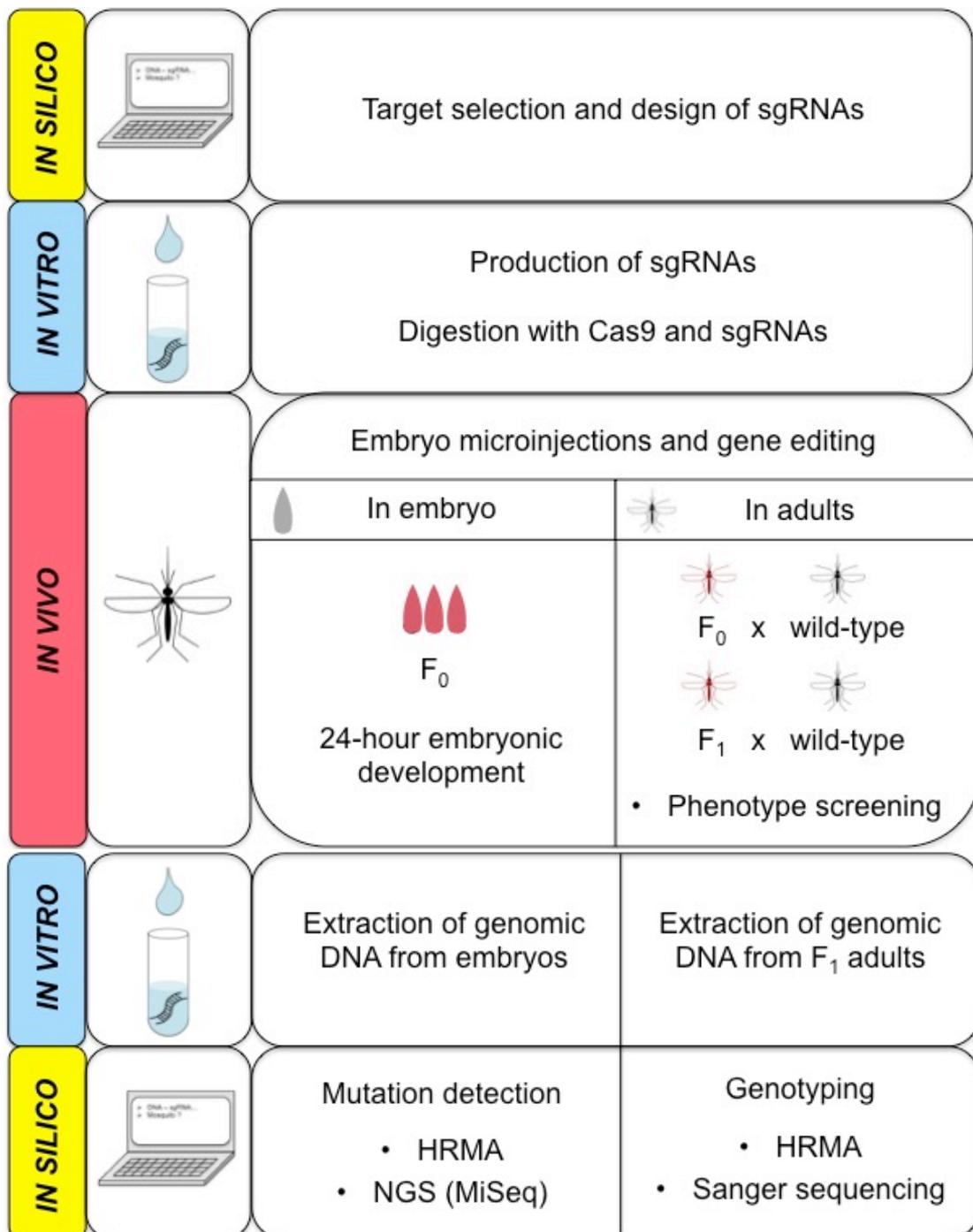
1. Strain background and rearing conditions

A wild-type strain of *Culex quinquefasciatus* mosquito (Pel) was used for microinjections with CRISPR/Cas9 components. The Pel strain was originally established from a large (>1000) sample of larvae collected in 1984 in Peliyagoda, Sri Lanka (Karunaratne et al., 1995). This mosquito strain had been maintained in our laboratory condition on a four-week generation cycle at a population size of 2,000-3,000 adult mosquitoes for over 15 generations before conducting the experimental work described in this chapter. On the day of pupation, mosquito pupae were randomly and evenly distributed between two insect rearing cages (32.5cm x 32.5cm x 32.5cm, Bugdorm, MegaView Science Co., Ltd., Taiwan) and allowed to emerge, sexually mature and mate. Collection of the egg rafts was performed over two gonotrophic cycles for both rearing cages and collected eggs were left to hatch *en masse*. In total, between 400-500 egg rafts were collected from each generation. On the day of hatching, first instar larvae (L1) were transferred into the rearing trays containing distilled water at the density of 500-600 individuals per 1 litre of water. Larvae were fed on crushed dry TetraMin fish food (Tetra GmbH, Germany) and adults on 10% glucose. Females were fed on defibrinated horse blood (TCS Biosciences Ltd., UK) using a *Hemotek* Insect Feeding System (Discovery Workshops, Accrington, UK) set at 37°C. The insectary was kept at 27°C (±1°C) and 70% (±10%) relative humidity with a 12-hour light/dark cycle. The same rearing conditions were applied to mosquito cohorts used in the experimental assays described in this chapter, unless stated otherwise.

2. 5′UTR characterisation of the Culex quinquefasciatus Actin-4 gene

Identification of a putative *C. quinquefasciatus* homologue of the *Aedes aegypti Actin-4* gene (Morales-Hojas & Fife, unpublished) provided the reference sequence in this study of targeted mutagenesis with the CRISPR/Cas9 system, therefore enabling identification and characterisation of *Actin-4* 5′ untranslated region (5′UTR) in this mosquito species. Experimental designs for the studies mentioned above were based on the genome sequence data deposited in VectorBase (Giraldo-Calderón et al., 2015) for the South African Johannesburg strain of *C. quinquefasciatus* (genome assembly CpipJ2), as well as the *Actin-4* gene sequence generated in this project from the Sri Lankan strain (Pel) of the same mosquito species.

In order to determine the transcription start site of the *C. quinquefasciatus Actin-4* gene, I studied the 5′ end of the Actin-4 RNA transcript, as described in the following subsections of this chapter (p. 61-65). My study focused on characterisation of the 5′UTR of the *C. quinquefasciatus Actin-4* gene, which consisted of several selected steps (see below) of molecular analysis based on Rapid Amplification of cDNA Ends (RACE) procedure for amplification of nucleic acid sequences from a messenger RNA (mRNA) template between a defined internal site and unknown sequences at the 5′-end of the mRNA (5′ RACE). Performance of the 5′ RACE procedure was based on the Instruction Manual (Life Technologies, 2004) attached to the 5′ RACE System for Rapid Amplification of cDNA Ends, version 2.0 kit (18374-058, Life Technologies). In this procedure (Life Technologies, 2004),

the first strand of cDNA is synthesised using a gene-specific antisense

oligonucleotide (GSP1) and SuperScript II reverse transcriptase. After first

strand cDNA synthesis, the original mRNA template is removed with the RNase

Mix (RNase H and RNase T1) and the first strand product is purified from

unincorporated deoxynucleotides (dNTPs), GSP1 primer and proteins. A

homopolymeric tail is then added to the 3′ ends of the cDNA using TdT

(Terminal deoxynucleotidyl Transferase) and deoxycytidine triphosphate

(dCTP). Tailed cDNA is subsequently amplified in PCR reactions using gene-

specific primers: GSP2, which anneals to GSP1 primer, and nested GSP3

primer. The combination of the complementary homopolymer-containing the

Abridged Anchor Primer (AAP) and corresponding adapter primer (Universal

Amplification Primer, UAP) permits amplification from the homopolymeric tail in

PCR. These PCR reactions allow amplification of unknown sequences between

the GSP2 primer and the 5′-end of the mRNA.

2.4. Isolation of RNA

By following the manufacturer's protocol (Appx. 2, p. 185), I extracted a

total RNA from 4 samples, containing 6 female or 6 male early pupae (< 15

hours old) per sample, using the TRIzol Reagent (15596018, Life Sciences)

method of RNA isolation - an approach which improves upon the existing

guanidine thiocyanate/acid-phenol method (Chomczynski & Sacchi, 1987).

TRIzol Reagent is a complete, ready-to-use monophasic solution of phenol and

guanidine thiocyanate used for the isolation of high-quality total RNA. This

reagent works by maintaining RNA integrity during tissue homogenisation,

allowing for the isolation of RNA, DNA and proteins in separate fractions upon the addition of chloroform. RNA is precipitated from the aqueous layer with isopropanol, washed to remove impurities and then resuspended for use in downstream applications.

Purity, as well as concentration of the isolated RNA, was checked using the NanoDrop 2000 spectrophotometer and the samples were stored at -80 °C.

2.5. PCR of dC-tailed cDNA

Following the 5′ RACE procedure, dC-tailed cDNA was directly amplified in a Polymerase Chain Reaction (PCR) using the Abridged Anchor Primer (a forward primer supplied with the kit):

5′- GGCCACGCGTCGACTAGTACGGGIIGGGIIGGGIIG-3′ (where 'I' is deoxyinosine), and the gene-specific (nested) reverse primer LA33 (CPIJ12572-GSP2): 5′-CCTTCGTAGATCGGGACGGTGTGG-3′.

Platinum *Taq* DNA polymerase High Fidelity (11304029, Life Technologies) and the product of the dC-tailing reaction were used in this PCR, with the following reaction conditions:

PCR reaction component	Volume for one 50-µl reaction
Sterilised, distilled H_2O	31.5 µl
10X High Fidelity PCR Buffer	5 µl
25 mM $MgCl_2$	3 µl
10mM dNTP mix	1 µl
Abridged Anchor Primer (10 µM)	2 µl
Nested GSP2 primer: LA33 (10 µM)	2 µl
dC-tailed cDNA	5 µl
Platinum *Taq* DNA polymerase (5 U/µl)	0.5 µl
TOTAL	50 µl

The PCR program with the Platinum *Taq* DNA polymerase for this reaction was as follows:

Step	Temperature	Time	Number of cycles
Initial denaturation	94 °C	2 minutes	1
Denaturation	94 °C	30 seconds	
Annealing	55 °C / 62 °C	30 seconds	35
Extension	68 °C	1 minute	
Final extension	68 °C	7 minutes	1
Hold	5 °C		

5 µl of the PCR reaction was run on a 1% agarose gel with a ready-made ethidium bromide solution (0.6 mg/ml) (30-30-06-DB, Severn Biotech Ltd.) to verify the size and specificity of amplification. Agarose gel electrophoresis was performed for 1 hour at 100V and the product was visualised on agarose gel with a UV transilluminator (Molecular Imager Gel Doc XR System, Bio-Rad).

2.6. Nested amplification

This step of the 5′ RACE procedure permits amplification of unknown sequences between the homopolymeric tail at the 5′-end of the mRNA and the gene specific GSP3 primer. This step of re-amplification of dilution of the original PCR product (generated with the Abridged Anchor Primer and the gene specific GSP2 primer) enables the generation of a specific product in a quantity that can be detected by gel staining.

1 µl of the 100-fold diluted product from the PCR of dC-tailed cDNA was subjected to a second PCR amplification using a pair of the forward UAP primer (Universal Amplification Primer, supplied with the kit): 5′-CUACUACUACUAGGCCACGCGTCGACTAGTAC-3′ and the reverse nested primer LA34 (CPIJ12572-GSP3): 5′-GGACAGCCTGGATGGCGACGTACACA-3′.

Platinum *Taq* DNA polymerase High Fidelity (11304029, Life Technologies) and the primary PCR product from the previous step were used in this PCR, with the following reaction conditions:

PCR reaction component	Volume for one 50-µl reaction
Sterilised, distilled H$_2$O	33.5 µl
10X High Fidelity PCR Buffer	5 µl
25 mM MgCl$_2$	3 µl
10mM dNTP mix	1 µl
UAP primer (10 µM)	1 µl
Nested GSP3 primer: LA34 (10 µM)	1 µl
100-fold diluted primary PCR product	5 µl
Platinum *Taq* DNA polymerase (5 U/µl)	0.5 µl
TOTAL	50 µl

The PCR program for this reaction was as follows:

Step	Temperature	Time	Number of cycles
Initial denaturation	94 °C	2 minutes	1
Denaturation	94 °C	30 seconds	
Annealing	55 °C / 65 °C	30 seconds	35
Extension	68 °C	1 minute	
Final extension	68 °C	7 minutes	1
Hold	5 °C		

50 µl of the PCR reaction was run on a 1% agarose gel with a ready-made ethidium bromide solution (0.6 mg/ml) (30-30-06-DB, Severn Biotech Ltd.) to verify the size and specificity of amplification. Agarose gel electrophoresis was performed for 1 hour at 100V.

The PCR product was visualised on agarose gel with a UV transilluminator (Molecular Imager Gel Doc XR System, Bio-Rad) and a single

band with the product was extracted from the gel with NucleoSpin Gel and PCR Clean-up kit (740609**,** Macherey-Nagel) according to the manufacturer's instructions (Appx. 3, p. 187). The use of this kit enables binding of dissolved DNA from a cut out agarose gel band to the silica membrane of the NucleoSpin Gel and PCR Clean-up column, in the presence of chaotropic salts. The bound DNA is further decontaminated in the following washing steps before being eluted under low salt conditions with a slightly alkaline elution buffer.

The purity, as well as concentration of the sample containing DNA template after elution, was checked with the NanoDrop 2000 spectrophotometer. The final PCR product obtained after purification was used in subsequent sequencing reactions. The sequencing service was provided by Source BioScience Ltd. (Nottingham, UK).

3. *Target selection and design of sgRNA*

The aim of this design was the complete disruption of *Culex quinquefasciatus* (Pel) *Actin-4* gene (*Cxq-A4*); therefore the strategy of selecting PAM (protospacer adjacent motif) sites located towards the 5′ end of the coding region was applied. Thus, the DNA sequence directly following the initiation site (ATG) and consisting of 350 bp of the second exon of the *Cxq-A4* gene (Pel) served as a reference for *in silico* identification of sgRNA target regions and for the subsequent design of sgRNAs. The reference DNA sequence was used for selecting the target sites for CRISPR/Cas9-directed mutagenesis.

In order to obtain target sites, the CHOPCHOP target prediction web tool (http://chopchop.rc.fas.harvard.edu/) was employed (Graham & Root, 2015). A list of candidate target sites generated with the CHOPCHOP tool was further checked for the presence of off-target effects through the use of the Basic Local Alignment Search Tool (BLAST) (Altschul et al., 1990) applied to the *C. quinquefasciatus* Johannesburg (JHB) strain genome sequence (CpipJ2, April 2014) and a visual analysis of the results was performed (VectorBase) (Giraldo-Calderón et al., 2015). This process included a review of the JHB strain genome sequence for the presence and complementarity of candidate target site sequences in other parts of the genome that were not the predicted target sites. The presence of the off-target sites may result in undesired genome modifications. This is a particularly unwanted situation, which can lead to the occurrence of off-target effects that can impact experimental assessment. Thus, genomic sequences that perfectly matched the final 12 nucleotides of the candidate target site (seed sequence) and the PAM sequence (5′-NGG-3′; where N is any nucleotide) were discarded at this design stage. Detection of Single Nucleotide Polymorphism (SNP) within 350bp of Pel *Actin-4* sequence was recorded upon comparing this sequence with the complementary part of the reference genome sequence (CpipJ2) and taken into account in the process of designing the target sites. The sgRNA design accounted for the selection of four target sites that were used in sgRNA production.

4. Production of sgRNA

4.1 sgRNA template generation

Linear double-strand DNA template for production of specific sgRNAs was generated using a template-free PCR-based system. In this approach, two partially overlapping oligonucleotides are used: forward primer CRISPRF and reverse primer sgRNAR (see the primer sequences below). The CRISPRF primer sequence includes upstream sequence for efficient *in vitro* transcription (underlined), promoter sequence for T7 RNA polymerase (highlighted in grey), unique GGN_{20} target site sequence and a portion of the sgRNA stem loop that is complementary to the part of the sgRNAR primer sequence (highlighted in bold). The GGN_{20} sequence consists of 20 unique nucleotides (N_{20}) - a unique string of bases that are complementary to the DNA target sequence and are required for the Cas9 recognition and cleavage. Additionally, the GGN_{20} sequence incorporates two guanines at its 5′ end, which are required for the T7 RNA polymerase promoter. Presence of complementary sequences allows the two primers to anneal. The reverse primer consists of the sequence encoding the remainder of sgRNA after the target site (Bassett et al., 2014).

The CRISPRF and the sgRNAR primer with highlighted sequence features are described in the text above:

- CRISPRF (5′ → 3′):

  GAAAT<mark>TAATACGACTCACTATA</mark>GGN$_{20}$**GTTTTAGAGCTAGAAATAGC**,

  where N is any nucleotide

- sgRNAR (5′ → 3′):

  AAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCC

  TTATTTTAACTT **GCTATTTCTAGCTCTAAAAC**

Four target sites selected in the previous step were used in the design of four different CRSPRF primers with their unique GGN$_{20}$ sequence and used in a template-free PCR reaction. These primers were synthesised and used in PCR with the reverse primer sgRNAR. The sequences (5′ → 3′) of the four CRISPRF primers used in separate PCR reactions with sgRNA primer were as follows:

1. CRISPRF_CQA4_1 (LA198)

   GAAATTAATACGACTCACTATA**GGAGCACTGGTCATTGACAA**GTTTTA GAGCTAGAAATAGC

2. CRISPRF_CQA4_2 (LA199)

   GAAATTAATACGACTCACTATA**GGGTCAAAAGGATGCCTACG**GTTTTA GAGCTAGAAATAGC

3. CRISPRF_CQA4_3 (LA200)

GAAATTAATACGACTCACTATA**GATGAAGCGCAATCGAAGAG**GTTTTA
GAGCTAGAAATAGC

4. CRISPRF_CQA4_4 (LA201)

GAAATTAATACGACTCACTATA**CGGTTGGACTTGGGATTCAA**GTTTTAG
AGCTAGAAATAGC

The PCR reactions with the forward CRISPRF primers and the reverse sgRNA primer were performed in order to produce full-length dsDNA templates for *in vitro* transcription of sgRNAs in the following step. Platinum *Pfx* DNA polymerase (11708-013, Life Technologies) from *Thermococcus* species *KOD* was used in this PCR, with the following reaction conditions:

PCR reaction component	Volume for one 50-µl reaction
10X Platinum buffer	5 µl
10 mM dNTP mixture	1 µl
50 mM MgSO$_4$	1 µl
1 µl CRISPR-F primer (100 µM) + 1 µl sgRNAR primer (100 µM) + 2 µl nuclease-free water	1 µl
Platinum *Pfx* DNA polymerase (2.5 U/µl)	0.5 µl
Nuclease-free water	41.5 µl
TOTAL	50 µl

The PCR program for this reaction was as follows:

Step	Temperature	Time	Number of cycles
Initial denaturation	94 °C	30 seconds	1
Denaturation	94 °C	10 seconds	
Annealing	60 °C	30 seconds	35
Extension	68 °C	15 seconds	
Final extension	68 °C	10 minutes	1
Hold	4 °C		

2 μl of the PCR reaction was run on a 2% agarose gel with SYBR Safe DNA stain (S33102, Life Technologies) to verify the size and specificity of amplification. Agarose gel electrophoresis was performed for 1 hour at 100V.

The remaining 48 μl of PCR reaction was purified with NucleoSpin Gel and PCR Clean-up kit (740609, Macherey-Nagel), according to the manufacturer's instructions (Appx. 4, p. 188). In this DNA purification procedure (as specified by the manufacturer), DNA product generated in a PCR binds to the silica membrane of the NucleoSpin Gel and PCR Clean-up column, in the presence of chaotropic salts. The bound DNA is further decontaminated in the following washing steps before being eluted under low salt conditions with a slightly alkaline elution buffer.

The purity, as well as concentration of the sample containing sgRNA template was checked using the NanoDrop 2000 spectrophotometer. The final PCR product after purification was used in *in vitro* transcription reaction (see the chapter subsection below).

4.2   In vitro transcription and purification of sgRNA

A purified DNA template for sgRNA was directly used for the *in vitro* transcription reaction with the MEGAscript T7 Kit (AM1334, Life Technologies) according to the manufacturer's instructions (Appx. 5, p. 189). T7 RNA polymerase is used in this reaction, which catalyses the formation of RNA in 5′ → 3′ direction. This RNA polymerase enables the synthesis of large amounts of RNA product from a double-stranded DNA template containing T7 RNA polymerase promoter located upstream of the transcribed DNA template.

After the final incubation step of the transcription reaction, the RNA samples were purified with the MEGAclear Kit (AM1908, Life Technologies) according to the manufacturer's instructions (Appx. 6, p.190). The transcription reactions were mixed with Binding Solution, passed over an RNA-binding filter and then washed. The elution step of the protocol allows for the release of purified synthetic RNA that can be directly used in downstream applications.

The purity, as well as the concentration of the sample containing recovered RNA was checked with the NanoDrop 2000 spectrophotometer. Each sgRNA was divided into 1 or 2.5 µg aliquots and stored at -80 °C.

Agarose gel electrophoresis was performed for aliquots of the purified product to verify the product sizes after transcription. The samples were run on a 1.5% agarose gel with SYBR Safe DNA stain (S33102, Life Technologies) for 45 minutes at 100V. The results of the electrophoresis were visualised on agarose gel with a UV transilluminator (Molecular Imager Gel Doc XR System, Bio-Rad).

5. Cas9 protein

A recombinant Cas9 protein, based on the *Streptococcus pyogenes* protein with the addition of a nuclear localisation signal (NLS) (CP01, PNA Bio), was used in microinjection based mutagenesis assays described in this chapter. The protein was divided into 1.5 µg aliquots and stored at -80 °C. Published recommendations for recombinant Cas9 highlight several advantages of using this protein instead of Cas9 mRNA (which is likely to be less stable). As reported by Kistler et al. (2015), Cas9 protein provides improved reproducibility, increased mutagenesis and better embryo survival after microinjections.

Cas9 nuclease from *S. pyogenes* (M0386, New England BioLabs Inc.) was used for *in vitro* digestion with sgRNA.

6. sgRNA in vitro digestion

For the purpose of sgRNA *in vitro* digestion with Cas9 nuclease, a DNA substrate generated in a single PCR was used. Oligonucleotides used for this PCR were: the forward primer LA348 (LAA4_CXCRISPR1_FWD): 5′-CTGTGCCGTCCAGCGATGAGTT-3′ and the reverse primer LA349 (LAA4_CXCRISPR1_REV): 5′- TCTGGGCAACGGAAACGCTC-3′. The expected PCR product size was 1098 bp based on *in silico* prediction for the Pel *Cxq-A4* gene region with Vector NTI 11.53 software (Invitrogen). DreamTaq DNA Polymerase (EP0701, Thermo Scientific) was used in this PCR, with the following reaction conditions:

PCR reaction component	Volume for one 25-µl reaction
10X DreamTaq buffer	2.5 µl
10 mM dNTP mixture	0.5 µl
Primer LA348 (10 µM)	0.5 µl
Primer LA349 (10 µM)	0.5 µl
DreamTaq DNA polymerase (5 U/µl)	0.2 µl
DNA template (176ng/µl)	1 µl
Nuclease-free water	19.8 µl
TOTAL	25 µl

The template used for the PCR reaction was DNA extracted from early (< 6 hours old) *C. quinquefasciatus* (Pel) pupae. PCR program for this reaction was as follows:

Step	Temperature	Time	Number of cycles
Initial denaturation	95 °C	1 minute	1
Denaturation	95 °C	30 seconds	35
Annealing	60 °C	30 seconds	
Extension	72 °C	2 minutes	
Final extension	72 °C	5 minutes	1
Hold	10 °C		

The PCR reaction was run on a 1.5% agarose gel with SYBR Safe DNA stain (S33102, Life Technologies) to verify the size and specificity of amplification. The agarose gel electrophoresis was performed for 1 hour at 90V. PCR product was visualised on agarose gel with a UV transilluminator (Molecular Imager Gel Doc XR System, Bio-Rad) and a single band with the

product was extracted from the gel with NucleoSpin Gel and PCR Clean-up kit (740609, Macherey-Nagel) (Appx. 3, p.187). The final PCR product after purification was used as a substrate in *in vitro* digestion with Cas9 and sgRNA.

The *in vitro* digestion procedure was based on New England BioLabs (NEB) manufacturer's protocol and performed as follows:

1. The double-stranded DNA substrate (1098 bp), as well the stock solution of each sgRNA (100 bp), were thawed on ice and diluted with nuclease-free water to the concentration of 30 nM and 300 nM, respectively, in order to maintain the efficient cleavage molar ratio of Cas9 and sgRNA per target (10:10:1).

2. After thawing additional components required for *in vitro* digestion, the reaction was assembled at room temperature in the following order:

Cas9 digestion reaction component	Volume for one 30-µl reaction
Nuclease-free water	20 µl
10X Cas9 Nuclease Reaction Buffer	3 µl
300nM sgRNA	3 µl (30nM final)
1 µM Cas9 nuclease	1 µl (~30nM final)
Pre-incubation for 10 minutes at 25 °C	
30 nM substrate DNA	3 µl
TOTAL	30 µl

Note: Pre-incubation step at 25 °C enables formation of Cas9 protein:sgRNA complex.

3. A 0.2 ml PCR tube containing 30 µl of the reaction was mixed thoroughly, puls-spun in a microfuge and incubated at 37 °C for either 2 or 4 hours.

4. After incubation, the samples were run on a 1.5% agarose gel with SYBR Safe DNA stain (S33102, Life Technologies) to verify digestion efficiency and the product sizes after digestion. The agarose gel electrophoresis was performed for 80 minutes at 80V followed by 10 minutes at 100V. Separated products were visualised on agarose gel with a UV transilluminator (Molecular Imager Gel Doc XR System, Bio-Rad) and results of the experiment were recorded.

7. Embryo microinjections

One of the ways to deliver nucleic acids into embryos in order to create genetically engineered organisms is microinjection. This sophisticated method, although complex and requiring a skilled operator, can be effectively adapted for the delivery of genome editing tools such as the CRISPR/Cas9 system.

7.1. Preparation of microinjection solution

The CRISPR/Cas9 system can be delivered to embryos in a microinjection solution that contains all of the components required for successful genome editing. In order to prepare the microinjection mixture, stock solutions of each sgRNA (four in total) were briefly thawed on ice together with

recombinant Cas9 protein (CP01, PNA Bio). Upon thawing, all four sgRNAs were mixed together to the final concentration of 40ng/µl (each) following addition of 4 µl (1.5 µg) of Cas9 protein (final concentration: 300 ng/µl) and topped up with nuclease-free water to the final reaction volume of 20 µl. After addition of all of the reagents and thorough but gentle mixing by pipetting, the microinjection solution was divided into two aliquots of 10 µl, quickly puls-spun and stored at -80 °C. On the day of injection, a tube with microinjection solution was thawed on ice, centrifuged for 5 minutes at 14,000 x *g* and used directly for microinjections.

7.2. *Embryo collection and handling*

Small insect rearing cages (3-5) (24.5cm x 24.5cm x 24.5cm, Bugdorm, MegaView Science Co., Ltd., Taiwan) with adult *C. quinquefasciatus* (Pel) males and females (100-150 individuals per cage) were used for embryo collection on the day of microinjection. Adult females (4-6 days post-emergence) were blood-fed and presented with an oviposition container 3-4 days after blood feeding. To provide a suitable place for egg laying, this oviposition container (100ml polypropylene specimen container) was filled with distilled water and put inside a mosquito cage. Additionally, in order to encourage the oviposition process, mosquito cages were placed in the dark. The oviposition containers were emptied every 45-60 minutes and egg rafts that were collected were used for microinjections. After oviposition, the egg rafts were moved to a microinjection room (room temperature), and allowed to mature for up to 30 minutes or until their colour changed from creamy white to

light grey. The change of the embryo colour corresponds to the gain in rigidity of the endochorion of the egg, therefore sufficient maturation time helps to minimise embryo susceptibility to damage by handling.

7.3. Microinjection procedure

The syncytial embryonic microinjection procedure was adapted from published protocols for *C. quinquefasciatus* (Allen et al., 2001) and for another mosquito species, *A. aegypti* (Crampton et al., 1997; Lobo et al., 2006), as well as based on previous hands-on experience with embryo microinjections for *A. aegypti* performed at Oxitec Ltd, Abingdon, UK.

