Development of a gene editing platform for sand fly vectors of Leishmania through CRISPR-Cas9 genetic modification targeting genes associated with olfaction

Rhodri Tomas Mawr Edwards

Thesis submitted in accordance with the requirements for the degree of

Doctor of Philosophy of the University of London

January 2023

Department of Infection Biology

Faculty of Infectious and Tropical Diseases

LONDON SCHOOL OF HYGIENE & TROPICAL MEDICINE

Funded by Medical Research Council

Research group affiliations:

Dr Matthew Yeo

Dr Thomas Walker

Dr Matthew Rogers

Declaration of own work

I, Rhodri Tomas Mawr Edwards, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

Current control strategies for mitigating the impact of Leishmaniasis via vector control do not provide a panacea. New strategies for vector control are required. This research aims to develop novel molecular tools to interrupt transmission of *Leishmania* by the manipulation of olfactory genes, affecting host seeking, and subsequently spreading traits rapidly through insect populations by the application of gene drive technologies.

CRISPR-Cas9 offers unparalleled capacity to manipulate eukaryotic genomes. The technology has been applied successfully to a variety of invertebrate species, initially to knockout genes, and more recently to insert exogenous DNA such as expression of anti-parasite peptides, and to target genes involved in fecundity. This method shows promise in the context of interrupting transmission of pathogens, and will be developed in phlebotomine sand flies.

The research has three main objectives. First, to identify and rationalise endogenous gene targets for proof-of-concept CRISPR-based modification in sand flies. Endogenous non-lethal phenotypic marker genes were identified, alongside olfactory genes involved in host seeking behaviour. Three key gene families were identified as playing an important role in olfaction and host detection in sand flies. Second, to develop a suite of modification tools to deliver CRISPR-Cas9 components using plasmid and non-plasmid based approaches, to demonstrate precise knockouts of these targeted genes. An *in vitro* cell line platform was developed and optimised for validation of these tools via chemo-transfection, prior to *in vivo* studies. Third, to attempt *in vivo* modification of sand flies. CRISPR tools targeting phenotypic marker genes and olfactory genes were delivered to sand fly embryos by conventional microinjection, and assessment of modifications was conducted using phenotypic observation, heteroduplex assessment, and Sanger sequence and computational methods. Overall, these objectives aimed to lay the foundations for the development of CRISPR-based tools towards novel gene drive control strategies for *Leishmania* vectors.

Acknowledgments

I would like to thank my funders the MRC and my supervisors Matt Yeo, Tom Walker and Matthew Rogers without whom this project would not have taken place. In particular, I would like to thank Matt for his encouragement throughout, and thesis suggestions in the final weeks. Beyond my primary supervisors, I particularly would like to thank Paco Olmo for his constant advice and support throughout the work, and Mojca Kristan for thesis advice. I would also like to thank the wider 347 lab group, especially Luke who has been a massive support right from the beginning as we both ventured into unknown territory with molecular biology, genetics and CRISPR. I am grateful also to the support from Lydia T in helping to get the project off the ground.

The project would not have been possible without the help of collaborators (and friends) in Prague, particularly Petr Volf, Erich, Barbora, and Lucie. The shear willingness and drive of this group was a privilege to be part of during my multiple visits, as well as the regular visits to Mrtvá Ryba.

I also have to thank 380. Harry and Richard have made me far poorer, yet sane, with many Store Street Espresso visits throughout the PhD, and Amy and Pepita who have provided endless support for little in return.

I would not have made it through without the support of my friends who have picked me up when most needed. In particular, Elliott for his constant presence and solid consistency, Lydia, Richard and Martin incredible housemates providing food and trashy TV in equal measures, Claire for always being there, Jen, Donal, Poppy, and Deji for always checking in, Anna and Ben for when I've needed an escape, and Leila.

Lastly, I must thank my parents and sister for their unwavering support and encouragement both from a scientific standpoint, and for keeping me going right to the end. I'm sure they will be glad it is over.

Contents

Abbreviations

bp	Base pair
Cas9	CRISPR associated protein 9
CRISPR	Clustered regularly interspaced short palindromic repeat
DAPI	4′,6-diamidino-2-phenylindole fluorescent stain
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DNTP	Deoxynucleoside triphosphates
DSB	Double stranded break
EAG	Electroantenography
ECFP	Enhanced cyan fluorescent protein
EGFP	Enhanced green fluorescent protein
FACS	Fluorescent-activated cell sorting
FSC-A	Forward scatter area
FCS	Foetal calf serum
G0	Generation zero
G1	Generation one
GFP	Green fluorescent protein
GR	Gustatory receptor
gRNA	Guide RNA
HDR	Homology directed repair
IC50	Half maximal inhibitory concentration

ICE	Inference of CRISPR Edits
Indel	Insertion/deletion
IR	Ionotropic receptor
IRS	Indoor residual spraying
kb	Kilobase
LB	Luria-Bertani agar
LLIN	Long-lasting insecticide treated net
mg	Milligram
mL	Millilitre
mM	Millimolar
mRNA	Messenger RNA
NHEJ	Non-homologous end joining
OBP	Odour binding proteins
Oligo	Oligonucleotide
OR	Olfactory receptor
ORN	Olfactory receptor neuron
PAM	Protospacer adjacent motif
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
Pen/strep	Penicillin/streptomycin
pgSIT	Precision guided sterile insect technique
RFP	Red fluorescent protein
RMCE	Recombination-mediate cassette exchange
RNA	Ribonucleaic acid

RNAi	Ribonucleaic acid interference
SSC-A	Side scatter area
sgRNA	Single guide RNA
SIT	Sterile insect technique
SNP	Single nucleotide polymorphism
TALEN	Transcription-activator-like effector nuclease
TCB	Tick Cell Biobank
µg	Microgram
µL	Microlitre
µM	Micromolar
VOC	Volatile organic compound
WGS	whole genome sequence
WHO	World Health Organisation
WT	Wildtype
ZFN	Zinc Finger Nucleases

Chapter 1

General Introduction

Epidemiology of Leishmaniasis

Leishmania spp, the causative agent of leishmaniasis, are protozoan parasites transmitted to humans by the bite of Phlebotomine sand flies in the Old and New World (Pigott et al., 2014). The parasites are both anthroponotic and zoonotic (depending *Leishmania species*), with reservoir hosts including domestic dogs, rats, and sylvatic animals such as opossums (Pigott et al., 2014). *Leishmania* is classified by the WHO as one of 20 neglected tropical diseases (NTDs) and it is considered to be the leading NTD for mortality and morbidity. 1-2 million cases are diagnosed annually in endemic countries. Of the > 80 endemic countries, only six are responsible for 90% of the cases; Brazil, Ethiopia, Sudan, South Sudan, India and Bangladesh (Pigott et al., 2014).

There are three main manifestations of the disease: cutaneous (CL), mucocutaneous (MCL) and visceral leishmaniasis (VL). Currently CL is endemic in 87 and VL is endemic in 75 countries. In the Americas, CL and MCL are found in 20 countries and are endemic in 18 countries, whilst VL is present in 12 countries (with 96% of all cases occurring in Brazil)(WHO, 2017).

There are approximately 700,000 to 1 million CL cases each year, with 90% of cases occurring in Afghanistan, Pakistan, Syria, Saudi Arabia, Algeria, Iran, Brazil and Peru (Burza, Croft, & Boelaert, 2018). Different parasites cause this form of the disease in the Old World and the New World; *Leishmania major* (*L. major*), *L. tropica*, and *L. aethiopica* are found in Africa and Asia, and *L. amazonensis*, *L. mexicana*, *L. braziliensis*, and *L. guyanensis* are the causative parasite of CL in Central and South America (Burza et al., 2018).

VL is the most severe form of the disease - leading to fatigue, weight-loss, the enlargement of the lymph nodes, and vital organs such as the spleen and liver. If untreated it is usually fatal within two years (Burza et al., 2018). Each year, there are ~700,000 to 1 million cases of VL in ~100 endemic countries. VL is caused by several parasites; *L. donovani* in Asia and Africa, and *L. chagasi/infantum* in the Middle East, Central and South America (Burza et al., 2018).

Leishmania parasite lifecycle

Leishmania parasites require two obligate hosts, sand flies and mammals, to complete their lifecycle (Figure 1). The parasites are not restricted to completing their lifecycle within humans; however, parasites transmission is influenced by the biting preference of the female sand fly species. When an infected sand fly bites, they inject *Leishmania* promastigotes (motile infective stage of the parasite), with saliva, into the mammalian host. Metacyclic promastigotes are phagocytosed by macrophages, or they actively invade these cells. Within the macrophages, the promastigotes develop into amastigotes (tissue stage of parasites) which multiply, before leaving infected cells and infecting further macrophages. Sand flies become infected by taking a blood meal containing amastigote-infected cells.

Once in the sand fly gut amastigotes transform into promastigotes. These migrate to the proboscis where the *Leishmania* parasite's lifecycle is complete.

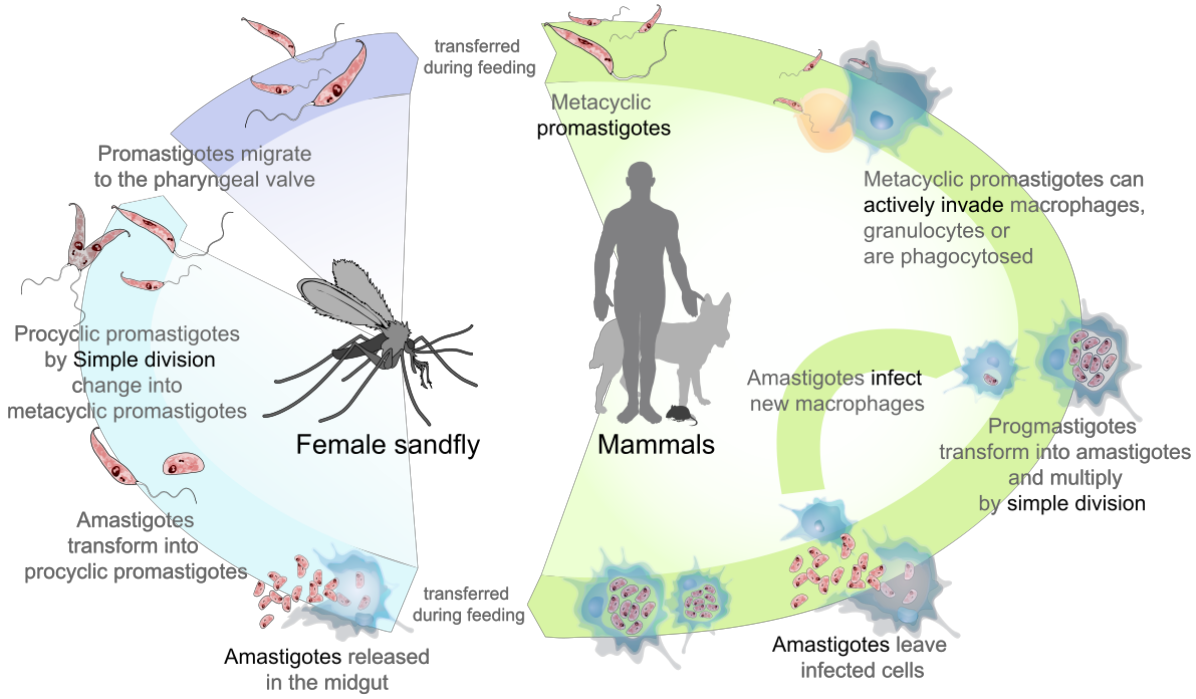


Figure 1. Lifecycle of *Leishmania* parasites, from the bite of an infected female sand fly, the development of the parasite within the mammalian host, and the development of the parasite within the sand fly once ingested in a blood meal. Image adapted from https://commons.wikimedia.org/wiki/File:Leishmaniasis_life_cycle_diagram_en.svg

Vectors of Leishmania

Phlebotomine sand flies are members of the Psycodidae family. Of five Psychodid subfamilies, only Phlebotominae have mouthparts that can pierce skin and take blood. Ninety-eight species with 56 *Lutzomyia* and 42 *Phlebotomus* species are identified as potential or incriminated vectors of human leishmaniasis. *Lutzomyia* are vectors of New World leishmaniases in North, Central, and South America. *Lutzomyia longipalpis* (*L. longipalpis*) is the species suspected of being a vector in the greatest number of countries (12) (**Figure 2**). *Phlebotomus* are vectors of Old World leishmaniases in Southern Europe, North Africa, East Africa, West Africa, the Sahel, Central Asia, South Asia and the Middle East. *Phlebotomus papatasi* (*P. papatasi*) is a suspected vector in 22 countries (Figure 2) (Maroli, Feliciangeli, Bichaud, Charrel, & Gradoni, 2013).

Sand flies are small, 1.5-3.5 mm, have long stilt-like legs, and have a dense covering of hairs on their wings and body. The wings have a characteristic venation, running parallel to the wing margins, and are held at 45° above the thorax in a V-shape. Both sexes have long filamentous antennae, and both have similar mouthparts, although males lack mandibles and have weaker stylets. Both sexes require carbohydrates (sugars), but only females take blood meals. Males are distinguished by the presence of claspers, part of the male genitalia, at the tip of the abdomen. In general, sand flies are difficult to identify to species, which can require identification of internal structure such as spermathaeca and cibarial armature.

Lifecycle

The lifecycle of sand flies consists of four stages. Thirty to 70 eggs are laid three to ten days after a bloodmeal. When eggs are freshly laid they are translucent, before melanisation to a dark brown or black colour. Eggs are elliptical, 0.3-0.5mm long and 0.1-0.15mm wide, and covered in sculptured ridges on the surface which can be used to differentiate species. Eggs are laid on moist substrates, and typically hatch 6-11 days post bloodmeal. The next stage in the lifecycle is the larval stage. First instar larvae are small, less than 1 mm, have a dark head capsule and a light body colour. The larvae are covered in characteristic matchstick hairs along the body, and two caudal setae. The second instar increases in size, and by the third instar the larvae are up to 3 mm in length and have an additional pair of caudal setae. In the fourth and final stage, larvae are up to 4 mm, and have a sclerotized anal plate visible on the dorsal surface. This stage stops feeding, before transforming into a 3-4 mm pupa (Lawyer et al., 2017). These pupae are identifiable by the presence of the larval skin (exuvium) at the posterior end of the pupae. Adults emerge from the puparium after 6-13 days.

Blood-feeding

Female sand flies use the mouthparts to tear the skin (with mandibles), leading to blood-pool formation. Their salivary secretions contain different components that aid bloodfeeding via vasodilation (Ribeiro, Katz, Pannell, Waitumbi, & Warburg, 1999), and anticlotting (Charlab, Valenzuela, Rowton, & Ribeiro, 1999).

Reproduction

Mating and reproductive behaviour varies depending on sand fly species. Male *Lutzomyia longipalpis* (*L. longipalpis)* form a mating aggregation (leks) on or near a host, which is likely to increase the chances of encountering females (Kelly & Dye, 1997). Females choose the mate they prefer, and mating often occurs on a host animal, where the female has taken a blood meal (Ready, 1979). Males produce sex pheromones that attracts females to the lek, and allows for discrimination between different members of a species complex (Morton & Ward, 1989). *Phlebotomus papatasi (P. papatasi)* females mate with multiple males, however factors involved in mating success are not known. Mate seeking involves male-produced pheromones that can attract females from a distance (I. Chelbi, Zhioua, &

Hamilton, 2011), and mating occurs without lekking taking place. Courting behaviour involves a sequence of contacts between the males and females, particularly with the ends of the legs and antennae, which potentially transfers cuticular hydrocarbons, which mainly function as a desiccation barrier, that may be used to discriminate between mates (Ifhem Chelbi, Bray, & Hamilton, 2012). Length of mating varies, and is conducted with males and female facing in opposite directions.

Host preference

Female *L. longipalpis* and *P. papatasi* (major vectors of leishmaniases in the New and Old world, selected as the models for this thesis research) bloodfeed on many hosts, including livestock, domestic animals, and sylvatic species. Sand flies are opportunistic feeders, however they demonstrate host preferences that can vary with habitat and location. In Brazil, *L. longipalpis* have no strong preference for a particular host, and bloodfeeding seems to be a function of the host size (Quinnell, Dye, & Shaw, 1992). Analysis of blood meals shows cows are the dominant source in Colombia. Lesser hosts include pigs, horses, humans, dogs, opossums, birds and reptiles. *L. longipalpis* are opportunistic feeders, however dogs and opossums are reservoir hosts of *L. chagasi*, making them important in the maintenance of *Leishmania* parasite lifecyles (Morrison, Ferro, Morales, Tesh, & Wilson, 1993). *P. papatasi* display a preference for humans (anthropophily), but are opportunistic towards cows, goats, dogs and birds (Palit, Bhattacharya, & Kundu, 2005).

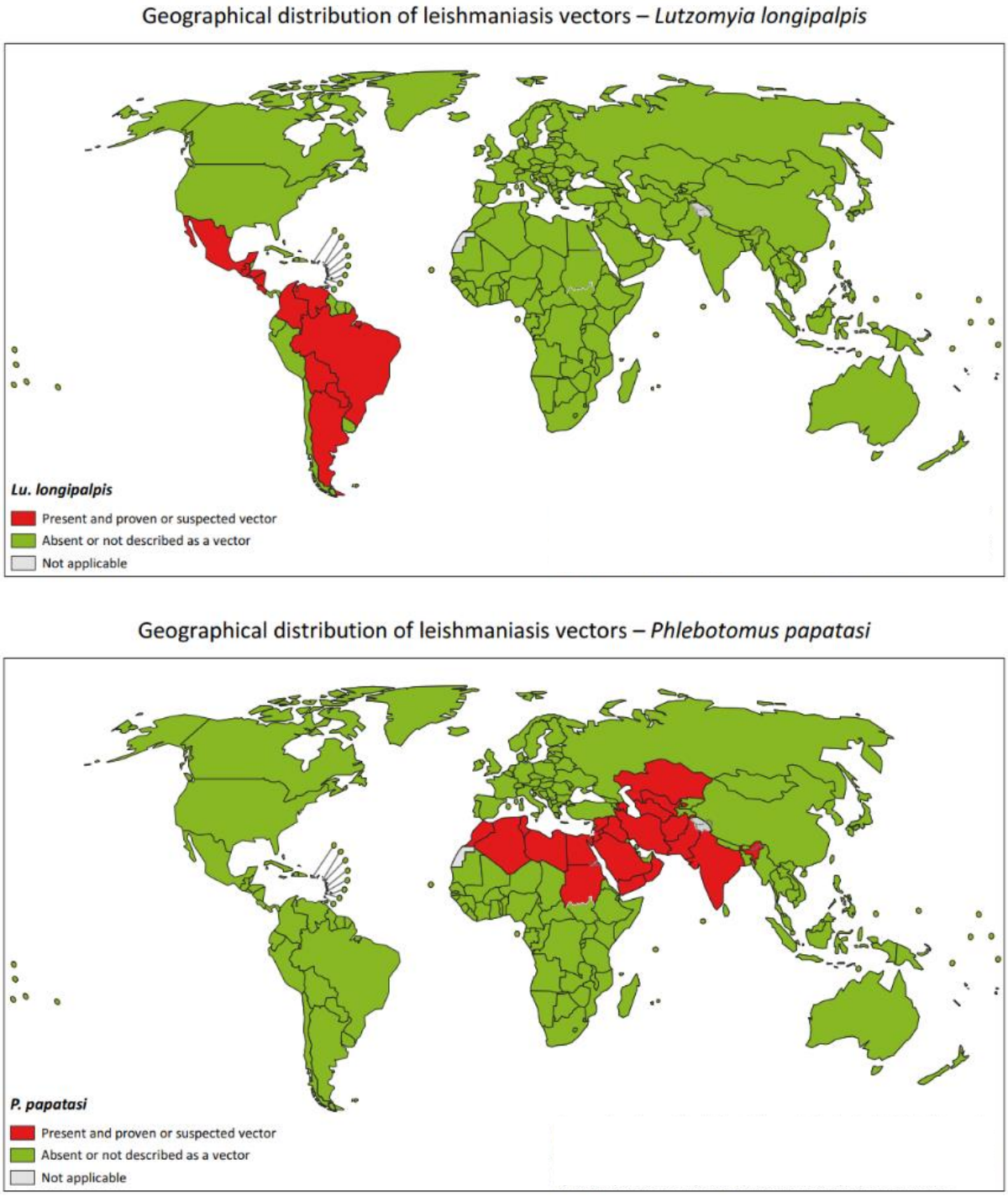


Figure 2. Geographical distribution of *L. longipalpis* (top) and *P. papatasi* (bottom) sand fly vectors of Leishmaniasis. Presence of vectors shown in red, absence shown in green. Data source: WHO Control of Leishmaniasis: report of a meeting of the WHO Expert committee on thee Control of Leishmaniases, Geneva, 22-26 March 2010. (WHO technical report series; no.949). Adapted from https://www.irycis.org/en/leishmaniasiscc-spain/.

Leishmaniasis control

Treatments and vaccines

Chemotherapeutic treatments for CL and VL, include Amphotericin B, antimonials, sitamaquine, pentamidine, paromomycin and miltefosine. However, these are often unavailable at point of need due to their cost, and have poor efficacy. Pentavalent antimonial monotherapy has been used for several decades to treat VL, however over time the efficacy has reduced in some locations leading to the requirement of higher doses (Burza et al., 2018). In addition, the antimonial drugs used can have serious side effects, and have to be administered intravenously (Sundar & Chakravarty, 2010). More recently oral Miltefosine has been used for CL, VL and MCL to reduce drug resistance and length of treatment regime, however there is now evidence of reduced efficacy (Dorlo, Huitema, Beijnen, & de Vries, 2012). For VL a third treatment option consists of Liposomal amphotericin B via intravenous infusion. Amphotericin B is very potent at reducing *Leishmania* infection , but is nephrotoxic inducing rapid deterioration in kidney function, as does sitamaquine (Bekhit et al., 2018).

Vaccines providing long-lasting immunity have proved hard to develop. Currently there are no registered leishmaniasis vaccines for humans, however there are approximately 16 candidate vaccine antigens for anti-parasite vaccines. A number of these candidates have been tested in animal models with promising results, suggesting that vaccines could contribute to future prevention efforts (Alvar et al., 2013). Different approaches to vaccine development have been attempted, including the use of sand fly salivary molecules to induce an antibody response (Kamhawi, Aslan, & Valenzuela, 2014), the use of inactivated parasites (Soudi, Hosseini, & Hashemi, 2011), recombinant proteins (Todolí et al., 2012), naked DNA (Khan, 2013; Todolí et al., 2012), and inoculation with live *L. major* to induce a single lesion (Carrión, Folgueira, Soto, Fresno, & Requena, 2011).

Vector control

Vector control requires a multi-faceted approach to be successful. Surveillance and control schemes are vital to assess the risk of transmission, and guide vector control. This includes monitoring the distribution and abundance of adult sand flies, often using a range of sampling methods such as sticky traps, light traps for females, and adult emergence traps (Alexander & Maroli, 2003).

Removing potential oviposition or larval sites such as organic waste, maintaining buildings by filling in cracks and holes, and removing reservoir hosts such as rats and mice have been attempted.

Insecticides have been used in different capacities. Spraying resting sites with malathion reduced sand fly density by 30% in Panama (Chaniotis, Parsons, Harlan, & Correa, 1982), temporarily eliminated the sand flies in Brazil with Organochlorine use (Ready, Arias, & Freitas, 1985), and reduced sand fly

density and human leishmaniasis cases in French Guiana (Esterre, Chippaux, Lefait, & Dedet, 1986). However, mass spraying measures are labour intensive and potentially damaging to the environment.

Attractive toxic sugar bait (ATSB) has been used to reduce sand fly populations in Israel, Morocco, and Iran. A range of different delivery methods have been used, with fermented fruit juice solution containing boric acid and spinosad sprayed on vegetation or fencing reducing *P. papatasi* numbers to 6-12% of that in control areas (Müller & Schlein, 2011). Similar results were achieved in Morocco, reducing populations by up to 83.2%, with limited impact on non-target insects (Qualls et al., 2015), and reducing incidence of cutaneous leishmaniasis after spraying (Saghafipour et al., 2017).

Prevention and control of leishmaniasis has mainly involved the use of insecticides. DDT was used effectively for control; however, this has ceased to be a long-term option because of environmental effects, and in some cases, the development of insecticide resistance (WHO, 2017). Residual Spraying has been used in domestic and peri-domestic environments, as well as spraying potential adult resting sites in sylvatic environments (WHO, 2017). Success relies on several factors; the class of insecticide used, the dosage, and the coverage. This method of control is logistically difficult, and there is emerging evidence of resistance to both DDT (Dinesh et al., 2010) and pyrethroids (Alexander et al., 2009) in both *L. longipalpis* and *P. papatasi* vectors.

The use of Long Lasting Insecticide Nets (LLINs) has proved inexpensive and effective at reducing contact between vectors and humans in some settings. An intervention trial using the Olyset net, led to a significant reduction in cases of cutaneous leishmaniasis, and reduced vector catches indoors over a two year trial period in Iran (Emami, Yazdi, & Guillet, 2009). An Olyset Plus net containing permethrin and a piperonyl butoxide (PBO) synergist successfully reduced CL incidence in Turkey from ~5% down to 0.37%, making the efficacy of net use above 90% (Gunay et al., 2014). Additionally, deltamethrin-treated nets demonstrated significant results against VL incidence (~47% reduction) in Bangladesh (Chowdhury et al., 2019). However, cluster randomized trials have indicated that the use of LLINs did not reduce incidence of VL or CL compared to indoor residual spraying of insecticides, suggesting this control method is not appropriate in all settings, as it is hard to maintain coverage and use compliance (Faraj et al., 2016; Picado et al., 2011).

Another control method is to target reservoir host species. Culling of dogs has been conducted in Brazil with mixed results in relation to the human incidences of leishmaniasis (Quinnell & Courtenay, 2009), however, more humane methods have been developed. In the domestic environment, deltamethrin impregnated dog collars have been used and have been found to be effective in Brazil and Italy, but not other locations (WHO, 2017). This control strategy is not universally effective, stray dogs are difficult to control, and collars must be replaced periodically - limiting the efficacy of this control strategy.

Overall, the WHO still considers bed nets the most effective method to control transmission of leishmaniasis within homes (WHO, 2017). Unfortunately, not all transmission occurs within the home as many vectors are peri-domestic and sylvatic.

Current control strategies are resource-intensive and are not universally effective. Novel approaches to control are needed to complement existing strategies, to interrupt transmission, and to reduce the burden of human suffering.

Genetic control of insect vectors

Multiple strategies for genetic control of insect vectors have been developed, and can be categorised with respect to the broad outcome to be achieved; either to supress an insect population size of incriminated vectors, or to replace an insect population with vectors that have a reduced vectorial capacity (transmission blocking). Further categorisation can be based on the degree of persistence of the control method within the environment. Some strategies can be classed as self-limiting, such as sterile insect techniques, whereby sustained release is required to maintain any effect. Alternatively, control can be self-sustaining, where modifications are spread within a population, such as gene drive, and persist indefinitely.

Multiple mechanisms have been developed to genetically modifying disease vectors to reduce transmission, and spread traits rapidly within populations. These genetic control strategies are contingent on mate-seeking behaviour, allowing them to be a highly-targeted species-specific control approaches.

Population suppression and population replacement

Genetic vector control strategies can broadly be defined by the outcome for the vector population; suppression, or replacement (strategies are described in Chapter 2, *CRISPR-based Gene Drives and Control strategies*). Population suppression involves reducing the vectorial capacity by releasing modified insects that spread transgenes resulting in infertility or reduced lifespan. Different mechanisms include sex ratio distortion (Galizi et al., 2016; Hall et al., 2015; Kyrou et al., 2018; Simoni et al., 2020), sterile insect technique (Chen et al., 2021; Li et al., 2021), or disruption of female-specific flight capacity (Navarro-Payá et al., 2020; O'leary & Adelman, 2020).

The alternative, population replacement, involves the release of genetically modified insects to supersede the wild-type population. Modified individuals are competitive with the wild-type population for food resources and mating opportunities, and spread pathogen refractory traits. One mechanism includes the release of transgenic insects that express parasite transmission-blocking effectors such as

antimalarial peptides (Dong, Simões, Marois, & Dimopoulos, 2018; Gantz et al., 2015), or antimicrobial peptides (Hoermann et al., 2022).

Self-limiting strategies

Release of Insects carrying a Dominant Lethal (RIDL)

RIDL is an extension of the sterile insect technique (SIT), which has been successful in some scenarios to control insect vectors and agricultural pests by the mass release of sterile males into local insect populations (Hendrichs, Franz, & Rendon, 1995; Krafsur, 1998; Vreysen et al., 2000). These insects compete for mates, and any successful mating results in non-viable offspring. Initially sterilisation approaches have included irradiation (Helinski, Parker, & Knols, 2006; Vreysen et al., 2000), and chemo-sterilisation (Grover et al., 1976), however these negatively impact on mating fitness. To overcome limitations genetic methods for sterilisation in an inherited manner have been vigorously pursued.

Typically RIDL involves genetic modification of a target species to carry a dominant lethal gene. When released, genetically modified males are homozygous for the lethal gene, and offspring resulting from mating with wildtype females are heterozygous for this dominant lethal, leading to death. For successful mass rearing of homozygous lethal males, the lethality is linked to a repression system. The repressor is only available when rearing in a laboratory environment and not in the natural environment. Professor Luke Alphey pioneered this method, with males expressing a tetracycline-repressible transactivator fusion protein (tTA) (Gong et al., 2005; Phuc et al., 2007; Thomas, Donnelly, Wood, & Alphey, 2000). In the presence of tetracycline (provided in the diet) tTA binds to the antibiotic, and prevents expression of a lethal effector. However, when tetracycline is removed from the diet the tTA can bind to tetracycline operators (tetO) leading to expression of the lethal effector (Figure 3) (Gong et al., 2005). Males are released *en masse*, transporting the transgene into the environment through mating with wildtype females. Resulting offspring possess the transgene, and succumb to the lethal effector.

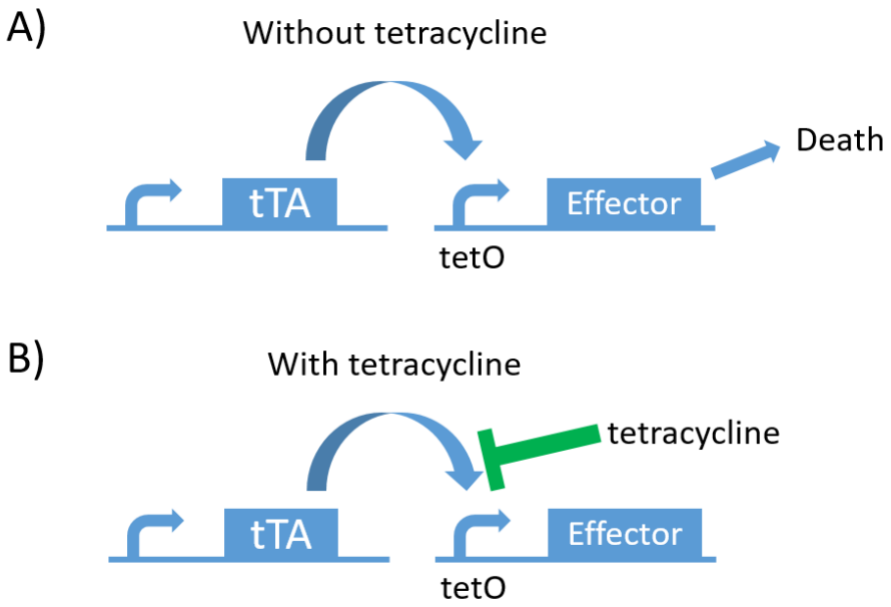


Figure 3. Illustration of the release of insects carrying dominant lethal (RIDL) mechanism. A) Without the presence of tetracycline tTA is expressed and binds with tetO, leading to the expression of a lethal effector molecule. B) In the presence of tetracycline the tTA is expressed, however this binds to tetracycline, preventing the expression of the lethal effector molecule. Adapted from Gong *et al*., (2005).

Release trials of *Aedes aegypti* (*Ae. aegypti*) in Brazil and the Cayman Islands have demonstrated that transgenic males have equal fitness compared to wildtype males (Harris et al., 2011), and that the repressible lethal genetic system functions to supress vector populations by 80-96% (Carvalho et al., 2015; Harris et al., 2012; Spinner et al., 2022)

The method has the potential to provide an effective and safe way to control the spread of insect-borne diseases, and it has been approved as a promising strategy for reducing the burden of diseases such as dengue fever, Chikungunya and Zika (Spinner et al., 2022).

Precision guided sterile insect technique (pgSIT)

A next generation CRISPR-based SIT (see Chapter 2 – Prospects and Opportunities for disease control, and Chapter 5 – Background to gene editing in insects for in depth discussion of CRISPR) with a mechanism involving simultaneous sexing and sterilization, has been developed in *Drosophila melanogaster* to control crop pests and vectors. This split-line system where two genetically modified homozygous insect lines are derived, one expressing Cas9 protein within the germline, and the other expressing multiple guide RNAs (gRNAs) targeting male fertility genes and female viability genes (Kandul et al., 2019). Remarkably, a strain containing gRNAs targeting *Beta tubulin* (male fertility gene), and *Sxl* (*sex lethal* female viability gene), crossed with Cas9-expressing lines, demonstrated 100% lethality to females and 100% sterility within males. Mating of individuals between both lines

results in the gRNAs knocking out both alleles of the two target genes, leading to conversion of all offspring to the desired phenotype.

Caveats to this approach are that gene edited individuals possess reduced mating competitiveness (22%); however, this finding is partially ameliorated by longer survival time than wildtype males, suggesting overall that pgSIT males fitness is not compromised.