After embryo maturation described in the previous section of this chapter, single egg rafts were placed on a piece of filter paper (Whatman) moistened with distilled water to reduce undesired embryo desiccation. Individual eggs were then separated from a portion of a raft and aligned in a straight line under the dissecting microscope (10X magnification) with their narrow posterior ends facing the same direction. Using a pair of fine forceps and a dissection needle, about 50-80 embryos were aligned within 5 minutes, after which the filter paper was briefly but thoroughly dried by pressing a piece of dry filter paper onto its edge. The embryos were subsequently transferred onto a glass cover slip covered with a strip of double-sided sticky tape (toupee tape) by gently pressing down on the coverslip down against the filter paper surface. Transferred embryos were desiccated at room temperature for up to 1 minute, or until the upper surface of the embryo began to bow slightly inwards. This step relieves the internal pressure of the embryo prior to a needle

puncture and creates space for an aliquot of the microinjection solution. Desiccated embryos were immediately covered with water-saturated halocarbon oil to prevent further loss of turgor. The oil solution used in this step was a mixture of halocarbon oil 700 (H8898, Sigma-Aldrich) and halocarbon oil 27 (H8773, Sigma-Aldrich) in 9:1 volume ratio, respectively. The cover slip with embryos attached was placed on a glass microscope slide and mounted onto the microinjections stage of an inverted microscope (BA410 compound microscope, Miotic). Microinjections were performed with a loaded and bevelled (BV-10, Microelectrode Beveler, Stutter Instrument) microcapillary needle mounted to the manual micromanipulator (Narishige) and connected to Femtojet microinjector (Eppendorf). Microinjection solution was injected horizontally into the posterior end of each embryo at an angle of 10-15°. The amount of injected solution corresponded to approximately 1-5% of the embryo volume. All injections were performed within 2 hours of oviposition.

Aluminosilicate microcapillaries with inner filament (OD: 1.0 mm, ID: 0.64 mm, Length: 10 cm; AF100-64-10, Sutter Instrument) were used for production of fine tip microinjection needles by pulling microcapillaries with a laser-based micropipette puller (P-2000, Sutter Instrument) set at the following parameter values: HEAT: 525, FIL: 4, VEL: 37, DEL: 250, PUL: 125 (program settings courtesy of Thomas Ant, Lancaster University). Microinjection needles were loaded by filling a pipette microloader tip (930001007, Eppendorf) with a microinjection solution and by introducing the drawn fluid into the open end of the microcapillary.

After injections, the cover slips with attached embryos were submerged in a microscope slide staining dish filled with distilled water in order to allow for the removal of the excess of halocarbon oil and left in this condition for

between 30 minutes to 5 hours. After the oil removal procedure, the cover slips were removed from the staining dish and a strip of filter paper with a slit for the lined up embryos was attached to the sticky tape fragment located directly below the anterior end of the embryos. The prepared cover slip was carefully submerged vertically in a well of a 6-well cell culture plate ('embryo recovery plate') filled with distilled water up to the level of the anterior end of the embryos (Fig. 7). A plate with cover slips holding the injected mosquito eggs was placed on a plastic tray, covered with a transparent plastic lid and kept in insectary conditions (70% (±10%) relative humidity, 12-hour light/dark cycle) for the time required for reaching an advanced embryonic developmental stage; or for complete embryonic development culminating in hatching.

Figure 7. Post-microinjection embryo recovery set up. Illustrated by the expanded circle, the cover slip with lined-up embryos attached to the double-sided tape and surrounded with a fragment of filter paper. Water level is indicated in the picture with a blue double-headed arrow.
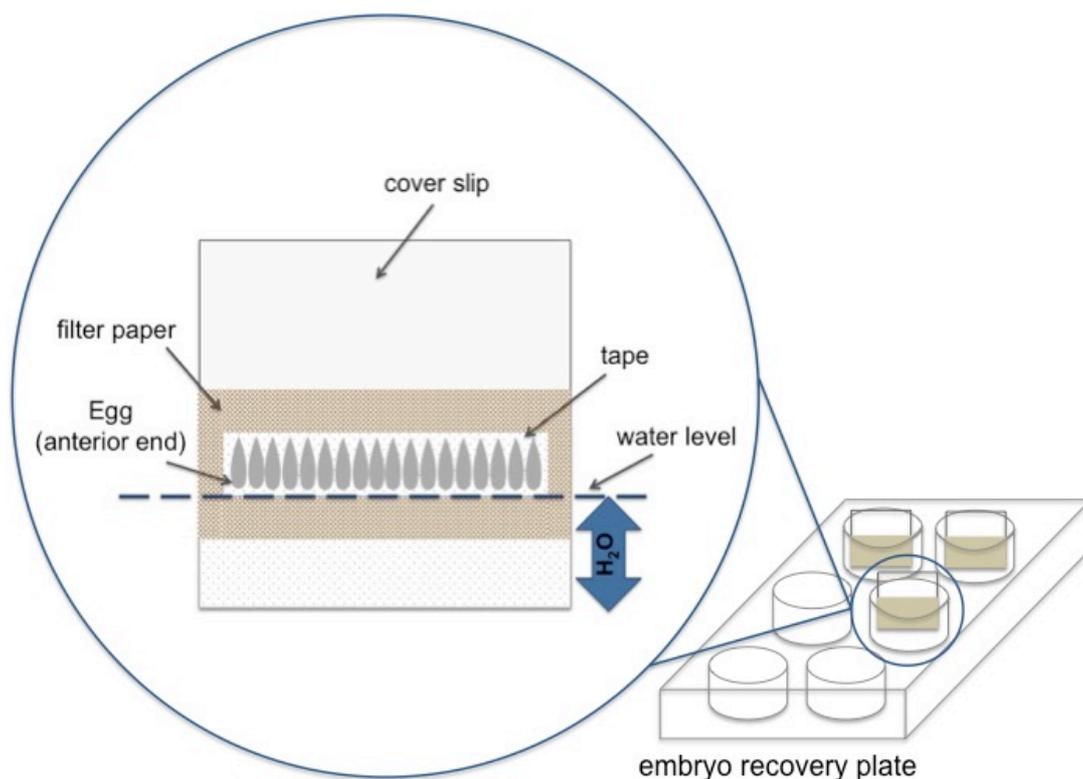
8. In vivo gene editing

8.1. Assay 1: Induced mutagenesis in embryos

The purpose of this experimental stage was to test the functionality of the synthesised sgRNAs *in vivo* in the presence of the Cas9 nuclease. This can be achieved by detecting the presence, type and frequency of genetic changes that were induced in a genome within the sgRNA target sites.

The experimental sample of embryos used in this assay included 6 cover slips with 80 to 128 lined up and microinjected embryos (see protocol for microinjection procedure, p. 77-80) and 6 controls – uninjected portions of egg rafts containing from 100 to 130 embryos each. On the first experimental day, I placed all injected and uninjected embryos in the insectary for 24 hours. 24 hours post-injection, I checked the embryos under a dissecting microscope for the presence of external, late developmental stage characteristics, such as segmentation of the embryo and ocelli pigmentation. I recorded the number of embryos exhibiting these developmental features. After recording the number of embryos in advanced developmental stage, I carefully transferred the embryos into a labelled 1.5 ml tube. Injected embryos were transferred from the cover slip with a pair of fine forceps. A thin paintbrush was used to transfer uninjected embryos from the water surface. After collection, all embryos were stored at -80 °C prior to extraction of genomic DNA (Appx. 7, p. 191). After genomic DNA isolation, I ran the samples on a 1.5% agarose gel with SYBR Safe DNA stain (S33102, Life Technologies) to verify the integrity of the gDNA. The agarose gel electrophoresis was performed for 50 minutes at 90V.

Separated products were visualised on agarose gel with a UV transilluminator (Molecular Imager Gel Doc XR System, Bio-Rad) and results of the experiment recorded. The isolated gDNA was further used in downstream applications (HRMA and NGS, see section 10.2 and 10.4, respectively).

8.2. Assay 2: Induced mutagenesis in adults

In this assay, mosquito eggs were injected (see protocol for microinjection procedure, p. 77-80) over 3 consecutive days (D1, D2 and D3). The number of microinjected embryos for each injection day was as follows: D1 – 204, D2 – 351 and D3 – 360 eggs (915 injected eggs in total). After microinjections, the embryos were placed in the insectary and allowed to hatch. After hatching (>36 hours), 1st instar larvae (L1) were counted and transferred into transparent polystyrene 8 oz. deli pots (Fabri-Kal, USA) with 100 ml of Liquifry baby fish food (Interpet Ltd, UK) freshly prepared solution composed of 1 droplet of Liquifry per 500 ml of distilled water. 24 hours later, the mosquito larvae were provided with dry, finely ground TetraMin fish food (Tetra, Germany) and fed on this food until pupation. On the day of pupation, mosquito pupae were counted, separated by sex and transferred into 100 ml polypropylene pots filled with distilled water. Female and male pupae were placed separately in insect rearing cages (24.5cm x 24.5cm x 24.5cm, Bugdorm, MegaView Science Co., Ltd., Taiwan) and allowed to emerge.

8.2.1. G_0 mosquito handling and genetic crosses

After 48 hours post-emergence, the number of eclosed G_0 virgin adults (females and males) was recorded. G_0 mosquitoes were separated individually (males) or pooled together (females) and outcrossed to their wild-type counterparts. Each individual G_0 male was crossed to the excess number of wild-type Pel females (15) and the pooled G_0 females were crossed to wild-type Pel males (number of males not exceeding the 1:1 sex ratio). Three days after establishing crosses, mosquitoes were blood-fed and after a further 3 days presented with an oviposition container. The oviposition pot was checked daily and newly laid egg rafts transferred to 8 oz. deli pots (Fabri-Kal, USA) with 100 ml of Liquifry baby fish food (Interpet Ltd, UK) solution (see p. 82 for more details).

8.2.2. G_1 mosquito handling and genetic crosses

After hatching, G_1 mosquito larvae were provided with dry, finely ground TetraMin fish food (Tetra, Germany) and fed on it until pupation. On the day of pupation, mosquito pupae were counted, separated by sex and transferred into 100 ml polypropylene pots filled with distilled water. Female and male pupae were then placed separately into insect rearing cages (24.5cm x 24.5cm x 24.5cm, Bugdorm, MegaView Science Co., Ltd., Taiwan) and allowed to emerge. After 48 hours post-emergence, the number of eclosed G_1 virgin adults (females and males) was recorded and all G_1 individuals were screened for the putative phenotype change in the ability to fly hypothesised to be a likely

consequence of a suitable mutation in the target region of the *Cxq-A4* coding

sequence of the gene. G_1 males were further outcrossed individually to 15 2-3-

day-old virgin wild-type females and the offspring reared in the same way as G_0

or G_1 cohort of mosquitoes. Genomic DNA (gDNA) from 3-4-day-old G_1 females

presenting the putative phenotype, as well as from all dead G_1 males from

single outcrosses, was extracted and used in downstream applications (Sanger

Sequencing, HRMA).

9. *Extraction of genomic DNA*

Genomic DNA from embryo (Assay 1), as well as adult (Assay 2)

samples was extracted using NucleoSpin Tissue kit (740952**,** Macherey-Nagel)

according to the manufacture's recommendations (Appx. 7, p.191). This

method of genomic DNA extraction is based on a lysis of the sample material in

a proteinase K/SDS solution, followed by creating the appropriate conditions for

DNA binding with the use of chaotropic salts and ethanol. The bound DNA is

further decontaminated in the following washing steps before being eluted

under low salt conditions with a slightly alkaline elution buffer.

The purity, as well as the concentration of the sample containing genomic

DNA, was checked with the NanoDrop 2000 spectrophotometer. To verify

genomic DNA integrity after extraction, the samples were run on a 1.5%

agarose gel with SYBR Safe DNA stain (S33102, Life Technologies). The

agarose gel electrophoresis was performed for 50 minutes at 90V. Results of

the electrophoresis were visualised on agarose gel with a UV transilluminator

(Molecular Imager Gel Doc XR System, Bio-Rad).

10. Detection of genetic variants

10.1. Phenotype screening

The phenotype screen consisted of a straightforward behavioural test for 2-day-old G_0 and G_1 individuals and was based on a visual assessment of the ability to fly in recovered adult mosquitoes. During this short test (lasting approximately 3 minutes), adult males and females were put in separate small insect rearing cages (24.5cm x 24.5cm x 24.5cm, Bugdorm, MegaView Science Co., Ltd., Taiwan) at low density (5 or less individuals per cage). The cages were agitated individually by gently shaking and tapping on the cage floor and walls. Adults were assumed to be able to fly if they could take off from the floor or a wall of the cage and stay in the air for a minimum of 3 seconds before landing. This visual function test was repeated 3 times on the same experimental day for each mosquito cage with a 2 hours resting period between observations. Mosquitoes that were unable to fly were individually placed in transparent polystyrene 8 oz. deli pots (Fabri-Kal, USA) and were tested a second time by following the same visual function test protocol.

10.2. High Resolution Melting Analysis (HRMA)

High Resolution Melting Analysis is a post-PCR method enabling to analyse genetic variation (SNPs, mutations, methylations) in PCR amplicons by the characterising the nucleic acid samples based on the DNA melting

(disassociation) behaviour. This rapid and cost-effective method for detection of unknown variation in a PCR amplicon was employed in this study prior to sequencing as an initial approach to detect CRISPR/Cas9-induced mutations in the experimental samples obtained from both mutagenesis assays (Assay 1 and Assay 2).

High Resolution Melting Analysis was based on the user guide for HRM Experiments (Life Technologies, 2010) by following the guidelines presented in the protocol for HRM Mutation Scanning Experiment (Chapter 4, p. 49-60). The workflow of the experimental process was adapted from the user guide for HRM Experiments (Life Technologies, 2010) and is presented in the Fig. 8.

Figure 8. Flow chart of the experimental process for the HRM Mutation Scanning Experiment.
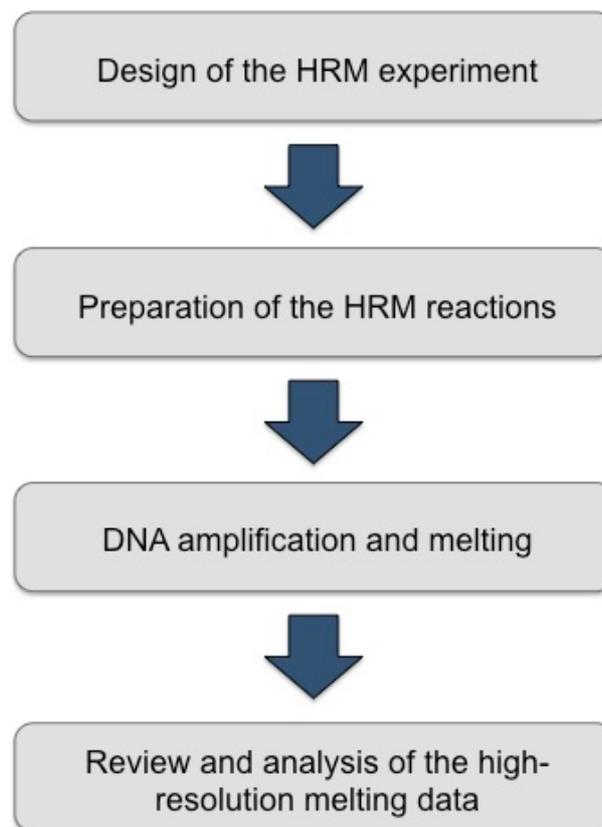
10.2.1.   HRM: experimental design

A critical step in setting up a successful HRM experiment is the design of a robust pair of primers to amplify a specific DNA fragment that spans the region targeted during mutagenesis. In this study, a set of primers was designed according to the guidelines presented in the user guide for HRM Experiments (Life Technologies, 2010) and with the use of Vector NTI 11.53 software (Invitrogen). Designed primer pairs were used in the amplification of sequences spanning over four sgRNA target sites. Design of the primers accounted for two primer pairs amplifying a single target site and one primer pair amplifying two target sites in one PCR reaction (Tab. 3). Each primer pair was tested for specificity to the target of interest, production of non-specific products and formation of primer-dimers. A PCR test for the synthesised HRM primer pairs (Sigma-Aldrich, UK) was performed with DreamTaq DNA Polymerase (EP0701, Thermo Scientific), with the following reaction conditions:

PCR reaction component	Volume for one 25-µl reaction
10X DreamTaq buffer	2.5 µl
10 mM dNTP mixture	0.5 µl
FWD primer (5 µM)	0.5 µl
REV primer (5 µM)	0.5 µl
DreamTaq DNA polymerase (5 U/µl)	0.2 µl
DNA template (176ng/µl)	1 µl
Nuclease-free water	19.8 µl
TOTAL	25 µl

The template used for the PCR reaction was the DNA extracted from early (< 6 hours old) *C. quinquefasciatus* (Pel) pupae. The PCR program for this reaction was as follows:

Step	Temperature	Time	Number of cycles
Initial denaturation	95 °C	3 minute	1
Denaturation	95 °C	1 minute	
Annealing	58 °C / 60 °C	30 seconds	30
Extension	72 °C	1 minute	
Final extension	72 °C	5 minutes	1
Hold	10 °C		

8 µl of PCR reaction was run on a 1.5% agarose gel with SYBR Safe DNA stain (S33102, Life Technologies) to verify the size and specificity of amplification. The agarose gel electrophoresis was performed for 50 minutes at 90V. The PCR product was visualised on agarose gel with a UV transilluminator (Molecular Imager Gel Doc XR System, Bio-Rad).

HRM experiments were designed for microinjected embryos from Assay 1 and for G_1 adults presenting the flightless phenotype (Assay 2). Each HRM experiment consisted of 3 assays corresponding to 3 pairs of primers used during the experiment. One negative control reaction (no DNA template present) was included in each assay (3 negative controls per experiment in total). There were 6 experimental samples and 4 wild-type controls in each assay of the experiment designed for microinjected embryos. Assays in the experiment designed for G_1 adults included 3 wild-type controls and 3 experimental samples (flightless positive phenotype) with 3 technical replicates for each sample.

Table 3. Primers for HRM experiments. The table presents the list of pairs of primers (forward – 'FWD' and reverse – 'REV') used in HRM experiments and includes the primer sequence, predicted PCR product size (bp) for each primer pair and the number of on-target sites within the amplicon.

Primer pair	Primer sequence (5′→3′)	Amplicon size (bp)	Number of amplified target sites
LA203 (MV_CRISPR_CXA4_FWD_2) LA254 (MV_CRISPR_CXA4_REV_22)	ACCCCCGAGGACCTAAAGCAG GGCACGTGGTGCATCATCAC	118	1
LA255 (MV_CRISPR_CXA4_FWD_5) LA256 (MV_CRISPR_CXA4_REV_5)	GATGATGCACCACGTGCCGT GGTAATGATACCGTGCTCGATCG	162	2
LA257 (MV_CRISPR_CXA4_FWD_6) LA258 (MV_CRISPR_CXA4_REV_6)	CGACATGGAGAAGATCTGGCATC AGAACGATACCGGTGGTACGACC	219	1

10.2.2. Preparation of HRM reactions

Reactions for each assay and experiment were prepared according to the protocol for the HRM mutation scanning experiment (Life Technologies, 2010) using the MeltDoctor Master Mix (4415440, Applied Biosystems), with the following components:

HRM reaction component	Volume for one 20-µl reaction
MeltDoctor HRM Master Mix	10 µl
Forward primer (5 µM)	1.2 µl
Reverse primer (5 µM)	1.2 µl
Genomic DNA (20 ng/µl)	1 µl
Nuclease-free water	6.6 µl
TOTAL	20 µl

Separate 1.5 ml microcentrifuge tubes with a HRM master mix (excluding DNA template) were vortexed and spun briefly before pipetting 19 µl aliquots of the HRM master mix into the appropriate wells of the optical 96-well reaction plate (MicroAmp Fast Optical 96-well plate, 4346906, Life Technologies). 1 µl of DNA template was added to each well with the HRM master mix before sealing the reaction plate with optical adhesive film (MicroAmp Optical Adhesive Film covers, 4360954, Life Technologies). The reaction plate was spun in order to mix and draw the liquid to the bottom of the wells before starting the run using a Real-Time PCR instrument.

10.2.3.   DNA amplification and melting

To amplify and melt the DNA and generate HRM fluorescence data, the reaction plate was run using 7500 Fast Real-Time PCR System with 7500 Software (Applied Biosystems). The real-time PCR program, combined with generation of a melting curve, was set according to the protocol for the HRM mutation scanning experiment (Life Technologies, 2010), and was as follows:

Stage	Step	Temperature	Time
Holding	Enzyme activation	95 °C	10 minutes
Cycling (40 cycles)	Denature	95 °C	15 seconds
	Anneal/extend	60 °C	1 minute
Melt curve	Denature	95 °C	10 seconds
	Anneal	60 °C	1 minute
	High resolution melting	95 °C	15 seconds
	Anneal	60 °C	15 seconds

10.2.4.   Post-HRMA quality control

The amplification plot generated after each experimental run was reviewed for normal characteristics, such as the level of fluorescence and the exponential increase in fluorescence using High Resolution Melt Software v3.0 (HRM Software). Additionally, the dissociation/melt curve was verified to ensure that the desired amplicon was detected, which is confirmed with the presence of a single melting peak. The presence of unexpected melting peaks

in a melt curve can be indicative of sample contamination or the presence of primer-dimers.

10.3. Sanger Sequencing

To screen for successful gene editing in selected sgRNA target regions of the *Cxq-A4* gene, genomic DNA samples obtained from Assay 2 (*in vitro* mutagenesis in adults) were used in Sanger sequencing (Sanger & Coulson, 1975) – a method based on *in vitro* DNA replication with selective incorporation of chain-terminating (lacking 3′-OH group), fluorescently labelled dideoxynucleotides (ddNTPs) by DNA polymerase and subsequent detection of these modified ddNTPs in automated DNA sequencing machines (Hunkapiller et al., 1991; Sanger et al., 1977). gDNA derived from individual G_1 adult mosquitoes was used to generate a PCR product that was subsequently used as a template in sequencing reactions with the forward LA388 (CxA4_Seq_FWD_2): 5′-CGTCCGAGTTGTTTGTGGATG -3′ and the reverse primer LA205 (MV_CRISPR_CXA4_REV_2): 5′-CGCTCGGTCAGGATCTTCAT-3′. Sequencing service was provided by Source BioScience Ltd. (Nottingham, UK).

10.3.1. Template preparation for sequencing

A PCR product incorporating the target region of the *Cxq-A4* gene was generated using the forward primer LA350 (LAA4_CXCRISPR2_FWD): 5′-

GACGCTCTCTGCCGCAGACTGTACT-3′ and the reverse primer LA351

(LAA4_CXCRISPR2_FWD): 5′-AGTCTCGTGGACACCGGTAGCTTCC-3′.

The expected amplicon size was 1234 bp based on *in silico* prediction for the

Pel *Cxq-A4* gene region with Vector NTI 11.53 software (Invitrogen).

Q5 High-Fidelity DNA Polymerase (M0491, New England BioLabs Inc.) was

used in this PCR, with the following reaction conditions:

PCR reaction component	Volume for one 25-µl reaction
5X Q5 Reaction buffer	5 µl
10 mM dNTP mixture	0.5 µl
Primer LA350 (10 µM)	1.25 µl
Primer LA351 (10 µM)	1.25 µl
Q5 High-Fidelity DNA polymerase (2 U/µl)	0.25 µl
DNA template (20-30ng/µl)	1 µl
Nuclease-free water	15.75 µl
TOTAL	25 µl

The PCR program for this reaction was as follows:

Step	Temperature	Time	Number of cycles
Initial denaturation	98 °C	30 seconds	1
Denaturation	98 °C	10 seconds	
Annealing	70 °C	30 seconds	35
Extension	72 °C	30 seconds	
Final extension	72 °C	2 minutes	1
Hold	12 °C		

The PCR product for each sample was visualised on agarose gel with UV transilluminator (Molecular Imager Gel Doc XR System, Bio-Rad) and a single band with the product was extracted from the gel with NucleoSpin Gel and PCR Clean-up kit (740609, Macherey-Nagel) by following the manufacturer's protocol (Appx. 3, p. 187).

5 µl of PCR product per sequencing reaction at a concentration of 1 ng/µl per 100bp was submitted for sequencing together with the sequencing primers: LA388 and LA205 (3.2 pmol/µl), following the Source BioScience Ltd. protocol for sample preparation.

10.4.   Next Generation Sequencing (NGS)

Samples from Assay 1 (*in vitro* mutagenesis in embryos) consisted of DNA obtained from the pools of microinjected embryos (between 80-124 embryos in each sample). In order to evaluate the level of somatic modification in the injected individuals, a deep sequencing method was applied.

To screen for successful gene editing (resulting in the native sequence modifications) in selected sgRNA target regions of the *Cxq-A4* gene, genomic DNA samples obtained from Assay 1 were subjected to the Illumina's MiSeq next-generation sequencing (NGS). The MiSeq application is an example of a high-throughput technique that is based on a short-read, sequencing by synthesis (SBS) method. In the SBS approach, after the addition of each dNTP, a fluorescently labeled reversible terminator is imaged and then cleaved to allow incorporation of the next base. The emission spectrum of the terminator indicates the identity of the base or bases complementary to

94

specific positions within the unknown nucleic acid template. Importantly, base calls are made directly from signal intensity measured during each cycle that enables reduction of the base-specific error rate during the sequencing process (Goodwin et al., 2016)

10.4.1.   Sample preparation and quality control

Purified gDNA extracted from injected embryos was used as a template in generating PCR products for each individual sample of pooled embryos. Oligonucleotides used in this PCR were the forward primer LA203 (MV_CRISPR_CXA4_FWD_2): 5′-ACCCCCGAGGACCTAAAGCAG-3′ and the reverse primer LA205 (MV_CRISPR_CXA4_REV_2): 5′-CGCTCGGTCAGGATCTTCAT-3′. The expected amplicon size was 618 bp based on *in silico* prediction for the Pel *Cxq-A4* gene region with Vector NTI 11.53 software (Invitrogen). Q5 High-Fidelity DNA Polymerase (M0491, New England BioLabs Inc.) was used in this PCR, with the following reaction conditions:

PCR reaction component	Volume for one 50-µl reaction
5X Q5 Reaction buffer	10 µl
10 mM dNTP mixture	1 µl
Primer LA203 (10 µM)	2.5 µl
Primer LA205 (10 µM)	2.5 µl
Q5 High-Fidelity DNA polymerase (2 U/µl)	0.5 µl
DNA template (20 ng/µl)	2.5 µl
Nuclease-free water	31 µl
TOTAL	50 µl

PCR program for this reaction was as follows:

Step	Temperature	Time	Number of cycles
Initial denaturation	98 °C	30 seconds	1
Denaturation	98 °C	10 seconds	
Annealing	70 °C	30 seconds	35
Extension	72 °C	30 seconds	
Final extension	72 °C	2 minutes	1
Hold	12 °C		

The PCR product for each sample was visualised on agarose gel with a UV transilluminator (Molecular Imager Gel Doc XR System, Bio-Rad) and a single band with the product was extracted from the gel with NucleoSpin Gel and PCR Clean-up kit (740609, Macherey-Nagel) by following the manufacturer's protocol (Appx. 3, p. 187). The purity, as well as concentration of the sample containing recovered DNA, was checked using the NanoDrop 2000 spectrophotometer. DNA samples were initially diluted to a concentration of 10 ng/µl and their concentration measured with the

NanoDrop 2000 spectrophotometer and the Qubit 3.0 fluorometer (Q33216, Life Technologies) with Qubit dsDNA BR Assay Kit (Q32850, Life Technologies). I repeated the measurements with the Qubit 3.0 fluorometer three times and calculated the mean for all 3 measurements. Further dilution to a working concentration of 0.2 ng/µl was applied to aliquots of DNA after the initial dilution.

Final concentration of DNA, as well as its integrity, was checked with the Agilent 2200 TapeStation System (G2964AA, Agilent Technologies) using D1000 ScreenTape (5067-5582, Agilent Technologies). Based on the outcomes from the quality control check, 5 samples with PCR product were selected and subsequently used in the process of sequencing library preparation for the Illumina MiSeq system.

10.4.2.   Library preparation and sequencing

Preparation of the libraries for the MiSeq sequencing includes fragmentation of a nucleic acid, followed by ligation of these fragments to a common adaptor set used for clonal amplification and sequencing. The amplicons were processed to construct sequencing libraries using the Nextera XT kit (Illumina), with an additional post-amplification quality control stage using a bioanalyser and the High Sensitivity DNA kit (Agilent) to characterise library length distribution after tagmentation (a modified transposition reaction enabling DNA fragmentation, tagging and adaptor ligation). Purified libraries were normalised and diluted to 12 pM using the established Nextera XT kit protocols and sequencing on an Illumina MiSeq using a 2x300 cycle paired-

end run on a version 3 chemistry MiSeq reagent cartridge. Raw data was outputted as fastq files.

11. Data analysis

11.1. High Resolution Melting Analysis (HRMA)

High Resolution Melt Software v3.0 (HRM Software) was used to perform high resolution melting analysis of the data to screen the samples for the presence of mutations. Software settings adjustment for HRMA was based on the protocol guidelines for the HRM mutation scanning experiment (Life Technologies, 2010).

Difference plots are the best way to visualise small differences between melt curves and allow for review of the individual software calls for each sample. In order to generate difference plots, melting profiles of the experimental samples were normalised against the control/reference sample. In the initial analysis, the aligned melt curves with their individual Tm, as well as difference plots, were generated. If experimental samples fell outside the range of variation in fluorescence seen in the control/wild-type samples, these samples were judged as indicative of mutations. Re-analysis of the data was performed in order to omit outliers or to change calls made by the software.

11.2.   Sanger Sequencing

Sequencing results for female and male samples generated after the nested amplification in the final step of the 5′ RACE procedure were aligned to reference sequence of *C. quinqefasciatus Actin-4* gene (CPIJ012572, CpipJ2, VectorBase) and analysed with the use of the ContigExpress assembly module (Vector NTI 11.53 software, Invitrogen).

Sequencing results for samples derived from G_1 individuals were screened for the presence of SNPs, indels, inversions and duplications in order to determine the sequences of the mutations. Although, direct Sanger sequencing for PCR amplified regions surrounding the target site is efficient, genomes of G_1 individuals that are heterozygous mutant carriers for an indel result in overlapping sequencing traces due to the presence of a mismatch region. In order to identify indels in G_1 individuals, an online tool called Poly Peak Parser (http://yosttools.genetics.utah.edu/PolyPeakParser/) (Hill et al., 2014) was applied, in addition to visual inspection of sequence traces. This tool enables an indel detection from a single sequence run performed directly on a PCR product by separating chromatogram data containing ambiguous base calls (including SNPs) into wild-type and mutant allele sequences.

11.3.   Next Generation Sequencing (NGS)

The Illumina MiSeq System was applied to examine the activity of the CRISPR/Cas9 system employed in this study. The 618 bp long, PCR-

generated amplicon that contained the target sequence for each sgRNA derived from 5 samples of the pooled G_0 embryos after microinjections and was subjected to deep sequencing. The generated sequencing reads were analysed for the presence of variants (especially insertions and deletions). The reads were aligned to the reference (wild-type) sequence for the genomic region of interest using the GEM mapper v. 3 with sensitive parameters (version 2 being described in (Marco-Sola et al., 2012)). The alignments were subsequently converted to BAM files and mpileup format using SAMtools (Li et al., 2009). SNPs and other variants (notably deletions that form the core target of this study) were called from BAM files or mpileup output using fast and robust Bayesian method based on the one described in (Raineri et al., 2012) due to the high and unknown number of genotypes present in the sample.

Results and Discussion

1. *Alternative splicing in 5′UTR of the Culex quinquefasciatus Actin-4 gene*

I performed analysis based on Rapid Amplification of cDNA Ends (RACE) in order to identify and characterise the 5′UTR of the *C. quinquefasciatus Actin-4* (*Cxq-A4*) gene. The final step of the RACE procedure was the nested amplification with the Universal Amplification

Primer (UAP) and the gene-specific CPIJ12572-GSP3 (LA34) primer. This reaction enabled amplification of unknown sequences between the homopolymeric tail at the 5′-end of the mRNA and the gene specific GSP3 primer. The result of this amplification was a product that differed in size between female and male samples, indicating the presence of sex-specific splice variants of the 5′UTR in *C. quinquefasciatus* early pupae. In concordance with findings from the previous studies, the detection of the alternatively spliced 5′UTR regions for the *C. quinquefasciatus* male and the female pupae has been also reported for the *Actin-4* homologues in two other mosquito species – *A. aegypti* (G. Fu et al., 2010) and *A. albopictus* (Labbé et al., 2012).

The difference in the product size for *C. quinquefasciatus* male and female pupae was clearly visible upon an analysis of the electrophoresis results (Fig. 9). The product size generated for the male samples was about 200 bp longer (length difference interpreted from the gel image above) than that generated for the female samples. Single bands of the same size for male (~ 800 bp) and female (~ 600 bp) samples were detected on the gel regardless of the primer annealing temperatures set for the PCR program, indicating the high specificity and robustness of the primers used in these PCR reactions.

The actual size difference of 210 bp between the male and female transcript was determined by analysing the sequencing results for the experimental samples. I aligned the generated sequence reads to the region located 5′ upstream of the first exon of the *Actin-4* gene (CPIJ12572, VectorBase) that was identified within the reference genome sequence

101

(CpipJ2, VectorBase). This step enabled the characterisation of the female

and male 5′UTRs with their introns and regions of complementarity. The

schematic organisation of the *Cxq-A4*, together with its 5′UTR, is presented in

Figure 10 (see also Appx. 8, p. 193 for the annotated gene structures in the

Cxq-A4 sequence).

Figure 9. Alternative splicing of the *C. quinquefasciatus Actin-4*. An

image of the ethidium bromide-stained 1% agarose gel with the PCR products

for male (♂ 55 °C and ♂ 65 °C) and female (♀ 55 °C and ♀ 65 °C) samples.

Samples after the PCR were run on the gel with 1kb DNA Hyperladder (BIO-

33053, Bioline Reagents).
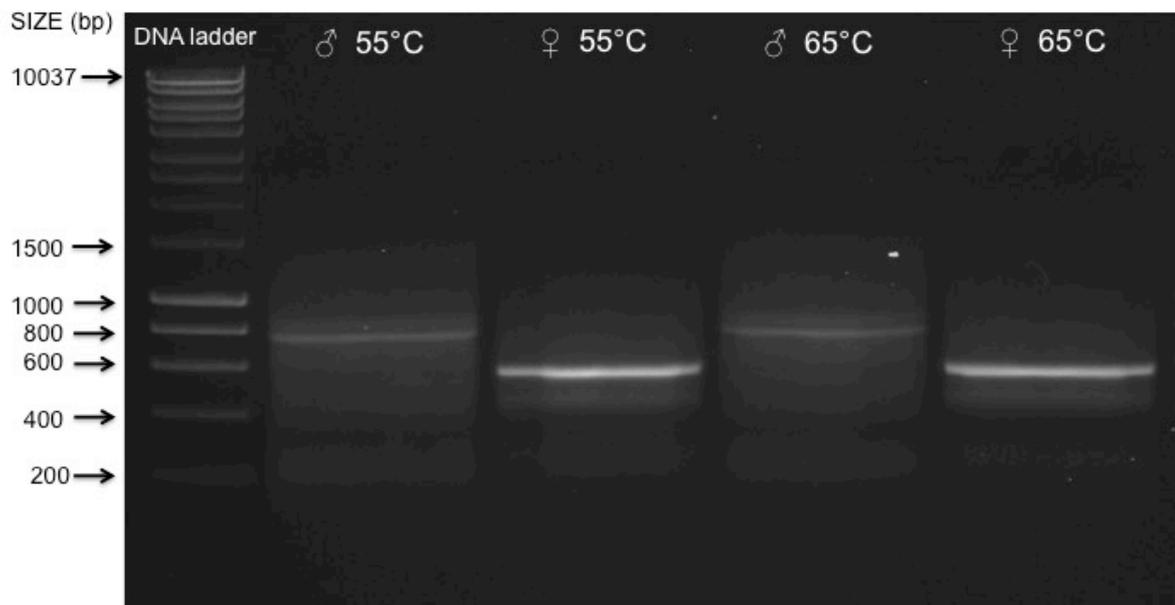
The 424 bp long DNA sequence of the *Cxq-A4* 5′UTR (Fig. 10A)

encoding female and male pre-mRNA transcript contains intron sequences of

328 bp and 118 bp, respectively. The total length of the complementary part

of the 5′UTR mRNA in both sexes is of 96 bp (Fig. 10B). Additionally, the

5′UTR of the male transcript contains 3 termination codons: one amber (UAG) and two opal (UGA) codons, as well as two start codons (AUG); which are located 5′ upstream of the identified male intron (118 bp). These stop and start codons may result in premature termination or initiation of translation that, in consequence, can lead to a reduced production of the functional Actin-4 protein, as previously suggested for *A. aegypti Actin-4* (G. Fu et al., 2010).

The *C. quinquefasciatus Actin-4* transcribed region, the 5′UTR and the alternatively spliced intron, appear to be somewhat shorter than their homologues in *A. aegypti* and *A. albopictus* (Labbé et al., 2012) mosquito species. Although there is a detectable variation in sizes of the coding and non-coding parts of *Actin-4* between these three species, the common feature of these homologous genes is that their 5′UTRs exhibit sex-specific RNA splicing. This sex-specific alternative splicing can be exploited along with the native *Actin-4* promoter specificity in the regulation of an engineered transgene expression at a particular stage during the mosquito development and serve in the development of a genetics-based pest insect control strategy (G. Fu et al., 2007; G. Fu et al., 2010; Labbé et al., 2012).

Figure 10. Schematic representation of the *C. quinquefasciatus Actin-4* gene with its alternatively spliced 5′UTR. Newly identified gene structures for male and female are coloured in blue and red, respectively. A number above the individual black double-headed arrows corresponds to the size of each structure of the gene. A broken line represents introns (splicing events). **(A)** The green arrow represents *Actin-4* promoter (CqAct4). 'Ex2' – Exon 2 and 'Ex3' – Exon 3 of the *Cxq-A4*. **(B)** Alternatively spliced 5′UTR of the *Cxq-A4. S*top and start codons are shown as short vertical lines facing upwards and downwards, respectively.
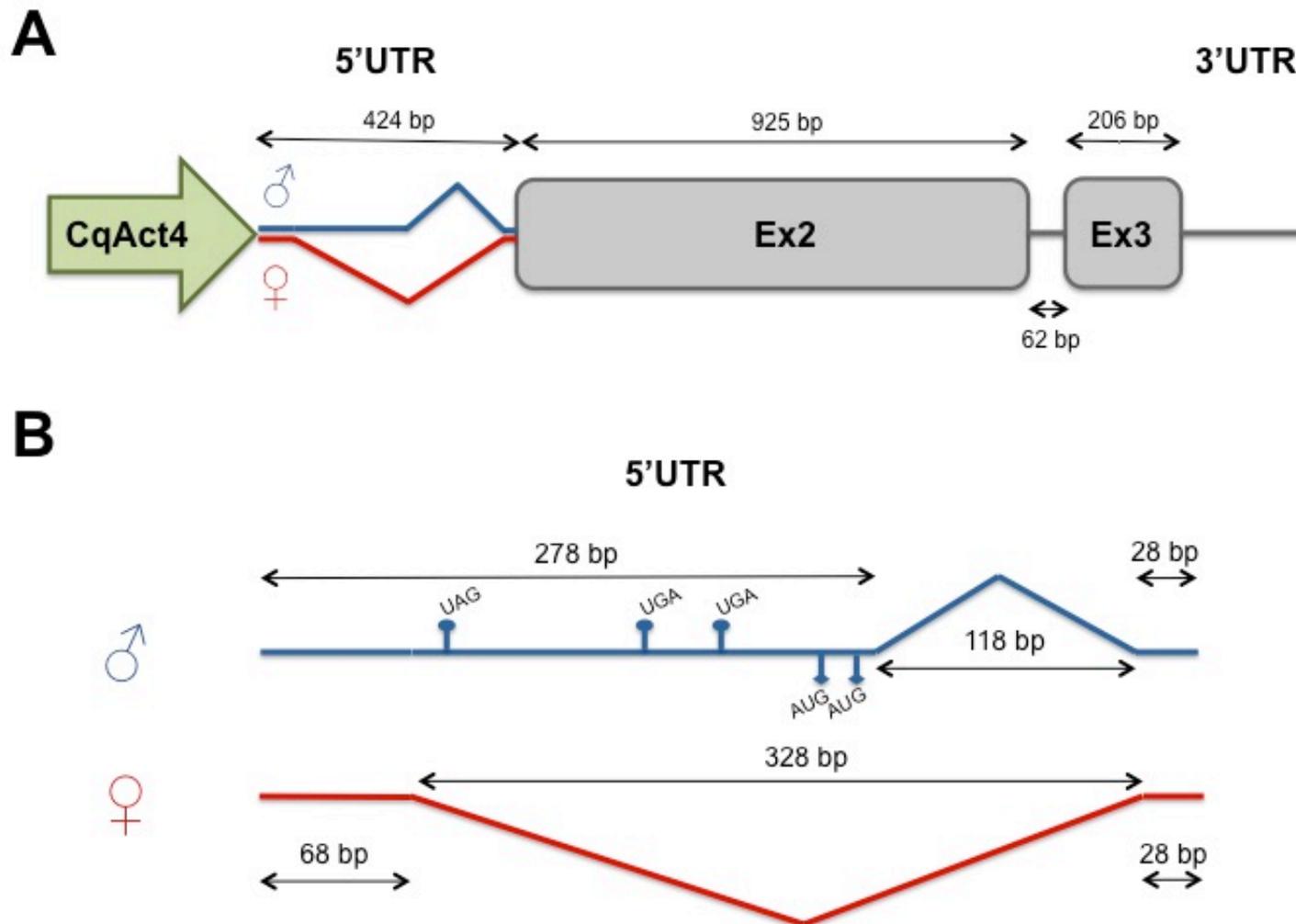
2. *Target selection and design of sgRNA*

The selection of target regions for the knock-out of the *Cxq-A4* (Pel) and the subsequent design of sgRNAs was based on the continuous DNA sequence directly following the initiation site (ATG) and consisted of 350 bp of the *Cxq-A4* second exon. I used this reference sequence for selecting the CRISPR/Cas9 target sites with the CHOPCHOP target prediction web tool (Graham & Root, 2015). Because the CHOPCHOP database does not include the *C. quinquefasciatus* genome, the reference sequence was screened with the CHOPCHOP tool for its closely related species, *A. aegypti*. A table, with a summary of the results generated after the search, contained the list of 25 candidate target sites (Tab. 4). I also checked the individual target sites from this list for the presence of the off-target effects in the *C. quinquefasciatus* reference genome (CpipJ2, VectorBase) (Giraldo-Calderón et al., 2015). Because the CHOPCHOP predictions were based on the *A. aegypti* genome, this additional check of the putative target sites was an inevitable part of the sgRNA design.

Table 4. CRISPR/Cas9 target sites based on the CHOPCHOP predictions. The table includes the first 15 (out of 25) candidate targets, which are the result of the *A. aegypti* genome screening using the CHOPCHOP web tool. The ranked sequences highlighted in bold were selected for the sgRNA *in vitro* production and used in the CRISPR/Cas9 study described in this chapter.

Rank	Target sequence	Genomic location	Exon	Strand	GC content (%)	Off-targets		
						0	1	2
1	**GATGAAGCGCAATCGAAGAGAGG**	sequence:170	1	-	52	0	0	0
2	**GGGTCAAAAGGATGCCTACGTGG**	sequence:145	1	-	57	0	0	1
3	TGAGTTGCGAGTCGCTCCGGAGG	sequence:280	1	-	65	0	0	1
4	TGATGATGATGCTGGAGCACTGG	sequence:7	1	-	52	0	0	0
5	GTGGGCGGCCAACGATGGACGGG	sequence:97	1	-	74	0	0	0
6	AGCAGCACTGGATGTTCCTCCGG	sequence:296	1	-	61	0	0	0
7	GCCAACGATGGACGGGAAGACGG	sequence:90	1	+	65	0	0	1
8	GTGGCGTGGGCGGCCAACGATGG	sequence:102	1	+	74	0	0	1
9	**CGGTTGGACTTGGGATTCAAGGG**	sequence:329	1	+	57	0	0	2
10	GAAGCCGGCTTTGCACATTCCGG	sequence:45	1	+	61	0	0	3
11	TGCATCATCACCAGCGAAGCCGG	sequence:60	1	-	61	0	0	3
12	GGAGCCTCAGTCAGCAGCACTGG	sequence:308	1	+	65	0	0	4
13	CGTGGGCGGCCAACGATGGACGG	sequence:98	1	+	74	0	1	0
14	**GGAGCACTGGTCATTGACAATGG**	sequence:20	1	-	52	0	1	1
15	CGTTGGCCGCCCACGCCACCAGG	sequence:106	1	-	78	0	1	1

I selected the first four target sites from the CHOPCHOP ranking above (Tab. 4) which showed complementarity only with the predicted target sites of the reference *C. quinquefasciatus* Johannesburg genome (CpipJ2) and used them in sgRNA production. The first three of them (counting downstream of the first ATG codon) targeted the sense strand of DNA and the fourth one targeted the antisense strand. These targets are underlined in the fragment of the *Cxq-A4* sequence (5′ → 3′) presented below. This sequence was used as the reference in the sgRNAs design.

ATGTGTGATGATGATGCT<u>GGAGCACTGGTCATTGACAA</u>TGGATCCGGAA
TGTGCAAAGCCGGCTTCGCTGGTGATGATGCACCACGTGCCGTCTTCCC
GTCCATCGTTGGCCGCCCACGCCACCAGGGTGTGATGGTCGGTAT<u>GGG</u>
<u>TCAAAAGGATGCCTACGTGGG</u>T<u>GATGAAGCGCAATCGAAGAGAGG</u>TATT
TTGACGTTAAAGTACCCGATCGAGCACGGTATCATTACCAACTGGGACG
ACATGGAGAAGATCTGGCATCACACGTTCTACAATGAGTTGCGAGTCGC
TCCGGAGGAACATCCAGTGCTGCTGACTGAGGCT<u>CCCTTGAATCCCAAG</u>
<u>TCCAACCG</u>

This sequence fragment derives from the Pel strain and includes three synonymous SNPs that are relative to the Johannesburg genome sequence. The detected sequence polymorphism was taken into account during the *in silico* sgRNA design. The SNPs are highlighted in the reference sequence above.

3. Production of sgRNA

3.1.   sgRNA template generation

I performed the PCR reactions with the four forward CRISPRF primers and the reverse sgRNAR primer in order to produce a full-length dsDNA template (Fig. 11) for *in vitro* transcription of sgRNAs in the following step.

Figure 11. dsDNA template for *in vitro* transcription of sgRNAs. An image of the SYBR Safe-stained 2% agarose gel with the PCR products generated using a pair of primers that included a unique CRISPRF and the common sgRNA primer. The gel lanes labeled with 'dsDNA_CRISPRF_1-4' correspond to the PCR reaction performed with one of the CRISPRF primers (1-4), as well as to the equivalent sgRNA target site number (1-4). Samples after the PCR were run on the gel with 1kb DNA Hyperladder (BIO-33053, Bioline Reagents).
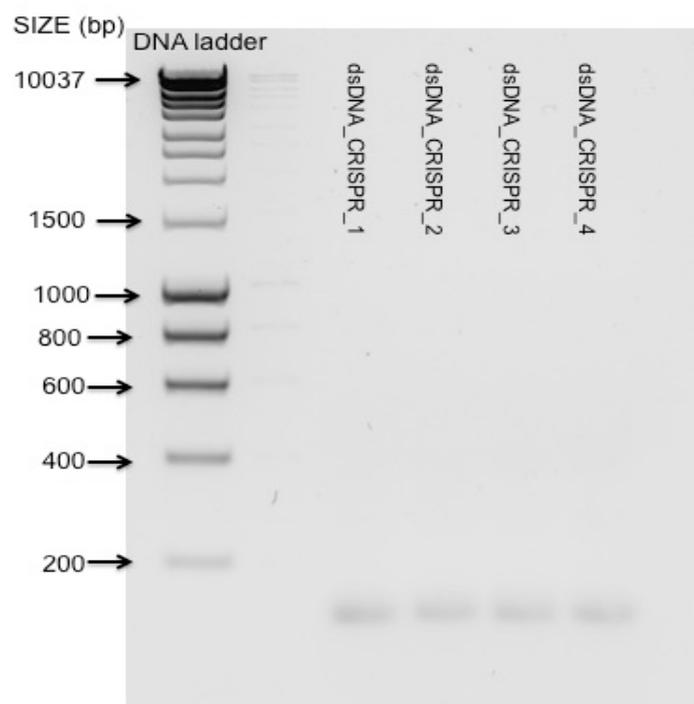
The expected size of the PCR product generated in separate PCRs for each primer pair was a dsDNA template of 122bp long. A single product of less than 200 bp was generated for each primer pair. Additionally, the similar lengths of all of the PCR products indicated the successful amplification of the expected 122bp long dsDNA template (Fig. 11). These PCR products were purified and their purity, as well as concentration, measured. The DNA concentration measured for each purified sample was as follows: 'CRISPR_1' – 138.1 ng/µl, 'CRISPR_2' – 106.6 ng/µl, 'CRISPR_3' – 100.9 ng/µl and 'CRISPR_4' – 132.6 ng/µl. Purified dsDNA was used as a template in the *in vitro* transcription reaction.

3.2. In vitro transcription reaction

The RNA concentration measured for the expected sgRNA product after *in vitro* transcription was as follows: 'sgRNA_1' – 653.6 ng/µl, 'sgRNA_2' – 635.8 ng/µl, 'sgRNA_3' – 637.4 ng/µl, 'sgRNA_4' – 171.9 ng/µl. The RNA concentration obtained for the sgRNA_4 was about 3.7-fold lower than for the other sgRNAs generated in the transcription reaction. This is not a surprising result, as the sense strand of the dsDNA template for the 'sgRNA_4' incorporates nucleotide C and G directly adjacent to the T7 RNA polymerase promoter. The nucleotide composition in this region of the dsDNA template is important because the sequence preference of this promoter is the presence of GG as the first 2 nucleotides of a transcript. According to a number of published studies (Cong et al., 2013; Hsu et al., 2013; Jinek et al., 2012), this region of sgRNA has a relatively minor influence on target specificity and

cleavage efficiency and can be relaxed if required. Although the first

nucleotide in the native target sequence of the sgRNA_4 – and therefore in its

sense strand of the dsDNA template – was not a guanidine, the transcription

reaction led to the synthesis of a compromised but sufficient amount of the

product required for the microinjection solution preparation.

   After each transcription reaction 300 ng of the product was visualised

on a gel (Fig. 12). Based on the *in silico* prediction, the successful

transcription reaction for each sample should lead to the generation of a 100

bp long RNA product (sgRNA). The results presented in Figure 12 indicated

the amplification of a single product of less than 200 bp long. In addition, the

size of the products for all of the samples was similar, which indicated the

successful amplification of the sgRNAs.

Figure 12. sgRNA generated through *in vitro* transcription. An image of

the SYBR Safe-stained 1.5% agarose gel with the products generated in the

in vitro transcription reactions. Samples after the transcription were run on the

gel with 1kb DNA Hyperladder (BIO-33053, Bioline Reagents).
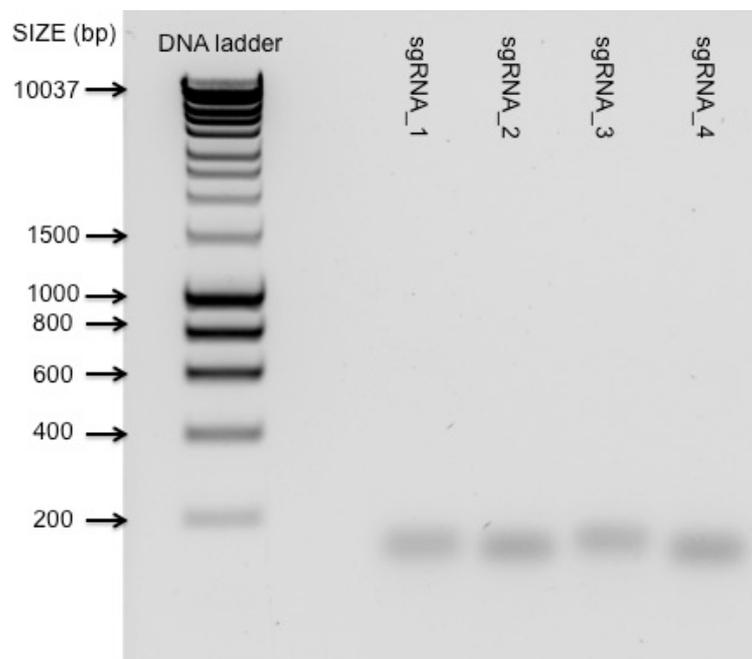
4. sgRNA in vitro digestion

I performed a PCR reaction in order to generate an amplicon that would be long enough to enable clear visualisation of the products on the agarose gel after digestion with Cas9 and sgRNA. The size of this amplicon was 1098 bp. It incorporated a part of the 5′UTR region of the *Cxq-A4* gene (Pel), as well as a portion of its second exon with the target sequences for the four synthesised sgRNAs. I used this PCR product as a substrate in *in vitro* digestion with Cas9 and sgRNA.

I performed two digestion experiments with different incubation periods (either 2 or 4 hours) at 37 °C (Fig. 13 and Fig. 14) in order to compare the cleavage efficiency of the Cas9:sgRNA complex for different exposure times. Negative controls – reactions that lacked the Cas9 nuclease, a sgRNA or both Cas9 and sgRNA together – were included in these experiments for comparison, allowing for separation of the effects of the experimental variable from the overall experimental conditions. The omitted component was substituted with the equivalent volume of nuclease-free water.

The first digestion experiment included two negative controls – one reaction without Cas9 protein and sgRNA and one reaction lacking sgRNA only. All samples in this experiment were incubated for 2 hours at 37 °C (Fig. 13). The manufacturer (New England BioLabs Inc.) recommends an incubation time of 1 hour for a digestion reaction, though extending this time in the first experiment did not lead to a complete digestion of the DNA template. Therefore, in the second digestion experiment an adjustment to the experimental conditions was applied by extending the time of incubation at 37

°C to 4 hours. The aim of this adjustment was to investigate the effect of the

extended exposure of the DNA template to the reaction reagents on the

overall cleavage efficiency.