Models simulating releases (weekly) over 6 months at a 10:1 adult ratio (pgSIT males : wildtype), or 200:1 egg ratio (eggs: wildtype adult) suggest more effective suppression of a population in comparison to traditional SIT methods, even when varying the male competitiveness and lifespan parameters (Figure 4).

pgSIT was originally developed in *Drosophila melanogaster* as proof of concept, and subsequently applied to the crop pest *Drosophila suzukii* (Kandul et al., 2022). Additionally, the method has been expanded successfully to vector control, targeting *Ae. aegypti* genes resulting in flightless females and sterile males (Li et al., 2021). *Ae. aegypti* and *Aedes. albopictus* have been identified as particularly strong targets for this method of control, as their eggs can diapause. This allows for the generation of large quantities of pgSIT eggs, for distribution to the field. Currently pgSIT has not been developed beyond laboratory cage studies, however it offers a self-limiting vector control alternative going forward (Li et al., 2021).

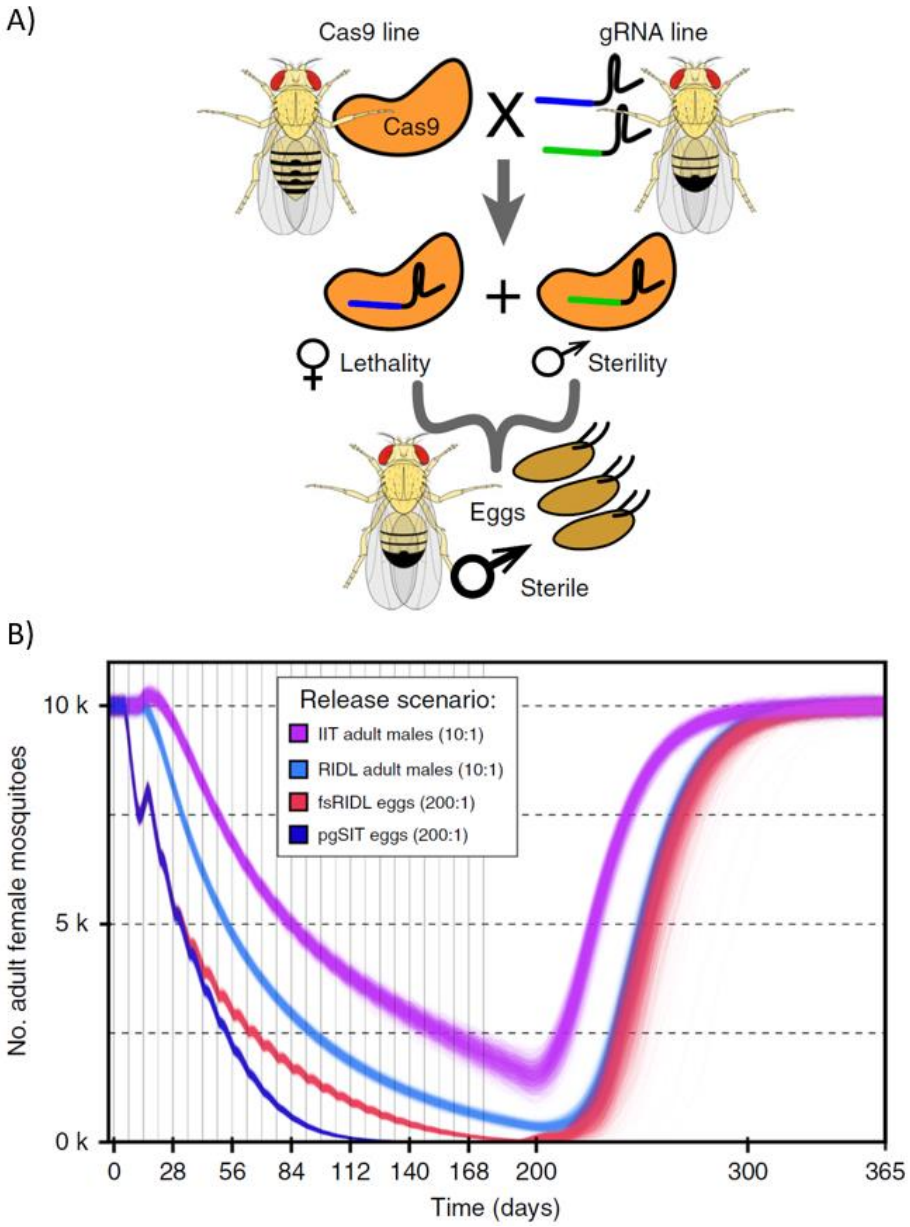


Figure 4. Precision guided sterile insect technique (pgSIT). **A)** An illustration of pgSIT using a split-line system of a Cas9 expressing line, and a line expressing gRNAs targeting female lethality, and male sterility. Homozygous lines are reared separately, then crossed leading to knockout of a female viability gene such as *sxl,* and a male fertility gene (*Beta tubulin*) resulting in survival of only F_1 sterile males. B) Model prediction of weekly pgSIT releases of *Ae. aegypti* over 6 months compared to releases of Wolbachia-based incompatible insect technique (IIT), RIDL, and female-specific RIDL. Release ratios (relative to wild adults) are shown in the key and *Ae. aegypti* population of 10,000 adult females was used. Adapted from Kandul *et al*., (2019).

Inherited female elimination by genetically encoded nucleases to interrupt alleles (Ifegenia)

A recent self-limiting approach developed in *An. gambiae* is termed Ifegenia (inherited female elimination by genetically encoded nucleases to interrupt alleles) (Smidler et al., 2022). The approach is CRISPR-based, disrupting the *femaleless* gene, (an essential gene) which leads to genetic sexing, by killing daughter offspring. Males modified by Ifegenia are still able to produce offspring, but spread the mutations of *femaleless* to the next generation (Figure 5).

In laboratory settings, population suppression is achieved by deriving separate lines expressing *femaleless*-targeting gRNAs, and another expressing Cas9 through the germline. When crossed, almost all offspring are male, indicative of female (larval) lethality; demonstrating a highly efficient sex-sorting method. Additionally, the *femaleless* transgene is inherited in subsequent generations (Smidler et al., 2022).

Modelling of this method suggests, that suppression of a population could be greater than 90% using a standard release ratio of 10:1 (Ifegenia:wildtype), requiring releases of 300 Ifegenia eggs for 22 weeks. Furthermore, longer-term releases are expected to have the potential for local elimination of vectors (Smidler et al., 2022).

The potential advantages of Ifegenia control are, that male fitness is not severely impacted allowing males to compete successfully for mates, easy mass rearing due to the split-line nature of the system, and the ability for releases to be performed at any stage in the lifecycle from eggs onwards - as sexing occurs genetically (Smidler et al., 2022). This method is yet to be tested beyond small caged trials, and has only been attempted in one vector (*An. gambiae*), however it is a promising method bridging classic SIT and gene drives.

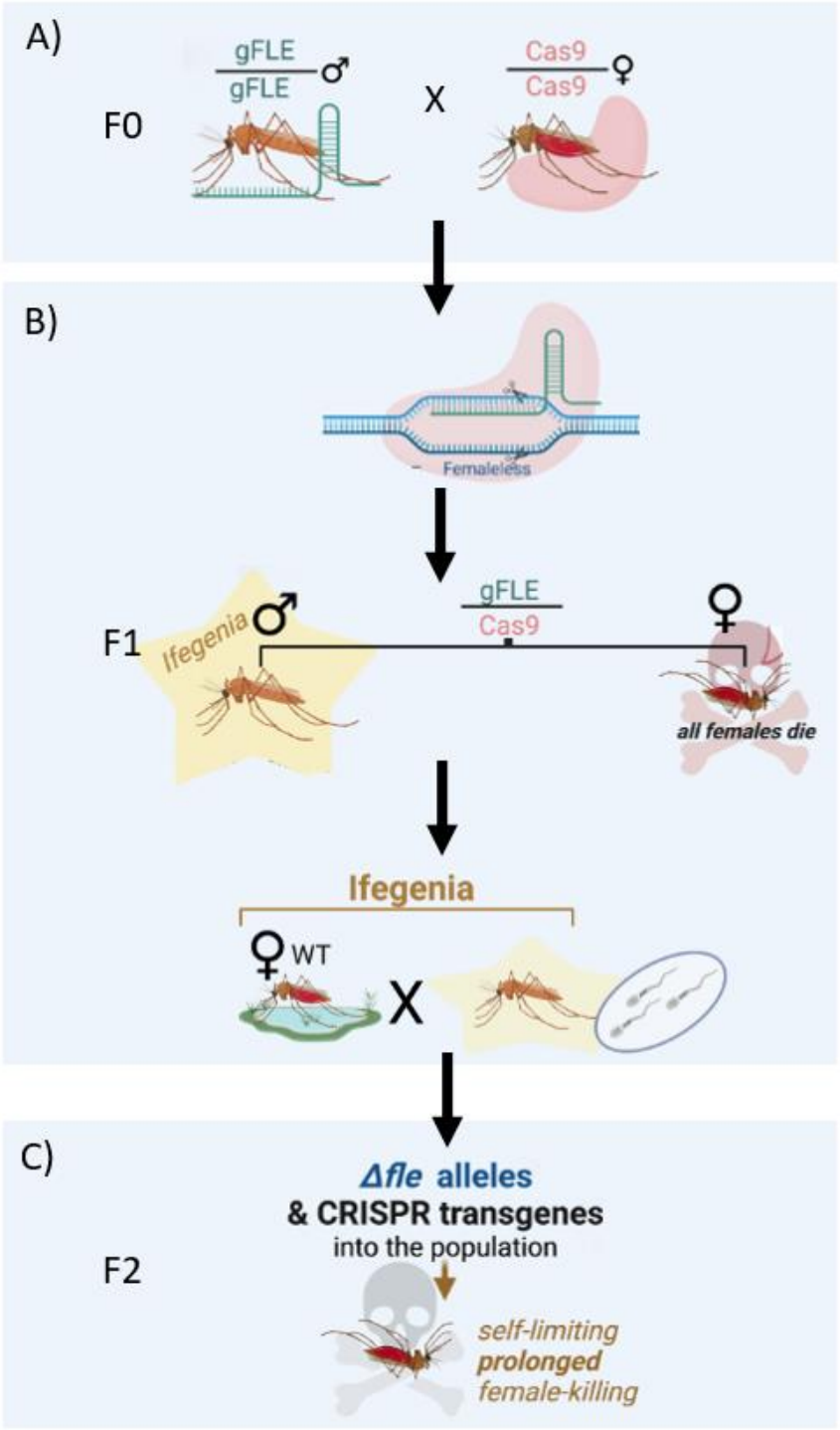


Figure 5. Illustration of Ifegenia self-limiting control. A) Ifegenia males and Cas9 females are mass-reared, and F0 crossing is conducted. B) F1 eggs undergo *femaleless* (*fle)* mutagenesis resulting in ~ 100% Ifegenia males. F1 males are then able to mate with wildtype females. C) The wild population is supressed, by lethality to females, and males able to spread the modified *fle* alleles, leading to multigenerational suppression. Using this strategy eggs can be released into the environment as the sterilisation and sexing effects are both genetic-based. Adapted from Smidler *et al*., (2022).

Self-sustaining strategies

Gene drives

Gene drives provide a mechanism to modify vectors and transmit modifications to offspring at a rate greater than the canonical Mendelian inheritance, allowing the rapid spread of genetic elements through a population. Many mechanisms have been designed over the past two decades, however the emergence of Clustered regularly interspaced short palindromic repeats (CRISPR) technology has led to a proliferation of research targeting vectors of disease and novel control methods.

The majority of drives are homing-based drives designed off the back of homing endonuclease genes (HEGs), which convert opposite chromosomes into identical copies by having endonucleases that recognise specific sequences and cut the opposite chromosomes forcing the cell to repair the break. Often the cellular repair mechanism in germline cells copies the undamaged chromosome, via homology directed repair (HDR), allowing the HEG to spread rapidly within germline cells (Figure 6).

This system has been appropriated for genetic control of vectors species, by either rapidly suppressing a population, or driving beneficial genes that reduce vectorial capacity. CRISPR-based homing drives have been developed in *Drosophila* (Gantz & Bier, 2015), and in mosquito vectors such as *An. gambiae* (Windbichler et al., 2011; Bernardini, Kriezis, Galizi, Nolan, & Crisanti, 2019; A. Hammond et al., 2016; Kyrou et al., 2018) and *An. stephensi* (Gantz et al., 2015).

In the malaria vector *An. stephensi*, a homing gene drive was designed to introduce anti-plasmodium effector genes (m2A10 and m1C3, induced by blood feeding). The HDR repair mechanisms functioned well within offspring from transgenic individuals expressing Cas9 within the germline, with the introduced traits frequencies exceeding mendelian inheritance (≥99.5% efficiency) (Gantz et al., 2015). However, non-homologous end joining (NHEJ) reduced inheritance of the effector genes (Gantz et al., 2015).

Multiple homing gene drives have been developed in the main vector of malaria, *An. gambiae*. One study targeted three genes that lead to recessive female sterility, with high expression in ovaries. These resulted in modified genes being transmitted through a caged population at a rate between 91-100% (A. Hammond et al., 2016). A second targeted the sex determination pathway, resulting in intersex females that were unable to mate with unaffected males. Caged studies demonstrated population collapse within seven to 11 generations (dependent on release frequencies) (Kyrou et al., 2018).

Overall, gene drives are likely to play a role in vector control and are being designed for a range of major disease vectors and settings. Currently, experimental trials of gene drives remain in small and medium-scale cages, however research is rapidly approaching field based trials. The next step requires a meticulous and cautious approach, considering the limitations of this technology.

Further description of Gene drive mechanisms are within Chapter 2, including sex ratio distortion, and transmission blocking. For comprehensive reviews of gene drives in vectors see Champer *et al.,* (2016), and Hammond and Galizi (2017).

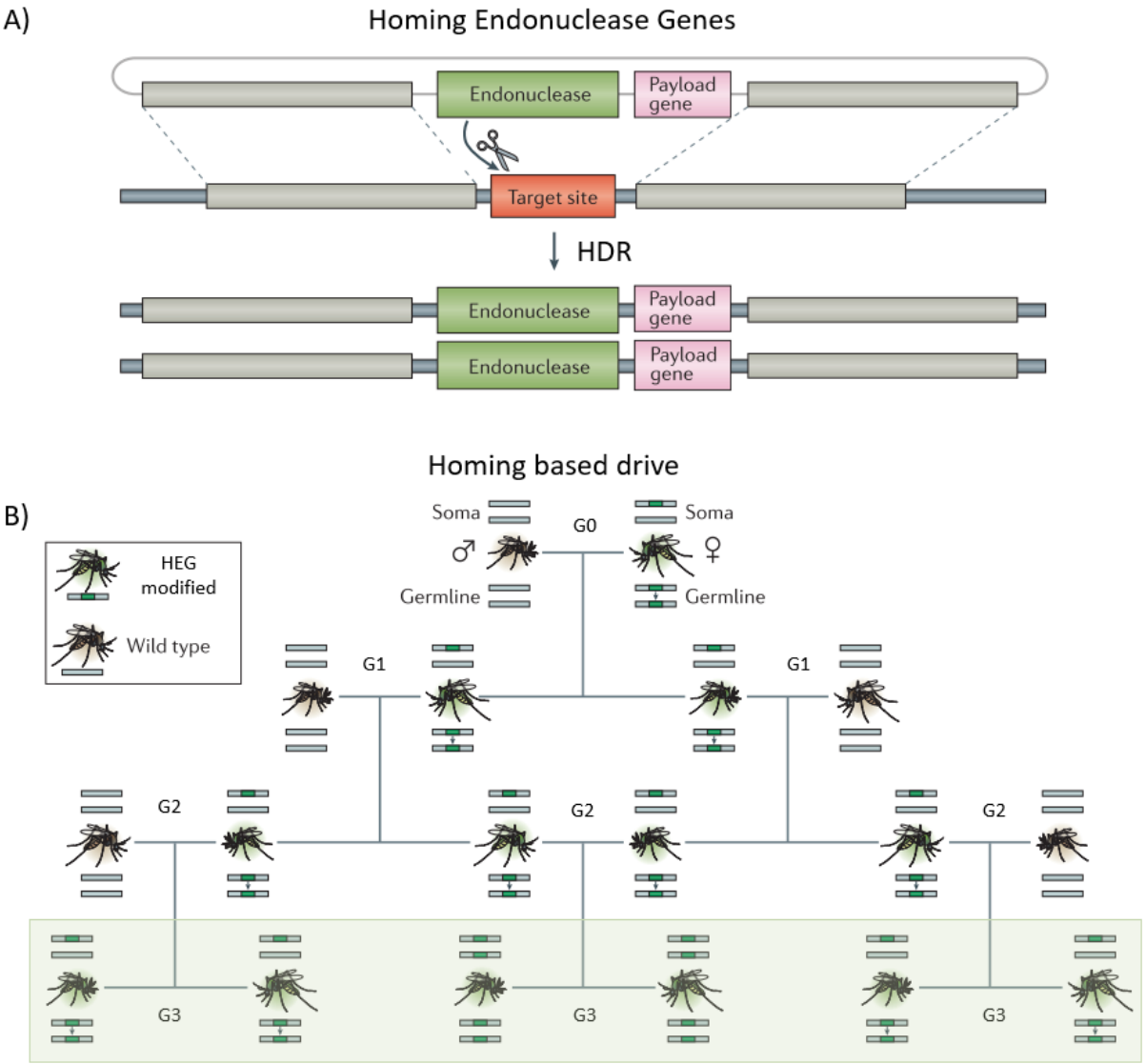


Figure 6. The mechanisms of homing endonuclease gene cleavage and repair, and spread of homing-based drives. A) homing endonuclease gene (HEG) cleave at a recognition site on the opposite chromosome of paired chromosomes. Cellular repair takes place and the presence of the intact chromosome provides a template for Homology directed repair (HDR), which leads to copying of the HEG to the homologous chromosome. B) The HEG mechanism can be used to develop a homing-based drive. The homing element is copied to both chromosomes, and inherited by offspring. Within relatively few generations the homing element can spread rapidly within the targeted population (green box). Adapted from Champer *et al.,* (2016).

Limitations

Despite their potential benefits, gene drives have limitations. An important consideration is the potential for the development of resistance to gene modifications. Over time, populations that are exposed to gene drive elements may evolve resistance to the gene drive through assortative mating, or unintended molecular repair mechanisms (NHEJ). In addition, alleles may already be present within wild vector populations containing single nucleotide polymorphisms (SNPs), which could prevent efficient homing of gene drive elements from occurring (A. M. Hammond & Galizi, 2017).

Although resistance is of concern, strategies have been suggested to limit the impact, such as multiplexing gRNAs to target multiple loci (Esvelt, Smidler, Catteruccia, & Church, 2014; Marshall, Buchman, Sánchez, & Akbari, 2017), or complex drive systems such as daisy chain gene drives which are unable to spread indefinitely within a population (Marshall et al., 2017; Noble et al., 2019). In addition, careful design of gene drive systems and analysis of experimental data, combined with modelling should be able to indicate if resistance is likely to develop, or have an impact on the functionality of the drive (Kyrou et al., 2018)

Another limitation of gene drives is the need for confinement and the ability to reverse the system. Because gene drives are designed to spread rapidly through a population, they can be difficult to contain or control once they are released into the environment. This can make it difficult to predict or control the long-term effects of gene drives, and it can raise safety concerns about the potential for harm to other species or the environment (Akbari et al., 2015).

To ensure responsible research guidelines have been developed regarding confinement, all gene drive work must follows the relevant biosafety regulations, research should be conducted within multiple levels of laboratory containment (Gantz & Bier, 2015), research should be conducted outside of the habitable range of the species prior to regulated field trials, and molecular separation of drive components should be pursued (Oye et al., 2014).

Reversal of gene drives is likely to be difficult once they are released into the population. While some gene drives are designed to be reversible, it is important to carefully consider the potential risks and benefits before implementing this technology. Other concerns include the potential to wipe out species, off-target effects and spread outside of intended areas (Knols, Bossin, Mukabana, & Robinson, 2007; Webber, Raghu, & Edwards, 2015).

Overall, researchers are working on a range of approaches to address the logistic and ethical concerns surrounding the use of gene drive technology for disease control. By developing new technologies and strategies for responsible use of this technology, researchers can help to ensure that it is used in a manner that is safe, effective, and ethical.

Novel approaches for control of sand fly vectors

As previously described, vector control for sand flies has primarily focussed on the use of insecticides via various modes of delivery (LLINs, IRS and ATSBs), however the emergence of resistance in Asia, South America, the Middle East, and Africa is of concern (Balaska, Fotakis, Chaskopoulou, & Vontas, 2021).

The recent capacity to precisely affect target genes, and the subsequent development of gene drives in other insects of medical importance presents a unique opportunity in the context of sand flies. Initial CRISPR-Cas9 gene editing has been attempted in *L. longipalpis* which has been unsuccessful to date (Martin-Martin, Aryan, Meneses, Adelman, & Calvo, 2018), but successful in a single publication with *P. papatasi*, inducing *Leishmania* refractoriness via immune pathway modification (Louradour, Ghosh, Inbar, & Sacks, 2019). These studies have established the opportunity for the development of novel genetic control in sand flies, including self-limiting strategies that supress populations, or alternatives that are self-sustaining. Here we present a pathway for gene editing in two sand fly species to facilitate progress towards the interruption of *Leishmania* transmission, which has so far been neglected.

Research Aims and Objectives

Aims

This project aims to develop a sand fly specific gene editing platform encompassing the bioinformatic process to identify gene targets to affect (non-lethal phenotypic markers and genes involved in olfaction), through to *in vitro* construction and assessment of gene editing components in cell lines. Finally translation of *in vitro* derived components taken forward to *in vivo* experiments in two sand fly species (*Lutzyomyia longipalpis* and *Phlebomtus papatasi*) at scale to demonstrate CRISPR-Cas and PiggyBac induced mutagenesis which has not been previously reported in these sand fly species. Results would be of considerable interest to the sand fly and insect research community.

Specific Objectives

Chapter 2

Review chapter of CRISPR-based gene editing methods applied to insect vectors with an emphasis of translation to sand fly vectors of *Leishmania* and prospects for affecting gene targets associated with olfaction. Produced during suspension of wet lab research due to the COVID pandemic.

Chapter 3

A Bioinformatics pipeline to identify and rationalise genes to target by CRISPR-Cas mediated gene editing. Chosen genes illicit a non-lethal phenotypic effect and genes involved in olfaction are identified for gene editing.

Chapter 4

Development of an *in vitro* platform to construct and validate CRISPR-Cas components and PiggyBac Plasmids in sand fly cell lines in readiness for *in vitro* experiments.

Chapter 5

In vivo mutagenesis of two sand fly species (*L. longipalpis* & *P. papatasi*) at scale by CRISPR-Cas and PiggyBac derived from Chapter 4.

References

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

Alexander, B., Barros, V. C., De Souza, S. F., Barros, S. S., Teodoro, L. P., Soares, Z. R., … Reithinger, R. (2009). Susceptibility to chemical insecticides of two Brazilian populations of the visceral leishmaniasis vector Lutzomyia longipalpis (Diptera: Psychodidae). *Tropical Medicine and International Health*, *14*(10), 1272–1277. https://doi.org/10.1111/j.1365-3156.2009.02371.x

Alexander, B., & Maroli, M. (2003). Control of phlebotomine sandflies. *Medical and Veterinary Entomology*, *17*(1), 1–18. https://doi.org/10.1046/j.1365-2915.2003.00420.x

Alvar, J., Croft, S. L., Kaye, P., Khamesipour, A., Sundar, S., & Reed, S. G. (2013). Case study for a vaccine against leishmaniasis. *Vaccine*, *31*(SUPPL2), B244–B249. https://doi.org/10.1016/j.vaccine.2012.11.080

Balaska, S., Fotakis, E. A., Chaskopoulou, A., & Vontas, J. (2021). Chemical control and insecticide resistance status of sand fly vectors worldwide. *PLoS Neglected Tropical Diseases*, *15*(8), 1–23. https://doi.org/10.1371/journal.pntd.0009586

Bekhit, A. A., El-Agroudy, E., Helmy, A., Ibrahim, T. M., Shavandi, A., & Bekhit, A. E. D. A. (2018). Leishmania treatment and prevention: Natural and synthesized drugs. *European Journal of Medicinal Chemistry*, *160*, 229–244. https://doi.org/10.1016/j.ejmech.2018.10.022

Bernardini, F., Kriezis, A., Galizi, R., Nolan, T., & Crisanti, A. (2019). Introgression of a synthetic sex ratio distortion system from Anopheles gambiae into Anopheles arabiensis. *Scientific Reports*, *9*(1), 4–11. https://doi.org/10.1038/s41598-019-41646-8

Burza, S., Croft, S. L., & Boelaert, M. (2018). Leishmaniasis. *The Lancet*, *392*(10151), 951–970. https://doi.org/10.1016/S0140-6736(18)31204-2

Carrión, J., Folgueira, C., Soto, M., Fresno, M., & Requena, J. M. (2011). Leishmania infantum HSP70-II null mutant as candidate vaccine against leishmaniasis: A preliminary evaluation. *Parasites and Vectors*, *4*(1), 1–10. https://doi.org/10.1186/1756-3305-4-150

Carvalho, D. O., McKemey, A. R., Garziera, L., Lacroix, R., Donnelly, C. A., Alphey, L., … Capurro, M. L. (2015). Suppression of a field population of Aedes aegypti in Brazil by sustained release of transgenic male mosquitoes. *PLoS Neglected Tropical Diseases*, *9*(7), 1–15. https://doi.org/10.1371/journal.pntd.0003864

Chaniotis, B. N., Parsons, R. E., Harlan, H. J., & Correa, M. A. (1982). A pilot study to control phlebotomine sand flies (Diptera: Psychodidae) in a neotropical rain forest. *Journal of Medical Entomology*, *19*(1), 1–5. https://doi.org/10.1093/jmedent/19.1.1

Charlab, R., Valenzuela, J. G., Rowton, E. D., & Ribeiro, J. M. C. (1999). Toward an understanding of the biochemical and pharmacological complexity of the saliva of a hematophagous sand fly Lutzomyia longipalpis. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(26), 15155–15160. https://doi.org/10.1073/pnas.96.26.15155

Chelbi, I., Zhioua, E., & Hamilton, J. G. C. (2011). Behavioral evidence for the presence of a sex pheromone in male Phlebotomus papatasi Scopoli (Diptera: Psychodidae). *Journal of Medical Entomology*, *48*(3), 518–525. https://doi.org/10.1603/ME10132

Chelbi, Ifhem, Bray, D. P., & Hamilton, J. G. C. (2012). Courtship behaviour of Phlebotomus papatasi the sand fly vector of cutaneous leishmaniasis. *Parasites and Vectors*, *5*(1), 1–9. https://doi.org/10.1186/1756-3305-5-179

Chen, J., Luo, J., Wang, Y., Gurav, A. S., Li, M., Akbari, O. S., & Montell, C. (2021). Suppression of female fertility in Aedes aegypti with a CRISPR-targeted male-sterile mutation. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(22), 1–8. https://doi.org/10.1073/pnas.2105075118

Chowdhury, R., Chowdhury, V., Faria, S., Akter, S., Dash, A. P., Bhattacharya, S. K., … Banu, Q. (2019). Effect of insecticide-treated bed nets on visceral leishmaniasis incidence in Bangladesh. A retrospective cohort analysis. *PLoS Neglected Tropical Diseases*, *13*(9), 1–14. https://doi.org/10.1371/journal.pntd.0007724

Dinesh, D. S., Das, M. L., Picado, A., Roy, L., Rijal, S., Singh, S. P., … Coosemans, M. (2010). Insecticide susceptibility of Phlebotomus argentipes in visceral leishmaniasis endemic districts in India and Nepal. *PLoS Neglected Tropical Diseases*, *4*(10), 1–5. https://doi.org/10.1371/journal.pntd.0000859

Dong, Y., Simões, M. L., Marois, E., & Dimopoulos, G. (2018). CRISPR/Cas9 -mediated gene knockout of Anopheles gambiae FREP1 suppresses malaria parasite infection. *PLoS Pathogens*, *14*(3), 1–16. https://doi.org/10.1371/journal.ppat.1006898

Dorlo, T. P. C., Huitema, A. D. R., Beijnen, J. H., & de Vries, P. J. (2012). Optimal Dosing of Miltefosine in Children and Adults with Visceral Leishmaniasis. *Antimicrobial Agents and Chemotherapy*, *56*(7), 3864–3872. https://doi.org/10.1128/aac.00292-12

Emami, M. M., Yazdi, M., & Guillet, P. (2009). Efficacy of Olyset long-lasting bednets to control transmission of cutaneous leishmaniasis in Iran. *Eastern Mediterranean Health Journal*, *15*(5), 1075–1083. https://doi.org/10.26719/2009.15.5.1075

Esterre, P., Chippaux, J. P., Lefait, J. F., & Dedet, J. P. (1986). Evaluation of a cutaneous leishmaniasis control program in a forest village of French Guyana. *Bulletin of the World Health Organization*, *64*(4), 559–565. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/3490925%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC2490890

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*(July2014), 1–21. https://doi.org/10.7554/eLife.03401

Faraj, C., Yukich, J., Adlaoui, E. B., Wahabi, R., Mnzava, A. P., Kaddaf, M., … Kleinschmidt, I. (2016). Effectiveness and cost of insecticide-treated bed nets and indoor residual spraying for the control of cutaneous leishmaniasis: A cluster-randomized control trial in Morocco. *American Journal of Tropical Medicine and Hygiene*, *94*(3), 679–685. https://doi.org/10.4269/ajtmh.14-0510

Galizi, R., Hammond, A., Kyrou, K., Taxiarchi, C., Bernardini, F., O'Loughlin, S. M., … Crisanti, A. (2016). A CRISPR-Cas9 sex-ratio distortion system for genetic control. *Scientific Reports*, *6*(April), 2–6. https://doi.org/10.1038/srep31139

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

Gantz, V. M., Tatarenkova, O., Macias, V. M., James, A. A., Fazekas, A., Bier, E., & Jasinskiene, N. (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. https://doi.org/10.1073/pnas.1521077112

Gong, P., Epton, M. J., Fu, G., Scaife, S., Hiscox, A., Condon, K. C., … Alphey, L. (2005). A dominant lethal genetic system for autocidal control of the Mediterranean fruitfly. *Nature Biotechnology*, *23*(4), 453–456. https://doi.org/10.1038/nbt1071

Grover, K. K., Curtis, C. F., Sharma, V. P., Singh, K. R. P., Dietz, K., Agarwal, H. V., … Vaidyanathan, V. (1976). Competitiveness of chemosterilised males and cytoplasmically incompatible translocated males of Culex pipiens fatigans wiedemann (Diptera, Culicidae) in the field. *Bulletin of Entomological Research*, *66*(3), 469–480. https://doi.org/10.1017/S0007485300006878

Gunay, F., Karakus, M., Oguz, G., Dogan, M., Karakaya, Y., Ergan, G., … Alten, B. (2014). Evaluation of the efficacy of Olyset® Plus in a village-based cohort study in the Cukurova Plain, Turkey, in an area of hyperendemic cutaneous leishmaniasis. *Journal of Vector Ecology*, *39*(2), 395–405. https://doi.org/10.1111/jvec.12115

Hall, A. B., Basu, S., Jiang, X., Qi, Y., Timoshevskiy, V. A., Biedler, J. K., … Tu, Z. (2015). A male-determining factor in the mosquito Aedes aegypti. *Science*, *348*(6240), 1268–1270. https://doi.org/10.1126/science.aaa2850

Hammond, A., Galizi, R., Kyrou, K., Simoni, A., Siniscalchi, C., Katsanos, D., … Nolan, T. (2016). A CRISPR-Cas9 gene drive system targeting female reproduction in the malaria mosquito vector Anopheles gambiae. *Nature Biotechnology*, *34*(1), 78–83. https://doi.org/10.1038/nbt.3439

Hammond, A. M., & Galizi, R. (2017). Gene drives to fight malaria: current state and future directions. *Pathogens and Global Health*, *111*(8), 412–423. https://doi.org/10.1080/20477724.2018.1438880

Harris, A. F., McKemey, A. R., Nimmo, D., Curtis, Z., Black, I., Morgan, S. A., … Alphey, L. (2012). Successful suppression of a field mosquito population by sustained release of engineered male mosquitoes. *Nature Biotechnology*, *30*(9), 828–830. https://doi.org/10.1038/nbt.2350

Harris, A. F., Nimmo, D., McKemey, A. R., Kelly, N., Scaife, S., Donnelly, C. A., … Alphey, L. (2011). Field performance of engineered male mosquitoes. *Nature Biotechnology*, *29*(11), 1034–1037. https://doi.org/10.1038/nbt.2019

Helinski, M. E. H., Parker, A. G., & Knols, B. G. J. (2006). Radiation-induced sterility for pupal and adult stages of the malaria mosquito Anopheles arabiensis. *Malaria Journal*, *5*, 1–10. https://doi.org/10.1186/1475-2875-5-41

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*(1–5), 371–377. https://doi.org/10.1111/j.1439-0418.1995.tb01303.x

Hoermann, A., Habtewold, T., Selvaraj, P., Del Corsano, G., Capriotti, P., Inghilterra, M. G., … Windbichler, N. (2022). Gene drive mosquitoes can aid malaria elimination by retarding Plasmodium sporogonic development. *Science Advances*, *8*(38), 1–9. https://doi.org/10.1126/sciadv.abo1733

Kamhawi, S., Aslan, H., & Valenzuela, J. G. (2014). Vector saliva in vaccines for visceral leishmaniasis: A brief encounter of high consequence? *Frontiers in Public Health*, *2*(AUG), 1–6. https://doi.org/10.3389/fpubh.2014.00099

Kandul, N. P., Liu, J., Buchman, A., Shriner, I. C., Corder, R. M., Warsinger-Pepe, N., … Akbari, O. S. (2022). Precision Guided Sterile Males Suppress Populations of an Invasive Crop Pest. *GEN Biotechnology*, *1*(4), 372–385. https://doi.org/10.1089/genbio.2022.0019

Kandul, N. P., Liu, J., Sanchez C, H. M., Wu, S. L., Marshall, J. M., & Akbari, O. S. (2019). Transforming insect population control with precision guided sterile males with demonstration

in flies. *Nature Communications*, *10*(1), 1–12. https://doi.org/10.1038/s41467-018-07964-7

Kelly, D. W., & Dye, C. (1997). Pheromones, kairomones and the aggregation dynamics of the sandfly Lutzomyia longipalpis. *Animal Behaviour*, *53*(4), 721–731. https://doi.org/10.1006/anbe.1996.0309

Khan, K. H. (2013). DNA vaccines: Roles against diseases. *Germs*, *3*(1), 26–35. https://doi.org/10.11599/germs.2013.1034

Knols, B. G. J., Bossin, H. C., Mukabana, W. R., & Robinson, A. S. (2007). Transgenic mosquitoes and the fight against malaria: Managing technology push in a turbulent GMO world. *American Journal of Tropical Medicine and Hygiene*, *77*(SUPPL. 6), 232–242. https://doi.org/10.4269/ajtmh.2007.77.232

Krafsur, E. S. (1998). Sterile insect technique for suppressing and eradicating insect population: 55 years and counting. *Journal of Agricultural and Urban Entomology*, *15*(4), 303–317.