Figure 13. 2-hour *in vitro* DNA cleavage with Cas9, *S. pyogenes*. An

image of the SYBR Safe-stained 1.5% agarose gel with the products of the

first *in vitro* digestion experiment. All reactions in the digestion experiment

contained a substrate of 1098 bp long linear dsDNA. After incubation,

samples were run on the gel with 1kb DNA Hyperladder (BIO-33053, Bioline

Reagents).
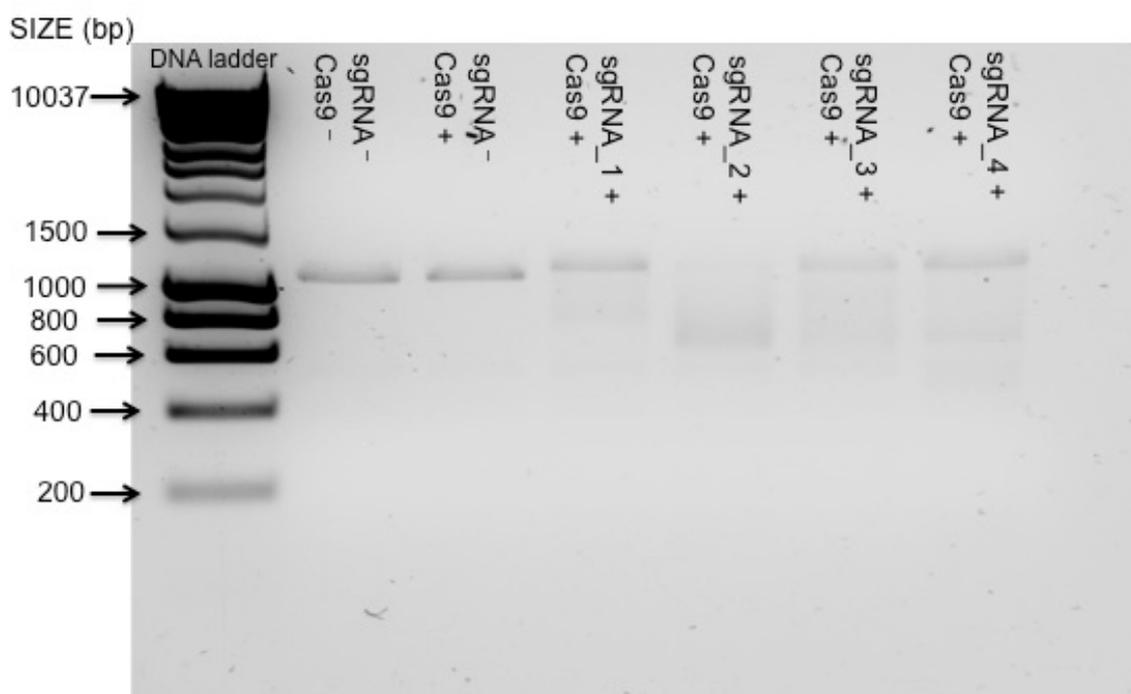
The second digestion experiment included four additional negative

controls compared to the first experiment. These additional control reactions

consisted of the individual sgRNAs, but were lacking the Cas9 protein. The

outcome of the experiment was a complete digestion of the dsDNA template

(Fig. 14). The sizes of the digestion products expected from *in silico* prediction (Vector NTI 11.53, Invitrogen) for the individual sgRNAs indicated a strong correspondence to the sizes of the visualised products on the gel. The predicted product sizes after digestion were as follows: 'sgRNA_1' – 744 and 354 bp; 'sgRNA_2' – 619 and 479 bp; 'sgRNA_3' – 594 and 504 bp; and 'sgRNA_4' – 653 and 445 bp. This prediction was based on the assumption that the Cas9 cleavage, leading to a DNA double-strand break, generally occurs 3-4 nucleotides upstream of the PAM sequence (Gasiunas et al., 2012; Jinek et al., 2012).
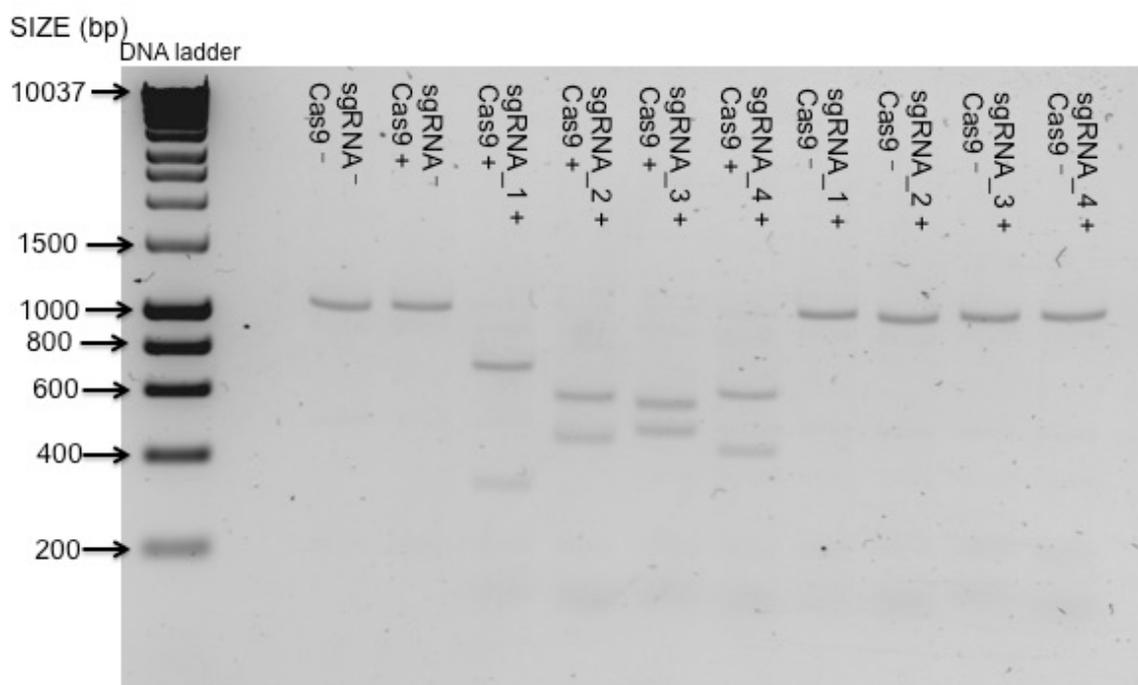
Figure 14. 4-hour *in vitro* DNA cleavage with Cas9, *S. pyogenes*. An image of the SYBR Safe-stained 1.5% agarose gel with the products of the second *in vitro* digestion experiment. All reactions in the digestion experiment contained a substrate of 1098 bp long linear dsDNA. After incubation, samples were run on the gel with 1kb DNA Hyperladder (BIO-33053, Bioline Reagents).

The results of the second digestion experiment showed that all of the designed sgRNAs are fully functional *in vitro*, under the controlled conditions as specified above.

The presence of uncut DNA substrate in the first digestion experiment, apart from suggesting potential issues with the experimental reaction set up; elucidated concerns regarding the quality of the reagents used for *in vitro* digestion. The integrity of sgRNA and DNA was essential for the reliable template cleavage *in vitro* and did not raise further concerns upon performing quality control checks (gel electrophoresis, spectrophotometry) for the nucleic acids used in the conducted experiments.

Although the Cas9 endonuclease efficiency proposed by the manufacturer can be questioned with regard to the cleaving efficiency that was determined through experiments presented here, it is important to remember the role of a sgRNA in facilitating the cleavage process. Because Cas9 requires a functional RNA for an efficient target sequence editing, the efficacy of sgRNA in this two-component system will inevitably affect the final outcome of any gene editing experiment.

So far, several factors have been proposed to be important for the design of efficient sgRNAs, such as (1) the GC-content of the sgRNA (ideally between 40% and 80%) (Gagnon et al., 2014), (2) the presence of guanine immediately 3′ of the PAM site (Doench et al., 2014), (3) guanine enrichment and adenine depletion as a major determinant of sgRNA stability (Moreno-Mateos et al., 2015). These factors have been incorporated as metrics in the online CRISPR/Cas9 target prediction tool – CHOPCHOP. The CHOPCHOP

tool was used to design optimal sgRNAs, the efficiency of which was evaluated in the CRISPR/Cas9 study presented in this chapter.

The partial digestion of the DNA template in the first experiment enabled the evaluation of *in vitro* efficiency of the generated sgRNAs. The results of this evaluation were compared to the CHOPCHOP prediction outcomes and are summarised in Table 5. The summary of results presented in Table 5 shows a major overlap of the sgRNA efficiency obtained from the CHOPCHOP tool with the efficiency that was evaluated experimentally. Interestingly, and in agreement with the findings of Doench et al. (2015), the sgRNAs ranked as the top two (sgRNA_2 and sgRNA_3) in the presented rankings incorporated guanine as the last nucleotide in their sequence (3′ end), opposed to the other two sgRNAs with adenine in the same position. The sgRNA_4 and sgRNA_1 with adenine at position 20 were ranked as the third (9th CHOPCHOP prediction) and the fourth (14th CHOPCHOP prediction) based on their cutting efficiency, respectively. Although only four sgRNAs were experimentally evaluated in the *in vitro* studies, their positions in both rankings indicate validity of the CHOPCHOP tool in the process of selecting the target sites for CRISPR/Cas9 experiments.

Table 5. On-target cleavage efficiency of the *Cxq-A4* sgRNAs. A comparison of the sgRNA rankings based on the *in silico* prediction ('CHOPCHOP ranking') and the experimental outcomes ('Cas9 digestion-based ranking') obtained for the four studied sgRNAs. The third column presents the sgRNA order based on the position of the target site in the *Cxq-A4* gene (5′→3′) and directly corresponds to the assigned sgRNA number (first column).

sgRNA	Target sequence (5′→3′)	Gene location (5′→3′)	CHOPCHOP ranking	Cas9 digestion-based ranking
sgRNA_1	GGAGCACTGGTCATTGACAA	1st	IV	IV
sgRNA_2	GGGTCAAAAGGATGCCTACG	2nd	II	I
sgRNA_3	GATGAAGCGCAATCGAAGAG	3rd	I	II
sgRNA_4	CGGTTGGACTTGGGATTCAA	4th	III	III

5. In vivo gene editing

5.1. Assay 1: Induced mutagenesis in embryos

5.1.1. Microinjections and quality control after DNA isolation

The total of 684 *C. quinquefasciatus* embryos (6 cover slips with 80 to 128 embryos each) were microinjected in this assay. Another 600 to 780 (6 portion of rafts with 100 to 130 embryos each) of uninjected embryos served as experimental controls. Of the 684 injected embryos, 309 (45%) exhibited features indicating the advanced developmental stage of embryogenesis on the day of sample collection for DNA isolation (24 hours post-injection). The advanced developmental stage was recorded for all of the samples with microinjected embryos and was detected in 32% to 64% of those embryos (Tab. 6).

The purity, as well as the concentration of the individual samples containing genomic DNA after isolation, was checked with the NanoDrop 2000 spectrophotometer and did not raise further concerns. Concentration of the DNA ranged between 124.3 to 168.7 ng/μl and between 125.3 to 182.8 ng/μl for the 'CRISPR' and 'Control' samples, respectively (Tab. 7).

Table 6. Microinjections for induced mutagenesis assay in embryos. The table presents the time recorded after microinjections performed on each set of embryos, the number of injected embryos and the number and percentage (%) of the embryos exhibiting late developmental stage characteristics ('number of developing embryos' column).

| Sample name	Microinjection time	Number of injected embryos	Number of developing embryos (%)
CRISPR_1	16:11	106	68 (64%)
CRISPR_2	16:48	122	69 (56.5%)
CRISPR_3	17:22	128	54 (42%)
CRISPR_4	19:33	124	40 (32%)
CRISPR_5	20:45	124	78 (63%)
CRISPR_6	NA	80	NA
TOTAL	NA	684	309 (45%)

Note: Sample 'CRISPR_6' was collected a few days earlier following the same protocol as for the other samples presented in this table. Neither time of injection nor 'number of developing embryos' was recorded for this sample (NA).

Table 7. DNA concentration of the samples with injected and uninjected embryos. A summary of the DNA concentrations measured for all experimental samples in Assay 1. The second column of the table presents the number of embryos in each sample.

Embryo sample	Number of embryos	DNA concentration (ng/µl)
CRISPR_1	106	162.2
CRISPR_2	122	130.8
CRISPR_3	128	146.4
CRISPR_4	124	150.4
CRISPR_5	124	168.7
CRISPR_6	80	124.3
Control_1	100-130	158.0
Control_2	100-130	167.7
Control_3	100-130	182.8
Control_4	100-130	160.7
Control_5	100-130	125.3
Control_6	100-130	146.4

A gel electrophoresis was performed for an aliquot (300-400 ng) of each DNA sample and the results were visualised on an agarose gel (Fig. 15). DNA visible on the gel resulted in the presence of tight bands with minimal smearing and had a molecular weight greater than 10 kb, thereby indicating a good quality of extracted DNA. After extraction, the gDNA was used in downstream application (HRMA and NGS).

Figure 15. gDNA integrity after isolation from samples with injected and uninjected embryos. An image of the SYBR Safe-stained 1.5% agarose gel with gDNA isolated from individual embryo samples from Assay 1. The image shows the results for 7 out of 12 experimental samples. The results for the other 5 samples were similar and are not presented here. Samples after the genomic extraction were run on the gel with 1kb DNA Hyperladder (BIO-33053, Bioline Reagents).
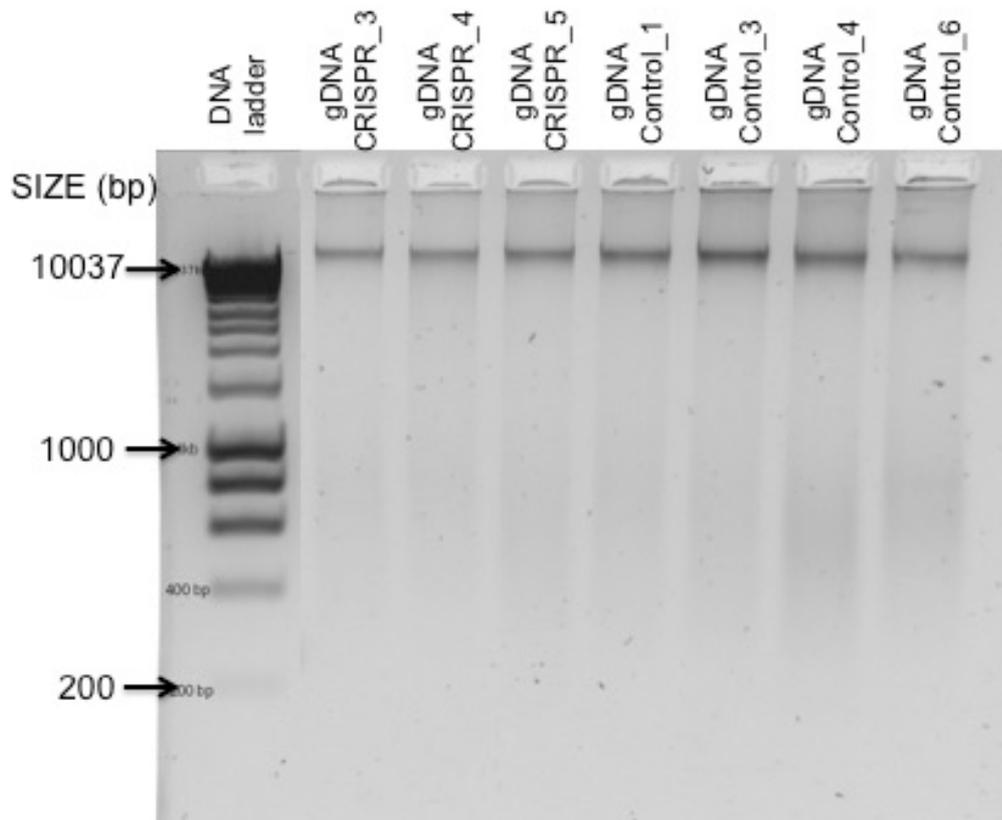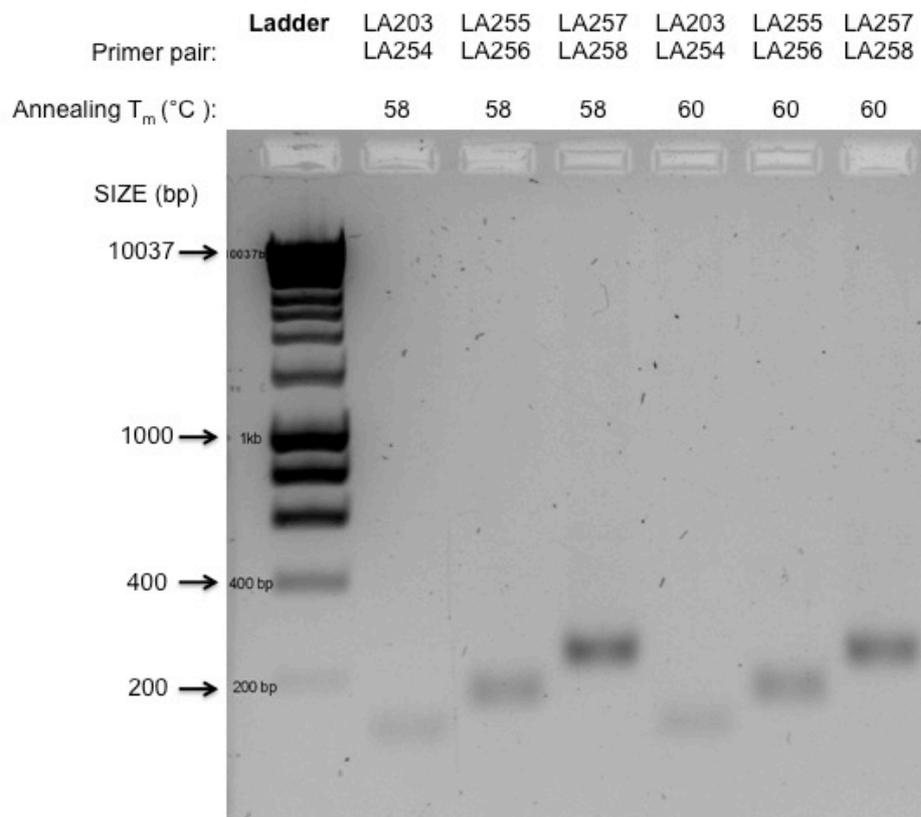
5.1.2. HRMA-based variant detection and characterisation in

embryos

5.1.2.1. Optimisation of the HRM primers

The success of a HRM experiment strongly depends on a robust pair of primers. I tested 3 gene-specific pairs of primers that were designed for the purpose of future HRM experiments.

The primers were tested in two PCR programs with different annealing temperatures (58 °C and 60 °C). The PCR reactions resulted in the generation of a single product regardless of the annealing temperature used (Fig. 16). The size of the products for each primer pair corresponded to that predicted *in silico* and was as follows: primer pair LA203 and LA254 – 118 bp, LA255 and LA256 – 162 bp, LA257 and LA258 – 219 bp. Neither primer-dimers nor non-specific products were detected on the agarose gel after electrophoresis. These results indicated a strong specificity of the tested primers to the target of interest. These primers were further used in all HRM experiments presented in this chapter.

Figure 16. Primer optimisation for HRM experiments. An image of the SYBR Safe-stained 1.5% agarose gel with the products of the PCR optimisation reactions performed with 3 primer pairs: LA203 – LA254, LA255 – LA256 and LA257 – LA258. Samples after the PCR were run on the gel with 1kb DNA Hyperladder (BIO-33053, Bioline Reagents).


5.1.2.2. Variant detection and characterisation

I used the DNA obtained from the embryo samples to screen for the presence of mutations in HRMA. I generated the aligned melt curve plots and the difference plots in order to review the HRM software calls for each sample (Fig. 17). The difference plots were normalised against the control/wild-type

sample. I scored the samples as indicative of mutations if they fell outside the range of variation in fluorescence seen in the control samples.

Out of three assays ('Assay 1: 203_254', 'Assay 2: 255_256', 'Assay 3: 257_258') two contained samples that were indicative of mutations (Tab. 8). I identified 3 non-wild-type variants in 5 samples ('CRISPR_1', 'CRISPR_2', 'CRISPR_4', 'CRISPR_5', 'CRISPR_6') of the 'Assay 1: 203_254' and one non-wild-type variant in two samples ('CRISPR_1' and 'CRISPR_5') of the 'Assay 2: 255_256'. The number of detected non-wild-type (putative mutant) variants decreased with the length of the amplicon obtained during the RT-PCR performed for each assay. There were no mutant variants identified in the 'Assay 2: 257_258' (the amplicon size of 219 bp). Conversely, the number of the wild-type variants increased with the size of the PCR amplicon. Two wild-type variants were detected in the 'Assay 1: 203_254' (the amplicon size of 118 bp), 3 in the 'Assay 2: 255_256' (the amplicon size of 162 bp), and 4 in the 'Assay 2: 257_258'. Although the recommended amplicon length for HRM analysis is from 100 bp to 300 bp, detection of subtle variations such as SNPs typically requires smaller amplicons (Applied Biosystems, 2009). Consequently, a substantial number of the artificially generated SNPs within a longer amplicon could have failed to be detected during this study.

Figure 17. HRMA-based variant detection in embryos. An example of the aligned melt curve plot (A) and the difference plots (B, C) generated in the HRM experiment. These plots are the alternative way of showing the change in fluorescence with the increase in temperature during the DNA melting step. Plot A and C show results for the wild type (WT) and the CRISPR samples and plot B depicts variant calls for the wild-type samples only. Variant types in the plots are coded with different colours.
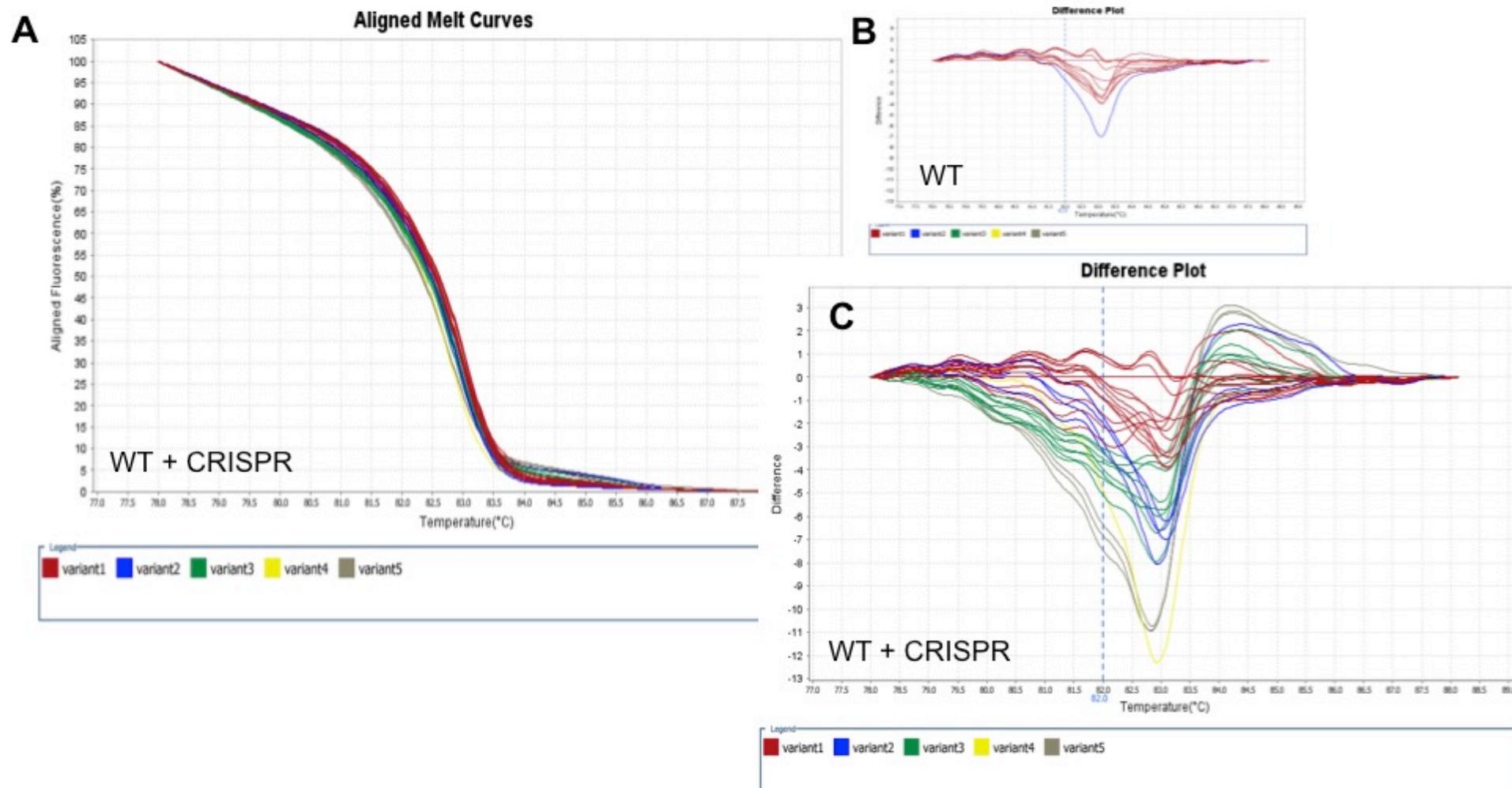
Under the assumption that the cleavage efficiency for sgRNA correlates with the number of variants detected within each target region, the outcome of this study suggests that the sgRNA_1 is the most efficient sgRNA, as it generates the largest number of mutant variants (3). Therefore, the lowest putative cleavage efficiency describes the sgRNA_4.

As expected, the sample replicates clustered together upon a variant call (detection of a sequence variant) (Fig. 17C), however, not many of these sample replicates were detected during this HRM experiment. In several cases, the individual replicates for a single experimental sample were detected within various variant call ranges; that also included the ranges that were identified for the wild-type samples. These undesirable outcomes could have several explanations, such as incorrect salt concentration in each reaction, low PCR efficiency or variation in the starting DNA concentration between samples. Additionally, some of the control wild-type samples were characterised by a wide spread in HRM curves, which might have been caused by a naturally occurring variation within the target region in Pel population.

In summary, the ambiguous results of the HRM experiment may not be surprising, particularly when taking into account the character of the experimental samples used in the HRMA. As each experimental sample consisted of the DNA obtained from 80 to 128 G_0 embryos (putative mosaics), detection of the pooled variation within a sample could have caused a bias in the true variant detection or dramatically reduce variant calling sensitivity.

Table 8. Number of variants scored for embryos in HRMA. The table presents the number of the wild-type ('WT') and non-wild-type ('Non-WT') variants detected during HRMA in each experimental assay for the embryo samples that were microinjected with the CRISPR/Cas9 components. Each assay was performed with a different pair of primers that was specific for the *Cxq-A4* gene.

	Assay 1: 203_254			Assay 2: 255_256			Assay 3: 257_258		
	WT variants	Non-WT variants	Total number of variants	WT variants	Non-WT variants	Total number of variants	WT variants	Non-WT variants	Total number of variants
	V1, V2	V3, V4, V5	5	V1, V2, V3	V4	4	V1, V2, V3, V4	-	4
Embryo sample									
CRISPR_1	V1	V3	1	V3	V4	2	V1, V2, V3	-	3
CRISPR_2	V2	V3	1	V1	-	1	V2, V3	-	2
CRISPR_3	V1, V2	-	2	V3	-	1	V1	-	1
CRISPR_4	V1	V3	2	V3	-	1	V4	-	1
CRISPR_5	V2	V4	2	V3	V4	2	V1, V2, V3	-	3
CRISPR_6	V2	V5	1	V1, V2	-	2	V2, V3	-	2

Therefore, taking into account the character of the experimental samples and the ambiguity of the obtained HRMA results, the CRISPR/Cas9 embryo samples were subjected to one of the methods of deep sequencing based on the MiSeq application, with the assumption that this approach can ensure more reliable variant detection and characterisation.

5.1.3. Deep sequencing-based variant detection and active sgRNA identification

5.1.3.1. Sample quality control

The sequencing method requires a pure nucleic acid template of a particular concentration in order to deliver good quality results; therefore, the quality control check of the template is required. I performed a quality control procedure for the purified and diluted PCR products of the *Cxq-A4* gene region using the Qubit 3.0 fluorometer and the Agilent 2200 TapeStation System (Tab. 9, Fig. 18). Combining these two methods of quality control was essential, as the limited sensitivity of the Qubit dsDNA BR Assay Kit did not provide a reliable and precise evaluation of the DNA concentration after the final dilution of the PCR product to a concentration below 2 ng/µl. The DNA visible on the gel image generated using the TapeStation appeared as tight bands below 700 bp in size (Fig. 18). The report generated after the sample run with the TapeStation indicated the presence of the expected single bands in all of the samples, with the DNA size of 661 to 672 bp. The estimated

concentration of the PCR amplicon for all of the samples was from 0.147 to 0.194 ng/µl (Tab. 9). Based on the outcomes from the quality control check using the TapeStation, 5 samples ('CRISPR_1', 'CRISPR_2', 'CRISPR_4', 'CRISPR_5' and 'CRISPR_6') with PCR product were selected and subsequently used in the process of sequencing library preparation for the Illumina MiSeq system.

Table 9. Post-dilution DNA concentration of the PCR product generated for injected embryo samples. A summary of DNA concentration measured after dilutions of the PCR product. Measurements for dilutions to about 10 ng/µl are shown for Nanodrop 2000 and Qubit 3.0 (3 consecutive measurements for each sample with the calculated mean). The last column presents TapeStation measurements for all of the CRISPR samples after further dilution to the final ~ 0.2 ng/µl.