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

Lawyer, P., Volf, P., Rowton, E., Rowland, T., Killick-Kendrick, M., Rowland, T., … Volf, P. (2017). Laboratory colonization and mass rearing of phlebotomine sand flies (Diptera, Psychodidae). *Parasite*, *24*, 42. https://doi.org/10.1051/parasite/2017041

Li, M., Yang, T., Bui, M., Gamez, S., Wise, T., Kandul, N. P., … Akbari, O. S. (2021). Suppressing mosquito populations with precision guided sterile males. *Nature Communications*, *12*(1). https://doi.org/10.1038/s41467-021-25421-w

Louradour, I., Ghosh, K., Inbar, E., & Sacks, D. L. (2019). CRISPR/Cas9 Mutagenesis in Phlebotomus papatasi: the Immune Deficiency Pathway Impacts Vector Competence for Leishmania major. *MBio*, *10*(4), 1–14.

Maroli, M., Feliciangeli, M. D., Bichaud, L., Charrel, R. N., & Gradoni, L. (2013). Phlebotomine sandflies and the spreading of leishmaniases and other diseases of public health concern. *Medical and Veterinary Entomology*, *27*(2), 123–147. https://doi.org/10.1111/j.1365-2915.2012.01034.x

Marshall, J. M., Buchman, A., Sánchez, C. H. M., & Akbari, O. S. (2017). Overcoming evolved resistance to population-suppressing homing-based gene drives. *Scientific Reports*, *7*(1), 1–12. https://doi.org/10.1038/s41598-017-02744-7

Martin-Martin, I., Aryan, A., Meneses, C., Adelman, Z. N., & Calvo, E. (2018). Optimization of sand fly embryo microinjection for gene editing by CRISPR/Cas9. *PLoS Neglected Tropical Diseases*, *12*(9), 1–18. https://doi.org/10.1371/journal.pntd.0006769

Morrison, A. C., Ferro, C., Morales, A., Tesh, R. B., & Wilson, M. L. (1993). Dispersal of the sand fly Lutzomyia longipalpis (Diptera: Psychodidae) at an endemic focus of visceral leishmaniasis in Colombia. *Journal of Medical Entomology*, *30*(2), 427–435. https://doi.org/10.1093/jmedent/30.2.427

Morton, I. E., & Ward, R. D. (1989). Laboratory response of female Lutzomyia longipalpis sandflies to a host and male pheromone source over distance. *Medical and Veterinary Entomology*, *3*(0), 219.

Müller, G. C., & Schlein, Y. (2011). Different methods of using attractive sugar baits (ATSB) for the control of Phlebotomus papatasi. *Journal of Vector Ecology*, *36*(SUPPL.1), 64–70. https://doi.org/10.1111/j.1948-7134.2011.00113.x

Navarro-Payá, D., Flis, I., Anderson, M. A. E., Hawes, P., Li, M., Akbari, O. S., … Alphey, L. (2020). Targeting female flight for genetic control of mosquitoes. *PLoS Neglected Tropical Diseases*, *14*(12), e0008876. https://doi.org/10.1371/journal.pntd.0008876

Noble, C., Min, J., Olejarz, J., Buchthal, J., Chavez, A., Smidler, A. L., … Esvelt, K. M. (2019). Daisy-chain gene drives for the alteration of local populations. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(17), 8275–8282. https://doi.org/10.1073/pnas.1716358116

O'leary, S., & Adelman, Z. N. (2020). Crispr/cas9 knockout of female-biased genes aeact-4 or myo-fem in ae. Aegypti results in a flightless phenotype in female, but not male mosquitoes. *PLoS Neglected Tropical Diseases*, *14*(12), 1–19. https://doi.org/10.1371/journal.pntd.0008971

Palit, A., Bhattacharya, S. K., & Kundu, S. N. (2005). Host preference of Phlebotomus argentipes and Phlebotomus papatasi in different biotopes of West Bengal, India. *International Journal of Environmental Health Research*, *15*(6), 449–454. https://doi.org/10.1080/09603120500392525

Phuc, H., Andreasen, M. H., Burton, R. S., Vass, C., Epton, M. J., Pape, G., … Alphey, L. (2007). Late-acting dominant lethal genetic systems and mosquito control. *BMC Biology*, *5*, 1–11. https://doi.org/10.1186/1741-7007-5-11

Picado, A., Singh, S. P., Rijal, S., Sundar, S., Ostyn, B., Chappuis, F., … Boelaert, M. (2011). Longlasting insecticidal nets for prevention of Leishmania donovaniinfection in India and Nepal: Paired cluster randomised trial. *Bmj*, *342*(7788), 92. https://doi.org/10.1136/bmj.c6760

Pigott, D. M., Bhatt, S., Golding, N., Duda, K. A., Battle, K. E., Brady, O. J., … Hay, S. I. (2014). Global distribution maps of the leishmaniases. *ELife*, *3*, 1–21. https://doi.org/10.7554/elife.02851

Qualls, W. A., Müller, G. C., Khallaayoune, K., Revay, E. E., Zhioua, E., Kravchenko, V. D., … Beier, J. C. (2015). Control of sand flies with attractive toxic sugar baits (ATSB) and potential impact on non-target organisms in Morocco. *Parasites and Vectors*, *8*(1), 1–9. https://doi.org/10.1186/s13071-015-0671-2

Quinnell, R. J., & Courtenay, O. (2009). Transmission, reservoir hosts and control of zoonotic visceral leishmaniasis. *Parasitology*, *136*(14), 1915–1934. https://doi.org/10.1017/s0031182009991156

Quinnell, R. J., Dye, C., & Shaw, J. J. (1992). Host preferences of the phlebotomine sandfly Lutzomyia longipalpis in Amazonian Brazil. *Medical and Veterinary Entomology*, *6*(3), 195–200. https://doi.org/10.1111/j.1365-2915.1992.tb00606.x

Ready, P. D. (1979). Factors affecting egg production of laboratory-bred Lutzomyia longipalpis (Diptera: Psychodidae). *Journal of Medical Entomology*, *16*(5), 413–423. https://doi.org/10.1093/jmedent/16.5.413

Ready, P. D., Arias, J. R., & Freitas, R. A. (1985). A pilot study to control Lutzomyia umbratilis (Diptera:Psychodidae), the major vector of Leishmania braziliensis guyanensis, in a peri-urban rainforest of Manaus, Amazonas State, Brazil. *Memórias Do Instituto Oswaldo Cruz*, Vol. 80, pp. 27–36. https://doi.org/10.1590/s0074-02761985000100005

Ribeiro, J. M. C., Katz, O., Pannell, L. K., Waitumbi, J., & Warburg, A. (1999). Salivary glands of the sand fly Phlebotomus papatasi contain pharmacologically active amounts of adenosine and 5'-AMP. *Journal of Experimental Biology*, *202*(11), 1551–1559. https://doi.org/10.1242/jeb.202.11.1551

Saghafipour, A., Vatandoost, H., Zahraei-Ramazani, A. R., Yaghoobi-Ershadi, M. R., Rassi, Y., Jooshin, M. K., … Akhavan, A. A. (2017). Control of zoonotic cutaneous leishmaniasis vector, Phlebotomus papatasi, using attractive toxic sugar baits (ATSB). *PLoS ONE*, *12*(4), 1–14.

https://doi.org/10.1371/journal.pone.0173558

Simoni, A., Hammond, A. M., Beaghton, A. K., Galizi, R., Taxiarchi, C., Kyrou, K., … Crisanti, A. (2020). A male-biased sex-distorter gene drive for the human malaria vector Anopheles gambiae. *Nature Biotechnology*, *38*(9), 1054–1060. https://doi.org/10.1038/s41587-020-0508-1

Smidler, A. L., Pai, J. J., Apte, R. A., Sánchez, H. M., Corder, R. M., Jeffrey Gutiérrez, E., … Affiliations, †. (2022). A confinable female-lethal population suppression system in the malaria vector, Anopheles gambiae. *BioRxiv*, 2022.08.30.505861. Retrieved from https://www.biorxiv.org/content/10.1101/2022.08.30.505861v1%0Ahttps://www.biorxiv.org/content/10.1101/2022.08.30.505861v1.abstract

Soudi, S., Hosseini, A. Z., & Hashemi, S. M. (2011). Co-administration of rectal BCG and autoclaved Leishmania major induce protection in susceptible BALB/c mice. *Parasite Immunology*, *33*(10), 561–571. https://doi.org/10.1111/j.1365-3024.2011.01318.x

Spinner, S. A. M., Barnes, Z. H., Puinean, A. M., Gray, P., Dafa'alla, T., Phillips, C. E., … Matzen, K. J. (2022). New self-sexing Aedes aegypti strain eliminates barriers to scalable and sustainable vector control for governments and communities in dengue-prone environments. *Frontiers in Bioengineering and Biotechnology*, *10*(October), 1–18. https://doi.org/10.3389/fbioe.2022.975786

Sundar, S., & Chakravarty, J. (2010). Antimony toxicity. *International Journal of Environmental Research and Public Health*, *7*(12), 4267–4277. https://doi.org/10.3390/ijerph7124267

Thomas, D. D., Donnelly, C. A., Wood, R. J., & Alphey, L. S. (2000). Insect population control using a dominant, repressible, lethal genetic system. *Science*, *287*(5462), 2474–2476. https://doi.org/10.1126/science.287.5462.2474

Todolí, F., Rodríguez-Cortés, A., Núñez, M. del C., Laurenti, M. D., Gómez-Sebastián, S., Rodríguez, F., … Alberola, J. (2012). Head-to-Head Comparison of Three Vaccination Strategies Based on DNA and Raw Insect-Derived Recombinant Proteins against Leishmania. *PLoS ONE*, *7*(12). https://doi.org/10.1371/journal.pone.0051181

Vreysen, M. J. B., Saleh, K. M., Ali, M. Y., Abdulla, A. M., Zhu, Z. R., Juma, K. G., … Feldmann, H. U. (2000). Glossina austeni (diptera: glossinidae) eradicated on the Island of Unguja, Zanzibar, using the sterile insect technique. *Journal of Economic Entomology*, *93*(1), 123–133. https://doi.org/10.1603/0022-0493-93.1.123

Webber, B. L., Raghu, S., & Edwards, O. R. (2015). Opinion: Is CRISPR-based gene drive a biocontrol silver bullet or global conservation threat? *Proceedings of the National Academy of Sciences of the United States of America*, *112*(34), 10565–10567. https://doi.org/10.1073/pnas.1514258112

WHO. (2017). WHO: Weekly epidemiological record: Global leishmaniasis update, 2006-2015, a turning point in leishmaniasis surveillance. *World Health Organization*, *92*(38), 557–572. https://doi.org/10.1186/1750-9378-2-15.Voir

Chapter 2

CRISPR-based modification of olfaction in insect vectors: Prospects for disease control and translation to sand fly vectors of *Leishmania*

Introduction

Olfaction in insects is implicated in fundamental behaviours spanning mating, host seeking, blood feeding, sugar feeding, and oviposition. Host seeking in hematophagous insects of medical importance is of particular interest, as it presents a potential target for interrupting transmission of associated pathogens through gene editing. An increasing body of evidence has identified different stimuli and cues important for host seeking in phylogenetically distinct classes of hematophagous insects (Carey & Carlson, 2011). Typically, blood feeding disease vectors including female mosquitoes and sand flies detect volatile compounds from the target host through receptors in the antenna and maxillary palps, which comprise the major olfactory organs. These relevant odour cues interact with olfactory appendages and receptors leading to host-seeking behaviour (Suh, Bohbot, & Zwiebel, 2014).

Recently the links between host seeking behaviours, specific host odour cues, and receptors have been elucidated with the aid of genetic modification strategies. The majority of this research alludes to the potential for these discoveries to inform vector control strategies via manipulation of sensory perception towards humans by the disruption of receptor genes (Konopka et al., 2021; McMeniman, Corfas, Matthews, Ritchie, & Vosshall, 2014; Sun, Liu, Ye, Baker, & Zwiebel, 2020; Y. Wang et al., 2022). Targeting of important olfactory genes in this way could lead to population replacement strategies that inhibit host detection, through gene drive, with the potential to interrupt disease transmission by reducing human-vector contact.

The application of highly targeted gene editing technologies including CRISPR-based approaches has broadened the potential to investigate gene function by knockout, insertion, and expression of specific gene targets including, pigmentation, wing development, fluorescence, sex determination, and insecticide resistance (Dong et al., 2015; Gratz et al., 2013; Hall et al., 2015; Hammond et al., 2016; Itokawa et al., 2016; Port et al., 2014). In the context of interruption of disease transmission, CRISPR-based gene drives have been developed that efficiently drive targeted traits through fast reproducing populations, with laboratory outcomes showing great promise (Galizi et al., 2016; Hammond et al., 2016, 2021; Hoermann et al., 2022; Kyrou et al., 2018; Simoni et al., 2020). In principle, these approaches are applicable to a wide variety of invertebrates of medical and veterinary importance including sand flies.

This review chapter summarises fundamental aspects of olfaction in insect vectors amenable to CRISPR-based approaches to target genes of interest, with the prospect of reducing disease transmission in sand fly vectors of Leishmania. In more detail, olfactory stimuli, olfactory structures, associated receptors, and candidate olfactory genes that might be affected by modification are described. Comment is made on the likely response to editing of genes implicated in the context of vector behaviour. Finally,

prospects for CRISPR strategies targeting olfaction and host-seeking in sand fly vectors of leishmaniasis implemented through gene drives are discussed.

Olfaction in insect vectors of medical importance

Vectors of medical importance are known to respond to different olfactory cues relating to breeding sites, food sources and host seeking. The genetic basis of stimuli influencing host-seeking is complex (DeGennaro et al., 2013; McBride et al., 2014). In anthropophilic hematophagous insects such as mosquitoes and sand flies, host-seeking cues comprise a human-specific combination of compounds emitted from the skin and via exhalation. More specifically, kairomones (chemical compounds released by an animal that evokes a behavioural response from the insect vector (Coutinho-Abreu, Riffell, & Akbari, 2021)) are produced when odourless sweat secreted by eccrine, apocrine, and sebaceous glands is converted by cutaneous microorganisms into a range of volatile organic compounds (Table 3). Key molecules implicated in attraction include CO_2, lactic acid, 1-octen-3-ol, and ammonia, with CO_2 being the major kairomone produced by vertebrates which activates resting mosquitoes and synergises with other volatile organic compounds (VOCs) (Andrade, Andrade, Dias, Pinto, & Eiras, 2008; Cork & Park, 1996; Coutinho-Abreu et al., 2021; Da Silva Tavares et al., 2018; Dormont, Bessière, & Cohuet, 2013; Lu et al., 2007; Magalhães-Junior et al., 2019; Mann, Kaufman, & Butler, 2009; Smallegange, Verhulst, & Takken, 2011; Takken & Knols, 1999). These odours are described in more detail below (see Stimuli, Receptors and Genes).

The main olfactory sensory organs in mosquitoes involved in detection of VOCs are antennae, maxillary palps and labella, which have common features including hair-like structures (sensilla) on their surface, housing dendrites of olfactory neurons surrounded by lymph (Konopka et al., 2021; Potter, 2014). The pattern of sensilla distribution across the sensory appendages depends on species, gender, age and, life stage, correlating to the importance of olfactory cues to different life history traits (Konopka et al., 2021; Suh et al., 2014). Sensilla interact with external stimuli via three main olfactory receptor types expressed on the dendrites: odorant receptors (ORs), ionotropic receptors (IRs), and gustatory receptors (GRs). Odorants enter sensilla via pores, and activate ORs, IRs, and GRs (Figure 8). Electrical signals are transmitted along the olfactory receptor neurons (ORNs) which join the antennal nerve and project to the antennal lobe, where olfactory information is processed in the brain. The antennal lobe consists of spheroid structures, glomeruli, which project neurons to higher brain areas including the mushroom body and lateral horn (linked to olfactory learning/memory, and innate olfactory behaviour respectively)(Masse, Turner, & Jefferis, 2009). Odour coding (encoding the identity of odours via the receptors and neurons involved) and Central processing are complex and comprehensively reviewed elsewhere (Guidobaldi, May-Concha, & Guerenstein, 2014).

Targeted editing in genes implicated in olfaction to modify, replace, and alter gene expression can drastically influence host seeking behaviour in mosquito vectors (described in more detail below)

(Carey & Carlson, 2011; DeGennaro et al., 2013; Lu et al., 2007; McMeniman et al., 2014). The majority of research has been conducted in mosquito species of medical importance, with far less research on olfaction in sand flies. However, both mosquitoes and sand flies are dipteran, displaying phylogenetic similarities with respect to key genes identified as important in host seeking behaviour (Figure 1), increasing the probability of common/comparable function in orthologues.

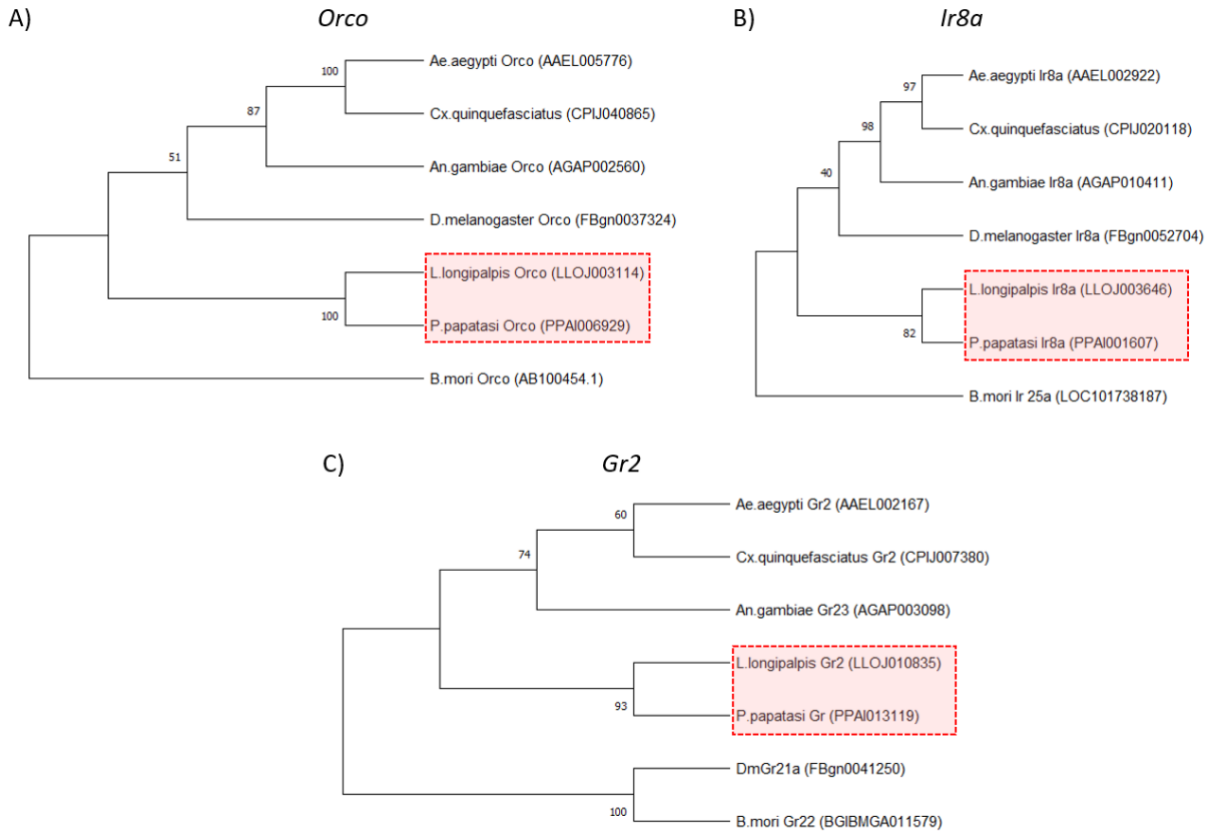


Figure 1. Orco (A), Ir8a (B) and Gr2 (C) orthologue phylogenies were estimated using *L. longipalpis*, *P. papatasi*, *Ae.aegypti*, *An. gambiae*, *Cx. quinquefasciatus*, *D. melanogaster* and *B. mori* protein sequences aligned with Muscle. Resulting phylogenies were visualised in MEGA X. Maximum likelihood (ML) method and JTT matrix-based model was used for tree estimation. The tree is rooted at the branch leading to Orco, Ir8a, and Gr2 (A, B, and C respectively). Branch support based on 500 bootstrap replications. Sequences were accessed from VectorBase, FlyBase and NCBI.

Olfactory structures and organs

Antennae

The antennae of *Anopheles* and *Culex* mosquitoes comprise 13 flagellomeres (segments) and the Johnson's organ at the junction with the head (McIver SB. 1982; Konopka *et al.* 2021). The antennae of females are covered in five types of sensilla with different functions (**Error! Reference source not found.**). In more detail, antennal sensilla Chaetica (sturdy hairs arising from sockets) are thought to have mechanosensory functions, sensilla Ampullacea (a peg located within a cuticular tube projecting inwards from the cuticle) have thermosensory function, alongside Coeloconic sensilla (a peg located in a cuticular pit), which have both thermosensory and olfactory functions (McIver, 1982), and are likely to express IRs. Trichoid sensilla are considered the main olfactory sensilla. They are the most abundant sensilla type, and mostly express ORs, and in some cases IRs (Riabinina et al., 2016). Grooved peg sensilla have two subtypes with olfactory and humidity sensing functions (Konopka et al., 2021)(Hill, Hansson, & Ignell, 2009).

In sand flies, gross antennal morphology differs substantially from mosquitoes**.** The structure and putative function of the antennae of *Lutzomyia longipalpis (L. longipalpis)* (Fernandes et al., 2008)*, Phlebotomus argentipes* (Ilango, 2000)*, and Phlebotomus duboscqi* (Bahia et al., 2021) have been elucidated primarily by Scanning Electron Microscopy (SEM). Broadly, antennal structure consists of a scape, a pedicel, and 14 flagellomeres that make up the antennae (Error! Reference source not found.). In *L. longipalpis* several types of sensilla have been identified: Chaetic sensilla, found on both male and female antennae are responsible for detection of olfactory stimuli for oviposition (in females) and semiochemicals such as sex pheromones. Grooved coeloconic sensilla (porous olfactory sensilla containing longitudinal grooves) detect volatile pheromones and kairomones. As in mosquitoes , sexual dimorphism relating to the number and distribution of sensilla has been described and is in part due to the requirement of different behavioural responses to different stimuli associated with the divergent feeding behaviour between the sexes (Fernandes et al., 2008).

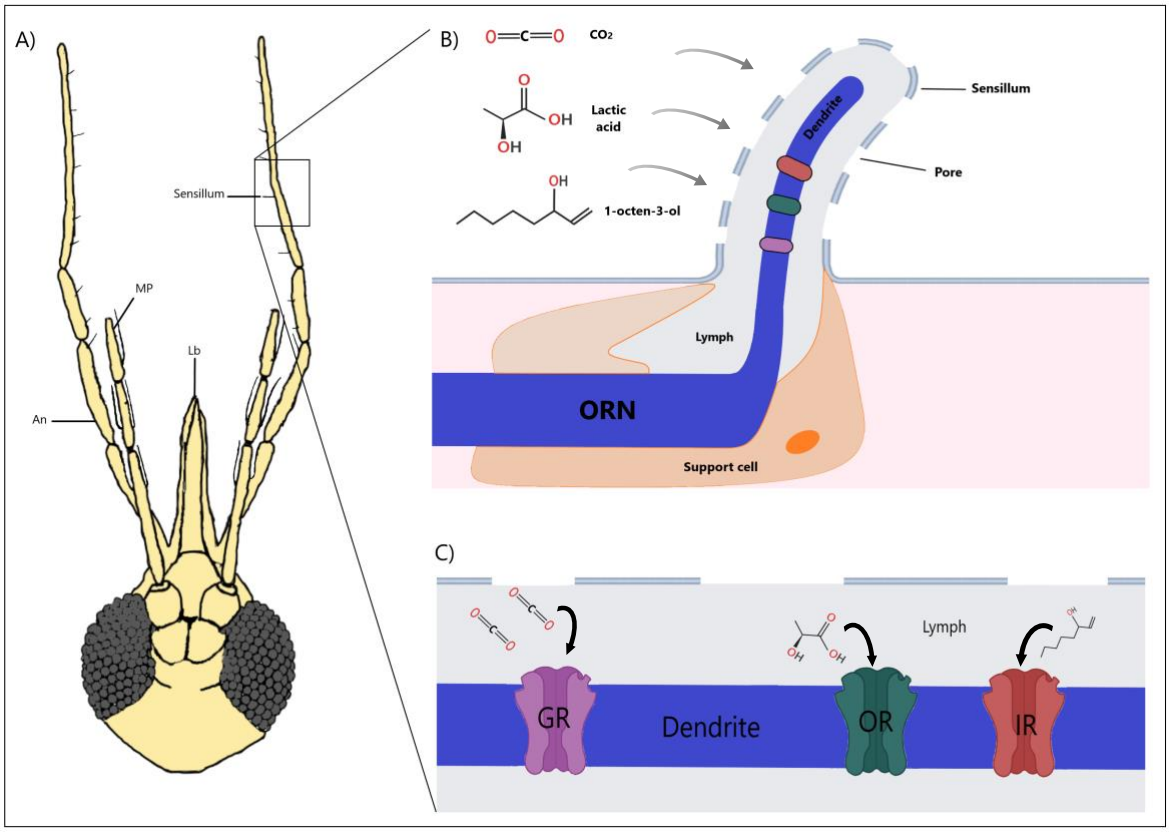


Figure 2. Illustration of vector olfaction from sensory appendages to receptors. A) A ventral schematic view of a sand fly head showing olfactory structures: Antennae (An), Maxillary palps (MP), and Labellum (Lb). Antennae and maxillary palps are covered in hair-like structure called sensilla (shown in box). B) Attractive odours such as CO_2, lactic acid and 1-octen-3-ol enter into the olfactory appendages through pores in the sensilla, crossing the lymph. C) Odours bind to specific receptor types on the dendrite of the olfactory receptor neuron (ORN). Three main receptor types are shown - Gustatory receptor (GR), Olfactory receptor (OR), and ionotropic receptor (IR).

Table 1. Comparison of Antennal sensilla in adult mosquito and sand fly vectors. Demonstrating the type, distribution, and function of the sensilla. Mosquito flagellomeres are denoted by F1-13, and sand fly flagellomeres are denoted by FI-XIV. Key: Short sharp-tipped Trichoid (SST), Long blunt-tipped Trichoid (LBT), Praying Hands Coeloconic (PHC), Medium pointed- tipped Trichoid (MPTT), Short prominent-base Trichoid (SPB), Short fixed-base Trichoid (SFB), Apical Trichoid (AT), Short Trichoid (ST), Medium Trichoid (MT), Long Pointed-tipped Trichoid (LPT).

Genus /species	segments	Sensilla types	Sensilla function/ Hypothesized	Sensilla organisation	Sexual dimorphism	References
Anopheles	13 flagellomeres (F1-13)	Chaetica Large Small	Mechanosensory		Male Olfactory sensilla located on two distal segments. Females have Olfactory sensilla on F2-13.	McIver SB. 1982 Konopka et al. 2021 Riabinina et al. 2016 Pitts and Zwiebel 2006
		Ampullacea	Thermosensory			
		Coeloconic Small Large	 Thermosensory Olfactory			
		Trichoid Sharp Blunt	Olfactory	 F1, F12, F13 F1-9		
		Grooved peg Subtype A Subtype B	Olfactory and/or Humidity	 F2-13 F3-13		
Culex quinquefasciatus	13 flagellomeres (F1-13)	Coeloconica	Hygrosensory Thermosensory	F13	N/A Only female observed	Hill, Hansson, & Ignell, 2009
		Chaetica Long Short	Mechanosensory	F1-13		
		Ampullacea	Hygrosensory Thermosensory	 F1-2		

		Trichoid LST SST SST-C SBT I SBT II	Olfactory	F1-13		
		Grooved pegs	Olfactory	F1-13		
Lutzomyia longipalpis	Scape Pedicel 14 flagellomeres (FI-XIV)	Chaetica	Olfactory	Scape:MT, LBT, Campaniform; Pedicel:MT, LBT, Squamiform; FI:ST,basiconic, LBT, Squamiform, Chaetic, PHC; FII:LBT, Chaetic, PHC; FIII:LBT, Chaetic, PHC; FIV:LBT, Chaetic; FV:LBT, Chaetic; FVI:LBT, Chaetic;	ST only on FIX, XI, XIII, XIV in males. Overall number of sensilla is greater in males.	Fernandes et al., 2008
		Trichoid ST MT LPT AT				
		Coeloconic				
		Grooved Peg Praying hands	Olfactory			
		Campaniform				
		Basiconic				
		Squamiform				

				FVII:LBT, Chaetic; FVIII:LBT, Chaetic; FIX: ST, LBT; FXI:ST, LBT; FXII:LBT, Grooved peg; FXIII:ST, LBT, LPT, Grooved peg; FXIV:LPT, Grooved peg, AT		
Phlebotomus argentipes	Scape Pedicel 14 flagellomeres (FI-XIV)	Trichoid ST LT	Olfactory	Scape: ST, Auricillicum; Pedicel:ST, Auricillicum; FI-XIV:ST, LT, Basiconica, Auricillicum (sparse), Coeloconic (increasing number towards	Different number and pattern. Males have more Coeloconica at distal flagellomeres. Different numbers of Chaetica.	Ilango, 2000
		Basiconic	Olfactory			
		Auricillic				
		Coeloconic	Olfaction Thermoreception			
		Chaetica	Mechanosensory			

				distal flagella)		
Phlebotomus duboscqi	Scape Pedicel 14 flagellomeres (FI-XIV)	Trichoid (7 types)	Olfactory	Scape:SST, LBT; Pedicel: SST, LBT, Squamiform, Campaniform; FI: LBT, SST, SBT, Basiconic, PHC, Chaetica, Squamiform; FII-III:LBT, PHC, Chaetica; FIV-VIII:LBT Chaetica; FIX: LBT, SPB, Chaetica; FX: LBT, MPT, Chaetica; FXI: LBT SPB, MPT, Chaetic; FXII:LBT, MPT,	N/A Only female observed.	Bahia et al, 2021
		Squamiform	Olfactory			
		Campaniform	mechanosensory			
		Basiconic	Olfactory			
		Coeloconic (2 types)	Olfactory			
		Chaetica				

		Grooved Coeloconic, Chaetica; FXIII:LBT, MPT, SPB,Grooved Coeloconic; FXIV: LBT, MPT, SFB, AT, Grooved Coeloconic.

Maxillary palps

Maxillary palps are also involved in insect vector olfaction. As a brief overview of morphology, the maxillary palps of *Anopheles* and *Aedes* comprise five segments, and are sexually dimorphic with respect to number and distribution of sensory structures, likely due to difference in attraction to food sources (Bohbot, Sparks, & Dickens, 2014; Mclver, 1982). The palps of *Anopheles* have three type of sensilla: the capitate peg sensillum (cp) have olfactory function, and Campaniform sensilla and Chaetica sensilla both have mechanosensory function (responsive to deformation of the cuticle, and to air currents or touch, respectively) (Lu et al., 2007; Mclver, 1982; Riabinina et al., 2016). Each cp has three olfactory neurons (cpA, cpB, and cpC) which respond to different compounds. cpA expresses GRs, responding to CO_2 (Lu et al., 2007), cpB and cpC express *Orco* and other tuning ORs, responding to 1-octen-3-ol, and other volatiles (Lu et al., 2007; Riabinina et al., 2016).

Aedes aegypti (*Ae.aegypti*) also have three types of sensilla: capitate sensilla basiconica (olfactory), non-porous Campaniform sensilla and Chaetica sensilla (mechanosensory). The capitate sensilla basiconica have three sensory neurons *A*, *B*, *C*. Neuron *A* responds to CO_2, via activation of *GR1* and *Gr3* receptors. Neuron *C* is activated by 1-octen-3-ol, via *OR8-Orco* receptors. The function of stimulus of neuron *B* are unknown (Bohbot et al., 2014).

Sand fly maxillary palps demonstrate similar morphology to mosquitoes. *L. longipalpis* maxillary palps also have five segments, containing a different combination of sensilla compared to mosquitoes: Campaniform sensilla, multiporous capitate peg sensilla, and trichoid sensilla. Capitate peg sensilla are

implicated in olfaction, and staining methods have indicated that the trichoid sensilla contain multiple pores, suggesting an olfactory function (Spiegel, Oliveira, Brazil, & Soares, 2005).

Currently there is sparse research with respect to sand fly maxillary palp involvement in olfaction compared to mosquito vectors. The identification of olfactory sensilla that express receptors involved in host detection is important to note, as there is homology in the sensilla types found in *L. longipalpis* and mosquito vector species. These similarities present potential targets for interruption of host-seeking behaviours via knockouts or alternative modifications.

Table 2. Maxillary palp sensilla in adult mosquito and sand fly vectors. Demonstrating the type, distribution, and function of the sensilla. Mosquito and sand fly maxillary palp segments are denoted by numerals I-V.

Species	segments	Sensilla types	Sensilla function/Hypothesized	Sensilla organisation	Sexual dimorphism	References
Anopheles gambiae	5 (I-V)	Capitate peg sensilla	Olfactory	Ventral on segments II, III, IV	Females have ~67 sensilla, males have ~14. Capitate peg sensilla only on segment IV in males	Lu T et al. 2007; Riabinina et al 2016; McIver 1982
		Capitate peg sensilla	Olfactory			
		Campaniform sensilla	Mechanosensory			
		Chaetica	Mechanosensory			
Aedes aegypti	5 (I-V)	Chaetica	Mechanosensory	Dorso-lateral segment IV		Bohbot et al 2014

		Capitate sensilla basiconica	Olfactory	Dorso-lateral segment IV		
		Campaniform	Proprioceptor	Distal end segment IV		
Lutzomyia longipalpis	5 (I-V)	Campaniform	Mechanosensory	Distal segment II;1-2 on segment III.	No sexual dimorphism	Spiegel et al 2005
		Capitate peg	Olfactory	1-6 segment II. Up to 25 distal end segment III.		
		Trichoid	Olfactory	1 on segment III; Many on segment IV and V.		

Proboscis (labella)

Another important structure with olfactory involvement include mouthparts, comprised of the labium (sheath covering the stylets, with labellar lobes at the tip), maxillae, mandibles, labrum (food canal), and hypopharynx. Of these, the Labellar lobes are the only feature that have been linked to olfaction in Anopheles mosquitoes (Kwon, Lu, Rützler, & Zwiebel, 2006; Raji et al., 2019; Riabinina et al., 2016; Saveer, Pitts, Ferguson, & Zwiebel, 2018).

Typically *Anopheles* labella have sensilla containing ~60 olfactory neurons per labellar lobe, with ~75% (45/60) expressing *Orco* (R. Jason Pitts, Fox, & Zwiebeil, 2004; Riabinina et al., 2016). Labella also have T2 olfactory sensilla (short thorn-shaped sensilla) containing olfactory neurons expressing IRs (Kwon et al., 2006; Saveer et al., 2018).

In sand flies, ~35 trichoid sensilla have been identified on the labellar lobes of *L. longipalpis*. These contain pores, suggesting a role in olfaction as Trichoid sensilla are normally used for short-range detection of sex pheromones/kairomones and food odours (Fernandes et al., 2008).

Research is limited regarding the labellar sensilla and receptors in mosquitoes and sand flies, and their olfactory behavioural contributions, due to difficulties in isolating them from the nearby maxillary palps and antennae. Some promising research has started to elucidate the function of the labellar receptors in *Culex* by ablation/inactivation of adjacent appendages (Choo, Buss, Tan, & Leal, 2015), which could be expanded to other sand fly and mosquito vectors. However, the presence of olfactory trichoid sensilla in sand flies presents potential opportunity for interrupting host-seeking behaviours.

Stimuli, Receptors and Genes

Host seeking behaviour is a complex interaction of stimuli and behaviours comprising activation (a change in behavioural state from quiescence to flight), odour-mediated long range attraction, hovering, landing, and feeding (Cardé, 2015). CO_2 and skin volatiles are important triggers, and visual cues help the insect to orient towards and land on the host assisted by temperature and humidity cues at close range. In mosquitoes, many odours and odour blends have been identified as inducing attractive behaviours, and to a lesser extent in sand fly vectors (Table 3). Key human odours implicated in attraction across mosquito vectors (*Anopheles*, *Aedes*, and *Culex*) and sand fly vectors (*Lutzomyia* and *Phlebotomus*) include CO_2, Ammonia, and lactic acid (CO2: Hinze, Lantz, Hill, & Ignell, 2021; Konopka et al., 2021; Mann et al., 2009; McMeniman et al., 2014; Pinto, Campbell-Lendrum, Lozovei, Teodoro, & Davies, 2001; Spitzen, Smallegange, & Takken, 2008; Syed & Leal, 2009), (Ammonia: Andrade et al., 2008; Bosch, Geier, & Boeckh, 2000; Smallegange, Qiu, van Loon, & Takken, 2005), (Lactic acid: McMeniman et al., 2014; Mukabana et al., 2012; Smallegange et al., 2005) (Table 3). Olfaction and associated attractive stimuli have been comprehensively reviewed by Coutinho-Abreu et al (2021) (Coutinho-Abreu et al., 2021), .

Briefly, host VOCs are produced by bacteria inhabiting the epidermal layer of the skin , that convert metabolites (salts, proteins, amino acids lipids and steroids) in sweat into odorants including short and long chain carboxylic acids, aldehydes, and ketones. Species-specific host odour profiles are likely due to different concentrations of sweat glands (previously described) and species composition of skin bacteria producing different ratios of VOCs including lactic acid and ammonia, in combination with exhaled CO_2, compared to other vertebrate hosts (Smallegange et al., 2011). The differences between anthropophilic and zoophilic species host-seeking behaviour is likely due to the greater abundance of particular chemicals in human odours (Sulcatone, geranylacetone, decanal), compared to those in animal odours (hexanal and heptanal) (McBride et al., 2014). Evidence suggests that Olfactory Receptor Neurons (ORNs) in anthropophilic mosquitoes are more sensitive to Sulcatone concentration (McBride et al., 2014).

Different levels of lactic acid and ammonia produced by humans impacts on attractiveness to different anthropophilic mosquito species (Acree, Turner, Gouck, Beroza, & Smith, 1968; Braks & Takken, 1999).

Table 3. Attractive host odours in sand fly species and mosquito vector species.

Species	Odour	Behavioural response	Notes	Reference(s)
L. whitmani *L. intermedia*	**CO_2**	Attraction	Human vs CO_2 baited traps. Human odours more attractive Relative attraction to CO_2 alone is greater in males than females. Increased concentration led to increased response	(Pinto et al., 2001)
L. shannoni	**CO_2** **1-octen-3-ol** 1-hexen-3-ol	Attraction Attraction Attraction	Baited traps with coloured LEDs	(Mann et al., 2009)
L. longipalpis *L. intermedia*	**1-octen-3-ol** **Lactic acid** **Hexanoic acid** **Ammonia**	Attraction Attraction Attraction Attraction	1-octen-3-ol increased attraction for Lu. Intermedia, not Lu.longipalpis	(Andrade et al., 2008)

L. longipalpis	Octanal Heptadecane **Nonanal** Decanal	Activation, Attraction ♂ Activation ♂ Activation ♂♀, Attraction ♂ Activation ♂♀, Attraction ♂	Only Decanal and Nonanal promote activation in females. VOC collected from dog hairs	(Magalhães-Junior et al., 2019)
L. intermedia	2-Phenylacetaldehyde 6-Methylhept-5-en-2-one (Sulcatone) Nonadecane Icosane	Activation ♀, Attraction ♀ Activation ♀, Attraction ♀ Activation ♀ Activation ♀, Attraction ♀	Human VOCs collected from leg hair samples	(Da Silva Tavares et al., 2018)
Ph. papatasi	CO_2 **1-octen-3-ol**	Attraction Neutral	Males and females show no response to 1-octen-3-ol alone	(Beavers et al., 2004)
An. gambiae	**Carbon dioxide** **Ammonia*** **L-(+) lactic acid** Carboxylic acids (C3-C8, C14) 1-Butanol 2,3-Butanedione 2-Methyl-1butanol 2-Methylbutanal	Attraction	Various odour blends with different combinations, concentrations or doses used. Several bioassay used including wind tunnels and trap catches. Ammonia was attractive as a	(Mukabana et al., 2012; Mweresa et al., 2016; Okumu et al., 2010; Qiu, van Loon, Takken, Meijerink, & Smid, 2006; Smallegange et al., 2005; Spitzen et al., 2008; van Loon et al., 2015)

	2-Methylybutanoic acid 2-Hydroxy-2-butanone 3-Methyl-1-butanol 3-Methylbutanal 3-Methylbutanoic acid benzene ethanol 2-Methylbutanoate butyl butyrate Butyl acetate butyl isobutyrate dimethylsulfide butan-1-amine 1-dodecanol		component of odour blend, and as individual odour.	
An. coluzzii	**CO_2 *** Pyridine* 2-methyl-2-thiazoline* Cyclohexanone* 4-methylthiazole* 4,5-dimethylthiazole*	Attraction	Neuron firing in a pattern similar to response to CO_2 , suggests these chemicals are mimics. All odours tested alone.	(F. Liu, Ye, Baker, Sun, & Zwiebel, 2020; Majeed, Hill, Birgersson, & Ignell, 2016)

	Cyclopentanone* Bis(methylthio)methane* Thiazole* Furfurylmercaptan* Allylisothiocyanate* Thiophene* trans-2-pentenol* 3-methylcyclohexanone* Cineole* **1-octen-3-ol**			
Ae. aegypti	**CO_2** **Ammonia** **L-(+) lactic acid*** Carboxylic acids (C1-C3, C5-C8) Acetone Dimethyl sulphide* Acetone Dichloromethane* Butanone dimethyl disulphide	Attraction	Various odour blends with different combinations, concentrations or doses used. Several bioassays used including wind tunnels, olfactometers and cage assays. Individually attractive odours.	(Allan, Bernier, & Kline, 2006; Bernier, Kline, Allan, & Barnard, 2007; Bernier et al., 2003; Bosch et al., 2000; Dekker, Geier, & Cardé, 2005; Majeed et al., 2016; McMeniman et al., 2014; Williams et al., 2006)

	allyl methyl sulphide Carbon disulphide* Chloroform Dimethyl sulphide Dichloromethane * 1,1,2-trichloromethane carbon tetrachloride tetrachloroethylene 1,1,1-trichloroethane Acetic acid* methyl sulphide* Methyl propyl disulphide* butanoic acid* 3-Methyl butanoic acid* heptanoic acid* Tetradecanoic acid* Hexadecanoic acid*			

	Octadecanoic acid* Benzoic acid* 2-Hydrobenzoic acid* **1-octen-3-ol**			
Cx. quinquefasciatus	**CO_2** **L-(+) lactic acid*** acetic acid* hexadecanoic acid* octadecanoic acid* Methyl sulphide* Methyl propyl disulphide* propanoic acid* **hexanoic acid*** heptanoic acid* octanoic acid* nonanoic acid* decanoic acid* undecanoic acid* tridecanoic acid* tetradecanoic acid*	Attraction	Various odour blends with different combinations, concentrations or doses used. Several bioassays used including wind tunnels, olfactometers and cage assays.	(Allan et al., 2006; Lacey & Cardé, 2011; Syed & Leal, 2009; Tauxe, Macwilliam, Boyle, Guda, & Ray, 2013)

	heptapentanoic acid (long range)* ethylene glycol* benzyl alcohol* cholesterol* Decanol (long range)* Heptanal* Propanal* **Nonanal*** Cyclopentanone*			

Asterisk (*) represents odours tested individually. Bold typeface represents odours identified as attractants in both Sand fly and mosquito vectors. Mosquito vector data adapted from Coutinho-Abreu *et al.*, (2021).

Very few studies have been conducted in sand flies to identify specific odours that elicit attractive behaviour. In more detail, attractive behaviours have been observed towards human skin odours without identifying the specific chemical composition, showing attraction of male and female *L. Longipalpis* from strains collected from different locations (Hamilton & Ramsoondar, 1994). Oviposition attractants (odours that stimulate egg laying by gravid females) have been identified in *L. longipalpis*, including 2-methyl-2-butanol and hexanal (Dougherty, Guerin, & Ward, 1995), and commercial lures emitting lactic acid, hexanoic acid, and ammonia illicit attractive behaviour(Andrade et al., 2008).

Additionally, VOCs collected from dogs and humans have been demonstrated to illicit a behavioural response in *Lutzomyia sp*. Specifically, Octanal, Heptadecane, Nonanal, and Decanal have been isolated from *Leishmania infantum* -infected dogs. Nonanal and Decanal activate females, but do not actively attract them. Conversely, in males Octanal, Nonanal, and Decanal are attractants. These are plant volatiles, suggesting that the response in related to search for sugars (Magalhães-Junior et al., 2019). A limited number of volatiles in human skin odours tested in wind tunnel assays on wild caught *L.*

Intermedia have shown attractive responses. In particular, 2-Phenylacetaldehyde, 6-methylhept-5-en-2-one (Sulcatone), Nonadecane and Icosane activated females. Of these Phenylacetaldehyde, Sulcatone and Icosane elicit attraction (Da Silva Tavares et al., 2018).

Positive attractive responses to CO_2, Ammonia, Lactic acid and Nonanal in sand flies are of particular interest as they elicit attractive responses in mosquitoes. Identification of the receptors and genes involved in the detection of these odours may allow exploration gene editing strategies in sand flies that could disrupt host seeking.

Receptors and Genes

The genetic basis of host-seeking behaviour is an area of increasing interest across vector species (Rinker, Zhou, Pitts, Consortium, & Rokas, 2013), as it allows an understanding of the interactions between host odours and vector response, with the potential to inform vector control strategies . Transcriptome profiling of the antennae of mosquito vectors has been used for different families of receptors, have been used to identify receptors within ORNs that bind odorants entering pores in the sensilla on the sensory organs. Three families of receptors have been identified as key in olfactory host detection; Odorant receptors (ORs), gustatory receptors (GRs), and Ionotropic receptors (IRs) (R. J. Pitts, Rinker, Jones, Rokas, & Zwiebel, 2011)..

ORs are transmembrane proteins that form ion channels (Sato et al., 2008) from the lymph of sensilla into ORNs composed of two subunits, the ligand-selective OR and an obligate odorant receptor co-receptor (*Orco*). Odorants traverse the lymph of the sensillum by interacting with Odour Binding Proteins (OBPs), and are released allowing attachment to the OR-Orco complex. The OR-Orco complex itself is involved in a wider general response to odorants and pheromones (for example 1-octen-3-ol, Indole, 2,3-butanedione) described by Carey *et al.,* 2010.

GRs are ligand-gated ion channels primarily involved in the detection of sugars, bitter compounds and pheromones (Montell, 2009). However, three GRs have been identified (*Gr1*, *Gr2*, *Gr3*, and orthologues thereof) as CO_2 receptors for host-seeking, and are highly conserved in mosquito species (Erdelyan, Mahood, Bader, & Whyard, 2012; Lu et al., 2007). GRs have been concisely reviewed elsewhere (Sparks & Dickens, 2017).

IRs are ligand-gated ion channels which are expressed in sensilla where ORs are not expressed, and specifically respond to amines and acid-based odorants (Lactic acid, butylamine, and carboxylic acids) components of human sweat (R Jason Pitts, Derryberry, Zhang, & Zwiebel, 2017). The ion channels are formed of odours-tuned IRs (~30 identified in *Ae. aegypti* antennae (Matthews, McBride, DeGennaro, Despo, & Vosshall, 2016)) and one of three co-receptors (*Ir8a*, *Ir25a* or *Ir76b*) (Abuin et al., 2011; Benton, Vannice, Gomez-Diaz, & Vosshall, 2009). Expression of *IR8a* is only in the antennae of *Ae. aegypti* (Matthews et al., 2016) and *An. gambiae* (R. J. Pitts et al., 2011; Rinker et al., 2013).

This suggests *Ir8a* in mosquitoes is purely for the detection of odours, compared to *Ir25a* and *Ir76b*, found in other sensory appendages (Abuin et al., 2011; Matthews et al., 2016).Co-receptors across the three main families have been identified as of particular importance. Without these, ligand-gated ion channels cannot form/are disrupted, no longer being able to allow movement of odour molecules to ORNs (Raji et al., 2019; Riabinina et al., 2016).

In *Ae. aegypti*, palp transcriptomes (olfactory organs) shows high levels of expression of a wide range of olfaction-related genes including *Gr1*, *Gr2*, *Gr3*, *Orco*, *Ir25a*, *Ir76b*, *Ir8a* (Bohbot et al., 2014). Specifically, *Gr1*, *Gr2,* and *Gr3* genes are involved in CO_2 detection in *Ae. Aegypti* and are highly conserved, with orthologues in *An. gambiae* and *D. melanogaster* (Kent, Walden, & Robertson, 2008). In female *An. gambiae* maxillary palps, 75 ORs, 61 GRs, and 46 IRs have been identified, and expression patterns of these genes suggest manipulation may affect host seeking (Rinker et al., 2013).

Until recently it was thought that there was only one receptor type per ORN, however recent work has shown this not to be the case (Task et al., 2022; Younger et al., 2022). There are in fact several receptors expressed per neuron from different receptor gene families. For example in Ae. *aegypti*, CO_2 receptor neurons also express *IR25a* receptors (Younger et al., 2022), and in *An. coluzzii* co-expression of *Orco* and *Ir25a* occurs within olfactory neurons (Task et al., 2022). The expression of multiple types of receptors in a single neuron may help adjust the behavioural responses to stimulation of these neurons.

With respect to sand fly receptors, little was known until a recent annotation and phylogenetic analysis of the *L. longipalpis* and *P. papatasi* genomes was conducted to identify genes involved in chemoreception (Hickner et al., 2020). 140 and 142 ORs were identified in *L. longipalpis* and *P. papatasi*, with more than 80% of these being only distantly related to those from other Dipteran vector species (*An. gambiae*, *Cx. quinquefasciatus*, *Ae. aegypti*). However, *Orco* orthologues were confirmed in both sand fly species. Additionally 82 and 77 GRs were identified in *L. longipalpis* and *P. papatasi* respectively. Only two of three major CO_2 receptor gene orthologues were present (*GR1* and *GR2*) in both sand fly species. The third, *Gr3*, (*Dmel Gr63a*/*AgamGr24* orthologues) was not conserved. Only 23 and 28 IRs were identified in *L. longipalpis* and *P. papatasi*, respectively, far fewer than in mosquitoes, however the major co-receptors (*Ir8a*, *Ir25a*, and *Ir76b*) were present (Hickner et al., 2020)**.**

More has been elucidated about the phylogenetic relatedness and conservation of sand fly genes involved in olfaction. Confirmation of conserved *Orco* orthologues in *L. longipalpis* and *P. papatasi* present putative targets for interruption of 1-octen-3-ol and nonanal detection amongst other odours. Conservation of two of the three major CO_2 receptor genes, *GR1* and *GR2*, provide prime targets to

disrupt CO_2 detection, and the presence of conserved *IR8a*, *IR25a*, and *IR76b* co-receptors in sand flies presents targets to modify host-seeking behaviour in response to lactic acid.

Gene Editing of Olfactory Genes

Genetic modification has been used to investigate the function of receptor genes and their responses to odours. To date a limited number of studies have demonstrated knockouts or temporary knockdown of gustatory, ionotropic and olfactory receptors/co-receptors in different mosquito species. GR knockouts have been achieved using CRISPR based approaches in *An. coluzzi* (F. Liu et al., 2020), Zinc Finger Nucleases (ZFNs) in *Ae. aegypti* (McMeniman et al., 2014), and RNA interference (RNAi) in *Ae. aeygpti* (Erdelyan et al., 2012). *Orco* olfactory knockouts have been achieved in *Ae. aegypti* using ZFNs (DeGennaro et al., 2013), in *Ae. albopictus* with RNAi (H. Liu et al., 2016), and CRISPR in *An. coluzzi* (Sun et al., 2020) and *An. sinensis* (Y. Wang et al., 2022) . IR co-receptors knockouts in *Ir8a* and *Ir76b* have been achieved in *Ae. aegypti* (Raji et al., 2019) and *An. coluzzi* (Ye et al., 2022) respectively. Outcomes are discussed in more detail below.

GR knockouts

Orthologues of the *D. melanogaster Gr21a* and *Gr63a* genes involved in CO_2 detection were identified in *Ae. aegypti* (*Gr1* and *Gr3*), along with a paralog, *Gr2*. RNAi knockdown of *Gr1* and *Gr3* reduces the ability of these mosquitoes to respond to CO_2 in olfactometer assays, however reduced response is not observed in *Gr2* knockdown (Erdelyan et al., 2012). Additionally, ZFN knockout of in *Gr3* in *Ae. aegypti* confirms lack of electrophysiological and behavioural responses to CO_2 alone, or in combination with lactic acid, skin odours or heat (McMeniman et al., 2014). However, *Gr3* knockouts are still attracted to individual human odorants, to live arm-in-cage studies and in semi-field cage experiments, when compared to wildtype. This suggests that at close-range, human odours with other close-range cues are sufficient for mosquitoes to identify human hosts, but that CO_2 is important for longer distance attraction to hosts (McMeniman et al., 2014).

CRISPR knockout of *Ae. aegypti Gr1*, *Gr2*, and *Gr3* orthologues in *An. coluzzii* (*Gr22*, *Gr23*, and *Gr24*) demonstrated an impact on responses to CO_2 and 1-octen-3-ol (Liu et al., 2020). Homozygous *Gr23* knockouts do not respond to CO_2 due to interrupted activation of the cpA neuron. However, they maintain a response to 1-octen-3-ol, which activates cpB and cpC neurons. Homozygous *Gr24* knockouts show similar results, suggesting *Gr23* and *Gr24* are essential components of CO_2 detection. Interestingly, homozygous *Gr22* knockouts retain low-level responses to CO_2, with altered function of the cpA neuron. The dynamics of this response have a knock-on effect and are likely to alter the host seeking ability of mosquitoes using CO_2 (F. Liu et al., 2020).

Overall, knockdown or knockout of both *Gr1* and *Gr3* in *Aedes* reduces response to CO_2, however in the presence of the whole human odour profile *Gr3* mutants still responds at close range, probably due to other cues and receptors being involved (vision/heat, other receptors picking up other odours). Interestingly *Gr2* knockouts do not respond to CO2 but do respond to Octenol, and Gr3 knockouts show similar responses. As discussed, *L. longipalpis* and *P. papatasi* have conserved orthologues for *Gr1* and *Gr2*, but not *Gr3* (Hickner et al., 2020). Therefore, the *Gr2* gene in sand flies is a promising target for knockout, that may result in reduced or ablated behavioural responses to CO_2 that has been demonstrated in other vector species.

Orco knockout

The obligate olfactory receptor co-receptor (*Orco*) (Vosshall & Hansson, 2011) identified in *D. melanogaster* (Larsson et al., 2004) provides a useful target to elucidate OR function in the context of host seeking. Homozygous *Orco* mutants generated by ZFN knockouts in *Ae.aegypti* show no attraction to honey in either sex, and females lose attraction to human skin odours in the absence of CO_2 (DeGennaro et al., 2013). However, when CO_2 is present, females respond to human skin odours in the same manner as wildtype females. This suggests that CO_2 primes mosquitoes for attraction via alternative receptor pathways, which compensate for the lack of responses in *Orco* knockouts. *Orco* mutants are still attracted to vertebrate hosts in host preference assays, however they have reduced ability to discriminate between human and animal hosts (DeGennaro et al., 2013). Furthermore, RNAi of the *Orco* gene in *Ae. albopictus,* resulted in more than a 50% reduction in transcript abundance, lower feeding rate, and reduced response to hexane, 3-methylindole, indole, 1-octen-3-ol, confirming the vital role of the *Orco* gene in human detection (H. Liu et al., 2016).

Additional studies in *Anopheles* mosquitoes strengthen the case for the importance of *Orco*. In *An. coluzzii, Orco* knockouts generated by CRISPR, show reduced activity to the majority of odour-evoked responses in adults and larvae. *Orco* mutants lack response to esters, ketones, alcohols and aldehydes. However, they do respond to amines and two carboxylic acids including L- (+)-lactic acid (Sun et al., 2020). The retained response to these odours suggest involvement of alternative receptors in the detection of these odours, much like the receptor redundancy identified by Younger *et al.,* (2020). Additionally, blood-feeding female *Orco* mutants have reduced attraction to human odours, non-human odours, or odour mimics, and gravid females are less responsive to oviposition attractants. This confirms the necessity of *Orco* in mosquito responses to ecologically significant odours which are relevant to host seeking or environmental behaviour (Sun et al., 2020). Further CRISPR knockout of *Orco* in *An. sinensis* demonstrated a reduced response to several human attractant odours (1-octen-3-ol, heptanal, nonanal , ethyl butyrate, hexanoic acid, heptanoic acid, nonanoic acid). When used in human-host proximity assays and olfactometer assays *Orco* knockouts were less attracted to humans, and had reduced preference for humans over other vertebrate hosts (Y. Wang et al., 2022).

These *Orco* knockout studies in *Aedes* and *Anopheles* vectors clearly demonstrate reduced responses to a range of human odours involved in host-seeking behaviours, such as 1-octen-3-ol. However, responses to CO_2 and L- (+)-lactic acid are not completely ablated in *Orco* mutants. Importantly, *Orco* orthologues have been identified in *L. longipalpis* and *P. papatasi* (Hickner et al., 2020), with a high degree of relatedness to the *Orco* genes modified (Figure 1), suggesting similar result may be achieved by knockout in sand flies.

Ir8a knockouts

IRs are important in the detection of amines, aldehydes, ketones, and acids produced as VOCs by skin bacteria, with expression of the co-receptor *IR8a* occurring uniquely in the antennae in *Ae. aegypti* and *An. gambiae* (Matthews et al., 2016; Rinker et al., 2013). This suggests that the *Ir8a* gene within mosquitoes are solely for the detection of odours due to their location. CRISPR knockout of *Ir8a* in *Ae. aegypti* ablates detection of lactic acid and carboxylic acids as individual odours, and attraction to human odours in the absence of CO_2. This suggests that CO_2 gates skin odour attraction in the IR8a pathway

IR8a knockout strains exposed to CO_2 mixed with Lactic acid respond like wildtype, however, response to human acid volatiles (Butyric acid, Heptanoic acid, Octanoic acid, Nonanoic acid and Lactic acid) is insignificant when using electroantenography (EAG). Mutants still respond to non-acid human odours (1-octen-3-ol, Octanal, Geranylacetone, Limonene, Nonanal, Sulcatone, 2-ethyl hexanol, Dodecanal, Linalool).

In human attraction assays, wildtype and *Orco* mutants show robust attraction, however *IR8a* mutants have significantly reduced attraction. Double *Ir8a* and *Orco* mutants maintain some level of attraction to human hosts (Raji et al., 2019). Double *Ir8a and Gr3* mutants respond weakly to CO_2, and in a similar manner to *Gr3* knockouts. therefore losses in response were not additive with respect to all receptor genes (Raji et al., 2019). The double knockouts results suggest alternate pathways exist for the detection of skin odours that can compensate for the loss of function in these specific combinations of two genes at the same time (Raji et al., 2019).

A range of gene editing methods (CRISPR, ZFNs, TALENs and RNAi) have been used to successfully knockout or knockdown genes from three major families (GRs, ORs, and IRs) of chemoreceptor genes in *Anopheles* and *Ae. aegypti*, however at the time of writing the number of studies are limited. Additionally, not all major mosquito vectors have had olfactory genes modified *in vivo*, and only *Ae. aegypti* have had all three of the major gene families described knocked out, thus providing scope for further studies.

Overall, the identification of key genes involved in host seeking olfaction, and the significant conservation of these genes across vector species is useful for control efforts. The knockdown/knockout of these in several species has demonstrated that behavioural responses also share similarities, opening the door for understudied vectors.

Relatively easy to use methods for genetic modification have allowed elegantly designed studies to link gene knockouts with electrophysiological and behavioural studies, to determine behaviour responses to odours. Importantly, an understanding that multimodal integration takes place in host seeking to identify a host (McMeniman et al., 2014) would suggest that targeting of at least two kinds of receptors simultaneously (CO_2/Heat/host odour), outside of the *GR3*/*Orco*/*IR8a* triad (Raji et al., 2019), would be required to overcome the multisensory nature of vector attraction to hosts.

A comprehensive literature search has revealed no evidence of research conducted to knockout any olfactory genes, including those described, in sand flies.

CRISPR – Prospects and opportunities for disease control

Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR), originally identified as part of the bacterial immune system, has been co-opted and applied to eukaryotic systems allowing precise genetic engineering (Doudna & Charpentier, 2014). Before its discovery other techniques were used to modify genes in insect vectors; these included zinc finger nucleases (ZFNs), transcription-activator-like effector nucleases (TALENs), and PiggyBac transposable elements (Gregory, Alphey, Morrison, & Shimeld, 2016).

The CRISPR- Cas9 system consists of a guide RNA (gRNA/sgRNA) and a CRISPR-associated endonuclease 9 (Cas9). The sgRNA is a short piece of RNA with a scaffold region, which binds to Cas9, and a 20-nucleotide region which is homologous to the target region of DNA to be affected. gRNAs are produced synthetically and can be modified to target different regions of DNA. A 3-base pair region, protospacer adjacent motif (PAM), located directly downstream of the unique sequence of the gRNA, facilitates the formation and binding of a functional Cas9-gRNA complex. The active complex binds to target DNA if there is homology between the gRNA sequence and the host DNA. Subsequently Cas9 endonuclease cleaves the target DNA, causing a double stranded break (DSB). A cellular repair mechanism fixes the damage by Non-homologous end joining (NHEJ) or Homology Directed Repair (HDR). NHEJ may result in insertions or deletions (indels) at the DSB cleavage site resulting in a knockout. HDR involves insertion of exogenous DNA when a homologous template is

provided, resulting in a knockin. HDR is less efficient but provides capacity to precisely modify genomes by expression of exogenous DNA (Figure 3).

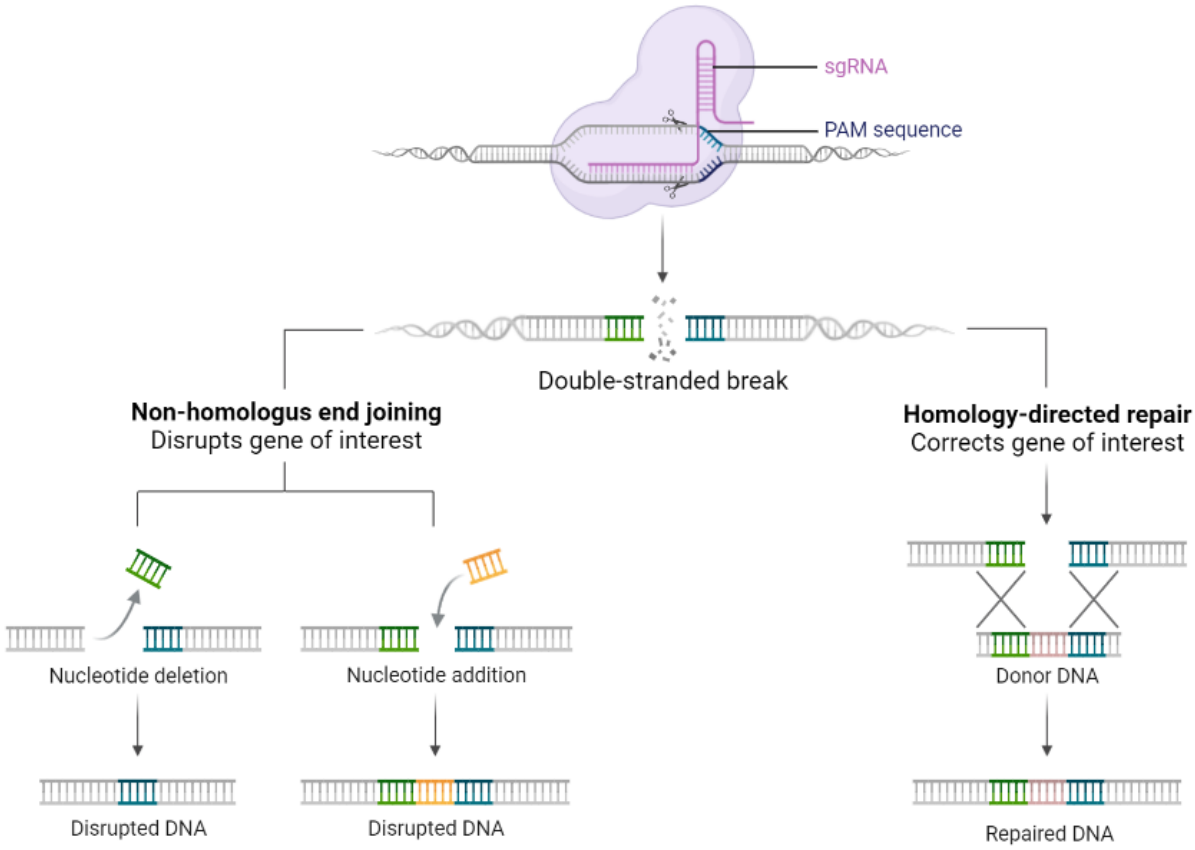


Figure 3. CRISPR and repair mechanisms. The formation of the CRISPR-Cas9 complex consisting of the Cas9 endonuclease and the sgRNA. The complex binds to complementary DNA before a DSB is made The DSB is repaired by NHEJ or by HDR if a homology repair template is provided. Figure from www.biorender.com.