Embryo sample	DNA concentration (ng/µl)					
	Nanodrop 2000	Qubit 3.0				TapeStation
		I	II	III	Mean	
CRISPR_1	9.0	5.20	4.84	4.72	4.9	0.173
CRISPR_2	6.3	3.44	3.27	3.27	3.3	0.183
CRISPR_3	9.4	5.28	4.96	4.92	5.0	0.167
CRISPR_4	7.7	5.08	4.88	4.84	4.9	0.147
CRISPR_5	10.6	5.08	4.84	4.80	4.9	0.194
CRISPR_6	7.1	3.39	3.31	3.26	3.3	0.162

Figure 18. Post-dilution TapeStation quality control for samples with PCR product generated for injected embryos. A gel image obtained from the analysis based on the use of a tape-based platform for electrophoresis. Each lane with the CRISPR sample shows the expected single band of a DNA fragment below 700 and above 500 bp in size.
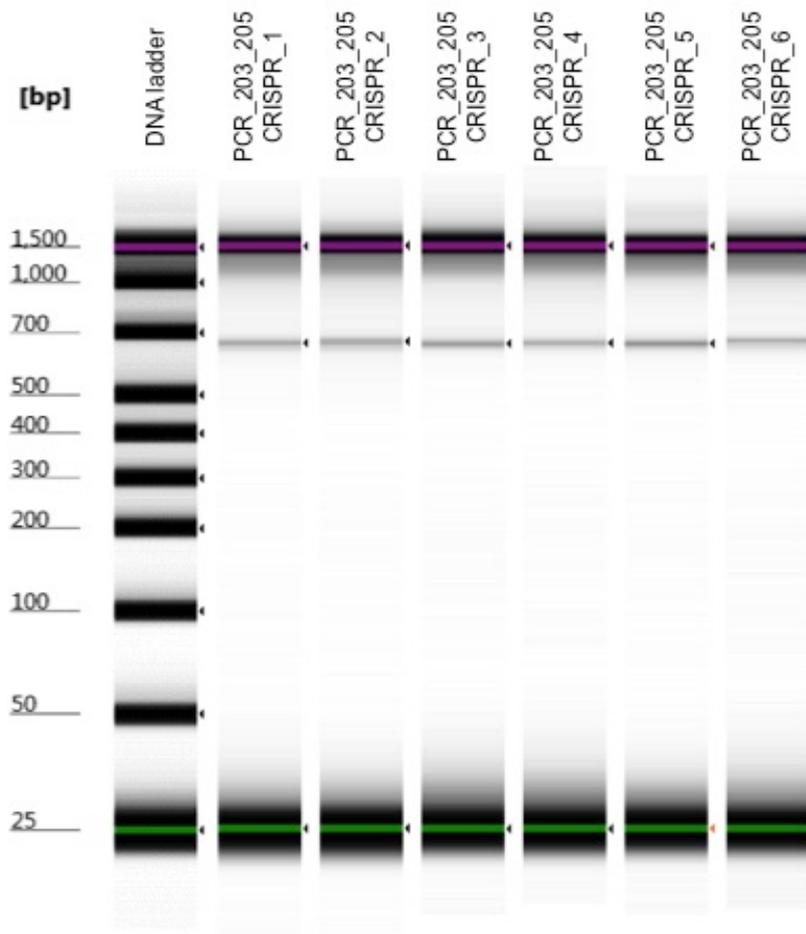
5.1.3.2. Post-sequencing variant detection and identification of the active sgRNAs

I applied a deep sequencing method (Illumina MiSeq) in order to examine the sgRNAs ability to induce somatic modification in the microinjected embryos from Assay 1 (*in vitro* mutagenesis in embryos). 5

samples derived from the pooled G_0 embryos ('CRISPR_1', 'CRISPR_2', 'CRISPR_4', 'CRISPR_5', 'CRISPR_6') contained the PCR-generated amplicon (618 bp) that incorporated four sgRNA target sites and were subjected to sequencing. The generated sequencing reads were aligned to the reference (wild-type) sequence and analysed for the presence of variants (especially insertions and deletions). The high average coverage (the number of reads that include a given nucleotide in the reconstructed sequence) was reported for all 5 samples and ranged between 15800 to 72500 reads.

The heatmap plots (Fig. 19A and Fig. 19B) generated for all 5 samples upon alignment of the reads to the reference sequence indicated the presence of variants (genotypes) within four sgRNA target sites. However, all the observed variants were of a low frequency (between 0.1% and 1% for most positions, with a maximum of less than 2%) and could not be distinguished from the noise generated by sequencing errors. The low frequency of the detected variants can be a direct result of a low level of mutagenesis in the embryos after injections. The embryos that exhibit a developmental arrest after injections will not exhibit sufficient proliferation of the cells carrying the DNA material with the artificially induced changes and therefore the ability to detect mutations after PCR is reduced.

Despite the observed low variant frequencies, a clear pattern can be observed in the prevalence and location of the detected deletions for the generated sequence reads. In particular, there are 4 clearly separated large sets of correlated deletions that originate from 4 distinct positions within the genomic fragment of interest (Fig. 19A and Fig. 19B). These 4 positions overlap tightly with the 4 sgRNA target sites and can be identified here as 'hot

spots' of the CRISPR/Cas9 activity. For each of these 'hot spots' the deletions appear to be aligned in a skewed line of a variable thickness. This is because of the different increasing lengths of the deletion for the particular nucleotide position in the sequence.

The presence of a large number of insertions for each sample weakly correlates with the 4 observed 'hot spots' of this CRISPR/Cas9 strategy. Additionally, the observed insertions are not always identified as duplications of the sequence flanking the sgRNA target site and in this stage of data analysis it is difficult to identify their origin. However, it has been shown that the DSBs can also be repaired through the insertion of DNA from distant regions of the genome (Onozawa et al., 2014).

The size of deletions and insertions varied between samples and ranged between 1 to 101 nucleotides and 1 to 116 nucleotides for deletions and insertions, respectively.

The results presented here are preliminary and require further analysis based on the application of adequate statistical models that allow for data analysis at the population level. This approach will result in reliable characterisation of the observed variants, as well as allow for quantitative analysis of the level of somatic modifications generated with the use of each sgRNA - a reasonable proxy for their editing efficiency *in vivo*.

Figure 19A. Heatmaps of observed genotypes (credit to Paolo Ribeca). The heatmap plots present the frequency distribution of the detected genotypes (i.e. all possible variants, including those present in the wild-type) at any given position of the 618 bp reference (wild-type) sequence. Axis x represents the nucleotide position within this sequence (5′ → 3′, L − R). Axis y represents the observed genotype for each nucleotide at any considered position within the reference sequence: labels in the form of a positive (+) number stand for insertions of corresponding size; labels in the form of negative (-) numbers stand for deletions of corresponding size. The frequencies of insertions or deletions of similar size have been accumulated (for instance "-6" shows the frequency of all deletions from size 6 to 10).
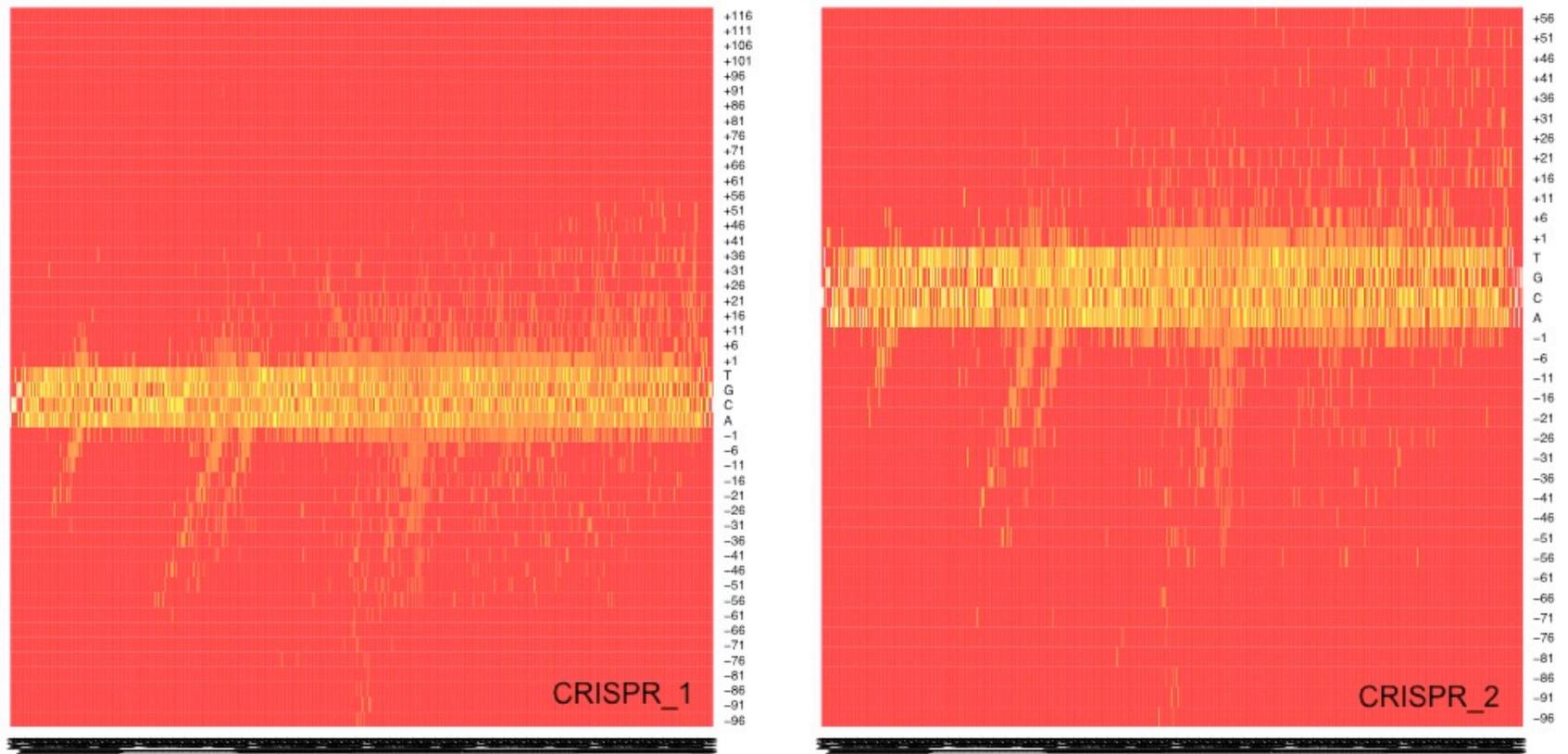
Figure 19B. Heatmap of observed genotypes (credit to Paolo Ribeca). The heatmap plots of the frequency distribution of the detected genotypes. Numbers (1, 2, 3 and 4) shown in the plot for the sample 'CRISPR_5' represent the CRISPR/Cas9 'hot spots' for each sgRNA: 1 – sgRNA_1, 2 – sgRNA_2, 3 – sgRNA_3 and 4 – sgRNA_4. Axis x and y represents the nucleotide position and the observed genotype, respectively (as described for Fig. 19A). The colour scale intensity in all of the heatmap plots (including Fig. 19A) was magnified in order to visualise the genotypes that occur with low frequencies.
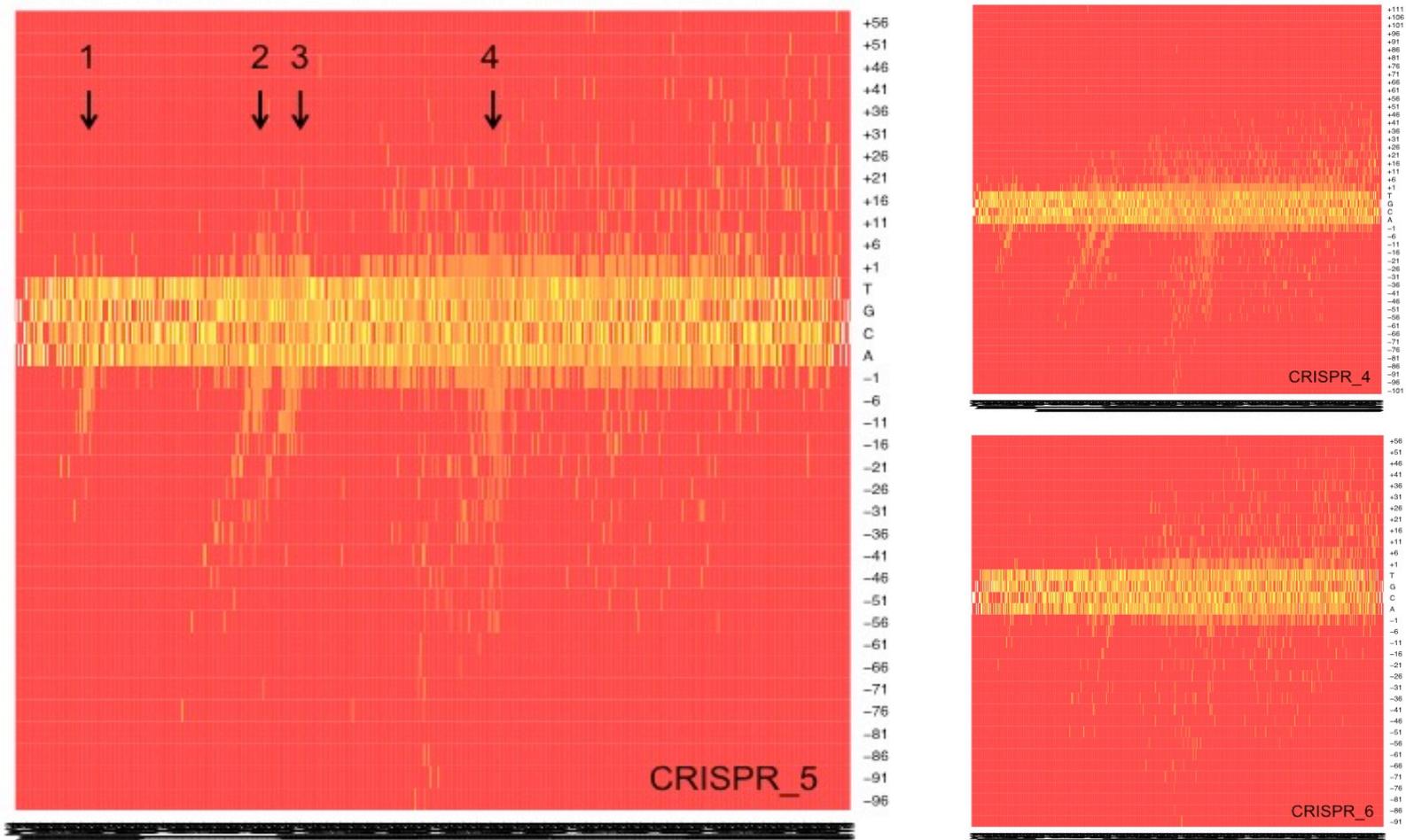
5.2. *Assay 2: Induced mutagenesis in adults*

5.2.1. Embryo microinjections and survival rate

I injected 915 mosquito embryos over 3 consecutive experimental days (D1, D2 and D3). The number of surviving larvae (L1), pupae and adults was recorded for the group of microinjected embryos from each day and is presented in Table 10.

Table 10. Survival in *C. quinquefasciatus* embryo after microinjections with CRISPR/Cas9 reagents. The table presents the number of mosquitoes that survived from embryo to the following distinct life cycle stages: larva (first instar, L1), pupa and adult. The number in brackets represents the percentage (%) of individuals surviving from the preceding to the focal life stage.

Day	Embryo	Larva (L1) (%)	Pupa (%)		Adult (%)	
			♀	♂	♀	♂
D1	204	4 (2)	0	1 (25)	0	1 (100)
D2	351	6 (1.7)	1 (16.7)	1 (16.7)	1 (100)	1 (100)
D3	360	7 (1.9)	1 (14.3)	2 (28.6)	1 (100)	2 (100)
TOTAL	915	17 (1.8)	2 (11.8)	4 (23.5)	2 (100)	4 (100)

Only 6 G$_0$ mosquitoes (2 females and 4 males) of 915 injected embryos emerged and were further outcrossed to wild-type individuals. Based on these outcomes, the survival rate from embryo to adulthood after injections with CRISPR/Cas9 components was 0.66%. The *C. quinquefasciatus* survival rate to adulthood was much lower than that obtained for *A. aegypti* mosquito (46.1 – 63.3%) in the CRISPR/Cas9 study, in which a similar concentration of Cas9 protein (333 ng/ul) and the same concentration of sgRNA (40 ng/µl) was used for the embryo microinjection solution (Kistler et al., 2015).

The higher survival rate to adulthood (~10%) after injections was also reported in the CRISPR/Cas9 study for *C. quinquefasciatus* species, with the use of mRNA for Cas9 and a concentration of sgRNA that was 3.75 times higher (150 ng/µl) than the one that I used for the purpose of my study (Itokawa et al., 2016). However, as reported by Itokawa et al. (2016) only about 13% of larvae survived to adulthood, compared to ~35% of the equivalent survival rate obtained from the study presented in this chapter.

The survival rate of the embryos after injection with CRISPR/Cas9 components was between 1.7 to 2% for the study presented in this chapter (Tab. 10), about 5 times lower than that obtained after microinjections with the *Hermes* transposon element used for transformation in the same mosquito species (Allen et al., 2001). The microinjection process is a complex, multi-step procedure that requires careful handling of the embryos, especially those that are very prone to desiccation, such as *C. quinquefasciatus* embryos. Thus, an overdesiccation of the embryos can be one of the factors affecting their survival rate after injections.

Another experimental factor contributing to the observed low survival rate in *C. quinquefasciatus* embryos after injections with CRISPR/Cas9 components could have been the toxic effect of the Cas9 protein. Although a higher concentration of the injected Cas9 protein typically results in an increase in mutagenic efficiency, it can, however, adversely affect survival rate of the embryo after injection. This phenomenon has been previously reported in CRISPR/Cas9 studies in *A. aegypti* (Kistler et al., 2015) and in mice and rats (Ménoret et al., 2015). Similar findings were reported by Bassett et al. (2014), based on the study in which mRNA for Cas9 was used instead of the Cas9 protein.

From the surviving 6 G_0 individuals, one outcross led to production of a single egg raft with a cohort of the viable G_1 larvae. 16 G_1 adults (8 males and 8 females) obtained from this larval cohort were screened for the presence of the putative flightless phenotype.

5.2.2. Flightless phenotype detection

The aim of the CRISPR/Cas9 design used in this study was the complete disruption of the *Cxq-A4* gene that could enable induction of the putative flightless phenotype in *C. quinquefasciatus* mosquitoes. 8 G_1 males and 8 G_1 females obtained from a single egg raft were screened for the presence of the putative flightless phenotype. 3 out of 8 females showed the flightless phenotype by failing to take off from the bottom of the cage in response to mechanical disturbance (gentle shakes and taps on the cage floor and the walls). A video showing this behaviour was recorded during my preliminary

136

behavioural testing and can be viewed here: https://vimeo.com/213975915 (password: CulexA4). All G_1 males were able to fly and were outcrossed to wild-type females.

Despite the addition of fresh virgin females to the experimental cages over time, none of the crosses with G_1 males and wild-type females led to the production of a G_2 cohort of mosquitoes. A variety of factors, also acting in combination, could have influenced the outcomes of the performed genetic crosses.

One of these factors was the observed over-engorgement behaviour of the wild-type females during a blood feeding that led to lethality of approximately 80% of the crossed females in each cage. I have recorded the over-engorgement behaviour during the standard rearing of Pel population for several generations in the past. The mortality rate, however, was never higher than 20% in one generation (data not shown).

Another explanation for the unsuccessful G_2 cohort generation could have been the quality of G_1 males. Upon replenishing cages with the fresh wild-type virgin females (7 days after setting up the initial crosses), the quality of the G_1 males and their mating performance could have been substantially reduced. As was shown for the *A. albopictus* mosquito, age can influence the sexual performance of males and have a negative influence on their potential reproductive fitness (Boyer et al., 2011).

Lastly, but of no less importance, the precopulatory reproductive behaviour of the *Culex* mosquito species involves the formation of a swarm. Male swarming seems imperative for mating and increases female insemination rate upon its formation (Gibson, 1985; Reisen et al., 1985). The

set-up of G_1 crosses - in which a single male was used - might have had a negative effect on the male mating behaviour and consequently led to a lack of, or substantial decrease in mating pair formation.

5.2.3. Sample preparation for HRMA and Sanger sequencing

I isolated DNA from the individual flightless G_1 females, as well as from the flying G_1 males, and measured its concentration (Tab. 11). DNA concentration ranged between 92.1 to 131.4 ng/µl and between 23.8 to 116.8 ng/µl for the female (CRISPR_fem_1-3) and the male (CRISPR_mal_1-8) samples, respectively. The higher DNA yield obtained from individual females was not surprising since female-biased sexual size dimorphism is observed in *C. quinquefasciatus* mosquitoes. The highest DNA concentration (116.8 ng/µl) was, however, obtained for a male sample ('CRISPR_mal_5'). This elevated value, in comparison to the other male samples, could have been an indicator of the sample contamination with a foreign DNA (e.g. bacterial and/or fungal load) after mosquito death, upon which a partial decomposition of its body was observed.

The DNA purity assessment, based on the 260/280 and 260/230 ratios, was indicative of 'pure' DNA in all of the experimental samples. Therefore, the extracted DNA was directly used in the downstream applications for mutation detection – HMRA and Sanger sequencing.

Table 11. gDNA concentration after isolation from experimental G$_1$ individuals. The table presents the concentration of the genomic DNA after isolation from adult G$_1$ mosquitoes (males 'mal' and females 'fem'). The concentrations were measured using the NanoDrop 2000 spectrophotometer together with 260/280 and 260/230 ratios for the sample purity assessment.

| Sample name	DNA concentration (ng/µl)	260/280	260/230
CRISPR_fem_1	123.4	2.09	2.34
CRISPR_fem_2	131.4	2.12	2.31
CRISPR_fem_3	92.1	2.06	2.21
CRISPR_mal_1	31.2	1.97	2.13
CRISPR_mal_2	37.1	2.07	2.27
CRISPR_mal_3	27.2	1.89	1.90
CRISPR_mal_4	54.4	1.99	2.27
CRISPR_mal_5	116.8	2.07	2.23
CRISPR_mal_6	44.0	1.96	2.21
CRISPR_mal_7	97.2	2.03	2.10
CRISPR_mal_8	23.8	1.90	1.80

5.2.4. HRMA-based variant detection and characterisation in adults

I used the DNA obtained from the individual wild-type and flightless females (G$_1$) to screen for the presence of mutations in HRMA. Similarly to

the HRM experiment approach for the samples obtained from the induced mutagenesis in embryos (Assay 1), I generated the aligned melt curve plots and the difference plots in order to review the HRM software calls for each sample. I normalised the difference plots against the control/wild-type sample and scored the samples as indicative of mutations if they fell outside the range of variation in fluorescence seen in the control samples (Fig. 20). 'Assay 5: 255_256' was the only assay out of 3 ('Assay 4: 203_254', 'Assay 5: 255_256', 'Assay 6: 257_258') performed during this HRM experiment that resulted in the generation of variants which were grouped into two apparent variant type clusters – one consisting purely of the variant calls for the wild-type samples (red) and another with the variant calls for the flightless female samples (blue) only. This clear difference in the melt curve shape between the variant clusters is strongly indicative of a putative mutation (Fig. 20B). The double peaks identified in the melt curves for the flightless female samples may suggest the presence of a larger mutation (insertion or deletion) induced in this target region of the *Cxq-A4.*

The number of variants detected for two other regions of the *Cxq-A4* was 4 for the 'Assay 4: 203_254' and 2 for the 'Assay 6: 257_258'. These variant calls might have indicated the presence of smaller changes (SNPs) that were induced in the mutagenesis assay, or that naturally occur in these regions in the Pel population. The largest number of variants within the 'sgRNA_1' region may indicate the highest putative cleaving efficiency of this sgRNA or reflect substantial polymorphism within this sequence region that describes Pel population, especially as similar findings were obtained for the HRM experiment with the pooled embryo samples (Assay 1).

Figure 20. HRMA-based variant detection in adults. The aligned melt curve

plot (A) and the difference plots (B) generated in the HRM experiment for

'Assay 5: 255_256'. Both plots present the combined melt curves for the wild

type and the flightless female samples. Plot B depicts variant calls that fall into

two variant clusters. The red and the blue curve correspond to the wild type

samples and the samples derived from the flightless females, respectively.
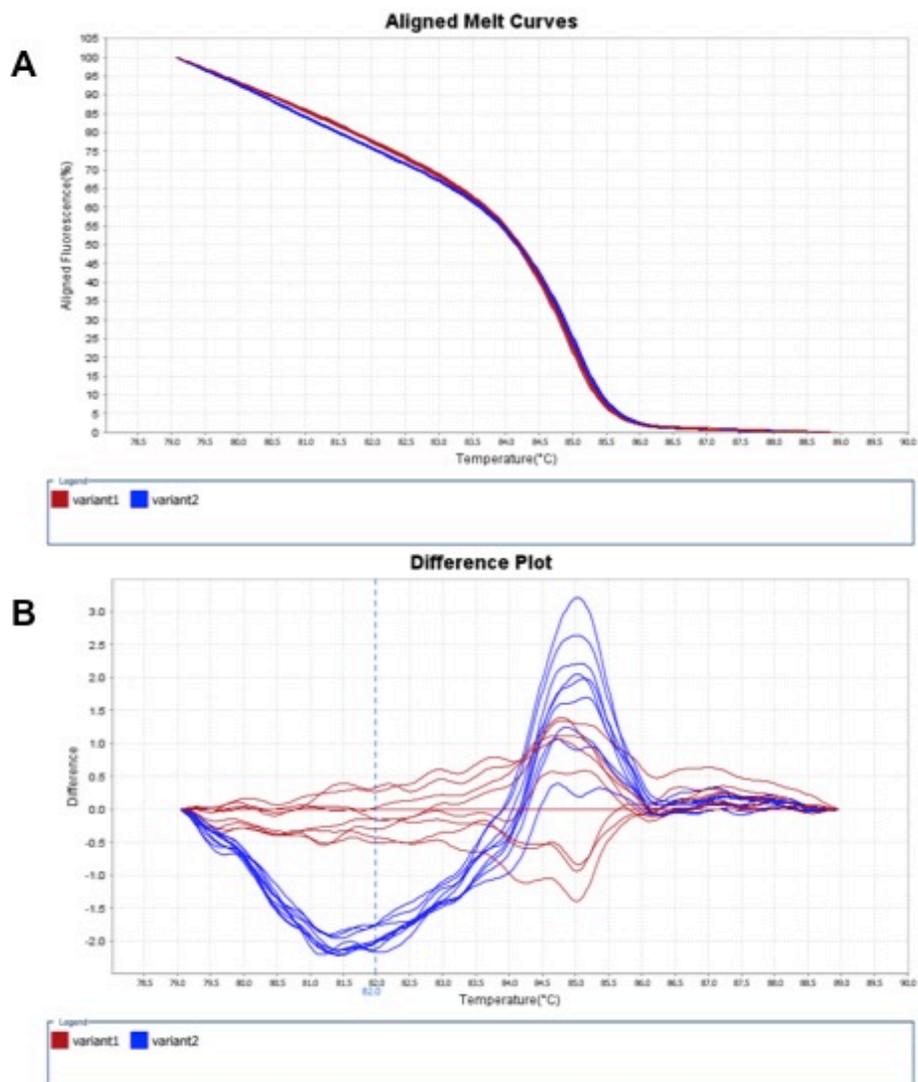
5.2.5. Variant detection through Sanger sequencing

Generation of the PCR product, incorporating four sgRNA target regions of the *Cxq-A4* gene, was successful for 8 out of 11 'CRISPR' samples with gDNA derived from G_1 individuals. Despite the efforts to optimise the PCR conditions by changing annealing temperatures or using a different set of primers, three male samples ('CRISPR_mal_1', 'CRISPR_mal_2' and 'CRISPR_mal_5') repeatedly failed to amplify and were discarded from further analysis. The male specimens were collected from a rearing cage up to 2 days after their death. This was, as observed upon sample collection, a sufficient time to enable a partial decomposition of the mosquito body, especially in those cases in which a dead mosquito specimen was found in the oviposition pot on the water's surface. Although gDNA extracted from all of the experimental samples did not show signs of degradation, a contamination of the original sample with a foreign DNA (bacterial and/or fungal load) could have been the reason for the observed amplification failure during PCR.

I performed sequencing reactions for 8 CRISPR samples containing a PCR product that was successfully generated with the LA388 and LA205 primers. Full coverage of the entire PCR product for both (forward and reverse) DNA strands enabled further sequence analysis of all of the samples. Comparison of the chromatograms generated from the bidirectional sequencing did not show any sequence discrepancies that could indicate possible inherent sequencing or PCR-associated errors.

The analysis, based on detection of mutations (SNPs, indels, inversions, duplications), involved the review of both sequences generated from each 'CRISPR' sample using the 4Peak sequence viewer (Nucleobytes, 2001-2017). The presence of the mismatch regions in the overlapping sequencing traces for G_1 individuals enabled identification of indels in the sgRNA target regions. I identified such mismatches in sequencing traces generated for all 3 flightless female samples ('CRISPR_fem_1', 'CRISPR_fem_2', 'CRISPR_fem_3') and one sample obtained from a flying male ('CRISPR_mal_8'). An example of these overlapping sequencing traces is presented in Figure 21.