Genetic modification mediated by CRISPR-Cas9 has now been applied to a wide range of organisms, including model insects (Gokcezade, Sienski, & Duchek, 2014) and insects of medical (Hammond et al., 2016; Kistler, Vosshall, & Matthews, 2015; Kyrou et al., 2018) and economic importance (Koutroumpa et al., 2016; Y. Li et al., 2016; Y. Y. Liu et al., 2014). In the context of vectors, much of the focus has been applying CRISPR-Cas9 gene drives to mosquitoes, to reduce reproductive capacity resulting in population elimination (Hammond et al., 2016; Kyrou et al., 2018), or to affect the ability of mosquitoes to transmit disease by blocking parasite development (Gantz et al., 2015).

The use of CRISPR in vectors has increased, with a range of milestones achieved in the three main mosquito genera (*Aedes*, *Anopheles*, and *Culex*), from gene knockouts and knockins, to development of gene drive tools. These are described below for each genus.

Aedes vector control

The first successful use of CRISPR-Cas9 for a vector species was achieved in *Ae. aegypti*, a known vector of arboviruses including yellow fever virus, dengue virus, chikungunya virus, and zika virus. A transgenic mosquito line expressing enhanced cyan fluorescent protein (*ECFP*) and *DsRed* with eye specific expression promotors was derived. Subsequently sgRNAs were developed to target the *ECFP* to disrupt function as a proof-of-concept (S. Dong et al., 2015). Here mutants could be identified as those expressing *DsRed* with loss of *ECFP* generating visual markers for efficiently screening transgenic organisms. A mutation rate of 5.5% was achieved, significantly lower than that seen in *Drosophila* using the similar approaches, demonstrating the difficulty in working with non-model insect species (S. Dong et al., 2015). *Ae. aegypti* CRISPR protocols were further optimised for inducing insertions or deletions (indels) using individual sgRNAs or combinations of two sgRNAs. The latter aims to delete large sequences within a target region in *Ae. aegypti*, allowing for simpler analysis of modification by gel electrophoresis. In addition, HDR protocols have been optimised successfully, resulting in >30% efficiency of HDR insertion of an *ECFP* cassettes leading to stable germline mutations (Kistler et al., 2015). These studies provided step-by-step guidelines for future modification in this vector.

Steps were taken towards targeting genes that could play a role in novel vector control strategies that distort sex ratios (described below). Hall *et al.,* (2015) identified a gene (*Nix*) ,linked to sexual differentiation in *D. melanogaster*, with involvement in in splicing *doublesex* (*dsx*) and *fruitless* (*fru*) genes (Salz & Erickson, 2010). They hypothesised that *Nix* may function as a male-determining factor (M-factor), which was confirmed by absence in females (Hall et al., 2015). CRISPR knockout of *Nix* resulted in feminised genetic males, and genetic females with male genitalia, concluding that *Nix* is essential to male development. CRISPR knockouts in somatic tissues lead to 69% of males showing deformities of feminisation of sexual organs, and 44% showing feminised antennae (Hall et al 2015). Such a result provides scope for the development of gene drive approaches spreading modifications that inhibit mating success, leading to population reduction of vectors.

An efficient method for generating genome edits using CRISPR is through the development of a Cas9 expressing line of insects, which facilitates assessment of gRNA following transfections. Li *et al.,* (2017) successfully generated several *Ae. aegypti* lines expressing Cas9 inherited through the germline using PiggyBac plasmids to introduce a Cas9 cargo. Utility of the Cas9 lines was determined via injections with sgRNAs targeting visual eye markers. High survival (> ~61%) and mutagenesis rate

(52-85%) in G0, and 58-66% mutation rate in generation 1 (G1) demonstrated the value of this tool (Ming Li et al., 2017).

More recently, CRISPR modifications targeting visible phenotypic marker genes (*kh* and *Yellow*) were successful in *Ae. albopictus*, with a *Kh* mutation rate of >51%, and germline transmission of ~78%. This line was bred for multiple generations with no adverse effects. Mutation efficiency was lower in the *Yellow* gene knock outs (32.92%)(T. Liu et al., 2019). This demonstrates successful translation of CRISPR to an invasive vector of chikungunya virus and dengue virus.

Anopheles vector control

Application of CRISPR methodologies to *Anopheles* species has also been a primary focus in the context of malaria transmission. Gantz *et al.,* (2015) published landmark results demonstrating gene drive in a vector species (*An. stephensi*, an Asian Malaria vector). They demonstrated that antimalarial peptides (m1C3 and m2A10) targeting *Plasmodium falciparum* ookinete protein chitinase 1, and circumsporozoite protein, could be expressed within the vector. Furthermore expression components were inherited at a frequency far above mendelian inheritance, an approach known as gene drive.

In this instance gene drive constructs comprised Cas9 promoted by *An. stephensi vasa*, *U6* promoter to drive gRNA targeting the *Kh* gene (a visible marker of Cas9-mediated mutagenesis), a *DsRed* marker promoted by *3xP3* eye promoter, and both anti parasite genes *m1C3* and *m2A10* (Gantz et al., 2015). Remarkably, by generation 3 (G3) males and females demonstrated 99.5% inheritance of the plasmid cargo (*DsRed*+). The extremely high level of allelic conversion demonstrates progress towards development of functional gene drives for vector control (Gantz et al., 2015).

Following on, Hammond *et al.,* (2016) demonstrated the first example of a gene drive in *An. gambiae*, the primary vector of Malaria in Africa using a different approach. In contrast to expressing anti-parasite peptides (above), genes that lead to recessive female sterility when modified (*AGAP005958, AGAP011377, and AGAP007280*) were targeted. Each of the three genes targeted produced different outcomes. *AGAP007280* homozygous knockout females laid eggs that did not hatch, and *AGAP011377* and *AGAP005958* homozygous knockouts females laid no eggs (Hammond et al., 2016). Modifications were achieved by inserting docking constructs containing green fluorescence protein (*GFP*) and docking elements into each locus, to disrupt function, allow for screening, and to facilitate integration of gene drive elements. Specifically, constructs delivered contained Cas9 promoted by germline specific promoter *vasa2*, gRNA (promoted by *U6*), and an *RFP* marker. Cage studies determined the spread of CRISPR alleles, by initiating a cage of wildtype and CRISPR modified individuals at 50:50 ratio. Over four generations the CRISPR-affected alleles rose from 50% in G1 reaching 97.3% in G5 for cages initiated with *AGAP011377* knockout heterozygotes crossed to wildtype (Hammond et al., 2016).

Culex vector control

With respect to the use of CRISPR in mosquitos *Culex* species which are important vectors of West Nile Virus and lymphatic filariasis, have been relatively neglected. The first use of CRISPR in *Culex* mosquitoes was applied to a strain known to be resistance to pyrethroids at the larval stage, attempting to reverse the insecticide resistance trait (Itokawa et al., 2016). The *CYP9M20* gene responsible for expression of cytochrome P450 was targeted by knockout, resulting in larvae from this previously resistant strain to revert to complete susceptibility to permethrin, like wildtype strains. This demonstrates the potential opportunity to use CRISPR to ameliorate the growing issue of insecticide resistance in vectors of disease be modifying resistance genes (Itokawa et al., 2016).

Further proof-of-concept CRISPR knockouts have been demonstrated in *Cx. quinquefasciatus*, targeting phenotypic markers that allow for rapid screening of modification events. Anderson *et al.,* (2019) developed a homozygous knockout of the *kmo* phenotypic eye marker gene using CRISPR. *Kmo* mutations are recessive, requiring both alleles to be mutated for a phenotypic effect. 2.7% of G1 possessed the white-eye phenotype (Anderson et al., 2019). Additionally, Li M *et al.,* (2020) targeted a recessive *white* gene (*CPIJ005542*) achieving generation 0 (G0) with knockout efficiency of up to 86% (including individuals with mosaic phenotypes). A multiplex approach was used whereby multiple gRNAs were injected to simultaneously target different regions of the white gene. This CRISPR approach was further refined by investigating heritability, an important component for development of gene drives. G0 males and females expressing mosaic phenotypes were crossed, and screened by identifying white-eye G1 individuals. The percentage of phenotypic mutants ranged from 61% for progeny resulting from single gRNA injections, to 86% when progeny resulted from co-injection with three gRNAs. Homozygous white-eye mutant lines were established, with no fitness cost detected in the assays performed (M. Li et al., 2020).

The more complicated approach of HDR knockin was successfully demonstrated in *Cx. quinquefasciatus* by Purusothaman DK *et al.,* (2021). A minimum integration rate of 1.6% was achieved using a construct containing *DsRed* and a *kmo* targeting gRNA. Integration of fluorescence markers at the *kmo* locus allowed rapid screening first by identification of white-eyed individuals, followed by expression of *DsRed* fluorescence to confirm integration and expression of exogenous DNA. The design of the HDR cassette, whereby exogenous DNA can be inserted, is an important step towards the development of gene drives in *Culex* in combination with germline specific promoters are utilised (Purusothaman, Shackleford, Anderson, Harvey-Samuel, & Alphey, 2021).

These approaches described in *Culex* demonstrate that CRISPR-based modification of phenotypic marker targets, along with genes that have acute relevance to vector control, can be achieved in a neglected vector species, and can facilitate the development of genetic control strategies.

Sand flies

At the time of writing, only two groups have attempted gene editing in sand fly vectors of leishmaniasis by application of CRISPR-based modification. Martin-Martin *et al.,* (2018) optimised the microinjection protocol for delivery of CRISPR components into *L. longipalpis* embryos targeting the *Yellow* gene for pigmentation change. An injection survival rate of 11.90-13.22% was achieved but without evidence of CRISPR gene editing (Martin-Martin, Aryan, Meneses, Adelman, & Calvo, 2018). However, Louradour *et al.,* (2020) successfully applied CRISPR-based approaches to knockout *Relish*, a gene involved in immunity (IMD pathway) within *P. papatasi*, to generate heterozygous individuals with increased permissiveness to *Leishmania major*. A fitness cost was observed making it difficult to maintain a homozygous line for more than a few generations due to the immune knockdown making them susceptible to microbial colonisation (Louradour, Ghosh, Inbar, & Sacks, 2019).The success of this approach confirms that CRISPR based approaches are achievable in principle.

CRISPR-based Gene Drives and Control strategies

Development of gene drives have huge potential to deliver a powerful new approach to interrupt disease transmission in fast reproducing invertebrate vectors of medical, veterinary and agricultural importance. In principle, genetically introduced traits and effectors can be inherited at high efficiencies resulting in a rapid spread through breeding populations in remarkably short timeframes. Examples of different approaches will be discussed in this section.

A gene drive system typically incorporates three components, a guide RNA, the Cas9 gene, and an insertional template incorporating homologous flanking arms targeting the desired insertional site, leading to autocatalytic incorporation of exogenous DNA template in the second allele. By implication, a mating between a transgenic individual and a wild type individual will result in most offspring possessing homozygous transgenic genotypes. Cas9 targeting is changed by altering the targeting sequence present in gRNA. Effectors can be driven by a range of promotors, for example cytoplasmic actin promotors, synthetic promotors (*3xP3)* (Schetelig & Handler, 2013), or promoters which are conserved across species and those targeting the germline to avoid somatic expression (*nanos, vasa* and *zpg*). Typically, CRISPR-Cas9 components are delivered to early stage embryos (Gantz et al., 2015) by microinjection, an inefficient process (typically 0.1% to 14%). Emergent insects are screened phenotypically either by knockout of non-lethal phenotypic markers such as eye or cuticle pigmentation, or by the introduction of fluorescent marker (Gantz et al., 2015).

CRISPR-Cas9 mediated gene drives have been rapidly adopted for modification of medically important insect vectors of malaria such as *An. gambiae* (Bernardini, Kriezis, Galizi, Nolan, & Crisanti, 2019; Hammond et al., 2016; Kyrou et al., 2018) and *An. stephensi* (Gantz et al., 2015). Here gene drives facilitate heritability of introduced traits at frequencies exceeding mendelian inheritance (≥99.5% efficiency in *An. stephensi*) and induce rapid population level transformation in fast reproducing insects (Gantz et al., 2015).

Different approaches utilising gene drives are currently pursued to drive desired traits into a population. Two main approaches are population suppression and population replacement strategies (Figure 4) both of which hold great potential.

Population suppression (Figure 4A) involves the release of genetically modified individuals that spread effectors impacting mating or fertility. This leads to a reduction in population numbers of the target vector species to below that required for disease transmission. Through gene drives small numbers of genetically modified insects spread traits through wild populations. Laboratory experiments have demonstrated remarkable efficiency in knocking down populations over few generations and short timeframes in *An. gambiae* (Hammond et al., 2016, 2021; Kyrou et al., 2018; Simoni et al., 2020).

In contrast, population replacement strategies (Figure 4B) focus on introducing and spreading traits that reduce vector competence. In this scenario, by using a gene drive individuals expressing effectors (such as anti-parasite peptides) produce offspring with the refractory traits. Over a number of generations the wildtype population is superseded by those carrying the desired traits, without reducing the insect population as a whole. One advantage of this approach is that potential ecological implications of removing pest populations is avoided, although the "biting nuisance" remains in the case of hematophagous invertebrates. Examples of both suppression and replacement strategies are explored in more detail in the following sections comprising Sex ratio distortion, CRISPR mediated sterile insect technique, transmission blocking and targeting female flight.

Under the population replacement and suppression strategies described (above), there are different approaches to achieve these aims such as sex ratio distortion, sterile insect technique, transmission blocking and targeting female flight.

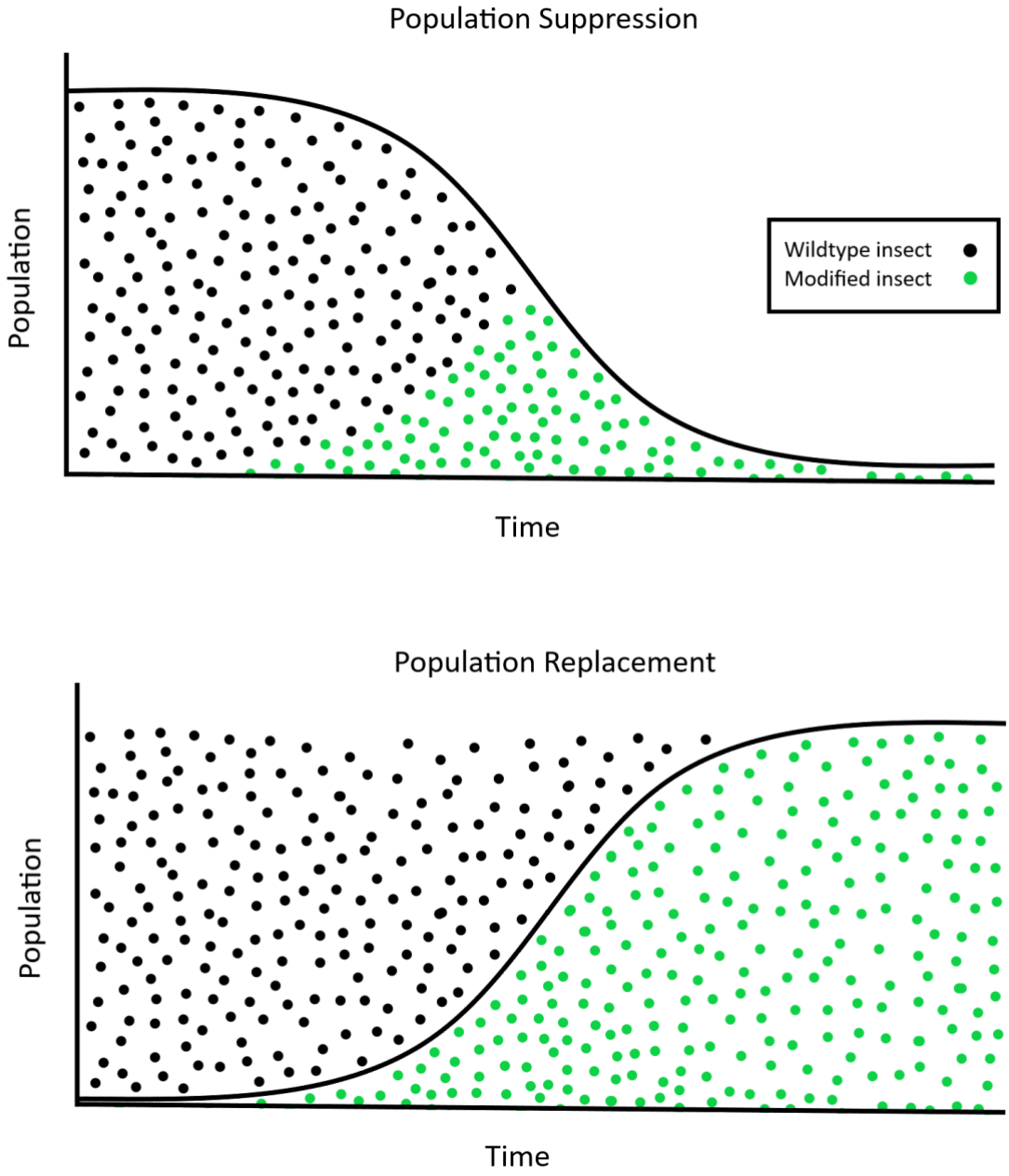


Figure 4. Population suppression and population replacement Gene drive strategies. In both cases the genetically modified insects are released at a low initial frequency and spread in the population over time (multiple reproductive generations). A) Population suppression strategies interfere with the mosquito fertility and fecundity, aiming to eliminate or suppress the population to levels that cannot support disease transmission. B) Alternatively, population replacement strategies spread introduced traits to rapidly replace wildtype vector population with insects inheriting transmission blocking traits , for example anti-parasite peptides.

Sex ratio distortion for population suppression

One method showing great promise in controlling insects is to distort the sex ratio of a population in favour of non-biting males, to reducing human biting, and overall population. By proxy fewer females in a given population results in producing fewer eggs leading to population decline. This has been achieved by targeting genes involved in sex determination (below in more detail) leading to either intersex individuals that are effectively sterile, or skewing a population towards males by targeting destruction of the X-chromosome.

In a landmark paper, Hall *et al.,* (2015) determined that the *Nix* gene in *Ae. aegypti* is an M-locus gene within the Y chromosome and is important in sex determination. CRISPR based knockout of *Nix* resulted in feminised genetic males, and/or genetic female with male genitalia. 69% of males possessed feminised sexual organs, and 44% showing feminised antennae (reduction of plumose antennal hairs). Furthermore , using plasmid-based HDR approaches, *Nix* expression under control of *AePUb* resulted in females possessing masculine (both external and internal features) or deformed genitalia (Hall et al., 2015). Remarkably this research demonstrated the potential for population suppression via sex distortion providing a targets for effective gene drive.

Another promising gene target is the *doublesex* gene (*dsx*) that plays an important role in insect sex differentiation, initially identified in *D. melanogaster* (Burtis & Baker, 1989). The *An. gambiae dsx* ortholog exists in a male and female transcript form (*dsxF* and *dsxM*, respectively). The difference between the two is the alternative splicing of an exon (5th exon) which is unique in females (Kyrou et al., 2018), leading to potential use in female-targeted control strategies such as sex ratio distortion, sterile insect technique, transmission blocking, or targeting female flight. This potential was confirmed by the generation of transgenic *An. gambiae* targeting the *dsx* gene. Heterozygous knockouts resulted in normal sex ratio; however *dsxF*-/- homozygotes resulted in individuals with male and female external feature, and internal reproductive structures alongside normal male phenotypes. The sex genotype of these intersex individuals was determined, showing that only females with the *dsx* homozygous mutation had intersex features. These females were unable to bloodfeed or produce eggs (Kyrou et al., 2018).

Having determined the importance of the *dsx* gene, a gene drive system was built to target this region. A cassette containing *zpg*::Cas9 (an efficient germline promotor), gRNA targeting *dsxF*, and an *RFP* marker (to confirm the cassette had replaced the 3xP3::*GFP* cassette) was inserted to knockout the *dsx* gene. This cassette acts as a homology template during meiosis, and is copied to the wildtype allele. Crossing heterozygote parents ($dsxF^{CRISPRh/+}$) with wildtype resulted in transmission of the $dsxF^{CRISPRh/+}$ alleles to between ~95.9-99.4% in offspring(Kyrou et al., 2018). Heterozygous female offspring had significantly reduced fecundity. Cage studies with a mix of 50% wildtype females , 25% wildtype

males, and 25% dsxF$^{CRISPRh/+}$ males were initiated to assess the spread of the drive resulting in an initial drive allele frequency of 12.5%. Interestingly, frequency of the drive allele increased rapidly to 100% in as little as 7 generations, leading to population collapse. This demonstrated an extremely effective gene drive for population suppression which matched predictive models (Kyrou et al., 2018).

Another variation of the population suppression approach is sex distortion through the "destruction" or shredding of the X-chromosome during sperm production. Galizi et al (2016) developed a CRISPR-based system whereby expression of Cas9 during meiotic spermatogenesis degrades the X-chromosome, resulting in a male bias as the female sex chromosome is exclusively targeted. The sex-distortion system (CRISPR^{SD}) in *An. gambiae* targets an X-linked ribosomal DNA sequence (also present in other Malaria vectors) with Cas9 promoted by the spermatogenesis-specific *β 2 tubulin* (*B2t*) promoter, active only during male meiosis. This results in a majority of Y-bearing sperm fertilizing eggs, leading to male progeny. From four lines developed, caged progeny resulted in 86.1-94.8% males within 5 generations, with no difference observed in the number eggs laid (compared to wildtype) (Galizi et al., 2016).

One consideration when applying this approach is that X-shredding sex-biased methods need multiple releases to suppress a population in the wild. To overcome this hurdle the sex-distortion system could be linked to the Y-chromosome meaning it would be present in all male offspring, allowing for self-perpetuation of this trait (Galizi et al., 2016). Another potential hurdle is that sex chromosomes become transcriptionally inactive during gametogenesis. Therefore elements meant for transcription during meiosis that have been inserted into the Y chromosome to shred the X chromosome will not be transcribed (in time to destroy/inactivate the X chromosome).This may prevent the successful development of a gene drive based on this approach. A potential solution is to design a sex distorter drive that is located on a non sex chromosome (autosomal chromosome) (Simoni et al., 2020).

Simoni *et al.,* (2020) did exactly this by designing a sex distorting gene drive that shreds the X-chromosome using *I-PpoI* endonuclease (promoted by male-specific germline promoter *B2t)*. The knockin construct also included Cas9 (*vasa* or *zpg* promoted, in male and female germlines) and a gRNA targeting a fertility gene (Simoni et al., 2020)**.** In mosquitoes with the transgene inserted into autosomal chromosomes, the homing element of the drive copies the transgene to the other autosomal chromosomes. When gametes are produced, each has undamaged X chromosome and autosomal chromosomes with the transgene. After mating, offspring produced have a 50:50 mendelian sex ratio, however homing of the transgene elements exceeds mendelian inheritance. For males with the transgene inserted into autosomal chromosomes, the homing element functions and copies over to other autosomal chromosomes, but additionally the sex distorter (I-Ppol) targets the X chromosome in females, shredding it. Overall, this results in gametes that only have Y chromosomes and the transgenic

autosomal chromosomes. When mating occurs with wildtype females, homing of the transgene elements exceeds mendelian inheritance , and the sex ratio is distorted to 95% males due to the X shredding (Simoni et al., 2020).

In developing this approach major early hurdles needed to be resolved. In early cage studies the X-shredding element (*I-Ppol*) was too highly expressed by the *B2t* promoter, resulting in male sterility. Modifying the promoter to be significantly less efficient, reduced the expression of the *I-Ppol*. Additionally, replacing the *vasa* Cas9 promoter with *zpg* reduced leaky expression of the Cas9, which in combination with the overexpressed *I-Ppol* had led to unviable embryos. Finally, a gRNA to target the 4-5 exon boundary of *dsx* gene was designed. The final construct (named SDGD^{dsx}) resulted in a ~93.1% male bias in offspring of heterozygous males, and inheritance of the construct was observed in 96-99.9% of progeny from both sexes. Using SDGD^{dsx} in cages (600 mosquitoes in total) at an allelic frequency of 2.5%, population elimination occurred by generation 9 and 13 across two cages (Simoni et al., 2020).

CRISPR-mediated sterile insect technique

Sterile Insect techniques (SIT) has long been a method for vector control involving the mass release of males that have been rendered sterile by exposure to irradiation or chemo-sterilisation agents (see Thesis Introduction). An improvement to the traditional radiation-based SIT , which incurs fitness cost, is a genetic approach targeting male fertility genes.

Chen *et al.,* (2021) identified an ortholog of the *D. melanogaster B2t* gene that is expressed in sperm and testes of *Ae. aegypti*. CRISPR knockout *B2t* completely disrupts male fertility, and also leads to suppressed wildtype female fertility, even when wildtype males are available for mating. Therefore releasing *B2t* knockout males may be a useful population suppression strategy.

A construct was developed to knockout the *B2t* gene via insertion of a GFP to facilitate screening. Promising outcomes in germline-expressing Cas9 *Ae. Aegypti* suggested *B2t* knockout males were as physically fit as wildtype males, however, when crossed with individual wildtype females fertility of homozygous *B2t* knockout males was 0% (no sperm was observed). Additionally, homozygous *B2t* knockout females were only slightly less fertile than wildtype females, which is consistent with the gene only being expressed in male testes (Chen et al., 2021).

Cage studies using homozygous *B2t* males were performed, introducing males for 24h followed by removal, and subsequent introduction of wildtype males. Egg production was suppressed from >90% to ~50% when only 3 *B2t* males had been present, and to 2.6% when 15 had been introduced, in spite

of wildtype males having mated with the females. This result indicated that pre-exposure to *B2t* males suppressed the female fertility. Of note, fertility suppression increased exponentially with the time of pre-exposure to the *B2t* males. Thirty minutes exposure had no impact on fertility, but saturated suppression at 24 hours of exposure. When females were exposed simultaneously to wildtype males and *B2t* males, there was a reduction in female fertility with an increasing ratio of *B2t* males:wildtype males. A 15:1 ratio resulted in wildtype female fertility of ~20% (Chen et al., 2021). Based on this approach an SIT strategy could be developed using this *B2t* sterile male to supress a wild population with repeated inundative releases. However, one obstacle to this method is the mass production of *B2t* sterile males in homozygous form that would be required (Chen et al., 2021).

In a variation on a theme Li M *et al.,* (2021) developed a novel precision guided sterile insect technique (pgSIT) in *Ae. aegypti*, applying CRISPR to target female viability and male fertility, with the only survivors being sterile males (method described in Thesis Introduction). Two separate transgenic strains of mosquitoes are required for this technique to function as a population suppression strategy: a Cas9 expressing strain and a gRNA expressing strain. Briefly, gRNAs were designed to target the male spermatogenesis gene (*βTub)*, and *myosin heavy chain* genes (*myo-fem*) that are crucial for female flight. Heterozygous lines were developed for these targets, in addition to a homozygous Cas9+ line. Crossing Cas9+ females with *βTub* + males results in offspring that are transheterozygous (carrying two different mutant alleles) that express Cas9 (Male gRNA$^{\beta Tub\ +:Cas9+}$, or Female gRNA$^{\beta Tub\ +:Cas9+}$). When transheterozygous males are crossed with wildtype females, sterility is incurred due to immotile sperm. However, transheterozygous females crossed with wildtype males has no fertility costs (Ming Li et al., 2021). Crossing transheterozygous *myo-fem* females (gRNA$^{myo\text{-}fem+:Cas9+}$) with wildtype males led to flightless females, but transheterozygous male crossed with wildtype females did not. The induced flightlessness resulted in loss of fitness (Ming Li et al., 2021).

Building on these transheterozygous lines, a double targeting strain of *βTub* with *myo-fem* was generated by a backcrossing strategy. These were crossed with Cas9+ lines leading to flightless females (pgSIT females) and sterile males (pgSIT males). Conducting caged studies with pgSIT males, high release ratios led to suppression/elimination by between generation 3-6 when adult pgSIT males were used. When pgSIT eggs were used, elimination was also achieved by generation six (G6) (Ming Li et al., 2021).

These impressive results demonstrate that variations of the SIT using genetic modification to disrupt male fertility alone, or in combination with female flight-targeting elements, could be a promising method for control.

Targeting genes implicated in flight

Another interesting method that has potential for use in control of mosquito vectors is the targeting of genes involved in flight, with a specific focus on female flight. Flight is vital for different aspects of mosquito behaviour including mating, blood feeding, and dispersal, therefore interfering with flight would essentially result in a sterile insect due to the inability to mate successfully.

To this end, an *Ae. aegypti* gene (*Act4*) involved in the indirect flight muscles was identified, with expression that is female-specific. The gene was targeted in *Ae. aegypti* and *Cx. quinquefasciatus* via CRISPR homology-directed repair methods with dominant mutations leading to flightless females in *Cx. quinquefasciatus*, and a recessive phenotype in *Ae. aegypti*. The deletion that caused the dominant flightless phenotype in *Cx. quinquefasciatus* was successfully replicated in *Aedes*, by designing a targeted mutagenesis (Navarro-Payá et al., 2020). This demonstrated that interference of the *AeAct4* gene and resultant protein impacted flight muscles of two important vector species.

Expanding on this, a further study used CRISPR to knockout *AeAct4* and *myo-fem* genes leading to 100% of females being flightless. Interestingly it did not affect the flight ability of knockout homozygous males, and did not prevent them from producing offspring (although reduced compared to wildtype in *AeAct-4* knockouts) (O'leary & Adelman, 2020).

The targeting of female flight as a strategy only affects adults, meaning that during the other developmental stages there is no interference thus allowing competition with wildtype individuals. Once flightless females emerge (adults), they are no longer able to reproduce or transmit pathogens. A potential gene drive could be developed by making adult males the carriers of the gene drive elements. These male-carriers would then mate with wildtype females, leading to flightless female offspring, and male offspring carrying the drive, that could mate with wildtype females (O'leary & Adelman, 2020).

Transmission blocking

A promising population replacement strategy for control of vector-borne disease is through the development of parasite transmission blocking**.** A range of anti-parasite effectors have been explored, alongside antimicrobial peptides (reviewed here (S. Wang & Jacobs-Lorena, 2013)), however integration of these molecules into vectors, with gene drive elements has only been described in a limited number of studies (Y. Dong, Simões, Marois, & Dimopoulos, 2018; Gantz et al., 2015; Hoermann et al., 2022).

As described previously, Gantz *et al.,* (2015) published landmark results demonstrating highly efficient gene drive in *An. stephensi.* Antimalarial peptides targeting *Plasmodium falciparum* ookinete protein chitinase 1, and circumsporozoite protein, were successfully expressed within the mosquito midgut,

however the activity of the antimalarial peptides against malaria parasites within the mosquitoes was not explored.

Dong *et al.,* (2018) went a step further by determining whether anti-parasite effectors actively functioned against parasites within the mosquito. CRISPR was used to knockout a gene *Fibrinogen-related protein 1* (*FREP1)*, a *Plasmodium* parasite agonist, to determine the impact on parasite transmission. *FREP1* is a member of a large family of genes found in *An. gambiae*, several of which have been shown to respond to challenge by pathogens. Specifically FREP1 plays a role in the parasite invading the mosquito midgut. Suppression studies demonstrate its impact on the development of plasmodium in mosquito midguts.

An *An. gambiae* strain was generated expressing gRNAs targeting the *FREP1* gene, which were crossed with *Vasa*::Cas9 expressing strain resulting in a homozygous *FREP1* mutant strain. The *FREP1* mutants were challenged with human and rodent malaria parasite, by feeding on parasite gametocyte cultures or infected mice (respectively) at biologically relevant gametocytemia levels. The *FREP1* mutants had parasite suppression resulting in zero oocysts, compared to 2-2.5 in control strains, and a decrease in the number of sporozoites found in the mosquito salivary glands. *FREP1* mutants fed on infected mice, showed up to 100% reduction in permissiveness to oocyst infection (Y. Dong et al., 2018). This provides a clear example of anti-parasite effectors functioning within mosquitoes to supress infection within the vector, however has not included the gene drive element demonstrated by Gantz *et al.,* (2015).

Recently the expression of anti-parasite effectors within vectors, and the gene drive elements have been combined in *An. gambiae*, providing a powerful transmission blocking gene drive tool to reduce disease transmission (Hoermann et al., 2022). Two anti-microbial peptides (magainin 2, and melittin) were integrated into the mosquito genome via CRISPR HDR approaches, which were effectively expressed within the midgut, which resulted in delayed emergence of malaria sporozoites. Modelling of the dynamics of this gene drive suggests its use, and that of similarly constructed drives will play a large role in reducing vector borne disease transmission (Hoermann et al., 2022).

To date most focus for vector control using a CRISPR-based gene drive has been on population suppression strategies, with little focus on population replacement strategies beyond pathogen transmission blocking. Substantial research has been conducted to elucidate the receptors and genes involved in host attraction for mosquitoes using CRISPR methods. Each study alludes to potential use for developing a vector control strategy, however, it appears that a gene drive strategy to modify olfactory genes and subsequently replace a population has not been pursued at this point. This is a possible omission from the toolbox to fight against vector-borne diseases.

Future prospects for Olfactory Gene modification in sand fly vectors

Sand fly vector control has historically relied on insecticides, mostly using indoor residual spraying and long lasting insecticide treated nets. However, these strategies may begin to fail due to the rise of insecticide resistance and new methods of control will be required (Balaska, Fotakis, Chaskopoulou, & Vontas, 2021). Modification of sand fly olfaction using CRISPR-based methods could add a potential control method to the tool-box, by changing the host seeking behaviour. This may steer sand flies away from hosts by interrupting their detection of specific odour profiles, reducing contact with humans and reducing disease transmission. To inform this approach significant knowledge gaps need to be resolved. Key to this is understanding which receptors and genes are involved in odour detection and subsequent attractive behaviour.

In the context of sand flies some odours have been identified as eliciting attractive response in female sand flies such as CO_2 alone (Quinnell, Dye, & Shaw, 1992), and in combination with the whole (unspecified) human odour profile (Pinto, M. C., Campbell-Lendrum, D. H., Lozovei, A. L., Teodoro, U. & Davies, C. R. 2001; Miranda, C. *et al.* 2018). Attractants specific to female *L. longipalpis* have been identified, including carboxylic acids, ketones and aldehydes (Dougherty, M. J., Guerin, P. M., Ward, R. D. & Hamilton, J. G. C. 1999), as well as odours found in commercial lures (lactic acid, caproic acid, ammonia, octenol, and nonanol). These responses to human odours and known mosquito attractants provide a starting point for understanding olfaction in sand flies. However they are yet to be backed up by electrophysiological responses which would help identify which receptors are involved (responses have been recorded for non-human host odours (Dougherty, M. J., Guerin, P. M., Ward, R. D. & Hamilton, J. G. C. 1999)), or knockdown or knockout studies for identification of the genes involved. This is a key omission which could be addressed using genetic modification strategies to disrupt key olfactory receptor genes, followed by assessment of behavioural responses to human odours.

Recent genome annotation of major sand fly vectors (*L. longipalpis* and *P. papatasi*) and identification of chemoreceptors related to important genes for host seeking in mosquitoes is an important milestone (Hickner et al., 2020). Of particular interest is the presence of two of three major CO_2 receptor genes (*GR1* and *GR2*), *Orco*, *Ir8a*, *Ir25a* and *Ir76b* in both sand fly species (Hickner et al., 2020). These have been the primary targets for understanding host-seeking in mosquitoes, and their presence and homology makes these prime targets for CRISPR knockouts in sand flies to demonstrate the proof-of-principle that behaviour change can be induced. In addition, a well-annotated genome provides an essential tool for selecting other potential targets for modification, such as those involved in female fertility and flight, or parasite transmission, and for a rational approach to designing CRISPR components.

The use of CRISPR-Cas9 in a range of non-model mosquito vectors to induce olfactory gene knockouts and subsequent incorporation into gene drives is encouraging for translation to sand fly vectors. To date microinjection of sand fly embryos is challenging and at the time of writing there have been limited attempts to microinject embryos with some minor success (Jeffries, Rogers, & Walker, 2018; Louradour et al., 2019; Martin-Martin et al., 2018), demonstrating the possibility of this technique going forward. Further adaptations to construct designs and modification protocols will allow this area of research to expand beyond the limited studies already completed, and towards the development of gene drives.

Challenges and opportunities for CRISPR mediated changes to olfaction in sand fly vectors

Sand fly species pose unique challenges and opportunities for modification using CRISPR-based approaches. A major challenge demonstrated in the chapter is the lack of knowledge of the olfactory system. There is limited information on the morphological features of olfactory organs, some behavioural studies in response to odours, and recently research identifying receptors from the three main chemosensory receptor families (GRs, ORs, and IRs) that are linked to olfaction in other mosquito vectors. This is a promising move towards better understanding of sand fly olfaction, however the research in this field lags far behind that of mosquito olfaction.

A second challenge is maintenance of sand fly colonies to study host-vector interaction, behaviour and control, something that has only been achieved by 35 laboratories worldwide (Volf & Volfova, 2011). Rearing sand fly colonies is complex and time-consuming when compared to rearing mosquitoes, with generation time taking between 1-3 months. Eggs usually hatch within 6-11 days, and can take as long as 30 days. Terrestrial larvae are small and fragile, taking between 20-25 days to grow to the final larval stage, when they pupate (Volf & Volfova, 2011). Larval stages can have high mortality if conditions including diet, moisture, and overabundance of fungal growth fall outside a suitable range. Too much food can result in fungal growth, and underfeeding can lead to cannibalism. Adults take a further 7-11 days to emerge after pupation (Lawyer et al., 2017).The complexity in rearing and length of time between developmental stages makes genetic modification more complicated and less high throughput in sand flies, when compared to the 7-14 days it takes for mosquitoes to reach adulthood. Additionally the small size and terrestrial rearing conditions for larvae make screening for modifications such as fluorescent markers cumbersome.

A third challenge is the delivery of genetic constructs, mRNA, or sgRNAs for modification (see Chapter 5). Eggs are small (0.3-0.5mm long and 0.1-0.15 mm wide), sticky, fragile, and laid directly onto a terrestrial surface making them hard to collect without damage, and difficult to process for microinjection. The majority of eggs are laid un-melanised (white soft chorion), however in some cases

they are laid fully melanised (dark brown/black firm chorion). This makes microinjection in sand flies less efficient than in mosquito species as not all eggs can be selected for microinjection. Fully melanised eggs cannot be injected as microinjection needles are unable to pierce the chorion. The collection process is laborious requiring delicate manipulation to move eggs from laying substrates to an injection surface, and to align them for efficient injections. Optimisation of this key step in genetic modification has taken place, however there are still opportunities to improve the process of sand fly egg manipulation.

For successful modification via editing of the germline cells, the microinjection needs to be conducted at the right location and the right time. This has been determined as the point where pole cells develop, at the pre-blastoderm stage, when no cell membranes have formed. Again there is limited research on the embryology of sand fly species beyond the identification of blastoderm formation by 60 hours (Abbassy, Helmy, Osman, Cope, & Presley, 1995). This provides a larger window than in *Ae. aegypti* (<8hours,), and Anopheles (<4 hours) (Juhn & James, 2006) for pre-blastoderm injections, however the limiting factor in sand flies is the degree of egg melanisation which can impede microinjection (Jeffries et al., 2018; Louradour et al., 2019; Martin-Martin et al., 2018).

In spite of the challenges described, to our knowledge three groups have attempted sand fly microinjections. One group successfully introduced *Wolbachia* from *D. melanogaster* into *L. longipalpis*, maintaining *Wolbachia* infections for four generations (Jeffries et al., 2018). Another developed a CRISPR protocol for injections of gRNAs and Cas9 mRNA into *L. longipalpis,* but were unsuccessful in achieving cuticular pigmentation modifications predicted from the gRNA target selected (Martin-Martin et al., 2018). A third group successfully achieved CRISPR-based knockout in *Ph. papatasi* to interrupt the *Relish* gene involved in immunity, and demonstrated increased susceptibility to Leishmania parasites as a result (Louradour et al., 2019). These recent studies provide the initial groundwork for further research opportunities with respect to CRISPR modification in sand fly vectors of leishmaniasis.

Conclusion

CRISPR-based modification of olfaction has been achieved in different mosquito species including *Ae. aegypti*, *Ae. albopictus*, *An. coluzzii* and *An. sinensis*, with vast potential to deliver new control strategies. These encompass population suppression and replacement approaches through modification of vector fertility, flight or ability to transmit pathogens. Research towards intervention based on olfaction is in its relative infancy, but provides a promising pathway to reduce disease transmission particularly with respect to behaviour change and host-seeking behaviour. Genetic tools have been used to elucidate the function of receptors and genes within insect olfactory system, and the advent of

CRISPR-Cas9 as a tool has allowed rapid progression of this research avenue due to the method's relative ease of use compared to ZFN and TALEN. In contrast to mosquito vectors, there is a dearth of research on sand fly olfaction, host detection and host seeking behaviour. This could be exploited by emulating the mosquito literature, conducting genetic knockout of olfactory genes followed by behavioural response assays. This research could then lead towards the development of novel gene drive approaches to replace populations with sand flies unable to locate human hosts. Together with more traditional vector control strategies, targeted gene editing within gene drive systems are likely to play an important future role in integrated vector management strategy for insect vectors of medical importance including sand flies.

References

Abbassy, M. M., Helmy, N., Osman, M., Cope, S. E., & Presley, S. M. (1995). Embryogenesis of the Sand Fly Phlebotomus papatasi (Diptera: Psychodidae): Organogenesis including Segmentation, Blastokinesis, Mouthparts, and Alimentary Canal. *Annals of the Entomological Society of America*, *88*(6), 815–820. https://doi.org/10.1093/aesa/88.6.815

Abuin, L., Bargeton, B., Ulbrich, M. H., Isacoff, E. Y., Kellenberger, S., & Benton, R. (2011). Functional Architecture of Olfactory Ionotropic Glutamate Receptors. *Neuron*, *69*(1), 44–60. https://doi.org/10.1016/j.neuron.2010.11.042

Acree, F. J., Turner, R. B., Gouck, H. K., Beroza, M., & Smith, N. (1968). L-Lactic Acid: A Mosquito Attractaut Isolated from Humans Abstract. *Science*, *161*(3848), 1346–1347. https://doi.org/10.1126/science.161.3848.1347

Allan, S. A., Bernier, U. R., & Kline, D. L. (2006). Attraction of mosquitoes to volatiles associated with blood. *Journal of Vector Ecology*, *31*(1), 71–78. https://doi.org/10.3376/1081-1710(2006)31[71:aomtva]2.0.co;2

Anderson, M. E., Mavica, J., Shackleford, L., Flis, I., Fochler, S., Basu, S., & Alphey, L. (2019). CRISPR/Cas9 gene editing in the West Nile Virus vector, Culex quinquefasciatus Say. *PLoS ONE*, *14*(11), 1–10. https://doi.org/10.1371/journal.pone.0224857

Andrade, A. J., Andrade, M. R., Dias, E. S., Pinto, M. C., & Eiras, Á. E. (2008). Are light traps baited with kairomones effective in the capture of Lutzomyia longipalpis and Lutzomyia intermedia? An evaluation of synthetic human odor as an attractant for phlebotomine sand flies (Diptera: Psychodidae: Phlebotominae). *Memorias Do Instituto Oswaldo Cruz*, *103*(4), 337–340. https://doi.org/10.1590/s0074-02762008000400004

Bahia, A. C., Barletta, A. B. F., Pinto, L. C., Orfanó, A. S., Nacif-Pimenta, R., Volfova, V., … Pimenta, P. F. P. (2021). Morphological characterization of the Antennal Sensilla of the afrotropical sand fly, phlebotomus duboscqi (Diptera: Psychodidae). *Journal of Medical Entomology*, *58*(2), 634–645. https://doi.org/10.1093/jme/tjaa247

Balaska, S., Fotakis, E. A., Chaskopoulou, A., & Vontas, J. (2021). Chemical control and insecticide resistance status of sand fly vectors worldwide. *PLoS Neglected Tropical Diseases*, *15*(8), 1–23. https://doi.org/10.1371/journal.pntd.0009586

Benton, R., Vannice, K. S., Gomez-Diaz, C., & Vosshall, L. B. (2009). Variant Ionotropic Glutamate Receptors as Chemosensory Receptors in Drosophila. *Cell*, *136*(1), 149–162. https://doi.org/10.1016/j.cell.2008.12.001

Bernardini, F., Kriezis, A., Galizi, R., Nolan, T., & Crisanti, A. (2019). Introgression of a synthetic sex ratio distortion system from Anopheles gambiae into Anopheles arabiensis. *Scientific Reports*, *9*(1), 4–11. https://doi.org/10.1038/s41598-019-41646-8

Bernier, U. R., Kline, D. L., Allan, S. A., & Barnard, D. R. (2007). Laboratory comparison of Aedes aegypti attraction to human odors and to synthetic human odor compounds and blends. *Journal of the American Mosquito Control Association*, *23*(3), 288–293. https://doi.org/10.2987/8756-971X(2007)23[288:LCOAAA]2.0.CO;2

Bernier, U. R., Kline, D. L., Posey, K. H., Booth, M. M., Yost, R. A., & Barnard, D. R. (2003). Synergistic Attraction of Aedes aegypti (L.) to Binary Blends of L-Lactic Acid and Acetone, Dichloromethane, or Dimethyl Disulfide. *Journal of Medical Entomology*, *40*(5), 653–656. https://doi.org/10.1603/0022-2585-40.5.653

Bohbot, J. D., Sparks, J. T., & Dickens, J. C. (2014). The maxillary palp of Aedes aegypti, a model of multisensory integration. *Insect Biochemistry and Molecular Biology*, *48*(1), 29–39. https://doi.org/10.1016/j.ibmb.2014.02.007

Bosch, O. J., Geier, M., & Boeckh, J. (2000). Contribution of fatty acids to olfactory host finding of female Aedes aegypti. *Chemical Senses*, *25*(3), 323–330. https://doi.org/10.1093/oxfordjournals.chemse.a014042

Braks, M. A. H., & Takken, W. (1999). INCUBATED HUMAN SWEAT BUT NOT FRESH SWEAT ATTRACTS THE MALARIA MOSQUITO Anopheles gambiae gambiae SENSU STRICTO. *Journal of Chemical Ecology*, *25*(3), 663–672. https://doi.org/10.1023/A

Burtis, K. C., & Baker, B. S. (1989). Drosophila doublesex gene controls somatic sexual differentiation by producing alternatively spliced mRNAs encoding related sex-specific polypeptides. *Cell*, *56*(6), 997–1010. https://doi.org/10.1016/0092-8674(89)90633-8

Cardé, R. T. (2015). Multi-Cue Integration: How Female Mosquitoes Locate a Human Host. *Current Biology*, *25*(18), R793–R795. https://doi.org/10.1016/j.cub.2015.07.057

Carey, A. F., & Carlson, J. R. (2011). Insect olfaction from model systems to disease control. *Proceedings of the National Academy of Sciences*, *108*(32), 12987–12995. https://doi.org/10.1073/pnas.1103472108

Carey, A. F., Wang, G., Su, C. Y., Zwiebel, L. J., & Carlson, J. R. (2010). Odorant reception in the malaria mosquito Anopheles gambiae. *Nature*, *464*(7285), 66–71. https://doi.org/10.1038/nature08834

Chen, J., Luo, J., Wang, Y., Gurav, A. S., Li, M., Akbari, O. S., & Montell, C. (2021). Suppression of female fertility in Aedes aegypti with a CRISPR-targeted male-sterile mutation. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(22), 1–8. https://doi.org/10.1073/pnas.2105075118

Choo, Y. M., Buss, G. K., Tan, K., & Leal, W. S. (2015). Multitasking roles of mosquito labrum in oviposition and blood feeding. *Frontiers in Physiology*, *6*(OCT), 1–11. https://doi.org/10.3389/fphys.2015.00306

Cork, A., & Park, K. C. (1996). Identification of electrophysiologically-active compounds for the malaria mosquito, Anopheles gambiae, in human sweat extracts. *Medical and Veterinary Entomology*, *10*(3), 269–276. https://doi.org/10.1111/j.1365-2915.1996.tb00742.x

Coutinho-Abreu, I. V., Riffell, J. A., & Akbari, O. S. (2021). Human attractive cues and mosquito host-seeking behavior. *Trends in Parasitology*, 1–19. https://doi.org/10.1016/j.pt.2021.09.012

Da Silva Tavares, D., Salgado, V. R., Miranda, J. C., Mesquita, P. R. R., De Medeiros Rodrigues, F., Barral-Netto, M., … Barral, A. (2018). Attraction of phlebotomine sandflies to volatiles from skin odors of individuals residing in an endemic area of tegumentary leishmaniasis. *PLoS ONE*, *13*(9), 1–14. https://doi.org/10.1371/journal.pone.0203989

DeGennaro, M., McBride, C. S., Seeholzer, L., Vosshall, L. B., Jasinskiene, N., James, A. A., … Dennis, E. J. (2013). orco mutant mosquitoes lose strong preference for humans and are not repelled by volatile DEET. *Nature*, *498*(7455), 487–491. https://doi.org/10.1038/nature12206

Dekker, T., Geier, M., & Cardé, R. T. (2005). Carbon dioxide instantly sensitizes female yellow fever mosquitoes to human skin odours. *Journal of Experimental Biology*, *208*(15), 2963–2972. https://doi.org/10.1242/jeb.01736

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), 1–13. https://doi.org/10.1371/journal.pone.0122353

Dong, Y., Simões, M. L., Marois, E., & Dimopoulos, G. (2018). CRISPR/Cas9 -mediated gene knockout of Anopheles gambiae FREP1 suppresses malaria parasite infection. *PLoS Pathogens*, *14*(3), 1–16. https://doi.org/10.1371/journal.ppat.1006898

Dormont, L., Bessière, J. M., & Cohuet, A. (2013). Human Skin Volatiles: A Review. *Journal of Chemical Ecology*, *39*(5), 569–578. https://doi.org/10.1007/s10886-013-0286-z

Doudna, J. A., & Charpentier, E. (2014). The new frontier of genome engineering with CRISPR-Cas9. *Science*, *346*(6213). https://doi.org/10.1126/science.1258096

Dougherty, M. J., Guerin, P. M., & Ward, R. D. (1995). Identification of oviposition attractants for the sandfly Lutzomyia longipalpis (Diptera: Psychodidae) in volatiles of faeces from vertebrates. *Physiological Entomology*, *20*(1), 23–32. https://doi.org/10.1111/j.1365-3032.1995.tb00797.x

Erdelyan, C. N. G., Mahood, T. H., Bader, T. S. Y., & Whyard, S. (2012). Functional validation of the carbon dioxide receptor genes in Aedes aegypti mosquitoes using RNA interference. *Insect Molecular Biology*, *21*(1), 119–127. https://doi.org/10.1111/j.1365-2583.2011.01120.x

Fernandes, F. D. F., Bahia-Nascimento, C., Pinto, L. C., Leal, C. D. S., Secundino, N. F. C., & Pimenta, P. F. P. (2008). Fine structure and distribution pattern of antennal sensilla of lutzomyia longipalpis (Diptera: Psychodidae) sand flies. *Journal of Medical Entomology*, *45*(6), 982–990. https://doi.org/10.1603/0022-2585(2008)45[982:FSADPO]2.0.CO;2

Galizi, R., Hammond, A., Kyrou, K., Taxiarchi, C., Bernardini, F., O'Loughlin, S. M., … Crisanti, A. (2016). A CRISPR-Cas9 sex-ratio distortion system for genetic control. *Scientific Reports*, *6*(April), 2–6. https://doi.org/10.1038/srep31139

Gantz, V. M., Tatarenkova, O., Macias, V. M., James, A. A., Fazekas, A., Bier, E., & Jasinskiene, N. (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. https://doi.org/10.1073/pnas.1521077112

Gokcezade, J., Sienski, G., & Duchek, P. (2014). Efficient CRISPR/Cas9 Plasmids for Rapid and Versatile Genome Editing in Drosophila . *G3: Genes|Genomes|Genetics*, *4*(11), 2279–2282. https://doi.org/10.1534/g3.114.014126

Gratz, S. J., Cummings, A. M., Nguyen, J. N., Hamm, D. C., Donohue, L. K., Harrison, M. M., … O'connor-Giles, K. M. (2013). Genome engineering of Drosophila with the CRISPR RNA-guided Cas9 nuclease. *Genetics*, *194*(4), 1029–1035. https://doi.org/10.1534/genetics.113.152710

Gregory, M., Alphey, L., Morrison, N. I., & Shimeld, S. M. (2016). Insect transformation with piggyBac: Getting the number of injections just right. *Insect Molecular Biology*, *25*(3), 259–271. https://doi.org/10.1111/imb.12220

Guidobaldi, F., May-Concha, I. J., & Guerenstein, P. G. (2014). Morphology and physiology of the olfactory system of blood-feeding insects. *Journal of Physiology Paris*, *108*(2–3), 96–111. https://doi.org/10.1016/j.jphysparis.2014.04.006

Hall, A. B., Basu, S., Jiang, X., Qi, Y., Timoshevskiy, V. A., Biedler, J. K., … Tu, Z. (2015). A male-determining factor in the mosquito Aedes aegypti. *Science*, *348*(6240), 1268–1270. https://doi.org/10.1126/science.aaa2850

Hamilton, J., & Ramsoondar, T. (1994). *Attraction of Lutzomyia longipalpis to human skin odours*. *8*(4), 375–380.

Hammond, A., Galizi, R., Kyrou, K., Simoni, A., Siniscalchi, C., Katsanos, D., … Nolan, T. (2016). A

CRISPR-Cas9 gene drive system targeting female reproduction in the malaria mosquito vector Anopheles gambiae. *Nature Biotechnology*, *34*(1), 78–83. https://doi.org/10.1038/nbt.3439

Hammond, A., Pollegioni, P., Persampieri, T., North, A., Minuz, R., Trusso, A., … Crisanti, A. (2021). Gene-drive suppression of mosquito populations in large cages as a bridge between lab and field. *Nature Communications*, *12*(1), 1–9. https://doi.org/10.1038/s41467-021-24790-6

Hickner, P. V., Timoshevskaya, N., Nowling, R. J., Labbé, F., Nguyen, A. D., McDowell, M. A., … Syed, Z. (2020). Molecular signatures of sexual communication in the phlebotomine sand flies. *PLoS Neglected Tropical Diseases*, *14*(12), 1–18. https://doi.org/10.1371/journal.pntd.0008967

Hill, S. R., Hansson, B. S., & Ignell, R. (2009). Characterization of antennal trichoid sensilla from female Southern house mosquito, culex quinquefasciatus say. *Chemical Senses*, *34*(3), 231–252. https://doi.org/10.1093/chemse/bjn080

Hinze, A., Lantz, J., Hill, S. R., & Ignell, R. (2021). Mosquito Host Seeking in 3D Using a Versatile Climate-Controlled Wind Tunnel System. *Frontiers in Behavioral Neuroscience*, *15*(March). https://doi.org/10.3389/fnbeh.2021.643693

Hoermann, A., Habtewold, T., Selvaraj, P., Del Corsano, G., Capriotti, P., Inghilterra, M. G., … Windbichler, N. (2022). Gene drive mosquitoes can aid malaria elimination by retarding Plasmodium sporogonic development. *Science Advances*, *8*(38), 1–9. https://doi.org/10.1126/sciadv.abo1733

Ilango, K. (2000). Morphological characteristics of the antennal flagellum and its sensilla chaetica with character displacement in the sandfly Phlebotomus argentipes Annandale and Brunetti sensu lato (Diptera: Psychodidae). *Journal of Biosciences*, *25*(2), 163–172. https://doi.org/10.1007/bf03404911

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*(April), 1–10. https://doi.org/10.1038/srep24652

Jeffries, C. L., Rogers, M. E., & Walker, T. (2018). Establishment of a method for Lutzomyia longipalpis sand fly egg microinjection: The first step towards potential novel control strategies for leishmaniasis. *Wellcome Open Research*, *3*(0), 55. https://doi.org/10.12688/wellcomeopenres.14555.2

Juhn, J., & James, A. A. (2006). oskar gene expression in the vector mosquitoes, Anopheles gambiae and Aedes aegypti. *Insect Molecular Biology*, *15*(3), 363–372. https://doi.org/10.1111/j.1365-2583.2006.00655.x

Kent, L. B., Walden, K. K. O., & Robertson, H. M. (2008). The Gr family of candidate gustatory and olfactory receptors in the yellow-fever mosquito Aedes aegypti. *Chemical Senses*, *33*(1), 79–93. https://doi.org/10.1093/chemse/bjm067

Kistler, K. E., Vosshall, L. B., & Matthews, B. J. (2015). Genome engineering with CRISPR-Cas9 in the mosquito aedes aegypti. *Cell Reports*, *11*(1), 51–60. https://doi.org/10.1016/j.celrep.2015.03.009

Konopka, J. K., Task, D., Afify, A., Raji, J., Deibel, K., Maguire, S., … Potter, C. J. (2021). Olfaction in Anopheles mosquitoes. *Chemical Senses*, *46*(April), 1–24. https://doi.org/10.1093/chemse/bjab021

Koutroumpa, F. A., Monsempes, C., François, M. C., De Cian, A., Royer, C., Concordet, J. P., & Jacquin-Joly, E. (2016). Heritable genome editing with CRISPR/Cas9 induces anosmia in a crop pest moth. *Scientific Reports*, *6*(February), 1–9. https://doi.org/10.1038/srep29620

Kwon, H. W., Lu, T., Rützler, M., & Zwiebel, L. J. (2006). Olfactory response in a gustatory organ of the malaria vector mosquito Anopheles gambiae. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(36), 13526–13531. https://doi.org/10.1073/pnas.0601107103

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

Lacey, E. S., & Cardé, R. T. (2011). Activation, orientation and landing of female Culex quinquefasciatus in response to carbon dioxide and odour from human feet: 3-D flight analysis in a wind tunnel. *Medical and Veterinary Entomology*, *25*(1), 94–103. https://doi.org/10.1111/j.1365-2915.2010.00921.x

Larsson, M. C., Domingos, A. I., Jones, W. D., Chiappe, M. E., Amrein, H., & Vosshall, L. B. (2004). Or83b encodes a broadly expressed odorant receptor essential for Drosophila olfaction. *Neuron*, *43*(5), 703–714. https://doi.org/10.1016/j.neuron.2004.08.019

Lawyer, P., Volf, P., Rowton, E., Rowland, T., Killick-Kendrick, M., Rowland, T., … Volf, P. (2017). Laboratory colonization and mass rearing of phlebotomine sand flies (Diptera, Psychodidae). *Parasite*, *24*, 42. https://doi.org/10.1051/parasite/2017041

Li, M., Li, T., Liu, N., Raban, R. R., Wang, X., & Akbari, O. S. (2020). Methods for the generation of heritable germline mutations in the disease vector Culex quinquefasciatus using clustered regularly interspaced short palindrome repeats-associated protein 9. *Insect Molecular Biology*, *29*(2), 214–220. https://doi.org/10.1111/imb.12626

Li, Ming, Bui, M., Yang, T., Bowman, C. S., White, B. J., & Akbari, O. S. (2017). Germline Cas9 expression yields highly efficient genome engineering in a major worldwide disease vector, Aedes aegypti. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(49), E10540–E10549. https://doi.org/10.1073/pnas.1711538114

Li, Ming, Yang, T., Bui, M., Gamez, S., Wise, T., Kandul, N. P., … Akbari, O. S. (2021). Suppressing mosquito populations with precision guided sterile males. *Nature Communications*, *12*(1). https://doi.org/10.1038/s41467-021-25421-w

Li, Y., Zhang, J., Chen, D., Yang, P., Jiang, F., Wang, X., & Kang, L. (2016). CRISPR/Cas9 in locusts: Successful establishment of an olfactory deficiency line by targeting the mutagenesis of an odorant receptor co-receptor (Orco). *Insect Biochemistry and Molecular Biology*, *79*, 27–35. https://doi.org/10.1016/j.ibmb.2016.10.003

Liu, F., Ye, Z., Baker, A., Sun, H., & Zwiebel, L. J. (2020). Gene editing reveals obligate and modulatory components of the CO2 receptor complex in the malaria vector mosquito, Anopheles coluzzii. *Insect Biochemistry and Molecular Biology*, *127*(August), 1–9. https://doi.org/10.1016/j.ibmb.2020.103470

Liu, H., Liu, T., Xie, L., Wang, X., Deng, Y., Chen, C. H., … Chen, X. G. (2016). Functional analysis of Orco and odorant receptors in odor recognition in Aedes albopictus. *Parasites and Vectors*, *9*(1), 1–10. https://doi.org/10.1186/s13071-016-1644-9

Liu, T., Yang, W. Q., Xie, Y. G., Liu, P. W., Xie, L. H., Lin, F., … Chen, X. G. (2019). Construction of an efficient genomic editing system with CRISPR/Cas9 in the vector mosquito Aedes albopictus. *Insect Science*, *26*(6), 1045–1054. https://doi.org/10.1111/1744-7917.12645

Liu, Y. Y., Ma, S., Wang, X., Chang, J., Gao, J., Shi, R., … Xia, Q. (2014). Highly efficient multiplex targeted mutagenesis and genomic structure variation in Bombyx mori cells using CRISPR/Cas9.

Insect Biochemistry and Molecular Biology, *49*(1), 35–42. https://doi.org/10.1016/j.ibmb.2014.03.010

Louradour, I., Ghosh, K., Inbar, E., & Sacks, D. L. (2019). CRISPR/Cas9 Mutagenesis in Phlebotomus papatasi: the Immune Deficiency Pathway Impacts Vector Competence for Leishmania major. *MBio*, *10*(4), 1–14.

Lu, T., Qiu, Y. T., Wang, G., Kwon, J. Y., Rutzler, M., Kwon, H. W., … Zwiebel, L. J. (2007). Odor Coding in the Maxillary Palp of the Malaria Vector Mosquito Anopheles gambiae. *Current Biology*, *17*(18), 1533–1544. https://doi.org/10.1016/j.cub.2007.07.062

Magalhães-Junior, J. T., Oliva-Filho, A. D. A., Novais, H. O., Mesquita, P. R. R., M. Rodrigues, F., Pinto, M. C., & Barrouin-Melo, S. M. (2019). Attraction of the sandfly Lutzomyia longipalpis to possible biomarker compounds from dogs infected with Leishmania infantum. *Medical and Veterinary Entomology*, *33*(2), 322–325. https://doi.org/10.1111/mve.12357

Majeed, S., Hill, S. R., Birgersson, G., & Ignell, R. (2016). Detection and perception of generic host volatiles by mosquitoes modulate host preference: Context dependence of (R)-1-octen-3-ol. *Royal Society Open Science*, *3*(11). https://doi.org/10.1098/rsos.160467

Mann, R. S., Kaufman, P. E., & Butler, J. F. (2009). Lutzomyia spp. (Diptera: Psychodidae) response to olfactory attractant- and light emitting diode-modified mosquito magnet X (MM-X) traps. *Journal of Medical Entomology*, *46*(5), 1052–1061. https://doi.org/10.1603/033.046.0512

Martin-Martin, I., Aryan, A., Meneses, C., Adelman, Z. N., & Calvo, E. (2018). Optimization of sand fly embryo microinjection for gene editing by CRISPR/Cas9. *PLoS Neglected Tropical Diseases*, *12*(9), 1–18. https://doi.org/10.1371/journal.pntd.0006769

Masse, N. Y., Turner, G. C., & Jefferis, G. S. X. E. (2009). Olfactory Information Processing in Drosophila. *Current Biology*, *19*(16), R700–R713. https://doi.org/10.1016/j.cub.2009.06.026

Matthews, B. J., McBride, C. S., DeGennaro, M., Despo, O., & Vosshall, L. B. (2016). The neurotranscriptome of the Aedes aegypti mosquito. *BMC Genomics*, *17*(1), 1–20. https://doi.org/10.1186/s12864-015-2239-0

McBride, C. S., Baier, F., Omondi, A. B., Spitzer, S. A., Lutomiah, J., Sang, R., … Vosshall, L. B. (2014). Evolution of mosquito preference for humans linked to an odorant receptor. *Nature*, *515*(7526), 222–227. https://doi.org/10.1038/nature13964

Mclver, S. B. (1982). *Review Article Sensilla of Mosquitoes (Diptera : Culicidae) 12*. *19*(5), 489–535.

McMeniman, C. J., Corfas, R. A., Matthews, B. J., Ritchie, S. A., & Vosshall, L. B. (2014). Multimodal integration of carbon dioxide and other sensory cues drives mosquito attraction to humans. *Cell*, *156*(5), 1060–1071. https://doi.org/10.1016/j.cell.2013.12.044

Montell, C. (2009). A taste of the Drosophila gustatory receptors. *Current Opinion in Neurobiology*, *19*(4), 345–353. https://doi.org/10.1016/j.conb.2009.07.001

Mukabana, W. R., Mweresa, C. K., Otieno, B., Omusula, P., Smallegange, R. C., van Loon, J. J. A., & Takken, W. (2012). A Novel Synthetic Odorant Blend for Trapping of Malaria and Other African Mosquito Species. *Journal of Chemical Ecology*, *38*(3), 235–244. https://doi.org/10.1007/s10886-012-0088-8

Mweresa, C. K., Mukabana, W. R., Omusula, P., Otieno, B., Van Loon, J. J. A., & Takken, W. (2016). Enhancing Attraction of African Malaria Vectors to a Synthetic Odor Blend. *Journal of Chemical Ecology*, *42*(6), 508–516. https://doi.org/10.1007/s10886-016-0711-1

Navarro-Payá, D., Flis, I., Anderson, M. A. E., Hawes, P., Li, M., Akbari, O. S., … Alphey, L. (2020).