Figure 21. Sanger sequencing-based indel detection in heterozygous G₁ individuals. An example of a sequencing trace with the overlapping sequences (a sequence string with 'double peaks') generated for the 'CRISPR_fem_3' sample of one of the flightless females (bottom). A chromatogram of the sequence for a wild-type individual is shown in the top panel. The rectangular selection indicates the location in the sequence traces where the overlap start was detected. The start of this overlap was in the same gene position for all of the flightless G_1 female samples and in one sample for a flying G_1 male.
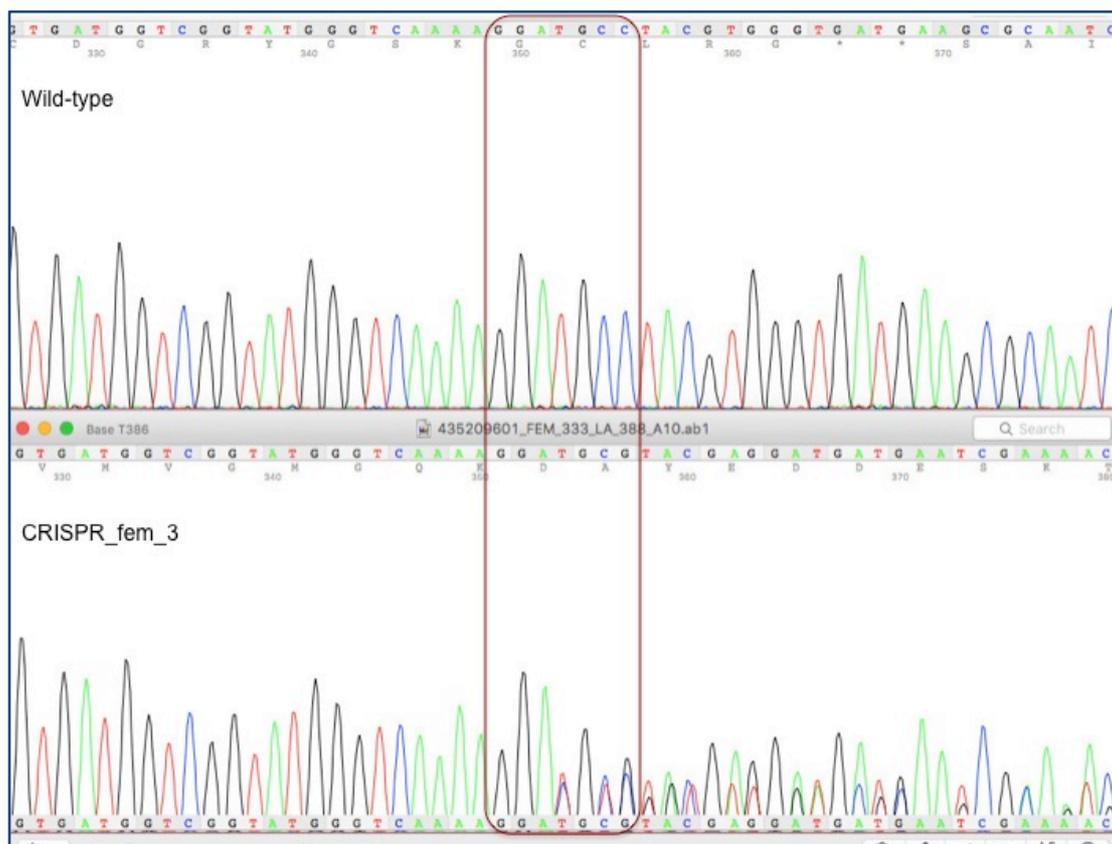
Mismatches detected in the overlapping regions that were determined with the use of the Poly Peak Parser tool (Fig. 21), were of the same type and

indicated the presence of a single 6-nucleaotide deletion (5′-TGCCTA-3′) in all of the G_1 individuals that were heterozygous mutant carriers. I identified this 6-nucleaotide deletion within the 'sgRNA_2' target region of the *Cxq-A4.* The 'sgRNA_2' was previously analysed *in silico* as the sgRNA with the second best cleaving efficiency, it was also ranked as the most efficient in the Cas9 digest-based *in vitro* experiment (Tab. 5, p. 115). Interestingly, the same 6-nucleotide deletion within the sgRNA_2 target region was also detected during an analysis of the results generated for the injected embryos after deep sequencing.

The modest sample size generated in the *in vivo* assay of the induced mutagenesis in adults gave a limited insight into the cleaving efficiency of the *Cxq-A4* sgRNAs. Nevertheless, the outcome of this assay is consistent with the combined results of the previous findings described in this chapter, which suggested that the 'sgRNA_2' performed as the most efficient *Cxq-A4* sgRNA.

Figure 21. Indel detection in the overlapping sequencing traces of the heterozygous G₁ individuals. An example of the alignment generated with the Poly Peak Parser tool. The sequence generated for the flightless female sample 'CRISPR_fem_1' with the forward LA388 primer was aligned to the reference sequence that was used in the sgRNA design step. The 'CRISPR_fem_1' chromatogram containing ambiguous base calls revealed a 6-nucleotide deletion upon separation of the overlapping traces.

Based on the nucleotide sequence of the native *Cxq-A4* gene, the identified 5′-TGCCTA-3′ deletion caused a replacement of 3 of the original amino acids in the translated Actin-4 protein: aspartic acid (GAT), alanine (GCC) and tyrosine (TAC), with a single amino acid – aspartic acid (GAC) (Fig. 22). This means that, upon deletion, the two amino acids (alanine and tyrosine) were missing from the native protein sequence after translation.

As shown experimentally, the identified deletion within the *Cxq-A4* open reading frame (ORF) is sufficient in causing the observed change in

phenotype in the heterozygous context, such that the females inheriting only one wild-type copy of the gene become flightless.

Figure 22. Deletion triggered changes in translated region of the *Cxq-A4*. The illustration presents changes in the native *C. quinquefasciatus* Actin-4 protein upon translation from the transcript, which incorporates the CRISPR/Cas9-induced mutation in the second exon of the *Cxq-A4*. The identified 6-nucleotide deletion (5′-TGCCTA-3′) in the second sgRNA target sequence (II) causes the replacement of the 3 original amino acids (represented by the grey circles): aspartic acid (Asp), alanine (Ala) and tyrosine (Tyr), with one alternatively coded Asp (shown with the pink circle). Four sgRNA target sites (I – IV) are depicted with the blue boxes.
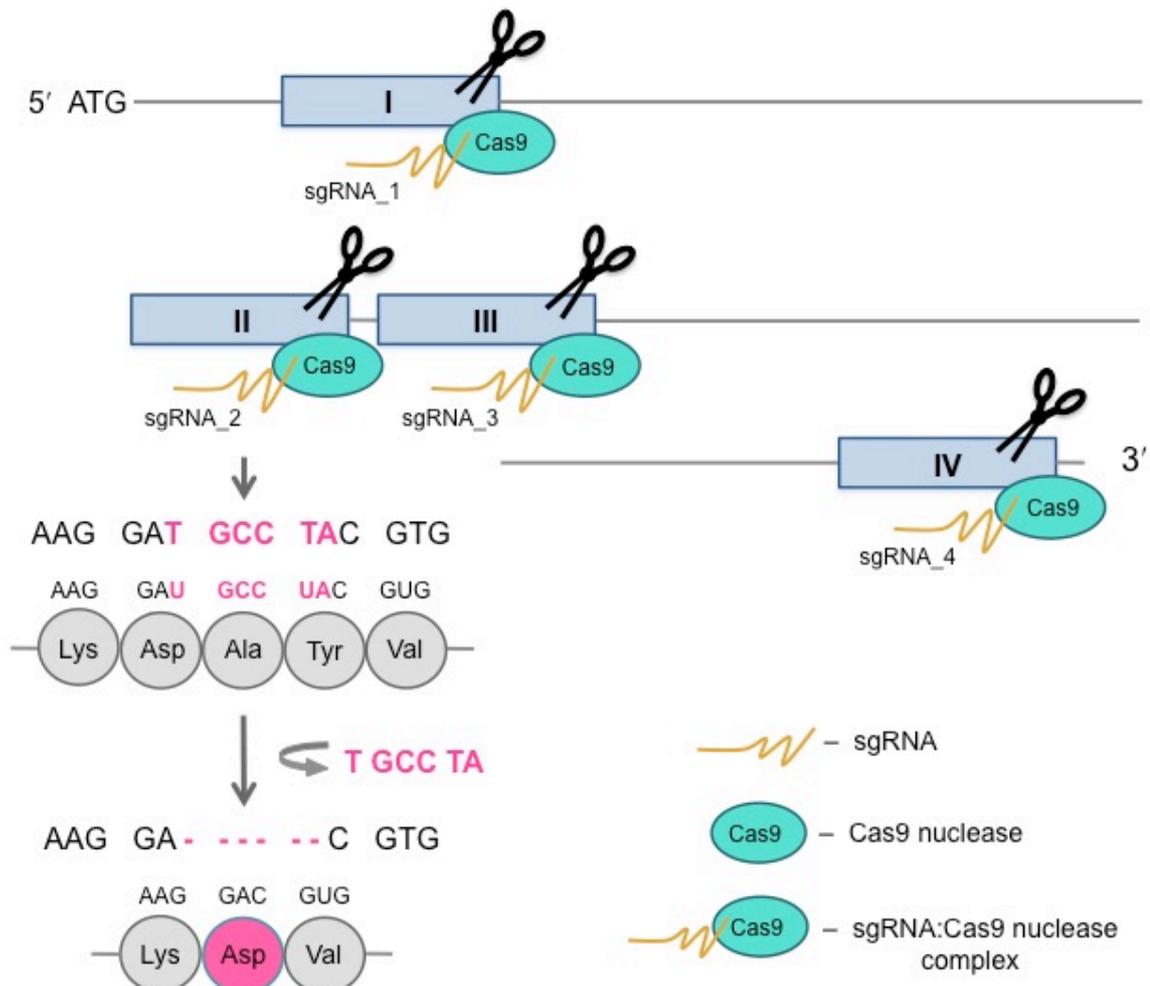
The artificially generated change in the coding sequence of the *Cxq-A4* may affect the Actin-4 protein expression and subcellular localization, its structure relevant to the protein folding and stability, protein-protein interactions, as well as its function, whereas the introduced mutation can result in destabilisation of the protein and complete loss of its specificity (Reva et al., 2011; Studer et al., 2013). In the case of loss of function, the level of the functional protein expressed from a single unchanged copy of the *Cxq-A4* may be insufficient for development of the functional indirect flight muscles in adult females. Thus, the putative loss-of-function mutation in the *Cxq-A4* gene may indicate gene haploinsufficiency, whereby a non-functional mutation in one allele of a gene leads to reduced production of the protein required for expression of the wild-type phenotype (Deutschbauer et al., 2005; Wilkie, 1994). The identified deletion may also cause a less extreme impact on the protein by compromising its function. The reduction in protein function can still, however, negatively affect development of the indirect flight muscles and consequently lead to flightlessness in females.

The gene haploinsufficiency is not, however, the only explanation for the observed change in female phenotype. The induced CRISPR/Cas9 deletion may have a dominant negative effect over the wild-type *Cxq-A4*. A dominant negative mutation antagonises activity of the wild-type allele by interfering with the simultaneously expressed wild-type protein and results in a non-functional phenotype. This phenomenon can be particularly apparent in multimeric proteins, such as actins, the activity of which depends on oligomerisation (Herskowitz, 1987). Numerous dominant negative mutant alleles have been identified and described for the *D. melanogaster* 88F actin

(*act88F)* that are expressed only in the IFM and impair flight ability without causing any adverse effect on the fly viability (An & Mogami, 1996; Hiromi & Hotta, 1985). Interestingly, in several cases, these single nucleotide substitution-triggered mutations were also described as hypomorphic (haploinsufficient) upon complete rescue of the flight abilities with an extra copy of the wild-type *act88F* (An & Mogami, 1996). As reported by An & Morgan (1996), the dual classification of the *D. melanogaster act88F* mutations into hypomorphic or antimorphic (dominant negative) can also correspond to the character of the changes that can be found upon mutagenesis within the coding region of the *Cxq-A4* gene. Therefore, in order to resolve the character of the mutation identified within the *Cxq-A4* during my study, further investigation is required.

As was suggested by the authors of the studies based on the use of the *Actin-4* homologue for mosquito population control (G. Fu et al., 2010; Labbé et al., 2012; Wise de Valdez et al., 2011), the dramatic change in phenotype, caused by malfunction of the female IFM, can lead to a substantial reduction in survival of the flightless females in the wild. Therefore, understanding the character of mutations induced within the *Cxq-A4* gene would be of particular importance in the development of novel population control strategies based on the *C. quinquefasciatus* genetics.

Conclusions

The results obtained from the study of the 5′UTR of the *C. quinquefasciatus Actin-4* (*Cxq-A4*) show that this gene region exhibits sex-

specific RNA splicing. Furthermore, the study employing targeted mutagenesis using the CRISPR/Cas9 genome editing system resulted in the generation of either antimorphic (dominant negative) or amorphic mutation. This loss-of-function mutation can be indicative of haploinsufficiency of the *Cxq-A4* gene. Additionally, all four sgRNA designed for the purpose of the targeted mutagenesis and tested in this study are active both *in vitro* and *in vivo*. The properties of the *Cxq-A4* that were identified and characterised in this chapter can be utilised as genetic tools for the purpose of future improvement or development of novel strategies for *C. quinquefasciatus* vector control.

Chapter 4: General Discussion

Many mosquito species are vectors of diseases that have a significant negative impact on human and animal health and well-being. Given the dramatic impact of vector-born diseases, interventions that enable the effective control of mosquito populations are strongly required today, and more so in the future given the on-going global spread of many invasive species of mosquitoes. Vector control can lead to the suppression or eradication of mosquito populations and consequently result in a decrease in disease transmission. Such control is essential for reducing the transmission of diseases for which an effective cure or preventive measures are not currently available. Vector control based on a genetics-based approach (e.g.

SIT, RIDL) has been previously developed and successfully implemented in the field, albeit only on a relatively small scale for mosquitoes thus far. In this thesis, I examined a novel genetic system design, as well as investigated the character and function of the genetic elements that could provide for future enhancement and developments of vector population control strategies.

In Chapter 2, I described work on a novel system design based on the late-acting lethality for the genetic control that can be engineered with the use of specific regulatory elements. I showed that the regulatory element (Os2 promoter) derived from the *A. aegypti Osiris* gene was able to induce a doxycycline-repressible, late-acting lethal phenotype when coupled to a tTAV/tetO effector system. Moreover, the lethality associated with transgene expression was predominately observed at the desired late developmental stage, enabling density-dependent competition between larvae. This design can strengthen the outcomes of genetic control by reducing the number of surviving wild-type mosquitoes in the field. However, the high mortality rate associated with the presence of two copies of the transgene in the homozygous OX5055 lines indicates the need for further assessment of this RIDL system design. Such assessment is crucial, since any downstream field use of transgenic mosquitoes of this type would require homozygous lines. Further evaluation of the doxycycline dose response of the transgenic OX5055 lines could lead to successful suppression of the lethal effect that is observed for the generated homozygous insertion lines. Also, individuals that are more resistant to the fitness costs associated with the transgene when reared on doxycycline can be selected in order to establish viable homozygous lines.

The Osiris design offers strong potential for the development of a novel approach to the RIDL method; therefore further research that can offer improvement to this design is strongly advisable. The reduction in toxicity of the transgene seems to be an apparent objective for future research. This can be achieved in several ways, such as the incorporation of insulators (Gaszner & Felsenfeld, 2006) into the original construct, the identification and characterisation of a weaker *Osiris* promoter that could replace Os2, or the use of a weaker effector in the transgenic construct.

The choice of the appropriate site for the insert integration – preferably outside the known transcription units – can lead to the elimination of the potential disadvantages that are incurred as the result of the position effect, or a reduction in the magnitude of this effect on the transcription of the transgene. One method that can help to achieve this is site-specific integration mediated by CRISPR/Cas9 and homology-directed DNA repair. This method has previously been successfully used for a precise knock-in of the exogenous gene cassette of interest into the zebrafish, *D. melanogaster* and *A. aegypti* genome (Gantz & Bier, 2015; Kawahara et al., 2016; Kistler et al., 2015; Lee et al., 2015).

The prospect of inducing lethality in a more stringent and precise manner during pupal development is also desirable. This can lead to a reduction in the observed variation for the stage at which lethality occurs. The availability of the complete developmental transcriptome for *A. aegypti* (Akbari, Antoshechkin, et al., 2013) can enable a more efficient search for genes that exhibit a strong expression at the specific developmental stage. This can advance the design process and facilitate future development of

transgenesis-based methods allowing for the engineering of the stage-specific expression of a lethal gene.

The *Os2* promoter has proven its ability in driving repressible, late-acting lethality, not only in the study presented in this chapter, but also in the previous studies conducted for *A. aegypti* and *A. albopictus* mosquitoes (Conway, 2014). This may not seem surprising, given that the *Os2* is derived from a well-conserved, insect-specific orthologous gene group (Shah et al., 2012). The *A. aegypti Osiris* gene, from which the *Os2* derives, also has its orthologue in other mosquito species, such as *C. quinquefasciatus* (*Osiris 11*) or the malaria vector *Anopheles gambiae* Giles (*Osiris 11*) (Shah et al., 2012). This suggests that the *Osiris* regulatory elements could be successfully employed in developing RIDL lines for these two vector mosquitoes. Additionally, evaluation of the transcription initiation sites from other genes belonging to this family may reveal more robust molecular tools for the future application of *Osiris* gene-based design in developing a more effective and versatile RIDL system approach for other pest insect species targeted in population control.

In Chapter 3, I showed that the 5′UTR region of the *C. quinquefasciatus Actin-4* (*Cxq-A4*) is alternatively spliced in females and males. This is in agreement with the findings from the previous studies, which reported the presence of the sex-specific RNA splicing of the *Actin-4* homologues in the *A. aegypti* (G. Fu et al., 2010; Muñoz et al., 2004) and *A. albopictus* (Labbé et al., 2012) mosquitoes. As was proposed in those studies, the sex-specific splicing of the 5′UTR of the *Actin-4* can be exploited in order to enable the sex-specific regulation of an engineered transgene expression.

This idea was previously employed in a genetic design for mosquito population control, e.g. female-specific RIDL (G. Fu et al., 2007; G. Fu et al., 2010; Labbé et al., 2012; Wise de Valdez et al., 2011) and can be also applied in the development of a similar vector control strategy for *C. quinquefasciatus*.

In addition, and similarly to the *Actin-4* homologues of the aforementioned mosquito species, the *Cxq-A4* shows expression at the pupal stage. However, further gene expression study is required in order to obtain an insight into the expression pattern and intensity of this gene during mosquito development. This is of particular importance if the purpose of a genetic design for vector control is the induction of the late-acting lethality in transgenic mosquitoes. Such a design, based on the use of a native promoter specificity for the regulation of an engineered transgene expression, has been previously developed for the RIDL system and led to the production of transgenic lines of mosquitoes carrying a late-acting, dominant-lethal gene (G. Fu et al., 2010; Labbé et al., 2012). The genetically engineered late-acting lethality, induced preferably at the late larval or pupal stage, is an important feature of a vector control design as far as the density-dependent competition among mosquito larvae is concerned (Phuc et al., 2007).

Presented in this thesis, the study of the *Cxq-A4* 5′UTR region revealed the existence of the start and stop codons in the male transcript. These codons are located upstream of the open reading frame (ORF) of the *Cxq-A4* and can lead to a reduction in the production of the functional Actin-4 protein (G. Fu et al., 2010). Furthermore, the identified and characterised knock-out mutation in the *Cxq-A4,* that was generated with the use of the

CRISPR/Cas9 genome editing system, resulted in the induction of the flightless phenotype only in the *C. quinquefasciatus* females. Taken together, these findings suggest that the female's ability to fly can depend on the production of the functional Actin-4 not only in the *C. quinquefasciatus*, but also in the two aforementioned species of mosquitoes - *A. aegypti* and *A. albopictus*, and that Actin-4 is an essential protein in development of the functional indirect flight muscles (IFMs) in the adult female mosquito. The role of actin (Actin III) in developing the IFMs has been previously shown in *D. melanogaster* through analysis of the mutants generated for the *act88F* that is expressed in the fruit fly's IFMs (An & Mogami, 1996; Hiromi & Hotta, 1985).

Additionally, further mutagenesis studies for the *C. quinquefasciatus* mosquito *Actin-4* can help with revealing the functional character of the *Cxq-A4* based on the relationship between its alleles. Since the mutation that was induced in *Cxq-A4* during my study is either antimorphic or amorphic in driving the observed flightless phenotype in the *C. quinquefasciatus* females, the study performed on the loss-of-function mutants can enable identification of the functional character of this gene (An & Mogami, 1996). This can be achieved by incorporation of the functional (wild-type) copy of the *Cxq-A4* into the genome of an *Actin-4* knock-out strain and can be mediated by a CRISPR/Cas9 and homology-directed DNA repair pathway.

The knowledge of the functional character of *Actin-4* based on the relationship between its alleles can be utilised in the development of a gene drive-based strategy for control of the vector mosquitoes, such as *C. quinquefasciatus*. A design based on exploiting haploinsufficiency of a cytoplasmic ribosomal gene in *D. melanogaster* has been previously

proposed as a component of the engineered underdominance gene drive system that can be used as a confinable and reversible population transformation system, enabling the introduction of refractory genes into wild populations (Akbari, Matzen, et al., 2013; Reeves et al., 2014). Additionally, it has been shown in mosquitoes that the homing drive based on the CRISPR/Cas9 system can provide very high homing frequencies. These frequencies enable the effective spread of the drive through a target population in a very short generation time and therefore serve as a potent gene drive system for vector population control (Gantz et al., 2015; Hammond et al., 2016). Although a CRISPR/Cas9 gene drive strategy for control of the *C. quinquefasciatus* mosquito is within reach of the scientific community, its implementation in the field will undoubtedly require careful consideration.

The CRISPR/Cas9 system has proven to be successful in inducing mutations in the *Cxq-A4*, both in *in vitro* as well as in *in vivo* assays, as presented in this thesis. Additionally, the *in vitro* editing efficiencies of the sequence-specific sgRNAs generated and examined in my studies strongly correspond to those predicted *in silico*. These results contribute to the growing number of findings that report the effective use of the CRISPR/Cas9-mediated system in successful targeted genome editing across a wide range of species.

The biological function of the sex-specifically spliced *Actin-4* in mosquitoes (including *Cxq-A4*) is still not known, i.e. why mosquitoes have sex-specific expression of IFM actin whereas Drosophila melanogaster, for example, does not. In order to reveal its function, further study on the *Cxq-A4* and its homologues identified in other species of mosquitoes is required. It has been previously reported that male mosquitoes detect and locate the

female's flight tone and that the female mosquito responds to male's flight tones (Gibson et al., 2010). Therefore, the observed sex-specific expression of the *Actin-4* in mosquitoes IFMs not only suggests the difference in flight muscle development between males and females but also indicates the role of selection in the evolution of the observed sexual dimorphism for expression of this gene. This suggests that the putative difference in developed male and female flight muscles might play an important role in mosquito sex recognition.

Overall, the results presented in this thesis offer new design approaches to the development of novel vector control strategies based on genetically engineered late-acting lethality. Additionally, I contribute to the existing knowledge on the character and function of the insect's *actin* genes. Moreover, I propose directions for future research related to my studies. My findings provide novel genetic tools that can be employed in the development of various genetics-based strategies for mosquito population control. This is of particular importance, as the realisation of novel strategies that enable the control of increasingly invasive vector species can help with safeguarding human health and in protecting endangered species.

List of references

Akbari, O. S., Antoshechkin, I., Amrhein, H., Williams, B., Diloreto, R., Sandler, J., & Hay, B. A. (2013). The Developmental Transcriptome of the Mosquito Aedes aegypti, an Invasive Species and Major Arbovirus Vector. *G3: Genes|Genomes|Genetics, 3*(9), 1493-1509. doi:10.1534/g3.113.006742

Akbari, O. S., Matzen, Kelly D., Marshall, John M., Huang, H., Ward, Catherine M., & Hay, Bruce A. (2013). A Synthetic Gene Drive System for Local, Reversible Modification and Suppression of Insect Populations. *Current Biology, 23*(8), 671-677. doi:10.1016/j.cub.2013.02.059

Allen, M. L., O'Brochta, D. A., Atkinson, P. W., & Levesque, C. S. (2001). Stable, Germ-Line Transformation of Culex quinquefasciatus (Diptera: Culicidae). *Journal of Medical Entomology, 38*(5), 701-710. doi:10.1603/0022-2585-38.5.701

Alphey, L. (2002). Re-engineering the sterile insect technique. *Insect Biochemistry and Molecular Biology, 32*(10), 1243-1247.

Alphey, L. (2014). Genetic control of mosquitoes. *Annual Review of Entomology, 59*, 205-224. doi:10.1146/annurev-ento-011613-162002

Alphey, L. (2016). Can CRISPR-Cas9 gene drives curb malaria? *Nat Biotech, 34*(2), 149-150. doi:10.1038/nbt.3473

Alphey, L., & Andreasen, M. (2002). Dominant lethality and insect population control. *Molecular and Biochemical Parasitology, 121*(2), 173-178.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology, 215*(3), 403-410.

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*(25), 3389-3402.

Amraoui, F., Atyame-Nten, C., Vega-Rúa, A., Lourenço-de-Oliveira, R., Vazeille, M., & Failloux, A. B. (2016). Culex mosquitoes are experimentally unable to transmit Zika virus. *Eurosurveillance, 21*(35), 30333. doi:10.2807/1560-7917.ES.2016.21.35.30333

An, H., & Mogami, K. (1996). Isolation of 88F Actin Mutants ofDrosophila melanogasterand Possible Alterations in the Mutant Actin Structures. *Journal of Molecular Biology, 260*(4), 492-505.

Applied Biosystems. (2009). A Guide to High Resolution Melting (HRM) Analysis. 1-20.

Asman, S. M., McDonald, P. T., & Prout, T. (1981). Field studies of genetic control systems for mosquitoes. *Annu. Rev. Entomol., 26*, 289–318.

Atkinson, C. T., & LaPointe, D. A. (2009). Introduced Avian Diseases, Climate Change, and the Future of Hawaiian Honeycreepers. *Journal of Avian Medicine and Surgery, 23*(1), 53-63.

Atkinson, M. P., Su, Z., Alphey, N., Alphey, L. S., Coleman, P. G., & Wein, L. M. (2007). Analyzing the Control of Mosquito-Borne Diseases by a

Dominant Lethal Genetic System. *Proceedings of the National Academy of Sciences of the United States of America, 104*(22), 9540-9545.

Babu, S., & Ramachandra, N. B. (2007). Screen for new mutations on the 2nd chromosome involved in indirect flight muscle development in Drosophila melanogaster. *Genome, 50*(4), 343-350. doi:10.1139/G07-012

Bachmair, A., Finley, D., & Varshavsky, A. (1986). In vivo half-life of a protein is a function of its amino-terminal residue. *Science, 234*(4773), 179-186. doi:10.1126/science.3018930

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D. A., & Horvath, P. (2007). CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes. *Science, 315*(5819), 1709-1712. doi:10.1126/science.1138140

Bassett, Andrew R., Tibbit, C., Ponting, Chris P., & Liu, J.-L. (2014). Highly Efficient Targeted Mutagenesis of Drosophila with the CRISPR/Cas9 System. *Cell Reports, 6*(6), 1178-1179.

Basu, S., Aryan, A., Overcash, J. M., Samuel, G. H., Anderson, M. A. E., Dahlem, T. J., Myles, K. M., & Adelman, Z. N. (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*(13), 4038-4043. doi:10.1073/pnas.1502370112

Benedict, M. Q., & Robinson, A. S. (2003). The first releases of transgenic mosquitoes: an argument for the sterile insect technique. *Trends in Parasitology, 19*(8), 349 - 355.

Bönisch, C., Temme, C., Moritz, B., & Wahle, E. (2007). Degradation of hsp70 and Other mRNAs in Drosophila via the 5′–3′ Pathway and Its Regulation by Heat Shock. *Journal of Biological Chemistry, 282*(30), 21818-21828. doi:10.1074/jbc.M702998200

Boyer, S., Gilles, J., Merancienne, D., Lemperiere, G., & Fontenille, D. (2011). Sexual performance of male mosquito Aedes albopictus. *Medical and Veterinary Entomology, 25*(4), 454-459. doi:10.1111/j.1365-2915.2011.00962.x

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

Carvalho, D. O., McKemey, A. R., Garziera, L., Lacroix, R., Donnelly, C. A., Alphey, L., Malavasi, A., & Capurro, M. L. (2015). Suppression of a Field Population of <italic>Aedes aegypti</italic> in Brazil by Sustained Release of Transgenic Male Mosquitoes. *PLoS Negl Trop Dis, 9*(7), e0003864. doi:10.1371/journal.pntd.0003864

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

Chevalier, B. S., & Stoddard, B. L. (2001). Homing endonucleases: structural and functional insight into the catalysts of intron/intein mobility. *Nucleic Acids Research, 29*(18), 3757-3774.