Targeting female flight for genetic control of mosquitoes. *PLoS Neglected Tropical Diseases*, *14*(12), e0008876. https://doi.org/10.1371/journal.pntd.0008876

O'leary, S., & Adelman, Z. N. (2020). Crispr/cas9 knockout of female-biased genes aeact-4 or myo-fem in ae. Aegypti results in a flightless phenotype in female, but not male mosquitoes. *PLoS Neglected Tropical Diseases*, *14*(12), 1–19. https://doi.org/10.1371/journal.pntd.0008971

Okumu, F. O., Killeen, G. F., Ogoma, S., Biswaro, L., Smallegange, R. C., Mbeyela, E., … Moore, S. J. (2010). Development and field evaluation of a synthetic mosquito lure that is more attractive than humans. *PLoS ONE*, *5*(1), 1–8. https://doi.org/10.1371/journal.pone.0008951

Pinto, M. C., Campbell-Lendrum, D. H., Lozovei, A. L., Teodoro, U., & Davies, C. R. (2001). Phlebotomine sandfly responses to carbon dioxide and human odour in the field. *Medical and Veterinary Entomology*, *15*(2), 132–139. https://doi.org/10.1046/j.1365-2915.2001.00294.x

Pitts, R. J., Rinker, D. C., Jones, P. L., Rokas, A., & Zwiebel, L. J. (2011). Transcriptome profiling of chemosensory appendages in the malaria vector Anopheles gambiae reveals tissue- and sex-specific signatures of odor coding. *BMC Genomics*, *12*(1), 271. https://doi.org/10.1186/1471-2164-12-271

Pitts, R. Jason, Fox, A. N., & Zwiebeil, L. J. (2004). A highly conserved candidate chemoreceptor expressed in both olfactory and gustatory tissues in the malaria vector Anopheles gambiae. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(14), 5058–5063. https://doi.org/10.1073/pnas.0308146101

Pitts, R Jason, Derryberry, S. L., Zhang, Z., & Zwiebel, L. J. (2017). Variant Ionotropic Receptors in the Malaria Vector Mosquito Anopheles gambiae Tuned to Amines and Carboxylic Acids. *Nature Publishing Group*, (September 2016), 1–11. https://doi.org/10.1038/srep40297

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

Potter, C. J. (2014). Stop the biting: Targeting a mosquito's sense of smell. *Cell*, Vol. 156, pp. 878–881. https://doi.org/10.1016/j.cell.2014.02.003

Purusothaman, D. K., Shackleford, L., Anderson, M. A. E., Harvey-Samuel, T., & Alphey, L. (2021). CRISPR/Cas-9 mediated knock-in by homology dependent repair in the West Nile Virus vector Culex quinquefasciatus Say. *Scientific Reports*, *11*(1), 1–8. https://doi.org/10.1038/s41598-021-94065-z

Qiu, Y. T., van Loon, J. J. A., Takken, W., Meijerink, J., & Smid, H. M. (2006). Olfactory coding in antennal neurons of the malaria mosquito, Anopheles gambiae. *Chemical Senses*, *31*(9), 845–863. https://doi.org/10.1093/chemse/bjl027

Quinnell, R. J., Dye, C., & Shaw, J. J. (1992). Host preferences of the phlebotomine sandfly Lutzomyia longipalpis in Amazonian Brazil. *Medical and Veterinary Entomology*, *6*(3), 195–200. https://doi.org/10.1111/j.1365-2915.1992.tb00606.x

Raji, J. I., Melo, N., Castillo, J. S., Gonzalez, S., Saldana, V., Stensmyr, M. C., & DeGennaro, M. (2019). Aedes aegypti Mosquitoes Detect Acidic Volatiles Found in Human Odor Using the IR8a Pathway. *Current Biology*, *29*(8), 1253-1262.e7. https://doi.org/10.1016/j.cub.2019.02.045

Riabinina, O., Task, D., Marr, E., Lin, C. C., Alford, R., O'Brochta, D. A., & Potter, C. J. (2016). Organization of olfactory centres in the malaria mosquito Anopheles gambiae. *Nature Communications*, *7*. https://doi.org/10.1038/ncomms13010

Rinker, D. C., Zhou, X., Pitts, R. J., Consortium, T. A. G. C., & Rokas, A. (2013). *Antennal*

transcriptome profiles of anopheline mosquitoes reveal human host olfactory specialization in Anopheles gambiae Antennal transcriptome profiles of anopheline mosquitoes reveal human host olfactory specialization in Anopheles gambiae.

Salz, H. K., & Erickson, J. W. (2010). Sex determination in Drosophila: The view from the top. *Fly*, *4*(1), 60–70. https://doi.org/10.4161/fly.4.1.11277

Sato, K., Pellegrino, M., Nakagawa, T., Nakagawa, T., Vosshall, L. B., & Touhara, K. (2008). Insect olfactory receptors are heteromeric ligand-gated ion channels. *Nature*, *452*(7190), 1002–1006. https://doi.org/10.1038/nature06850

Saveer, A. M., Pitts, R. J., Ferguson, S. T., & Zwiebel, L. J. (2018). Characterization of Chemosensory Responses on the Labellum of the Malaria Vector Mosquito, Anopheles coluzzii. *Scientific Reports*, *8*(1), 1–10. https://doi.org/10.1038/s41598-018-23987-y

Schetelig, M. F., & Handler, A. M. (2013). A functional comparison of the 3xp3 promoter by recombinase-mediated cassette exchange in drosophila and a tephritid fly, anastrepha suspensa. *G3: Genes, Genomes, Genetics*, *3*(4), 687–693. https://doi.org/10.1534/g3.112.005488

Simoni, A., Hammond, A. M., Beaghton, A. K., Galizi, R., Taxiarchi, C., Kyrou, K., … Crisanti, A. (2020). A male-biased sex-distorter gene drive for the human malaria vector Anopheles gambiae. *Nature Biotechnology*, *38*(9), 1054–1060. https://doi.org/10.1038/s41587-020-0508-1

Smallegange, R. C., Qiu, Y. T., van Loon, J. A., & Takken, W. (2005). Synergism between ammonia, lactic acid and carboxylic acids as kairomones in the host-seeking behaviour of the malaria mosquito Anopheles gambiae sensu stricto (Diptera: Culicidae). *Chemical Senses*, *30*(2), 145–152. https://doi.org/10.1093/chemse/bji010

Smallegange, R. C., Verhulst, N. O., & Takken, W. (2011). Sweaty skin: An invitation to bite? *Trends in Parasitology*, *27*(4), 143–148. https://doi.org/10.1016/j.pt.2010.12.009

Sparks, J. T., & Dickens, J. C. (2017). Mini review : Gustatory reception of chemicals affecting host feeding in aedine mosquitoes. *Pesticide Biochemistry and Physiology*, *142*, 15–20. https://doi.org/10.1016/j.pestbp.2016.12.009

Spiegel, C. N., Oliveira, S. M. P., Brazil, R. P., & Soares, M. J. (2005). Structure and distribution of sensilla on maxillary palps and labella of Lutzomyia longipalpis (Diptera: Psychodidae) sand flies. *Microscopy Research and Technique*, *66*(6), 321–330. https://doi.org/10.1002/jemt.20180

Spitzen, J., Smallegange, R. C., & Takken, W. (2008). Effect of human odours and positioning of CO2 release point on trap catches of the malaria mosquito Anopheles gambiae sensu stricto in an olfactometer. *Physiological Entomology*, *33*(2), 116–122. https://doi.org/10.1111/j.1365-3032.2008.00612.x

Suh, E., Bohbot, J. D., & Zwiebel, L. J. (2014). Peripheral olfactory signaling in insects. *Current Opinion in Insect Science*, *6*, 86–92. https://doi.org/10.1016/j.cois.2014.10.006

Sun, H., Liu, F., Ye, Z., Baker, A., & Zwiebel, L. J. (2020). Mutagenesis of the orco odorant receptor co-receptor impairs olfactory function in the malaria vector Anopheles coluzzii. *Insect Biochemistry and Molecular Biology*, *127*(October). https://doi.org/10.1016/j.ibmb.2020.103497

Syed, Z., & Leal, W. S. (2009). Acute olfactory response of Culex mosquitoes to a human- and bird-derived attractant. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(44), 18803–18808. https://doi.org/10.1073/pnas.0906932106

Takken, W., & Knols, B. G. J. (1999). Odor-mediated behavior of Afrotropical malaria mosquitoes. *Annual Review of Entomology*, *44*(May 2014), 131–157. https://doi.org/10.1146/annurev.ento.44.1.131

Task, D., Lin, C. C., Vulpe, A., Afify, A., Ballou, S., Brbic, M., … Potter, C. J. (2022). Chemoreceptor co-expression in Drosophila melanogaster olfactory neurons. *ELife*, *11*, 1–69. https://doi.org/10.7554/eLife.72599

Tauxe, G. M., Macwilliam, D., Boyle, S. M., Guda, T., & Ray, A. (2013). XTargeting a dual detector of skin and CO2 to modify mosquito host seeking. *Cell*, *155*(6), 1365–1379. https://doi.org/10.1016/j.cell.2013.11.013

van Loon, J. J. A., Smallegange, R. C., Bukovinszkiné-Kiss, G., Jacobs, F., De Rijk, M., Mukabana, W. R., … Takken, W. (2015). Mosquito Attraction: Crucial Role of Carbon Dioxide in Formulation of a Five-Component Blend of Human-Derived Volatiles. *Journal of Chemical Ecology*, *41*(6), 567–573. https://doi.org/10.1007/s10886-015-0587-5

Volf, P., & Volfova, V. (2011). Establishment and maintenance of sand fly olonies. *Journal of Vector Ecology*, *36*(SUPPL.1), 1–9. https://doi.org/10.1111/j.1948-7134.2011.00106.x

Vosshall, L. B., & Hansson, B. S. (2011). A unified nomenclature system for the insect olfactory coreceptor. *Chemical Senses*, *36*(6), 497–498. https://doi.org/10.1093/chemse/bjr022

Wang, S., & Jacobs-Lorena, M. (2013). Genetic approaches to interfere with malaria transmission by vector mosquitoes. *Trends in Biotechnology*, *31*(3), 185–193. https://doi.org/10.1016/j.tibtech.2013.01.001

Wang, Y., He, X., Qiao, L., Yu, Z., Chen, B., & He, Z. (2022). CRISPR/Cas9 mediates efficient site-specific mutagenesis of the odorant receptor co-receptor (Orco) in the malaria vector Anopheles sinensis. *Pest Management Science*, (April). https://doi.org/10.1002/ps.6954

Williams, C. R., Bergbauer, R., Geier, M., Kline, D. L., Bernier, U. R., Russell, R. C., & Ritchie, S. A. (2006). Laboratory and field assessment of some kairomone blends for host-seeking Aedes aegypti. *Journal of the American Mosquito Control Association*, *22*(4), 641–647. https://doi.org/10.2987/8756-971X(2006)22[641:LAFAOS]2.0.CO;2

Ye, Z., Liu, F., Sun, H., Ferguson, S. T., Baker, A., Ochieng, S. A., & Zwiebel, L. J. (2022). Discrete roles of Ir76b ionotropic coreceptor impact olfaction, blood feeding, and mating in the malaria vector mosquito Anopheles coluzzii. *Proceedings of the National Academy of Sciences of the United States of America*, *119*(23). https://doi.org/10.1073/pnas.2112385119

Younger, M. A., Herre, M., Goldman, O. V., Lu, T., Caballero-Vidal, G., Qi, Y., … Vosshall, L. B. (2022). Non-Canonical Odor Coding in the Mosquito. *BioRxiv*, 1–86.

Chapter 3

Identification and rationalisation of phenotypic marker genes, and olfactory gene targets to affect by gene editing

Introduction

Prior to attempting CRISPR mutagenesis within sand fly species *L. longipalpis* and *P. papatasi*, appropriate genes must be selected as targets. With respect to proof-of-concept studies, phenotypic markers are often selected to allow rapid assessment of mutagenesis. Typically gene targets are non-lethal, and cause limited fitness cost to survivors.

Genes associated with pigmentation characteristics provide useful markers that provisionally do not require molecular analysis for identification of transgenesis, facilitating confirmation or estimation transgenic individuals.

Many phenotypic marker genes have been identified in *Drosophila melanogaster* and also in mosquito species that have been successfully modified to result in a phenotypic change. Some common examples are cuticle pigmentation (*y,* and *e,*), eye pigmentation (*W*, *sc*, *cin*, *kh*, and *kmo*), and wing morphology (*Vg*, *r*, *cysu*, and *Ct*) (Table 1) as clear indicators of successful modification via CRISPR gRNAs or constructs. These provide initial targets for identification of orthologues within sand flies for consideration as candidate targets for downstream demonstration of CRISPR mediated trangenesis. Gene editing of phenotypic genes has not previously been demonstrated in *L. longipalpis* or *P. papatasi.*

Beyond proof-of-concept targets of modification, genes involved in fundamental behaviours provide potential targets for control strategies. In particular, olfactory genes are important in host-seeking behaviour within insect vectors, (see Chapter 2), which are prime targets for disruption and potential gene-drive-mediated population replacement. Published research on sand fly olfaction is sparse, however the model organism *D. melanogaster* and mosquito vectors from two key orders (*Anophelinae* and *Culicinae*) have been studied extensively.

The genomes of *L. longipalpis* and *P. papatasi* sand flies have been sequenced (see Methods, *Sand fly genome assemblies*), and recently annotated with respect to olfactory genes (Hickner et al., 2020). Chemoreceptors related to important genes for host-seeking in mosquitoes have also been identified within sand flies (see Chapter 2 for full description), sharing phylogenetic similarities to *Drosophila* and mosquito species, allowing for orthologue comparisons between the annotated olfactory genes in these model organisms and sand fly species.

Once potential target genes have been identified and evaluated, CRISPR tools (gRNAs for in vitro transcription or for plasmid construction) are typically designed prior to construction. The design requires adherence to several parameters, a process which has been simplified by the availability of genome databases and bioinformatics webtools (outlined below, and in Methods, *Bioinformatic identification and design of gRNAs* and *Oligonucleotide Design for CRISPR constructs*).

Design of gRNAs

CRISPR-Cas9 modification requires a gRNA sequence (~17-20 base pairs) to target a Cas9 nuclease to a specific genomic location for cleavage (due to complementarity with the genomic DNA) (Figure 1). In addition, a 2-6 base pair protospacer adjacent motif (PAM) is required, 3' of the gRNA for Cas9 nuclease recognition prior to cleaveage. The PAM sequence is specific to the bacterial species origin of the Cas9 (Wu, Kriz, & Sharp, 2014); in the *S. pyogenes* CRISPR system (commonly used), the PAM sequence is most efficient in the form 5'-NGG-3' (X. Zhang, Koolhaas, & Schnorrer, 2014).Therefore the design of gRNAs must incorporate these feature.

For *in vivo* transcription of gRNAs in eukaryotic cells, RNA polymerase III (Pol III) promoters are required, such as the U6 promoter (within the CRISPR pDCC6 plasmid, see Chapter 4). This promoter only initiates transcription of short RNA sequences at a guanine (*G*) base, therefore, gRNA sequence design must take this constraint into account.

Different bioinformatics webtools have been developed to simplify the process of gRNA design for CRISPR-Cas approaches, to ensure relevant feature for transcription are included. Popular software developed for gRNA optimisation include CRISPick (Doench et al., 2016), Cas-OFFinder (Bae, Park, & Kim, 2014), and ChopChop (Labun et al., 2019) amongst others. These tools use algorithmic methods to select candidate gRNA sequences to maximise on-target activity, and provide a rank to aid selection. Multiple features can be optimised, including selection of the genome and genes of interest to be targeted, selection of the required parameters such as the 3' PAM site structure (NGG), the length of gRNA sequence (~17-20), additional 5' requirements (guanine nucleotide), and restriction of off-target sites within the genome to prevent unintended Cas9 cleavage.

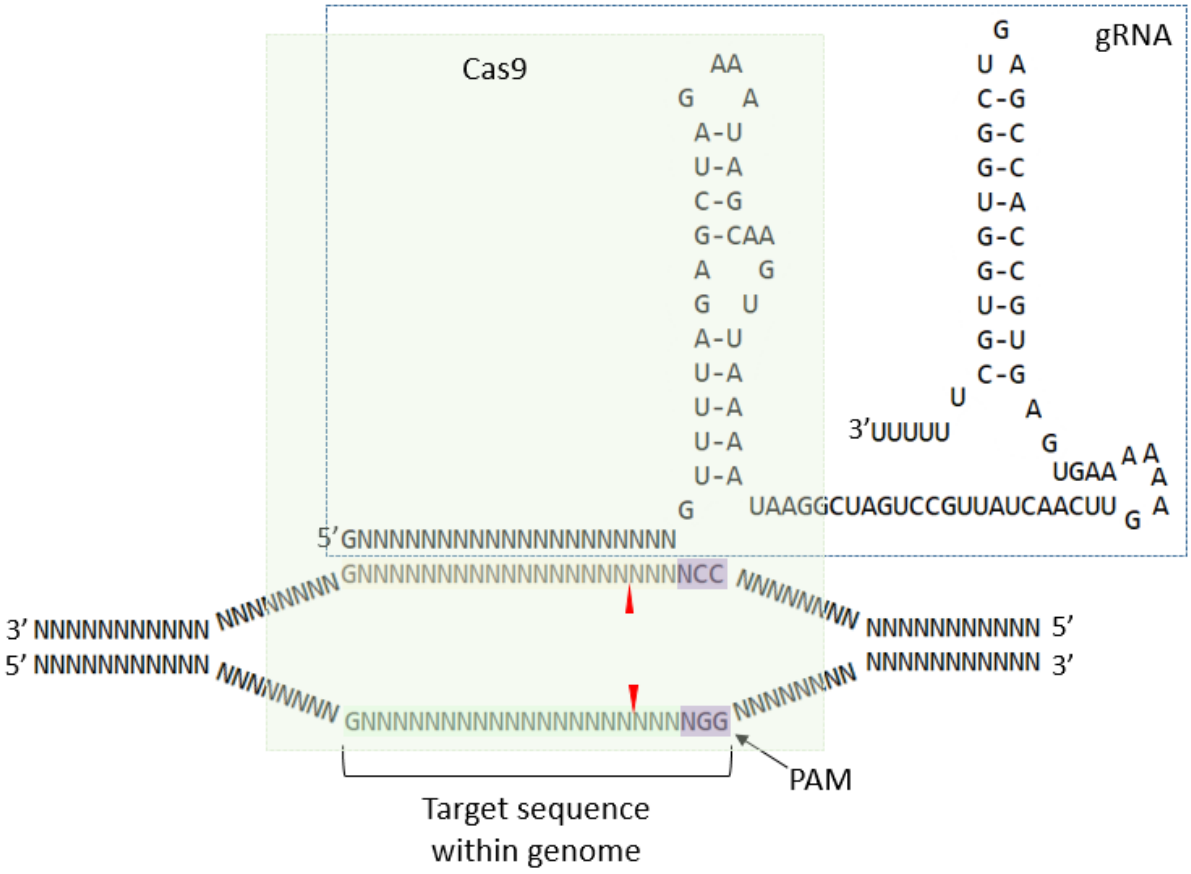


Figure 1. Illustration of Cas9 protein interacting with gRNA to direct cleavage of genomic DNA at a target site. The gRNA sequence and target sequence within the genome are homologous, and contains a PAM site (NGG, purple) facilitating Cas9 induced cleavage. The gRNA (blue box) consists of the recognition sequence (GN^{17-20}), alongside Cas9 recognition sequences (hairpins/loops). Cas9 cleavage occurs 3 – 4 nucleotides 5' of the PAM (red arrow). Adapted from DiCarlo *et al.* 2013.

Single and Multiple gRNA targeting strategies (gRNA multiplexing)

Commonly single gRNAs are applied to affect gene knockouts. However, indels can be small and difficult to detect, for example by gel electrophoresis necessitating the use of sequence-based approaches. The use of multiple gRNAs inducing deletion of a larger fragment can make visualisation easier by PCR amplification followed by a gel electrophoresis (Gratz et al., 2013; Kondo & Ueda, 2013) (Figure 2). This approach can also be applied to target different genes simultaneously (H. Wang et al., 2013), or to modify a gene by offset-nicking (Mali et al., 2013; Ran et al., 2013).

Attempting a large deletion within a gene is appropriate for proof-of-concept studies, allowing rapid PCR screening particularly when no phenotype is visible. Gratz *et al.,* (2013) utilised this method to target the 5' and 3' ends of the *Yellow* locus in *Drosophila*. A 4.6 kilobase (Kb) region was successfully deleted, and confirmed by both yellow mosaicism of the cuticle in adults and PCR. In addition, sequence results identified non-homologous end joining (NHEJ) as the repair mechanism at the two cut sites,

inducing small indels. Furthermore, a plasmid approach has been used to target the *White* gene in *Drosophila*. A plasmids containing two gRNAs promoted by U6, targeted a 1.6 Kb region, successfully resulting in phenotypic mutants, 14% of which had deletions confirmed by PCR (Kondo & Ueda, 2013),

Synchronous delivery of multiple *in vitro* transcribed sgRNAs has been demonstrated in *Cx. quinquefasciatus*. Again, the *White* gene with 1-3 sgRNAs, showed increased efficiency of mutations when two of three sgRNAs were co-injected, with G0 mutagenesis up to 86%. Additionally germline mutations and subsequent inheritance of mutations in *Cx. quinquefasciatus* G1 were more than 79% (M. Li et al., 2020).

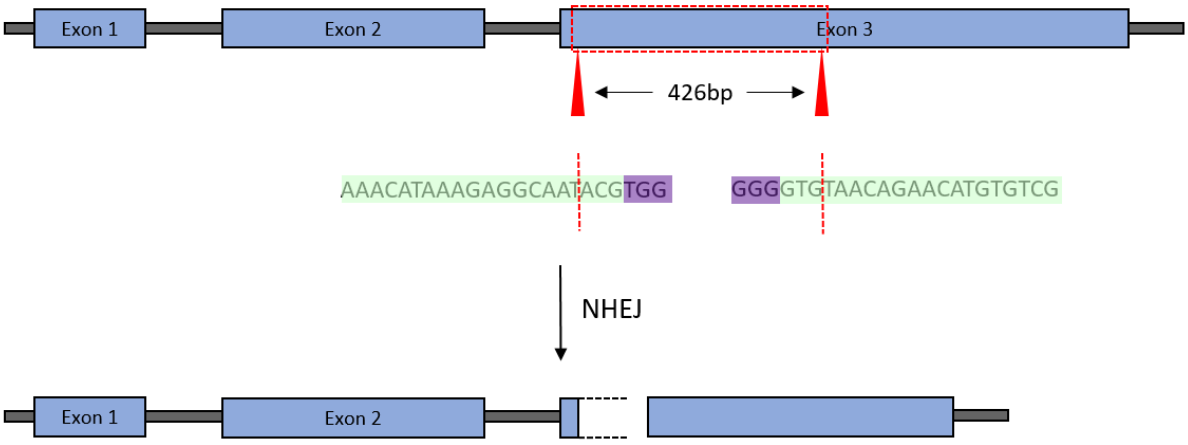


Figure 2. Illustration of two gRNA targeting of a gene to induce a large knockout to interrupt a gene. Designing gRNAs that target the same gene with a separation of several hundred base pairs may successfully result in a large portion of the gene being knocked out, both disrupting the function of the gene and enabling easy screening via PCR. Red arrows indicate the location of the cut site (red hashed lines) within the target sequence, 3-4 nucleotides upstream of the PAM site (purple). Cellular repair is likely to result in non-homologous end joining (NHEJ), truncating the gene.

In summary, the availability of annotated genomes with associated functions in model insects, in tandem with gRNA generating software, now allows identification of orthologous genes and an informed translation to sand flies. The steps to identify and rationalise targets to affect are described below.

Aims

This chapter aimed to develop a pipeline for identification of putative target genes in model organisms and insect vectors, followed by the identification of orthologous genes in two sand fly species. Subsequently, gRNAs are transcribed for downstream assessment (Chapter 4 and Chapter 5) as part of a CRISPR mediated strategy. In more detail, we first identify target genes of interest for CRISPR mutagenesis using different approaches (see Methods below) based on model dipteran insects. We prioritise non-lethal phenotypic marker genes (cuticle pigmentation, eye pigmentation, and wing development for proof-of-concept studies in *L. longipalpis* and *P. papatasi*, in addition to a selection of olfactory gene targets for modification via CRISPR knockouts in *L. longipalpis*. The process derives sand fly specific gRNAs for CRISPR constructs, for application to *in vitro* (Chapter 4) and *in vivo* studies (Chapter 5).

Specific objectives:

- Review of phenotypic and olfactory candidate gene targets
- Selection of target genes from candidate loci
- Rationalisation of phenotypic and olfactory candidate gene targets
- Design of gRNAs to target phenotypic and olfactory genes for CRISPR mediated gene editing

Methods

Sand fly genome assemblies

The *Lutzomyia longipalpis* and *Phlebotomus papatasi* genomes used in this research are hosted on VectorBase (https://vectorbase.org/vectorbase/app/). VectorBase is a bioinformatics Resource for invertebrate vectors of human disease hosting a range of bioinformatics tools, and genome assemblies. The *Lutzomyia longipalpis* genome was assembled by the Baylor College of Medicine (BCM) Sand Fly Genome Project (https://www.hgsc.bcm.edu/arthropods/sand-fly-genome-project) using three whole genome sequences (WGS). In total BCM used 22.6 million reads to generate the WGS, with 38.9x coverage. The genome was built using Celera CABOG assembler, and further assemblies were conducted using BCMs ATLAS-link.gaps, ATLAS-gapfil, and ATLAS-gapmerge to complete the genome. The genome is 154.23Mbp, and consists of 11,539 scaffolds, and 10,765 genes. The version used in this thesis for gene identification, and sequence alignment is labelled LLonJ1.5.

The *Phlebotomus papatasi* genome was assembled by The Genome Institute, Washington University School of Medicine (St. Louis, USA) from laboratory reared females and hosted on VectorBase. The genome was built using the Newbler assembler with 22.5x whole genome coverage. Gaps in the genome were filled by PyGap, which merged overlapping contigs. The genome is 363.77Mbp, and consists of 11,835 genes. The version used in this thesis for gene identification and sequence alignment is labelled Ppap1.6.

Phylogenetic analysis for rationalisation of candidate genes

Candidate genes were identified through a review of the published literature and communication with experts in the field of insect vector olfaction. Protein sequences for phenotypic and olfactory gene orthologues were identified using the FlyBase and VectorBase databases by inputting gene IDs. Peptide sequences were aligned in MEGA X (version10.1.8) using the ClustalW multiple sequence alignment function . Once aligned, phylogenetic analysis was performed to identify conserved genes that are likely to have similar function. The evolutionary relationship between the genes was inferred by using the Maximum Likelihood method and JTT matrix-based model (Jones, Taylor, & Thornton, 1992). The bootstrap consensus trees were inferred from 500 replicates and taken to represent the evolutionary history of the taxa analysed (Felsenstein, 1985). The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (500 replicates) are shown next to the branches (Felsenstein, 1985). Initial trees for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the JTT model, and then selecting the topology with superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites. Fewer than 5% alignment gaps, missing data,

and ambiguous bases were allowed at any position. Evolutionary analyses were conducted in MEGA X (Kumar, Stecher, Li, Knyaz, & Tamura, 2018).

Bioinformatic identification and design of gRNAs

The ChopChop webtool (https://chopchop.cbu.uib.no/) (Labun et al., 2019) was used to identify potential gRNAs for targeting genes of interest (LLOJ007802, LLOJ008326, LLOJ001311, LLOJ001495 , LLOJ009814, LLOJ009695, and LLOJ009278) and associated primers to amplify the regions of interest flanking gRNA sequences.

Stock settings for inducing CRISPR knock-out frameshift mutations in the genes of interest were applied. This generated gRNA sequence with 20nt and 3' *NGG* PAM site. The gRNAs were checked for self-complementarity, which can reduce cutting efficiency (Thyme, Akhmetova, Montague, Valen, & Schier, 2016). For *in vitro* transcription of gRNAs additional guanine nucleotides (*G or GG*) were added at the 5' end of the sequence to allow for T7 transcription (see Chapter 4 Methods, *In vitro transcription of gRNAs*). Off-targets sites were identified within the genome, and efficiency scores calculated in relation to the presence of a guanine at position 20, proximal to the PAM (Doench et al., 2016). Primers were chosen that minimised off-target amplification, and were optimised for size, melting temperature, and amplicon size suggested by the software. Primers that matched the required specifications are generated automatically through on a schematic diagram of the gene in tandem with relevant information, including off-target mismatches.

Oligonucleotide Design for CRISPR constructs

To transcribe gRNAs and insert into CRISPR plasmid (see Chapter 4, *Generation of Plasmid constructs via Gibson Assembly*) constructs additional features are required (Figure 3). Briefly, a backbone plasmid (pDCC6, containing a Cas9 encoding cargo sequence and gRNA spacer promoted by the U6 promotor) is digested using enzymes creating sticky ends, and a double stranded gRNA oligonucleotide is ligated into the gap. To ensure correct ligation of gRNAs into the pDCC6 plasmid backbone, four nucleotide overhangs (AATT) are added to the 5' end of each gRNA sequence and the 3' end (AAAC) of the reverse complement sequence. An additional guanine nucleotide (*G*) is added to the beginning of the forward gRNA sequence if not already present.

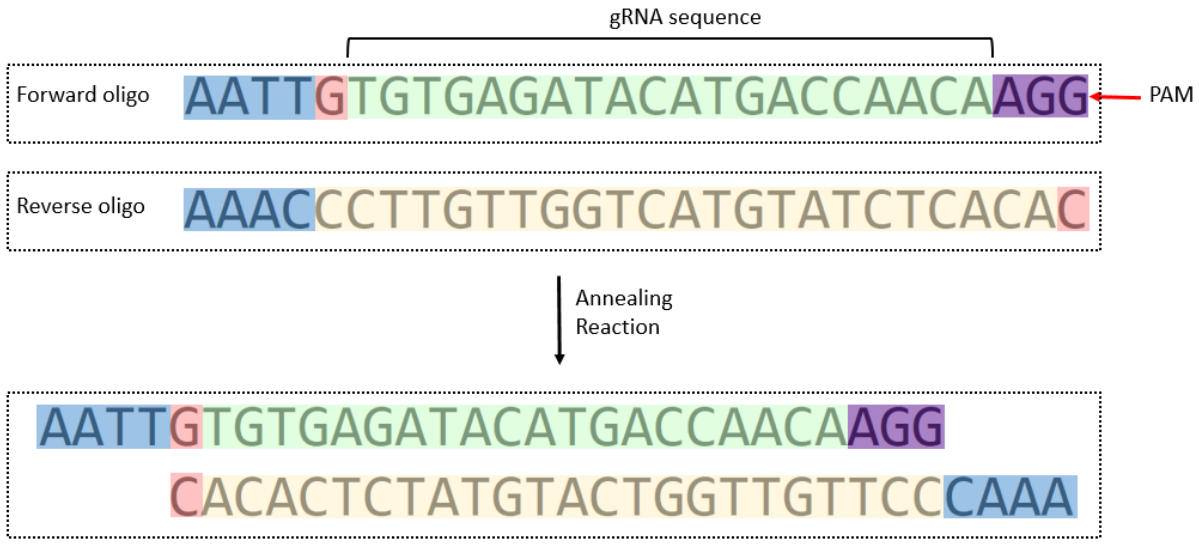


Figure 3. gRNA oligonucleotide design for construction of pDCC6 CRISPR plasmids. Overhangs (blue) are required for the forward and reverse oligos for insertion of the gRNA sequence (green) into the plasmid backbone. The protospacer adjacent motif (PAM) sequence (purple) takes the format NGG, and in some cases an additional guanine nucleotide is required (red) if gRNA sequence does not begin with a guanine (*G)* base. Forwards and reverse oligos are annealed using T4 ligase to derive a double stranded oligo which is subsequently ligated into the plasmid.

Results

The results described cover the objectives set out for this chapter. First, candidate genes are reviewed, and key candidates selected, followed by rationalisation phylogenetic analysis. Phenotypic candidate genes are presented first, followed by olfactory candidate genes. Next, the design of gRNAs for CRISPR mediated gene editing is presented for both phenotypic and olfactory candidates.

Review of phenotypic candidate genes

Initially a thorough review of the literature was performed (using PubMed database) to identify putative phenotypic targets to affect from other dipterans (Table 1). Two melanin-associated (*Yellow* and *Ebony*) pigment genes, and three eye pigmentation genes (*White*, *Scarlet* and *Cinnabar*) were identified as promising candidates as phenotypic markers to affect. Additionally, two wing development genes (*Vestigial* and *Rudimentary*) were identified. Three olfactory genes (*Orco*, *Ir8a*, and *Gr2*) involved in host-detection in mosquitoes, with putative orthologues in sand flies, were identified (see Chapter, *Gene Editing of Olfactory Genes*). Functional characteristics of each pre-screened gene of interest, derived from literature, are described in more detail below.

Table 1. **Putative Phenotypic gene targets identified for potential use in CRISPR based modification of *L. longipalpis* and *P. papatasi* sand fly vectors.** Asterisk (*) identifies genes that were investigated further and gRNAs synthesised.

Gene name	*Symbol*	*FlyBase/VectorBase gene*	*L. longipalpis orthologue*	*P. papatasi orthologue*	*Putative Function*
*Yellow**	y	FBgn0004034	LLOJ007802	PPAI006879	cuticle pigment
*Ebony**	e	FBgn0000527	LLOJ008326	PPAI005863	cuticle pigment
*White**	w	FBgn0003996	LLOJ001311	PPAI002409	Eye pigment
*Scarlet**	sc	FBgn0003515	LLOJ001495	PPAI001306	Eye pigment
*Cinnabar**	cin	AAEL008879	LLOJ009814	PPAI011008	Eye pigment
Kynurenine 3-monooxygenase/ Kynurenine hydroxylase	Kh	AAEL008879	LLOJ009814	PPAI006618	Eye pigment
Aristaless	Al	FBgn0000061	LLOJ001010	PPAI009675	antenna
Antennaless	Ant	FBgn0004473	n/a	n/a	antenna
Deformed	Dfa	FBgn0000438	n/a	n/a	antenna
Kruppel	Kr	FBgn0001325	LLOJ002951	PPAI003484	malformed segmentation
Rotund	Rn	FBgn0267337	LLOJ000842	PPAI000502	legs, wings, eyes
*Vestigial**	Vg	FBgn0003975 / FBpp0086898	LLOJ009695	PPAI008343	wing shape
*Rudimentary**	r	FBgn0003189	LLOJ009278	PPAI007465	wing shape
Wingless	Wg	FBgn0284084	LLOJ009102	PPAI002492	wing shape
Target of wingless	tow	FBgn0035719	LLOJ005852	n/a	wing shape
Giant	Gt	FBgn0001150	LLOJ005637	PPAI000977	body size
Curly su	cysu	FBgn0038511	LLOJ006480	PPAI009865	wing shape
Nubbin	nub	FBgn0085424	LLOJ001969	PPAI005907	wing shape
Cut	Ct	FBgn0004198	LLOJ004256	PPAI008545	wing shape
Miniature	m	FBgn0002577	LLOJ009922	PPAI010695	wing shape
Serrate	ser	FBgn0004197	LLOJ007958	PPAI003965	wing shape
Apterous	ap	FBgn0267978	LLOJ000480	PPAI001922	wing shape
curled	cu	FBgn0261808	LLOJ000587	PPAI003353	wing shape

Melanin-associated genes

Yellow gene *(y)*

The *yellow* gene is implicated in melanin pigmentation patterns in adult *Drosophila* cuticles, and also pigmentation of larval cuticles. Insects synthesise and secrete melanin precursors throughout their imaginal epidermis during formation of the cuticular exoskeleton. Regulation and function of *Yellow* and *Ebony* genes in *Drosophila* are important in the molecular mechanisms that generate pigmentation

patterns. *Yellow* protein is required to produce melanin (Black) and *Ebony* protein is required to supress melanin (Wittkopp, True, & Carroll, 2002). Mutants have either partial loss of pigmentation, so show a mosaic pattern, or complete loss of pigmentation (Figure 4). Interestingly, male mating ability is reduced in hemizygous *Yellow* individuals, due to reduced movement and abnormal courtship behaviour (Burnet & Wilson, 1980; Wilson, Burnet, Eastwood, & Connolly, 1976).