Chomczynski, P., & Sacchi, N. (1987). Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Analytical Biochemistry, 162*(1), 156-159.

Christophers, S. R. (1960). *Aedes aegypti: the yellow fever mosquito*: CUP Archive.

Coates, C. J., Jasinskiene, N., Pott, G. B., & James, A. A. (1999). Promoter-directed expression of recombinant fire-fly luciferase in the salivary glands of Hermes-transformed Aedes aegypti. *Gene, 226*(2), 317-325.

Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R.et al. (2013). Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science, 339*(6121), 819-823. doi:10.1126/science.1231143

Conway, M. (2014). *Improving transgenic approaches to mosquito population control.* (DPhil), University of Oxford.

Cooper, G. M. (2000). Structure and Organization of Actin Filaments. *The Cell: A Molecular Approach.* (2nd ed.). Sunderland (MA): Sinauer Associates.

Crampton, J. M., Beard, C. B., & Louis, C. (1997). *The molecular biology of insect disease vectors : a methods manual* (J.M. Crampton, C.B. Beard, & C. Louis Eds. 1st ed.). London ; New York : Chapman and Hall.

Curtis, Z., Matzen, K., Neira Oviedo, M., Nimmo, D., Gray, P.et al. (2015). Assessment of the Impact of Potential Tetracycline Exposure on the Phenotype of Aedes aegypti OX513A: Implications for Field Use. *PLOS Neglected Tropical Diseases, 9*(8), e0003999. doi:10.1371/journal.pntd.0003999

Davis, N. C. (1935). An Investigation of Possible Vectors of Wuchereria bancrofti (Cobbold) in Bahia, Brazil. *The Journal of Parasitology, 21*(1), 21-26. doi:10.2307/3271791

Deltcheva, E., Chylinski, K., Sharma, C. M., Gonzales, K., Chao, Y., Pirzada, Z. A., Eckert, M. R., Vogel, J., & Charpentier, E. (2011). CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature, 471*(7340), 602-607.

Deutschbauer, A. M., Jaramillo, D. F., Proctor, M., Kumm, J., Hillenmeyer, M. E., Davis, R. W., Nislow, C., & Giaever, G. (2005). Mechanisms of Haploinsufficiency Revealed by Genome-Wide Profiling in Yeast. *Genetics, 169*(4), 1915.

Diagne, C. T., Diallo, D., Faye, O., Ba, Y., Faye, O.et al. (2015). Potential of selected Senegalese Aedes spp. mosquitoes (Diptera: Culicidae) to transmit Zika virus. *BMC Infectious Diseases, 15*, 492. doi:10.1186/s12879-015-1231-2

DiCarlo, J. E., Chavez, A., Dietz, S. L., Esvelt, K. M., & Church, G. M. (2015). Safeguarding CRISPR-Cas9 gene drives in yeast. *Nat Biotech, 33*(12), 1250-1255. doi:10.1038/nbt.3412

Doench, J. G., Hartenian, E., Graham, D. B., Tothova, Z., Hegde, M.et al. (2014). Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nat Biotech, 32*(12), 1262-1267. doi:10.1038/nbt.3026

Dong, S., Lin, J., Held, N. L., Clem, R. J., Passarelli, A. L., & Franz, A. W. E. (2015). Heritable CRISPR/Cas9-Mediated Genome Editing in the

Yellow Fever Mosquito, Aedes aegypti. *PLOS ONE, 10*(3), e0122353. doi:10.1371/journal.pone.0122353

Dorer, D. R., Rudnick, J. A., Moriyama, E. N., & Christensen, A. C. (2003). A Family of Genes Clustered at the Triplo-lethal Locus of Drosophila melanogaster Has an Unusual Evolutionary History and Significant Synteny With Anopheles gambiae. *Genetics, 165*(2), 613-621.

Dyck, V. A., Hendrichs, J., Robinson, A. S., & Eds. (2005). *Sterile Insect Technique: Principles and Practice in Area-Wide Integrated Pest Management.* Dordrecht, The Netherlands: Springer.

Dye, C. (1982). Intraspecific competition amongst larval *Aedes aegypti*: food exploitation or chemical interference? *Ecological Entomology, 7*, 39–46.

Dye, C. (1984a). Competition amongst larval *Aedes aegypti*: the role of interference. *Ecological Entomology, 9*, 355–357.

Dye, C. (1984b). Models for the population dynamics of the yellow fever mosquito, *Ae. aegypti. . Journal of Animal Ecology, 3*, 247–268.

Eastwood, G., Kramer, L. D., Goodman, S. J., & Cunningham, A. A. (2011). West Nile Virus Vector Competency of Culex quinquefasciatus Mosquitoes in the Galápagos Islands. *The American Journal of Tropical Medicine and Hygiene, 85*(3), 426-433. doi:10.4269/ajtmh.2011.10-0739

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

Facchinelli, L., Valerio, L., Ramsey, J. M., Gould, F., Walsh, R. K.et al. (2013). Field Cage Studies and Progressive Evaluation of Genetically-Engineered Mosquitoes. *PLOS Neglected Tropical Diseases, 7*(1), e2001. doi:10.1371/journal.pntd.0002001

Farajollahi, A., Fonseca, D. M., Kramer, L. D., & Marm Kilpatrick, A. (2011). "Bird biting" mosquitoes and human disease: a review of the role of Culex pipiens complex mosquitoes in epidemiology. *Infect Genet Evol, 11*(7), 1577-1585. doi:10.1016/j.meegid.2011.08.013

Ferguson, N. M., Rodríguez-Barraquer, I., Dorigatti, I., Mier-y-Teran-Romero, L., Laydon, D. J., & Cummings, D. A. T. (2016). Benefits and risks of the Sanofi-Pasteur dengue vaccine: Modeling optimal deployment. *Science, 353*(6303), 1033-1036. doi:10.1126/science.aaf9590

Fernandes, J., Bate, M., & Vijayraghavan, K. (1991). Development of the indirect flight muscles of Drosophila. *Development, 113*(1), 67-77.

Fortini, L. B., Vorsino, A. E., Amidon, F. A., Paxton, E. H., & Jacobi, J. D. (2015). Large-Scale Range Collapse of Hawaiian Forest Birds under Climate Change and the Need 21st Century Conservation Options. *PLoS ONE, 10*(10), e0140389. doi:10.1371/journal.pone.0140389

Fu, G., Condon, K. C., Epton, M. J., Gong, P., Jin, L., Condon, G. C., Morrison, N. I., Dafa'alla, T. H., & Alphey, L. (2007). Female-specific insect lethality engineered using alternative splicing. *Nat Biotechnol, 25*(3), 353-357. doi:10.1038/nbt1283

Fu, G., Lees, R. S., Nimmo, D., Aw, D., Jin, L.et al. (2010). Female-specific flightless phenotype for mosquito control. *Proc Natl Acad Sci U S A, 107*(10), 4550-4554. doi:10.1073/pnas.1000251107

Fu, Y., Sander, J. D., Reyon, D., Cascio, V. M., & Joung, J. K. (2014). Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat Biotech, 32*(3), 279-284. doi:10.1038/nbt.2808

Fyrberg, E. A., Fyrberg, C. C., Biggs, J. R., Saville, D., Beall, C. J., & Ketchum, A. (1998). Functional Nonequivalence of Drosophila Actin Isoforms. *Biochemical Genetics, 36*(7), 271-287. doi:10.1023/a:1018785127079

Gagnon, J. A., Valen, E., Thyme, S. B., Huang, P., Ahkmetova, L.et al. (2014). Efficient Mutagenesis by Cas9 Protein-Mediated Oligonucleotide Insertion and Large-Scale Assessment of Single-Guide RNAs. *PLOS ONE, 9*(5), e98186. doi:10.1371/journal.pone.0098186

Gantz, V. M., & Bier, E. (2015). The mutagenic chain reaction: A method for converting heterozygous to homozygous mutations. *Science, 348*(6233), 442-444. doi:10.1126/science.aaa5945

Gantz, V. M., Jasinskiene, N., Tatarenkova, O., Fazekas, A., Macias, V. M., Bier, E., & James, A. A. (2015). Highly efficient Cas9-mediated gene drive for population modification of the malaria vector mosquito Anopheles stephensi. *Proceedings of the National Academy of Sciences, 112*(49), E6736-E6743. doi:10.1073/pnas.1521077112

Gasiunas, G., Barrangou, R., Horvath, P., & Siksnys, V. (2012). Cas9–crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proceedings of the National Academy of Sciences, 109*(39), E2579–E2586. doi:10.1073/pnas.1208507109

Gaszner, M., & Felsenfeld, G. (2006). Insulators: exploiting transcriptional and epigenetic mechanisms. *Nat Rev Genet, 7*(9), 703-713.

Gibson, G. (1985). Swarming behaviour of the mosquito Culex pipiens quinquefasciatus: a quantitative analysis. *Physiological Entomology, 10*(3), 283-296. doi:10.1111/j.1365-3032.1985.tb00049.x

Gibson, G., Warren, B., & Russell, I. J. (2010). Humming in Tune: Sex and Species Recognition by Mosquitoes on the Wing. *JARO: Journal of the Association for Research in Otolaryngology, 11*(4), 527-540. doi:10.1007/s10162-010-0243-2

Giraldo-Calderón, G. I., Emrich, S. J., MacCallum, R. M., Maslen, G., Dialynas, E.et al. (2015). VectorBase: an updated bioinformatics resource for invertebrate vectors and other organisms related with human diseases. *Nucleic Acids Research, 43*(Database issue), D707-D713. doi:10.1093/nar/gku1117

Gong, P., Epton, M. J., Fu, G., Scaife, S., Hiscox, A.et al. (2005). A dominant lethal genetic system for autocidal control of the Mediterranean fruitfly. *Nat Biotech, 23*(4), 453-456.

Goodwin, S., McPherson, J. D., & McCombie, W. R. (2016). Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet, 17*(6), 333-351. doi:10.1038/nrg.2016.49

Gossen, M., Freundlieb, S., Bender, G., xfc, ller, G., Hillen, W., & Bujard, H. (1995). Transcriptional Activation by Tetracyclines in Mammalian Cells. *Science, 268*(5218), 1766-1769.

Graham, D. B., & Root, D. E. (2015). Resources for the design of CRISPR gene editing experiments. [http://chopchop.rc.fas.harvard.edu/]. *Genome Biology, 16*, 260. doi:10.1186/s13059-015-0823-x

Gratz, N. G. (1999). Emerging and resurging vector-borne diseases. *Annual Review of Entomology, 44*, 51-75.

Gu, J., & Lieber, M. R. (2008). Mechanistic flexibility as a conserved theme across 3 billion years of nonhomologous DNA end-joining. *Genes & Development, 22*(4), 411-415. doi:10.1101/gad.1646608

Guha, T. K., Wai, A., & Hausner, G. (2017). Programmable Genome Editing Tools and their Regulation for Efficient Genome Engineering. *Computational and Structural Biotechnology Journal, 15*, 146-160.

Guo, X. X., Li, C. X., Deng, Y. Q., Xing, D., Liu, Q. M.et al. (2016). Culex pipiens quinquefasciatus: a potential vector to transmit Zika virus. *Emerg Microbes Infect, 5*(9), e102. doi:10.1038/emi.2016.102

Hall, A. B., Basu, S., Jiang, X., Qi, Y., Timoshevskiy, V. A.et al. (2015). A male-determining factor in the mosquito Aedes aegypti. *Science, 348*(6240), 1268-1270. doi:10.1126/science.aaa2850

Hammond, A., Galizi, R., Kyrou, K., Simoni, A., Siniscalchi, C.et al. (2016). A CRISPR-Cas9 gene drive system targeting female reproduction in the malaria mosquito vector Anopheles gambiae. *Nat Biotech, 34*(1), 78-83. doi:10.1038/nbt.3439

Handler, A. M. (2002). Use of the piggyBac transposon for germ-line transformation of insects. *Insect Biochemistry and Molecular Biology, 32*(10), 1211-1220.

Harris, A. F., McKemey, A. R., Nimmo, D., Curtis, Z., Black, I.et al. (2012). Successful suppression of a field mosquito population by sustained release of engineered male mosquitoes. *Nature Biotechnology, 30*(9), 828-830. doi:10.1038/nbt.2350

Harris, A. F., Nimmo, D., McKemey, A. R., Kelly, N., Scaife, S., Donnelly, C. A., Beech, C., Petrie, W. D., & Alphey, L. (2011). Field performance of engineered male mosquitoes. *Nature Biotechnology, 29*(11), 1034-1037.

Hartberg, W. K. (1971). Observations on the mating behaviour of Aedes aegypti in nature. *Bulletin of the World Health Organization, 45*(6), 847-850.

Heinrich, J. C., & Scott, M. J. (2000). A repressible female-specific lethal genetic system for making transgenic insect strains suitable for a sterile-release program. *Proc Natl Acad Sci U S A, 97*(15), 8229-8232. doi:10.1073/pnas.140142697

Hendrichs, J., Franz, G., & Rendon, P. (1995). Increased effectiveness and applicability of the sterile insect technique through male-only releases for control of Mediterranean fruit flies during fruiting seasons. *Journal of Applied Entomology, 119*, 371–377.

Herskowitz, I. (1987). Functional inactivation of genes by dominant negative mutations. *Nature, 329*(6136), 219-222.

Hickey, W. A., & Craig, G. B. (1966). Genetic Distortion of Sex Ratio in a Mosquito, AEDES AEGYPTI. *Genetics, 53*(6), 1177-1196.

Hill, J. T., Demarest, B. L., Bisgrove, B. W., Su, Y.-c., Smith, M., & Yost, H. J. (2014). Poly Peak Parser: Method and software for identification of unknown indels using Sanger Sequencing of PCR products. *Developmental dynamics : an official publication of the American Association of Anatomists, 243*(12), 1632-1636. doi:10.1002/dvdy.24183

Hiromi, Y., & Hotta, Y. (1985). Actin gene mutations in Drosophila; heat shock activation in the indirect flight muscles. *The EMBO Journal, 4*(7), 1681-1687.

Hsu, P. D., Scott, D. A., Weinstein, J. A., Ran, F. A., Konermann, S.et al. (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotech, 31*(9), 827-832. doi:10.1038/nbt.2647

Hunkapiller, T., Kaiser, R. J., Koop, B. F., & Hood, L. (1991). Large-Scale and Automated DNA Sequence Determination. *Science, 254*(5028), 59-67.

Hwang, W. Y., Fu, Y., Reyon, D., Maeder, M. L., Tsai, S. Q., Sander, J. D., Peterson, R. T., Yeh, J. R. J., & Joung, J. K. (2013). Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nat Biotech, 31*(3), 227-229. doi:10.1038/nbt.2501

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

Itokawa, K., Komagata, O., Kasai, S., Ogawa, K., & Tomita, T. (2016). Testing the causality between CYP9M10 and pyrethroid resistance using the TALEN and CRISPR/Cas9 technologies. *Scientific Reports, 6*, 24652. doi:10.1038/srep24652

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science, 337*(6096), 816-821. doi:10.1126/science.1225829

Juliano, S. A. (2007). Population Dynamics. *Journal of the American Mosquito Control Association, 23*(2 Suppl), 265-275.

Karunaratne, S. H. P. P., Hemingway, J., Jayawardena, K. G. I., Dassanayaka , V., & Vaughan, A. (1995). Kinetic and Molecular Differences in the Amplified and Non-amplified Esterases from Insecticide-resistant and Susceptible Culex quinquefasciatus Mosquitoes. *Journal of Biological Chemistry, 270*(52), 31124-31128. doi:10.1074/jbc.270.52.31124

Kawahara, A., Hisano, Y., Ota, S., & Taimatsu, K. (2016). Site-Specific Integration of Exogenous Genes Using Genome Editing Technologies in Zebrafish. *International Journal of Molecular Sciences, 17*(5), 727. doi:10.3390/ijms17050727

Kistler, Kathryn E., Vosshall, Leslie B., & Matthews, Benjamin J. (2015). Genome Engineering with CRISPR-Cas9 in the Mosquito Aedes aegypti. *Cell Reports, 11*(1), 51-60.

Klassen, W. (2005). *Area-Wide Integrated Pest Management and the Sterile Insect Technique In Sterile Insect Technique: Principles and Practice in Area-Wide Integrated Pest Management.* (V. A. Dyck, J. Hendrichs, A. S. Robinson, & Eds. Eds.). Dordrecht, The Netherlands: Springer, pp. 39-68.

Kraemer, M. U. G., Sinka, M. E., Duda, K. A., Mylne, A. Q. N., Shearer, F. M.et al. (2015). The global distribution of the arbovirus vectors Aedes aegypti and Ae. albopictus. *eLife, 4*, e08347. doi:10.7554/eLife.08347

Labbé, G. M., Scaife, S., Morgan, S. A., Curtis, Z. H., & Alphey, L. (2012). Female-specific flightless (fsRIDL) phenotype for control of *Aedes*

albopictus. PLoS Neglected Tropical Diseases, 6(7), e1724. doi:10.1371/journal.pntd.0001724

Lawrence, P. A. (1982). Cell lineage of the thoracic muscles of drosophila. *Cell, 29*(2), 493-503.

Lee, J. S., Kallehauge, T. B., Pedersen, L. E., & Kildegaard, H. F. (2015). Site-specific integration in CHO cells mediated by CRISPR/Cas9 and homology-directed DNA repair pathway. *Scientific Reports, 5*, 8572. doi:10.1038/srep08572

Legros, M., Lloyd, A. L., Huang, Y., & Gould, F. (2009). Density-Dependent Intraspecific Competition in the Larval Stage ofAedes aegypti(Diptera: Culicidae): Revisiting the Current Paradigm. *Journal of Medical Entomology, 46*(3), 409-419. doi:10.1603/033.046.0301

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J.et al. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics, 25*(16), 2078-2079. doi:10.1093/bioinformatics/btp352

Liao, W., Atkinson, C. T., LaPointe, D. A., & Samuel, M. D. (2017). Mitigating Future Avian Malaria Threats to Hawaiian Forest Birds from Climate Change. *PLOS ONE, 12*(1), e0168880. doi:10.1371/journal.pone.0168880

Life Technologies. (2004). *5' RACE System for Rapid Amplification of cDNA Ends, Version 2.0*.

Life Technologies. (2010). *HRM Experiments Using MeltDoctorTM HRM Reagents and High Resolution Melt Software v3.0 User Guide*.

Lobo, N. F., Clayton, J. R., Fraser, M. J., Kafatos, F. C., & Collins, F. H. (2006). High efficiency germ-line transformation of mosquitoes. *Nat. Protocols, 1*(3), 1312-1317. doi:10.1038/nprot.2006.221

Lofgren, C. S., Dame, D. A., Breeland, S. G., Weidhaas, D. E., Jeffery, G., Kaiser, R., Ford, H. R., Boston, M. D., & Baldwin, K. F. (1974). Release of Chemosterilized Males for the Control of Anopheles Albimanus in El Salvador. *The American Journal of Tropical Medicine and Hygiene, 23*(2), 288-297.

Marco-Sola, S., Sammeth, M., Guigo, R., & Ribeca, P. (2012). The GEM mapper: fast, accurate and versatile alignment by filtration. *Nat Meth, 9*(12), 1185-1188.

McDonald, P. T., Hausermann, W., & Lorimer, N. (1977). Sterility introduced by release of genetically altered males to a domestic population of *A. aegypti* at the Kenya coast. *American Journal of Tropical Medicine and Hygiene, 26*, 553–561.

Ménoret, S., De Cian, A., Tesson, L., Remy, S., Usal, C.et al. (2015). Homology-directed repair in rodent zygotes using Cas9 and TALEN engineered proteins. *Scientific Reports, 5*, 14410. doi:10.1038/srep14410

Miller, A. (1950). The internal anatomy and histology of imago of *Drosophila melanoagster.* . In M. Demerec (Ed.), *Biology of Drosophila.* (pp. 424-442). New York: Wiley.

Mogami, K., Fujita, S. C., & Hotta, Y. (1982). Identification of Drosophila Indirect Flight Muscle Myofibrillar Proteins by Means of Two-Dimensional Electrophoresis 1. *The Journal of Biochemistry, 91*(2), 643-650. doi:10.1093/oxfordjournals.jbchem.a133736

Mogami, K., & Hotta, Y. (1981). Isolation of Drosophila flightless mutants which affect myofibrillar proteins of indirect flight muscle. *Molecular and General Genetics MGG, 183*(3), 409-417. doi:10.1007/bf00268758

Moreira, L. A., Edwards, M. J., Adhami, F., Jasinskiene, N., James, A. A., & Jacobs-Lorena, M. (2000). Robust Gut-Specific Gene Expression in Transgenic Aedes aegypti Mosquitoes. *Proceedings of the National Academy of Sciences of the United States of America, 97*(20), 10895-10898.

Moreno-Mateos, M. A., Vejnar, C. E., Beaudoin, J.-D., Fernandez, J. P., Mis, E. K., Khokha, M. K., & Giraldez, A. J. (2015). CRISPRscan: designing highly efficient sgRNAs for CRISPR-Cas9 targeting in vivo. *Nat Meth, 12*(10), 982-988. doi:10.1038/nmeth.3543

Msangi, A., Saleh, K., Kiwia, N., II, M., Mussa, W.et al. (2000). Success in Zanzibar: eradication of Tsetse. In K. Tan (Ed.), *Area-wide control of fruit flies and other insect pests.* (pp. 57-66). Penerbit Universiti Sains Malaysia, Penang.

Muñoz, D., Jimenez, A., Marinotti, O., & James, A. A. (2004). The AeAct-4 gene is expressed in the developing flight muscles of female Aedes aegypti. *Insect Molecular Biology, 13*(5), 563-568. doi:10.1111/j.0962-1075.2004.00519.x

Niu, Y., Shen, B., Cui, Y., Chen, Y., Wang, J.et al. Generation of Gene-Modified Cynomolgus Monkey via Cas9/RNA-Mediated Gene Targeting in One-Cell Embryos. *Cell, 156*(4), 836-843. doi:10.1016/j.cell.2014.01.027

Nongthomba, U., & Ramachandra, N. B. (1999). A Direct Screen Identifies New Flight Muscle Mutants on the Drosophila Second Chromosome. *Genetics, 153*(1), 261-274.

Nucleobytes, B. V. (2001-2017): Nucleobytes B.V.

Onozawa, M., Zhang, Z., Kim, Y. J., Goldberg, L., Varga, T., Bergsagel, P. L., Kuehl, W. M., & Aplan, P. D. (2014). Repair of DNA double-strand breaks by templated nucleotide sequence insertions derived from distant regions of the genome. *Proceedings of the National Academy of Sciences of the United States of America, 111*(21), 7729-7734. doi:10.1073/pnas.1321889111

Patterson, R. S., Weidhaas, D. E., Ford, H. R., & Lofgren, C. S. (1970). Suppression and Elimination of an Island Population of Culex pipiens quinquefasciatus with Sterile Males. *Science, 168*(3937), 1368-1370.

Phuc, H. K., Andreasen, M. H., Burton, R. S., Vass, C., Epton, M. J.et al. (2007). Late-acting dominant lethal genetic systems and mosquito control. *BMC Biol, 5*, 11. doi:10.1186/1741-7007-5-11

Pratt, T. K. (2009). *Conservation biology of Hawaiian forest birds : implications for island avifauna*. New Haven: Yale University Press.

Raineri, E., Ferretti, L., Esteve-Codina, A., Nevado, B., Heath, S., & Pérez-Enciso, M. (2012). SNP calling by sequencing pooled samples. *BMC Bioinformatics, 13*, 239-239. doi:10.1186/1471-2105-13-239

Reed, J. M., DesRochers, D. W., VanderWerf, E. A., & Scott, J. M. (2012). Long-Term Persistence of Hawaii's Endangered Avifauna through Conservation-Reliant Management. *BioScience, 62*(10), 881-892. doi:10.1525/bio.2012.62.10.8

Reeves, R. G., Bryk, J., Altrock, P. M., Denton, J. A., & Reed, F. A. (2014). First Steps towards Underdominant Genetic Transformation of Insect Populations. *PLOS ONE, 9*(5), e97557. doi:10.1371/journal.pone.0097557

Reisen, W. K., Knop, N. F., & Peloquin, J. J. (1985). Swarming and Mating Behavior of Laboratory and Field Strains of Culex tarsalis (Diptera: Culicidae). *Annals of the Entomological Society of America, 78*(5), 667-673. doi:10.1093/aesa/78.5.667

Reva, B., Antipin, Y., & Sander, C. (2011). Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic Acids Research, 39*(17), e118-e118. doi:10.1093/nar/gkr407

Richards, S. L., Anderson, S. L., & Lord, C. C. (2014). Vector competence of Culex pipiens quinquefasciatus (Diptera: Culicidae) for West Nile virus isolates from Florida. *Tropical medicine & international health : TM & IH, 19*(5), 610-617. doi:10.1111/tmi.12277

Robinson, A. S. (2005). *Genetic Basis of the Sterile Insect Technique In Sterile Insect Technique: Principles and Practice in Area-Wide Integrated Pest Management.* (V. A. Dyck, J. Hendrichs, A. S. Robinson, & Eds. Eds.). Dordrecht, The Netherlands: Springer, pp. 95-114.

Rodems, S. M., & Friesen, P. D. (1993). The hr5 transcriptional enhancer stimulates early expression from the Autographa californica nuclear polyhedrosis virus genome but is not required for virus replication. *Journal of Virology, 67*(10), 5776-5785.

Roeder, R. G. (1996). The role of general initiation factors in transcription by RNA polymerase II. *Trends in Biochemical Sciences, 21*(9), 327-335.

Rogers, D. J., Wilson, A. J., Hay, S. I., & Graham, A. J. (2006). The Global Distribution of Yellow Fever and Dengue. In A. G. Simon I. Hay & J. R. David (Eds.), *Advances in Parasitology* (Vol. Volume 62, pp. 181-220): Academic Press.

RStudio Team. (2015). RStudio: Integrated Development for R (Version 0.98.1102). Boston, MA RStudio, Inc. Retrieved from http://www.rstudio.com/

Samuel, M. D., Woodworth, B. L., Atkinson, C. T., Hart, P. J., & LaPointe, D. A. (2015). Avian malaria in Hawaiian forest birds: infection and population impacts across species and elevations. *Ecosphere, 6*(6), 1-21. doi:10.1890/ES14-00393.1

Sanger, F., & Coulson, A. R. (1975). A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *Journal of Molecular Biology, 94*(3), 441-448.

Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences, 74*(12), 5463-5467.

Sardelis, M. R., Turell, M. J., Dohm, D. J., & O'Guinn, M. L. (2001). Vector competence of selected North American Culex and Coquillettidia mosquitoes for West Nile virus. *Emerg Infect Dis, 7*(6), 1018-1022. doi:10.3201/eid0706.010617

Schaffner, F., & Mathis, A. (2014). Dengue and dengue vectors in the WHO European region: past, present, and scenarios for the future. *The*

Lancet Infectious Diseases, 14(12), 1271-1280. doi:10.1016/S1473-3099(14)70834-5

Shah, N., Dorer, D. R., Moriyama, E. N., & Christensen, A. C. (2012). Evolution of a Large, Conserved, and Syntenic Gene Family in Insects. *G3: Genes|Genomes|Genetics, 2*(2), 313-319. doi:10.1534/g3.111.001412

Simmons , C. P., Farrar , J. J., van Vinh Chau , N., & Wills , B. (2012). Dengue. *New England Journal of Medicine, 366*(15), 1423-1432. doi:10.1056/NEJMra1110265

Studer, Romain A., Dessailly, Benoit H., & Orengo, Christine A. (2013). Residue mutations and their impact on protein structure and function: detecting beneficial and pathogenic changes. *Biochemical Journal, 449*(3), 581.

Thomas, D. T., Donnelly, C. A., Wood, R. J., & Alphey, L. S. (2000). Insect Population Control Using a Dominant, Repressible, Lethal Genetic System. *Science, 287*(5462), 2474-2476. doi:10.1126/science.287.5462.2474

Triteeraprapab, S., Kanjanopas, K., Suwannadabba, S., Sangprakarn, S., Poovorawan, Y., & Scott, A. L. (2000). Transmission of the Nocturnal Periodic Strain of Wuchereria bancrofti by Culex quinquefasciatus: Establishing the Potential for Urban Filariasis in Thailand. *Epidemiology and Infection, 125*(1), 207-212.

van Riper, C., van Riper, S. G., Goff, M. L., & Laird, M. (1986). The Epizootiology and Ecological Significance of Malaria in Hawaiian Land Birds. *Ecological Monographs, 56*(4), 327-344. doi:10.2307/1942550

Vega-Rúa, A., Zouache, K., Girod, R., Failloux, A.-B., & Lourenço-de-Oliveira, R. (2014). High Level of Vector Competence of *Aedes aegypti* and *Aedes albopictus* from Ten American Countries as a Crucial Factor in the Spread of Chikungunya Virus. *Journal of Virology, 88*(11), 6294-6306. doi:10.1128/jvi.00370-14

Warner, R. E. (1968). The Role of Introduced Diseases in the Extinction of the Endemic Hawaiian Avifauna. *The Condor, 70*(2), 101-120.

Washburn, J. O. (1995). Regulatory factors affecting larval mosquito populations in container and pool habitats: implications for biological control. *Journal of The American Mosquito Control Association, 11*(2), 279–283.

WHO Technical Report Series. (2009). Dengue: Guidelines for Diagnosis, Treatment, Prevention and Control: New Edition. Geneva: World Health Organization. *Geneva: WHO*.

Wilkie, A. O. (1994). The molecular basis of genetic dominance. *Journal of Medical Genetics, 31*(2), 89.

Wise de Valdez, M. R., Nimmo, D., Betz, J., Gong, H.-F., James, A. A., Alphey, L., & Black, W. C. (2011). Genetic elimination of dengue vector mosquitoes. *Proceedings of the National Academy of Sciences, 108*(12), 4772-4775. doi:10.1073/pnas.1019295108

Wittkopp, P. J., & Kalay, G. (2012). Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. *Nat Rev Genet, 13*(1), 59-69.

Wyss, J. H. (2000). Screwworm Eradication in the Americas. *Annals of the New York Academy of Sciences, 916*(1), 186-193. doi:10.1111/j.1749-6632.2000.tb05289.x

Yakob, L., Alphey, L., & Bonsall, M. B. (2008). *Aedes aegypti* Control: The Concomitant Role of Competition, Space and Transgenic Technologies. *Journal of Applied Ecology, 45*, 1258-1265. doi:10.HH/j.1365-2664.2008.01498.x

Zhou, L., Jiang, G., Chan, G., Santos, C. P., Severson, D. W., & Xiao, L. (2005). Michelob_x is the missing inhibitor of apoptosis protein antagonist in mosquito genomes. *EMBO Reports, 6*(8), 769-774. doi:10.1038/sj.embor.7400473

Appendices

Appendix 1.S1 – S3. Mortality in OX5055 lines. Barplots of average proportions for 'death' categories as well as functional flying individuals ('flying adults') scored during the lethality (Appx. 1.S1 and Appx. 1.S2) and the homozygous viability assay (Appx. 1.S3) for transgenic and wild-type mosquitoes from four OX5055 lines (D, F, G and I) after being reared in the absence ('OFF DOX', Appx. 1.S1) or in the presence of doxycycline ('ON DOX', Appx. 1.S1, Appx. 1.S3). The 'death' categories includes dead pupae, 'partially eclosed' individuals failing to emerge from the pupal carapace, adults dead on the water surface ('dead imago water'), and adults dead on the cage floor ('dead imago floor'). Error bars show standard error of the mean (SEM) of three replicates, each comprising a cohort of 200 individuals.
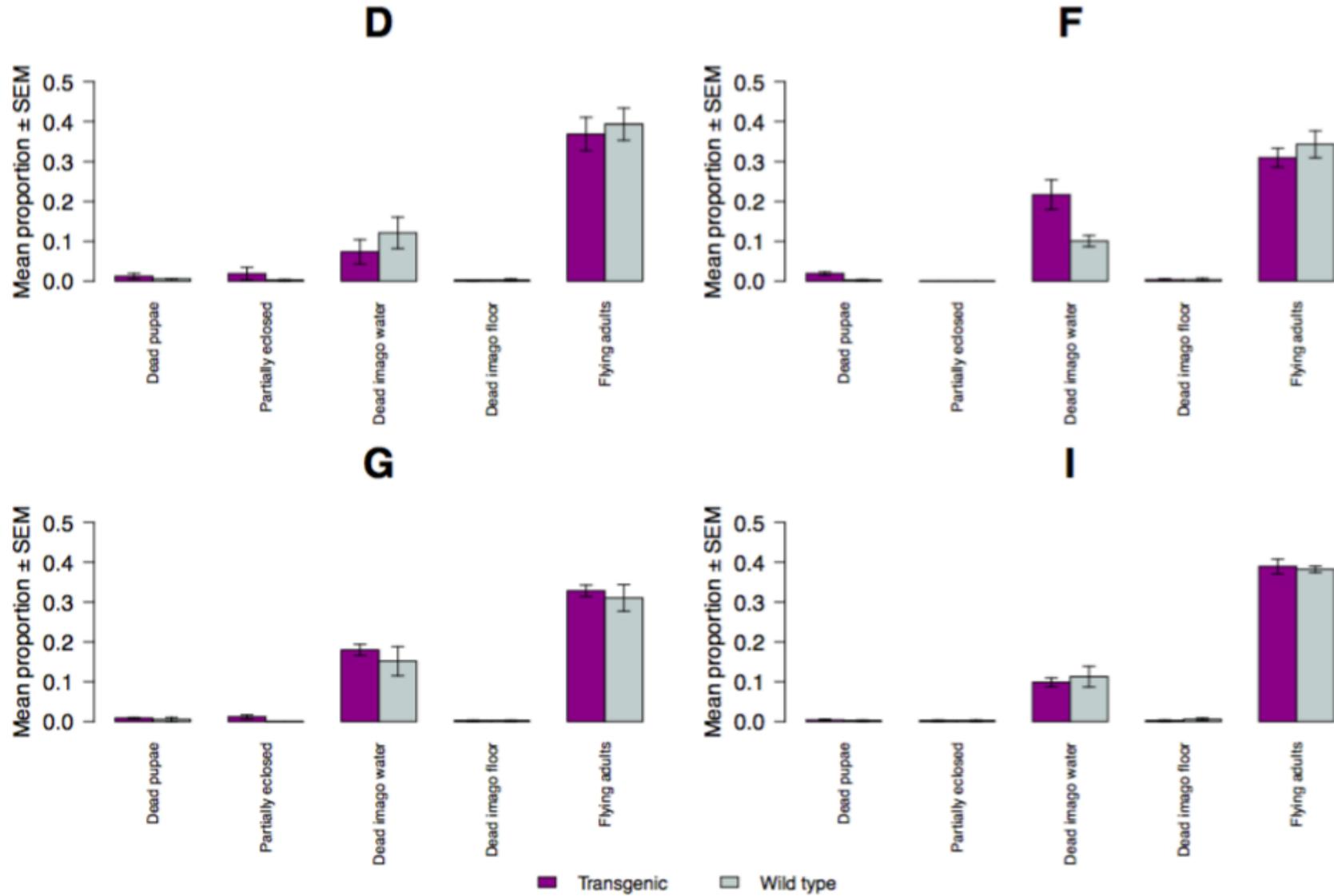
Mortality in OX5055 lines: OFF DOX
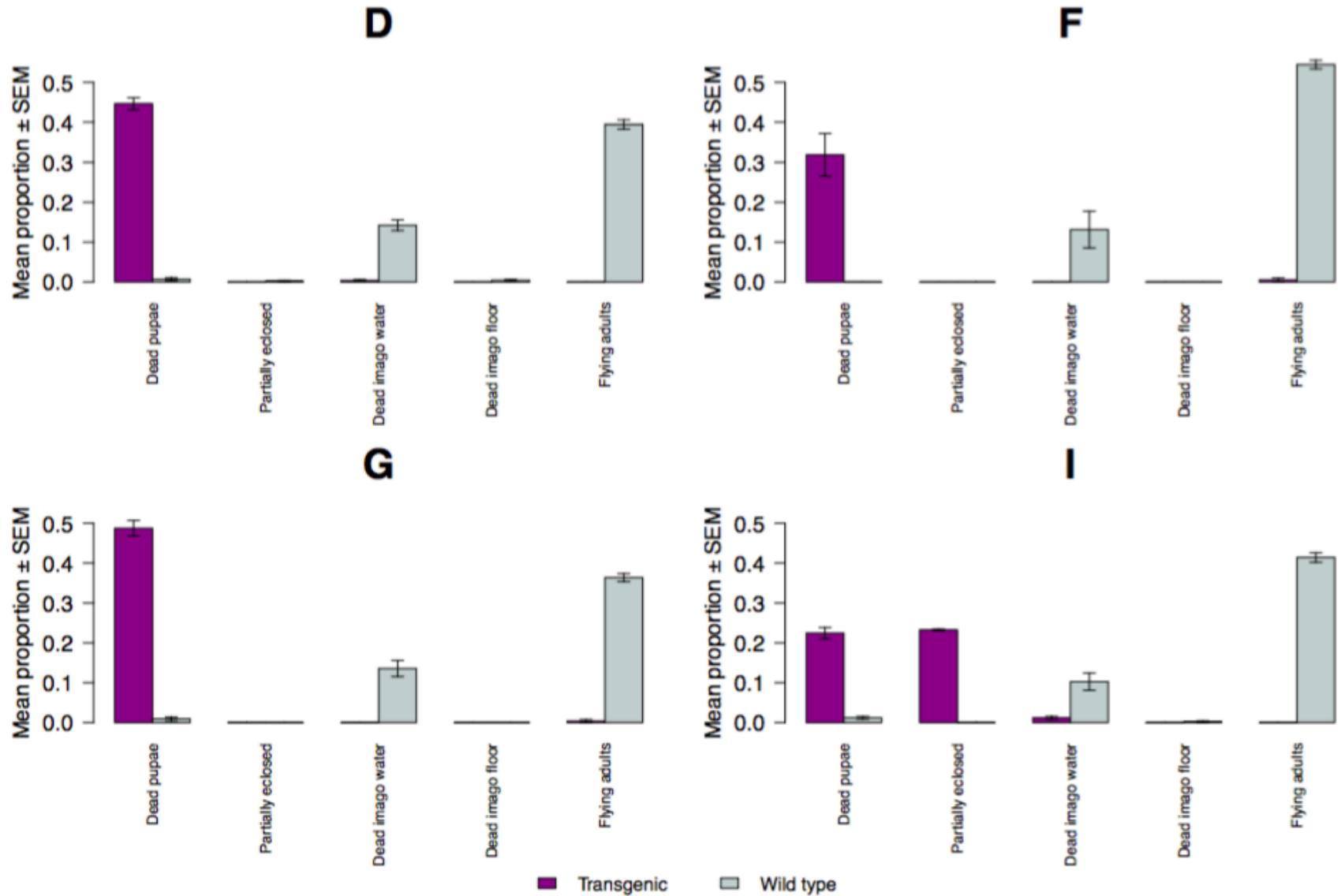
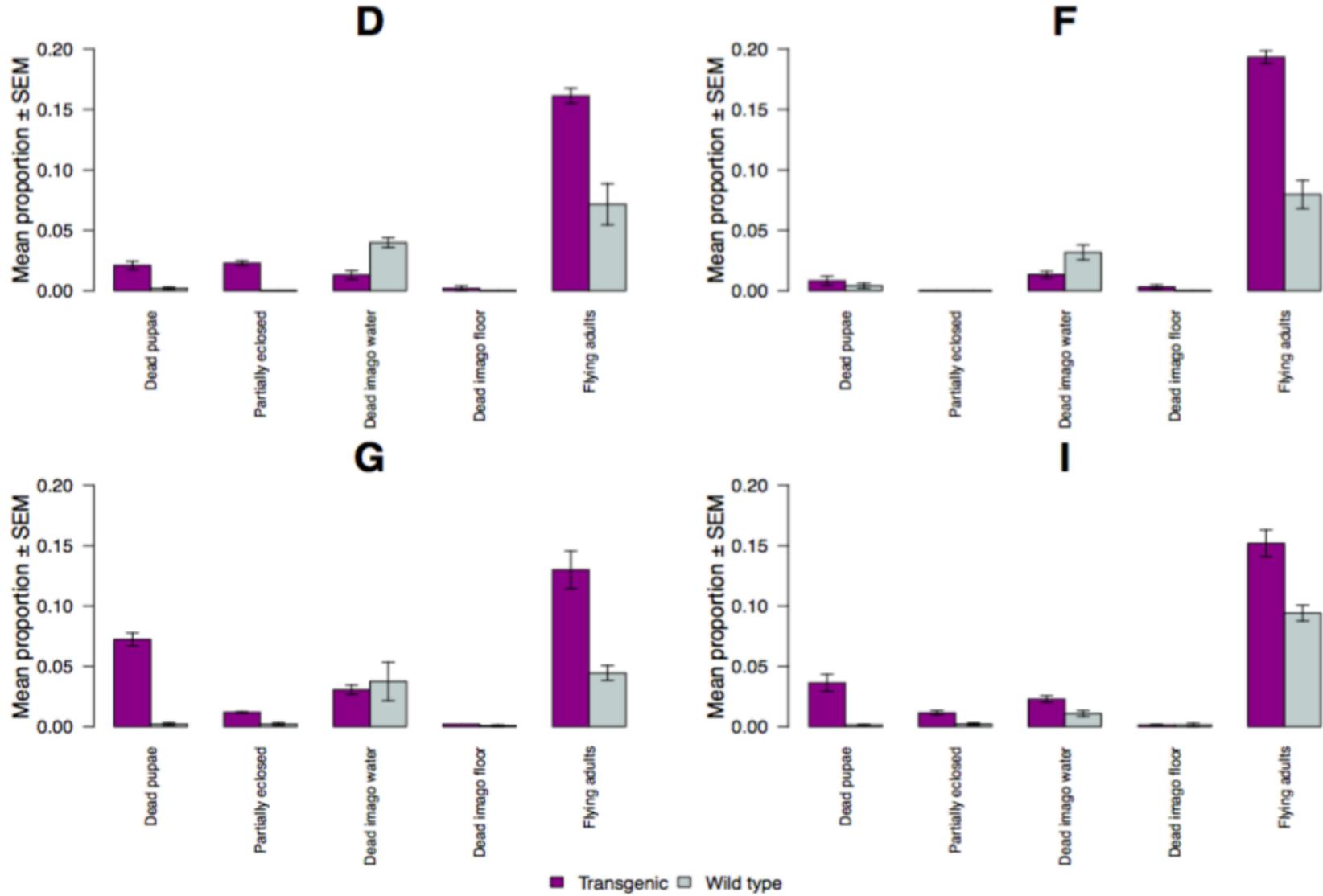Mortality in OX5055 lines: ON DOX

Mortality in intercrossed OX5055 lines: ON DOX

Appendix 2. Manufacturer's protocol for TRIzol Reagent (15596018, Life Sciences) method of RNA isolation.

1. Tissue material was lysed and homogenized in 1 ml of TRIzol Reagent using 1.5 ml RNase-free microcentrifuge tubes and a sterile polypropylene pellet pestle (749521-1590, Kimble Chase Life Science).

2. After 5-minute incubation at room temperature (RT), 0.2 ml of chloroform was added and the samples additionally incubated at RT for 3 minutes.

3. The samples were centrifuged for 15 minutes at 12,000 × g at 4°C and the aqueous phase containing the RNA was transferred to a new 1.5 ml tube by angling the original tube at 45° and pipetting the solution out.

4. 0.5 ml of isopropanol was added to the aqueous phase. The samples were incubated for 10 minutes (RT) and centrifuged for 10 minutes at 12,000 × g at 4°C.

5. The supernatant was discarded by pipetting, leaving a white gel-like pellet (RNA precipitate) at the bottom of the tube. This pellet was resuspended in 1 ml of freshly prepared 75% ethanol (Molecular Biology Grade absolute ethanol, Fisher BioReagents).

6. The samples were vortexed briefly and centrifuged for 5 minutes at 7,500 × g at 4°C.

7. After centrifugation, the supernatant was discarded and the RNA pellet air-dried for 10 minutes and resuspended in 50 µl of RNase-free water.

8. The samples were incubated in a heat block set at 60 °C for 15
    minutes.

Appendix 3. Manufacturer's protocol for DNA extraction from the gel with NucleoSpin Gel and PCR Clean-up kit (740609, Macherey-Nagel).

1. Gel extracted slice was placed in new 1.5 ml microcentrifuge tube and its weight determined.

2. For each 100 mg of agarose gel 200 µl of Buffer NTI was added and incubated in a heat block at 50 °C and vortexed briefly every 3-4 minutes until the gel slice was completely dissolved.

3. A NucleoSpin Gel and PCR Clean-up column was placed in a collection tube (2 ml), loaded with the sample and centrifuged for 30 seconds at 11,000 x g.

4. After the DNA binding (previous step), the column was washed with 700 µl of Buffer NT3, centrifuged for 30 seconds at 11,000 x g and the flow-through discarded. This step was performed twice.

5. To dry silica membrane of the spin column and remove residual ethanol, a collection tube with the column was centrifuged for 1 minute at 11,000 x g following a heat block incubation for 5 minutes at 70 °C.

6. 20 µl of Buffer NE was added to the column after transferring it into a new 1.5 ml microcentrifuge tube and incubated at room temperature for 1 minute.

7. After incubation the sample was centrifuged for 1 minute at 11,000 x $g.$

Appendix 4. Manufacturer's protocol for PCR product purification with NucleoSpin Gel and PCR Clean-up kit (740609, Macherey-Nagel).

1. In order to adjust DNA binding condition 2 µl of nuclease-free water and 100 µl of Buffer NTI was added to 48 µl of PCR reaction.

2. DNA sample was placed on a NucleoSpin Gel and PCR Clean-up column and centrifuged for 30 seconds at 11,000 x g.

3. After the DNA binding the column was washed with 700 µl of Buffer NT3 and centrifuged for 30 seconds at 11,000 x g and the flow-through discarded. This step was performed twice.

4. To dry silica membrane of the spin column and remove residual ethanol, a collection tube with the column was centrifuged for 1 minute at 11,000 x g following a heat block incubation for 5 minutes at 70 °C.

5. 20 µl of Buffer NE was added to the column after transferring it into a new 1.5 ml microcentrifuge tube and incubated at room temperature for 1 minute.

6. After incubation the sample was centrifuged for 1 minute at 11,000 x g.

Appendix 5. Manufacturer's protocol for *in vitro* transcription reaction with the MEGAscript T7 Kit (AM1334, Life Technologies).

1. 10X Reaction Buffer and 4 ribonucleotides (ATP, CTP, GTP, UTP) were briefly thawed on ice and vortexed. Once thawed, ribonucleotides were stored on ice and 10X Reaction Buffer was kept at room temperature while assembling the reaction.

2. Reaction components were mixed gently but thoroughly by pipetting up and down and centrifuged briefly after being added to 200 µl PCR tube in following the order:

In vitro transcription reaction components	Volume for one 20-µl reaction
Nuclease-free water	0 or 1 µl
Free ribonucleotides (ATP, CTP, GTP, UTP)	1 µl each (4 µl in total)
10X Reaction Buffer	1 µl
sgRNA template (390-425 ng/µl)	3 or 4 µl
T7 Enzyme Mix	1 µl
TOTAL	10 µl

3. PCR tube with transcription reaction components was incubated for 16 hours at 37 °C.

4. After 16-hour incubation 1 µl of TURBO DNase (2 U/µl) was added into the tube, contents of the tube mixed well and incubated at 37 °C for additional 15 minutes.

Appendix 6. Manufacturer's protocol for RNA purification with the MEGAclear Kit (AM1908, Life Technologies).

1. RNA sample was brought to 100 µl with Elution Solution (89.5 µl) and mixed gently but thoroughly.

2. 350 µl of Binding Solution Concentrate was added to the sample and mixed gently by pipetting.

3. 250 µl of Molecular Biology Grade absolute ethanol (Fisher BioReagents) was added to the sample and mixed gently by pipetting.

4. RNA mixture was pipetted onto the Filter Cartridge placed into a 1.5 ml collection tube and centrifuged for 1 minute at 14,000 x *g*. The flow-through was discarded and collection tube used in the following washing steps.

5. The RNA sample was washed twice with 500 µl of Washing Solution applied on the filter and centrifuged for 1 minute at 14,000 x *g*. After discarding the flow-through the Filter Cartridge was additionally centrifuged for 30 seconds at 14,000 x *g* and placed into a new collection tube.

6. 50 µl of Elution Solution was applied to the center of the Filter Cartridge incubated in a heat block set to 70 °C for 10 minutes. After incubation the tube was centrifuged for 1 minute at 14,000 x *g*. This step was repeated twice and the flow-through collected to the same collection tube.

Appendix 7. Manufacturer's protocol for genomic DNA extraction using NucleoSpin Tissue kit (740952, Macherey-Nagel).

1. All samples prior to DNA extraction were stored at -80 °C, placed on ice immediately and briefly but thoroughly homogenised using a sterile polypropylene pellet pestle (749521-1590, Kimble Chase Life Science). Further tissue disruption was performed upon adding 40 µl of Buffer T1 and by subsequent addition of 140 µl of Buffer T1 with 25 µl of Proteinase K (lysis solution).

2. Homogenised samples were incubated overnight in a heat block set to 56 °C and vortexed periodically during the first few hours of incubation and upon its completion.

3. 200 µl of Buffer B3 was added to each sample, vortexed vigorously and incubated in a heat block set to 70 °C for 10 minutes.

4. After incubation, the samples were vortexed briefly, centrifuged for 2 minutes at 14,000 x *g* and the supernatant was transferred to a new microcentrifuge tube.

5. To adjust DNA binding conditions, 200 µl of Molecular Biology Grade absolute ethanol (Fisher BioReagents) was added and the content of each sample was vortexed vigorously.

6. Each sample was applied to a NucleoSpin Tissue Column, placed in a Collection Tube and centrifuged for 1 minute at 11,000 x *g*. The flow-through was discarded and the column placed in a new Collection Tube.

7. Two washings were performed for each sample using 500 µl of Buffer BW (1st wash) and 600 µl of Buffer B5 (2nd wash). The flow-through after each washing was discarded, the column placed back into the Collection Tube and centrifuged for 1 minute at 11,000 x *g*.

8. The NucleoSpin Tissue Column was placed into a new 1.5 ml microcentrifuge tube and 50 µl of Buffer BE was added onto the column, incubated at room temperature for 1 minute and centrifuged for 1 minute at 11,000 x *g*.

Appendix 8. The annotated gene structures in the *Cxq-A4* sequence. The newly identified 5′UTR region is annotated separately for males and females. The 5′UTR region transcribed in both sexes is indicated by the underlined sequence. The alternatively spliced male and female intron within the 5′UTR of the *Cxq-A4* is represented by the bold sequences in red and blue, respectively. The 3-nucleotide sequence in bold indicates start and stop codons of the male transcript.

Female 5′UTR:

5′ – <u>CAATCACTCACTAATCAGGATTCTTGACGCTCTCTGCCGCAGACTGT ACTTGTTTTGTTCTCCTAAAA</u>**GGTCCGTGCCGCGTGGAGTAGTTGGTGG TTAGCCGTCCTGTGCCGTCCAGCGATGAGTTCCAGAGCTGTTGTCCCG TCCTGTGGCGTCGATTGTGGTGGTGTCTTCTGTTGTGAATTGTGCCGTC CAGTGCCGTCCGAGTTGTTTGTGGATGAGTCCGTCCTTGTCCCTGTGAT CGTCTTTCCATGGGTTCTTTATGTGACATGAGTTATGTAAGTGAAAAGT GTTGAAAAGGCTTGCATGTTTTGCATACGAGATTGTGTAACTTTTATGTT TTGAGAAAATGAAAAGGACTTTGTGAATCTCACGAGATTTTTTTTATAAT TTAGG**<u>ACCCCCGAGGACCTAAAGCAGCGCCAAG</u> – 3′

Male 5′UTR:

5′ – <u>CAATCACTCACTAATCAGGATTCTTGACGCTCTCTGCCGCAGACTGT ACTTGTTTTGTTCTCCTAAAA</u>GGTCCGTGCCGCGTGGAG**TAG**TTGGTGG TTAGCCGTCCTGTGCCGTCCAGCGATGAGTTCCAGAGCTGTTGTCCCGT CCTGTGGCGTCGATTGTGGTGGTGTCTTCTGTTG**TGA**ATTGTGCCGTCC AGTGCCGTCCGAGTTGTTTGTGGA**TGA**GTCCGTCCTTGTCCCTGTGATC GTCTTTCC**ATG**GGTTCTTT**ATG**TGACATGAGTTAT**GTAAGTGAAAAGTGT TGAAAAGGCTTGCATGTTTTGCATACGAGATTGTGTAACTTTTATGTTTT GAGAAAATGAAAAGGACTTTGTGAATCTCACGAGATTTTTTTTATAATTT AGG**<u>ACCCCCGAGGACCTAAAGCAGCGCCAAG</u> – 3′

Exon 2

5′ – ATGTGTGATGATGATGCTGGAGCACTGGTCATTGACAATGGATCCG
GAATGTGCAAAGCCGGCTTCGCTGGTGATGATGCACCACGTGCCGTCTT
CCCGTCCATCGTTGGCCGCCCACGCCACCAGGGTGTGATGGTCGGTAT
GGGTCAAAAGGATGCCTACGTGGGTGATGAAGCGCAATCGAAGAGAGG
TATTTTGACGTTAAAGTACCCGATCGAGCACGGTATCATTACCAACTGGG
ACGACATGGAGAAGATCTGGCATCACACGTTCTACAATGAGTTGCGAGT
CGCTCCGGAGGAACATCCAGTGCTGCTGACTGAGGCTCCCTTGAATCCC
AAGTCCAACCGTGAGAAGATGACCCAGATCATGTTTGAGACGTTCGCTT
CTCCGGCTGTGTACGTCGCCATCCAGGCTGTCCTGTCCCTGTACGCTTC
CGGTCGTACCACCGGTATCGTTCTGGATTCCGGAGATGGTGTCTCCCAC
ACCGTCCCGATCTACGAAGGTTATGCCCTTCCCCATGCCATCCTCCGTA
TGGACTTGGCTGGTCGCGATCTTACGGACTACCTGATGAAGATCCTGAC
CGAGCGCGGTTACTCCTTCACCACCACCGCTGAGCGTGAAATCGTTCGC
GATATCAAGGAAAGCTGTGCTACGTCGCCCTGGACTTTGAGCAGGAAA
TGCAAAGTTGCCCAATTTTTGTCTTTTGCTGAGAAGTCCTATGAACTTCC
CGGCGGACAGGTCATCCCCATCGGCAACGAACGTTTCCGTTGCCCCGA
GGGCTTGTTCCAGCCTTCCTTCCTGGGCATGGAAGCTACCGGTGTCCAC
GAGACTGTCTACAACTCGATCATGCGTTGCGACGTTGACATCCGTAAGG
ATCTGTACGCCAACAGCGTCTTGTCCGGTGGTACCACCATGTACCCAG –
3′

Intron between exon 2 and 3

5′ – GTACATAACATCTCTCACTCCTCCTCCAAACCCAAATATTAACTCAAA
CCTCCCCTCTCCAG – 3′

Exon 3

5′ – GTATTGCCGATCGTATGCAAAAGGAAATCACCTCCCTCGCCCCGTC
CACGATCAAGATCAAGATCATCGCCCCGCCCGAGCGTAAGTACTCGGTC
TGGATCGGCGGCTCCATCCTGGCCTCGCTGTCCACCTTCCAGACGATGT
GGATCTCGAAGCAGGAGTATGACGAGGGTGGCCCAGGAATCGTCCACC
GCAAGTGCTTCTAAGTGCATGCTTACTCTAATCGATGACATTGCCAGTAA

CAACCGCAAAAACAGCACAAGCAAAGTACAAACGACAGCAACTTTAGCA
ACAACAACAACTCAACTGCCAGCAACAACTAGCTTAGTTCGGGCTAATAA
CAAGCAAGGGAGAAGAACTCGTGTGGACTCGTCCGGGGCCGGAGGAAA
TCATCACCAGCAGCAACAACAACCAAAAAAGGCTTGCGCTTTACTTGTTT
TAGAAGTTTTGCGTTCGACATCTCAAATCCATCCAATCCATCATTGTGTT
CTCGTCGTTGGAGGCAAACGGGGCGAGAAGCAGTTCTTA – 3′