The *Yellow* gene has been used as a phenotypic marker in *Drosophila* and mosquitoes, particularly during CRISPR modifications. Targeting the *Yellow* gene of *Drosophila* using sgRNAs alongside Cas9 protein has resulted in 66% of injected males with mosaic of yellow patches, suggesting deletion of the gene within somatic cells (Gratz et al., 2013). Additionally, crossing individuals expressing Cas9 (promoted by *actin5C* or *vasa)*, with those expressing *Yellow*-targeting gRNAs under control of the U6:3 promoter resulted in all adults with yellow cuticles (Port, Chen, Lee, & Bullock, 2014).

In *Ae. albopictus* the 1-to-1 orthologue of the *Drosophila* version of the *Yellow* gene was targeted via injections of sgRNAs. G0 survivors had mosaicism for pigmentation, and crosses of G0 survivors lead to homozygous *Yellow* knockouts, producing adults with yellow pigmented cuticle (T. Liu et al., 2019). *Ae. aegypti* injected with Cas9 protein and sgRNAs targeting the 1:1 orthologue gene, resulted in G0 individuals with both yellow bodies and mosaic bodies. Additionally G1 mutants had obvious yellow cuticle in all life stages (Ming Li et al., 2017). *Yellow* gene knockouts have been attempted in sand flies targeting LLOJ007802, however no results were reported beyond survivors of injections (Martin-Martin, Aryan, Meneses, Adelman, & Calvo, 2018).

Ebony* gene *(e)

The *Ebony* gene and *Yellow* gene are closely related. *Ebony* encodes a protein that links beta-alanine to biogenic amines including dopamine or histamine. It controls the amount of free biogenic amine, dopamine, in cuticle formation, and of histamine in visual signal. The *Ebony* gene is expressed in cells that are both melanised and un-melanised in the cuticle; therefore, expression is not spatially regulated. Essentially a low level of expression is required for pigmentation. *Ebony* protein is required to produce tan pigment. N-beta-alanyl dopamine (NBAD) is the tan coloured polymer that is produced by the biochemical pathway (Wittkopp et al., 2002).

Melanin is the major class of pigmentation in insect cuticles. A biochemical pathway is involved in converting pigmentation precursors into pigment. Tyrosine is converted to dopa by the enzyme tyrosine hydroxylase and Dopa is converted to dopamine by dopa-decarboxylase. This can be converted to black or brown melanin by the *Yellow* protein. Dopamine also combines with Beta-alanine, which happens due to interaction with an enzyme produced by the *Ebony* gene. This produces NBAD.

Ebony's role in pigmentation is conserved in different insect species. Without *Ebony* expression melanisation increases, and with the expression of *Ebony*, melanin formation is reduced (Figure 4). In

Bombyx mori, *Ebony* knockouts have darker pigmentation in early developmental stages due to an inability to convert dopamine to NBAD. As a result, alternative pathways are utilised to generate NBAD, giving darker pigment (Futahashi et al., 2008). In several butterfly species (*Vanessa cardui*, *Junonia coenia* and *Bicyclus anynana*) CRISPR knockouts of the *Ebony* gene resulted in G0 with mosaic phenotypes, confirmed by sequencing at the site of modification. In particular, wings showed dark phenotypes, in late stage pupae. However, these were not seen in adults, potentially suggesting mortality related to these modifications (L. Zhang & Reed, 2017).

The *Ebony* gene is useful for visual identification of transgenic individuals, and has the potential to be used in gene drive systems for this purpose, particularly in insects that are light in colour such as *L.longipalpis* and *P.papatasi* sand flies. Port *et al* (2014) utilised Cas9-expressing *Drosophila* lines, and generated an *Ebony* gRNA-expressing line, which were then crossed. All offspring of *act*-Cas9 and *Ebony*-gRNA lines resulted in mosaic pigmentation, with large portions demonstrating dark pigmentation. This showed high levels of biallelic targeting of the *Ebony* gene within somatic cells. Similar results were seen in crosses with *vasa*-Cas9 lines. Similar results have been achieved in *Ae. aegypti* using a transgenic cas9-expressing lines and injecting these with sgRNAs targeting the *Ebony* gene, with 82% of G0 survivors having dark pigmentation as adults, which was also inherited (Ming Li et al., 2017).

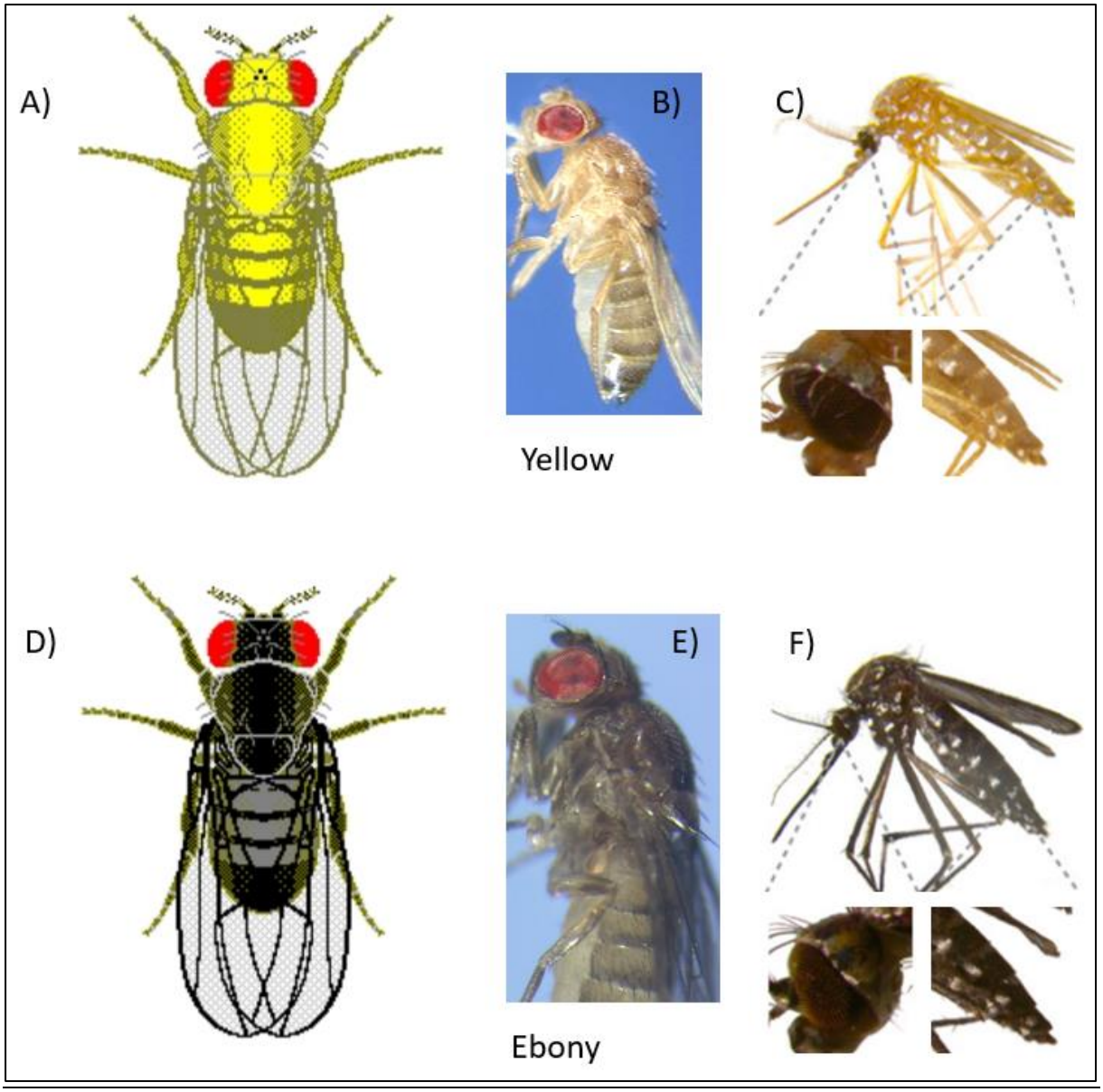


Figure 4. Yellow and Ebony gene knockout phenotypes. A) Expected phenotype from gene knockout in *D. melanogaster* (diagram), and *in vivo* (B). Yellow gene knockout in *Ae. aegypti* via CRISPR showing the whole body and close up images of the eye and abdomen (C). Expected phenotype from Ebony knockout (diagram) (D), and in vivo €. Ebony knockout in *Ae. aegypti* via CRISPR (F). *Drosophila* images from https://cgslab.com/phenotypes and https://www.biologie.uni-halle.de/entwicklungsgenetik/lehre/studenten/drosophila/mutanten. *Aedes* images are from *Li et al (2017)*.

Eye pigmentation genes

Insect eye pigments are subcellular pigment granules that are transported to the required cells as pigment precursors. All insect eye colour is derived from two classes of pigment; tryptophan ommochromes (brown pigments), and guanine pteridines (red pigments), however in some insects such as mosquitoes, only ommochromes play a role (Khan, Reichelt, & Heckel, 2017). *White*, *Cinnabar*, and *Scarlet* genes

are members of the ATP-binding cassette (ABC) transporters which move ommochrome precursors into pigment granules in the eyes where they are deposited. Modification of eye colour gene via CRISPR-Cas9 knockout may provide a useful marker for successful transgenesis (Figure 5).

White* gene *(W)

The *White* gene encodes a subunit of ABC trans-membrane transporters, which combines with *Scarlet* to transport tryptophan ommochromes, and the *Brown* gene to transport guanine pteridines into the eyes where they are deposited. The *White* gene transporter subunit must function fully for these transporters complexes to deliver pigment (Mackenzie et al., 1999). Mutations of the *White* gene can result in a range of eye colour phenotypes depending on which part of the *White* protein is affected by the mutation. Guanine transport, and therefore red pigment deposition is affected by mutations of the extracellular portion of the trans-membrane transporter, and mutations in the ATP-binding domain affect both guanine and tryptophan transporters equally, leading to light coloured eyes in *Drosophila* (Mackenzie et al., 1999).

Eye modifications have been observed in insect vector species when the *White* gene is mutated. In *Cx. quinquefasciatus* CRISPR knockout of the *White* gene (CJI005542) achieved somatic mutagenesis (mosaic white eye) using single and multiple sgRNAs targeting the gene with G0 mutation rates up to 86%, and G1 rates of more than 79% with completely white eyes (M. Li et al., 2020). A shortened version of the gene, *mini-white*, has most of the first intron removed and has been widely used as a selectable marker in transformation constructs.

Cinnabar* gene *(cin) / Kyneurenine hydroxylase* gene *(kh) / Kynurenine-3-monooxygenase* gene *(kmo)

The *Cinnabar* gene is involved in ommochrome biosynthesis, by producing the enzyme kynurenine hydroxylase which is key for the development of dark eye pigmentation. Several name are used for this gene within insects including *Kyneurenine hydroxylase* (*Kh*) and *Kynurenine-3-monooxygenase* (*Kmo*). The kynurenine hydroxylase enzyme is an intermediate product in the formation of brown pigment within the eyes of *Drosophila* and *Ae. aegypti*. When the *cinnabar* gene is knocked out eye colour becomes lighter in multiple insect orders including Diptera (Sethuraman & O'Brochta, 2005), and in hymenoptera (*Nasonia vitripennis*) (Chaverra-Rodriguez et al., 2020). Further confirmation of the genes function have come from CRISPR-based targeting of the gene in *Ae. aegypti* (AAEL008879) (Basu et al., 2015; Ming Li et al., 2017), *Ae. albopictus* (AALF006786) (T. Liu et al., 2019), and in *Cx. quinquefasciatus* (Anderson et al., 2019).

In *Ae. aegypti,* G0 mosaic phenotypes were observed in 49% of adults with sections of lighter ommatidia within the eyes, and in G1 progeny completely pale-eyed individuals were (32% mutants) observed (Basu et al., 2015). Even higher rates of mutatgenesis were observed by Li et al (2017) when targeting the same gene, giving pale-eyed larvae, pupae and adults in the G1 survivors. In *Ae.*

albopictus, G0 were observed with mosaicism within the eyes of pupae and adults, and a germline transmission rate of 78% in G1 showing completely white-eyed phenotypes (T. Liu et al., 2019).

Cinnabar has been identified as cell autonomous in *Ae. aegypti*, meaning a cell that contain a mutation in this gene displays mutant phenotypes within an organisms that has cell populations that do not contain the genetic mutation. Cell autonomous mutations are particularly useful as genetic modification markers (Sethuraman & O'Brochta, 2005).

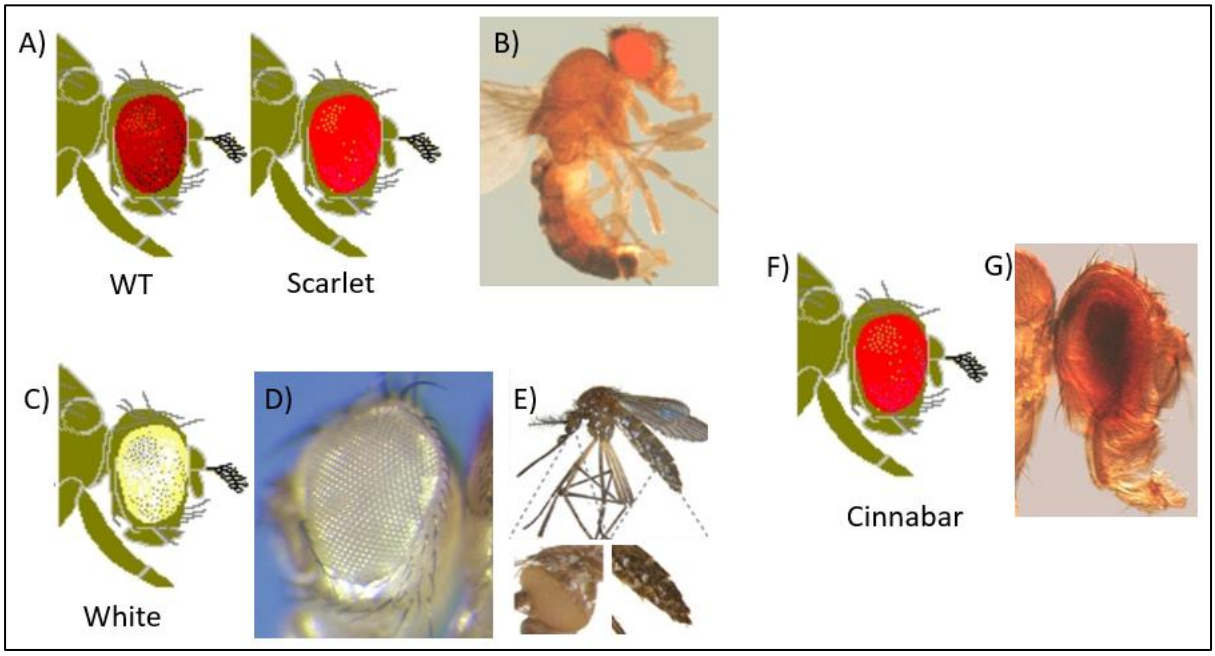


Figure 5. *Scarlet*, *White* and *Cinnabar* gene knockout phenotypes. A) Wildtype and expected phenotype from *Scarlet* gene knockout in *D. melanogaster* (diagram), and *in vivo* (B). *White* gene knockout expected phenotype (C), and *in vivo* within *D. melanogaster* (D), and *Ae. aegypti* via CRISPR showing the whole body and close up images of the eye and abdomen€(E). Expected phenotype from *Cinnabar* knockout (diagram) (F), and *in vivo* within *D. melanogaster* (G). *Drosophila* images from cgslab.com/phenotypes and www.biologie.uni-halle.de/entwicklungsgenetik/lehre/studenten/drosophila/mutanten. *Aedes aegypti* images are from *Li et al (2017)*.

Scarlet* gene *(sc)

The *Scarlet* gene has been implicated in eye pigmentation across several insect species and other non-insect species as an ommochrome transporter. Ommochromes are pigments that are generated by the processing of tryptophan, and are eye pigments within insect species (Grubbs, Haas, Beeman, & Lorenzen, 2015). In *Drosophila* pigment precursors are moved into appropriate cells for processing by ATP-binding cassette (ABC) transmembrane transporters (Dreesen, Johnson, & Henikoff, 1988; Mackenzie et al., 1999; Mount, 1987; Tearle, Belote, McKeown, Baker, & Howells, 1989).

The *Scarlet* transport protein provides a brown colour, and therefore when not functioning the result is brightly coloured red eyes in *Drosophila*. However, mutations of the *Scarlet* orthologue in other species have a different outcome. In *T. castaneum*, *B. mori*, *An. gambiae and Ae. aegypti* mutations lead to white-eyed phenotypes (Benedict, Besansky, Chang, Mukabayire, & Collins, 1996; Cornel, Q. Benedict, Salazar Rafferty, Howells, & Collins, 1997; Grubbs et al., 2015; Tatematsu et al., 2011). In Cotton bollworm (*Helicoverpa armigera*) CRISPR knockouts of the *Scarlet* gene result in larvae with no pigment, and adults with yellow eyes, compared to the wildtype green eyes (Khan et al., 2017). The function of the *Scarlet* gene is highly conserved, as RNAi and CRISPR knockout of the gene in *Daphnia magna* (Crustaceans) lead to complete loss of black eye pigmentation (Ismail, Kato, Matsuura, & Watanabe, 2018).

***Wing morphology gene: Vestigial* gene *and Rudimentary* gene**

Wing development commences during embryogenesis, with wing growth requiring the compartmentalisation of cells. The *Vestigial* gene is involved in the development of wing margin; leading to expression of proteins during embryogenesis, where the wing and halteres begin to develop (Delanoue et al., 2004). More specifically, the vestigial gene is a selector gene that helps determine the fate of cells in forming differentiated structures. It functions as a transcriptional activator in concert with the products of the *Scalloped* gene, to regulate proliferation of cells (Kim et al., 1996; Klein & Arias, 1999; Pimmett et al., 2017). In addition indirect flight muscle formation has been linked to the gene, with loss of the gene resulting in cell fates changing from indirect muscle, to direct muscle in *Drosophila* (F. Bernard et al., 2003; Frédéric Bernard et al., 2009).

Modification of the *Vestigial* gene has been attempted in a range of species, with largely the same phenotypic outcomes. In ladybird crop pests, RNAi of the gene reduced the wing size in *Henosepilachna vigintioctopunctata* and *Harmonia axyridis* rendering adults flightless (Ohde et al., 2009). RNAi knockdown of the *Vestigial* gene in *Tribolium castaneum* results in varying degrees of elytra and hindwing formation, with early knockdown having a greater reduction in size (Clark-Hachtel, Linz, & Tomoyasu, 2013). In *Ae. aegypti* knockout of the *Vestigial* gene using crosses of transgenic Cas9 and sgRNA-expressing strains resulted in phenotypes with 75% of adults survivors having pronounces defects in the wings, and halteres in G0 (Figure 6C) (Ming Li et al., 2017).

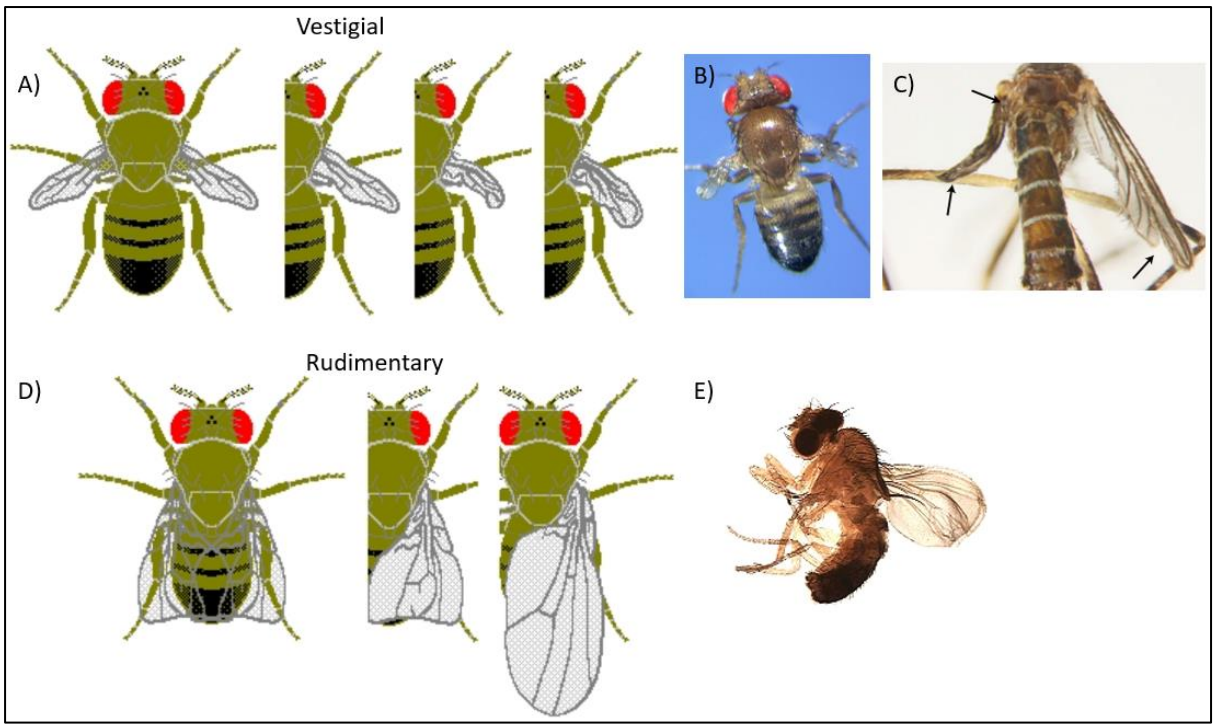


Figure 6. *Vestigial* and *Rudimentary* gene knockout phenotypes. A) Expected phenotype from *Vestigial* gene knockout in *D. melanogaster* (diagram), and *in vivo* within *D. melanogaster* (B), and *Ae. aegypti* with arrows indicating phenotype (C). *Rudimentary* gene knockout expected phenotype (D), and *in vivo* within *D. melanogast€er* (E). *Drosophila* images from https://cgslab.com/phenotypes and https://www.biologie.uni-halle.de/entwicklungsgenetik/lehre/studenten/drosophila/mutanten. *Aedes aegypti* images are from *Li et al (2017).*

The *Rudimentary* gene encode for enzymes involved in the pyrimidine biosynthesis pathway, including CPSase, ATCase and DHOase (CAD proteins). In larvae, there are high levels of activity for these enzymes, that result in adults with truncated wing phenotypes (Figure 6E) (Tsubota, Ashburner, & Schedl, 1985).

In *Drosophila Rudimentary* mutants, a range of phenotypes have been observed, from normal shaped wings, to wings with wrinkles and blisters that are shorter than the abdomen of the fly. In particular, it has been observed that homozygous females and hemizygous males have truncated wings, with the wing-cells being small compared to wildtype (Fausto-Sterling & Hsieh, 1976).

Rationalisation of Phenotypic marker genes

Identification and phylogenetic analysis of phenotypic marker genes was performed for five pigmentation and two wing development genes derived from *D. melanogaster Ae. aegypti*, *An. gambiae*, and *Cx. Quinquefasciatus*. Orthologues in *L. longipalpis* and *P. papatasi* were identified via Clustal Omega multiple protein sequence alignment using VectorBase. Phylogenetic trees were constructed in MEGA X for all seven genes (

Figure 7).

For the *D. melanogaster yellow* gene (FBgn0004034), a *L. longipalpis* orthologue (LLOJ007802) was identified consisting of four exons, located at JH689934: 30,092-37,432 on the forward strand, consisting of 512 amino acid residues. Although VectorBase identified four exons, sequencing of this region identified only three exons, due to repetition of sequence regions (Martin-Martin et al., 2018). Phylogenetic analysis (

Figure 7A) shows identity between the *yellow* gene (FBgn0004034) and *L. longipalpis* and *P. papatasi* orthologues (LLOJ007802, and PPAI006879). These genes are closely related phylogenetically, likely to have a conserved function. In addition, the orthologous genes in mosquitoes have been confirmed as pertaining to cuticular pigmentation via gene knockouts (Ming Li et al., 2017; T. Liu et al., 2019), strengthening the likelihood that LLOJ007802 has similar function within sand flies.

Ebony gene FBgn0000527 (polypeptide FBpp0083505) was used to identify sand fly orthologues, with LLOJ008326 consisting of seven exons, located at JH690011: 88,200-97,732 on the reverse strand, and consisting of 418 amino acids. For *P. papatasi*, PPAI005863 was identified, consisting of three exons, located at JH661316.1: 33,602-34,503 in the reverse strand, and consisting of 231 amino acids. LLOJ008326 and PPAI005863, exhibit high identities, and share similarity with *D. melanogaster* and *Cx. quinquefasciatus Ebony* genes (FBgn0000527 and CQUJHB011653, respectively) (

Figure 7B). They were more distantly related to *An. gambiae* and *Ae. aegypti Ebony* genes (AGAP002291 and AAEL005793-PA/PB).

One-to-one orthologues of *Kynurenine 3-monooxygenase* (*Kmo*) gene (FBgn0010482) were identified in sand flies. LLOJ009814 has 29 exons, located at JH689413:143,742-170,419 in the reverse strand, consisting of 2,206 amino acid residues. The phylogenetic tree (

Figure 7C) demonstrates that the *L. longipalpis* and *P. papatasi* orthologues (LLOJ009814 and PPAI006618) are closely related to the *Kmo* gene (FBgn0010482). There is also conserved function across more distantly related *Kmo* genes from mosquito species, suggesting that the function of the *Kmo* in sand flies is also conserved.

The *scarlet* gene, FBgn0003515, from *D. melanogaster* has an orthologue in *L. longipalpis* (LLOJ001495), located at JH689450: 49,171-55,138 in the reverse strand. This gene has three exons and 643 amino acid residues. Orthologues were also identified in *Cx. quinquefasciatus*, *An. gambiae,* and *Ae. aegypti*, that were more distantly related, however with conserved function (

Figure 7D).

For the wing development genes, *Rudimentary* (FBgn0003189) consists of three polypetides (Polypeptide IDs: r-PA: FBpp0088675, r-PD: FBpp0309141, and r-PE: FBpp0309142). An orthologue (LLOJ009278) in *L. longipalpis* was identified, with 14 exons, located at JH690161: 14,771- 41,354 on the reverse strand, and consisting of 3,707 amino acids, along with a *P. papatasi* orthologue (PPA007465). These were most closely related to each other, followed by the major mosquito vector orthologues (AAEL009475, CPIJ007153, and AGAP000300), with the *An. gambiae* orthologue identified as the most related (

Figure 7E).

The *Vestigial* wing development gene FBgn0003975, (poplypeptide sequence FBpp0086898) was used to identify sand fly orthologues. LLOJ009695 has nine exons, located at JH689411– 28,819 - 71,175 on the reverse strand, consisting of 444 amino acids. *Vestigial* or *vestigial*-like proteins were identified in several species, With *Cx. quinquefasciatus* and *An. gambiae* genes being more closely related to the sand fly genes than the *D. melanogaster* orthologue, *Aedes* orthologues were more distantly related to the sand fly genes than the *D. melanogaster* genes were (

Figure 7F).

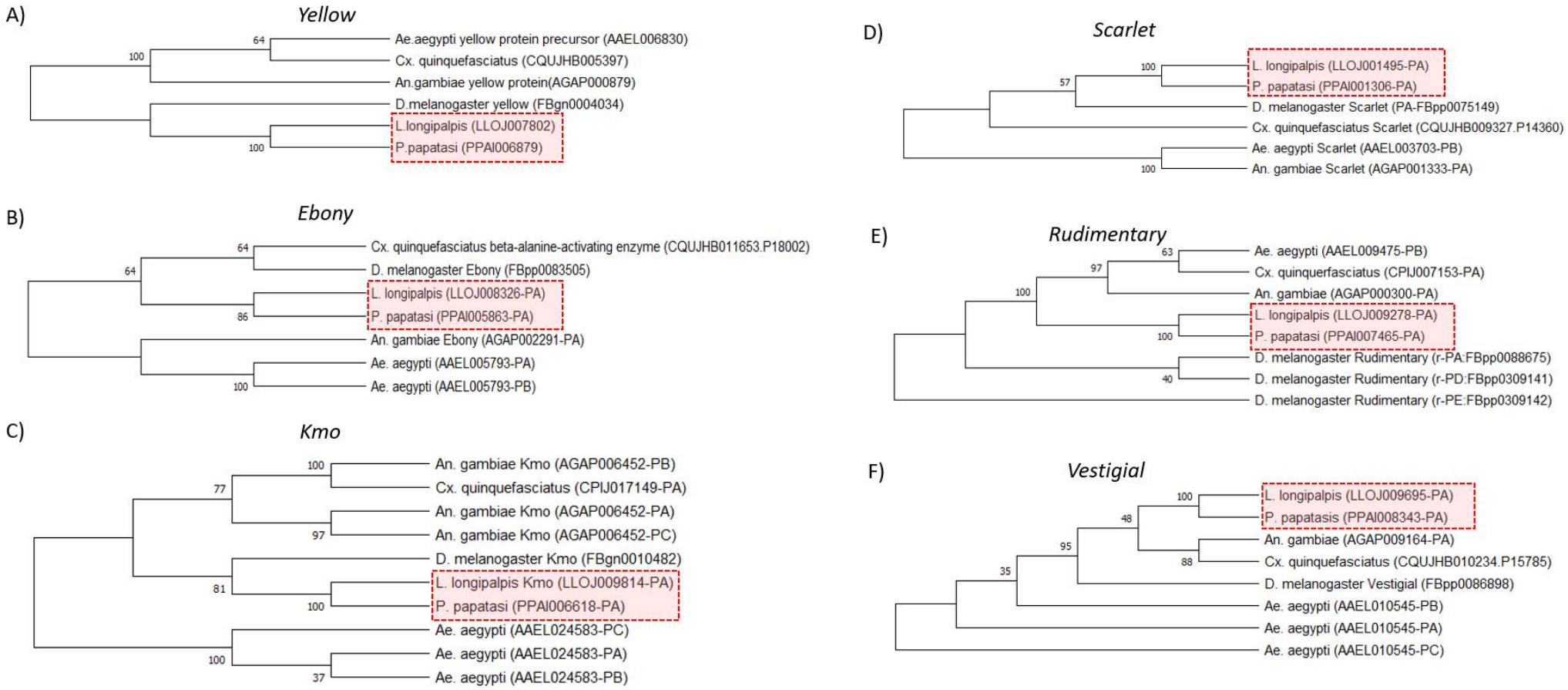


Figure 7. Phylogenetic analysis of phenotypic marker genes. A) *Yellow*, B) *Ebony*, C) *Kmo*, D) *Scarlet*, E) *Rudimentary*, F) *Vestigial*. Bootstrap consensus values are at each node. Orthologue phylogenies were estimated using *L. longipalpis*, *P. papatasi*, *A.aegypti*, *A. gambiae*, *C.quinquefasciatus*, and *D. melanogaster* protein sequences aligned with Clustal W. Resulting phylogenies were visualised in MEGA X. Maximum likilhood (ML) method and JTT matrix-based model was used for tree estimation. Branch support based on 500 bootstrap replications. Sequences were accessed from VectorBase, FlyBase and NCBI.

Identification of Olfactory genes for targeted CRISPR knockout

To identify olfactory gene targets of interest a rationalisation pipeline was followed, with an extensive literature search (PubMed database) for Olfaction in insects, with a particular focus on hematophagous vector species. Gene families were identified in addition to human kairomones (chemicals emitted, that when detected by an organism of a different species, leads to an advantage for this second organism) attractive to hematophagous insects (see Chapter 2, *Stimuli, Receptors and Genes*). Orthologous genes were identified in VectorBase which provided comparative sequence information and gene function to identify genes that were highly conserved in the sand fly, with likely similarities in olfactory function. Protein sequence alignment was performed (see Methods).

Initially a large number of ORs, gRs and IRs (see Methods, *Phylogenetic analysis for rationalisation of candidate genes*) were identified in *D. melanogaster* (50) and *An. gambiae* (138), in addition to being cross-referenced with *Aedes* and *Culex* mosquito species. The list was reduced to 25 genes by identification of those involved in the response to known kairomones (see Chapter 2) attractive to host seeking mosquitoes (Table 2). Orthologues of these genes in *L. longipalpis* were identified as described previously. The list was further narrowed to eight genes informed by VectorBase alignment to compare conservation of function between genes of different species. As a final confirmation of these chosen targets, individual research academics in the field of insect olfaction were contacted for discussion of potential targets. A final choice of targets was discussed, with a unanimous consensus agreement (Prof. Leslie Vosshall (Rockerfeller University), Ass.Prof. Conor McMeniman (Johns Hopkins), and Dr Matthew DeGennaro (Florida International University)) resulting in three final target choices, *Orco*, *Gr2*, and *IR8a*.

Table 2. Olfactory genes identified from Diptera spp. demonstrating orthologues with associated function present within *L. longipalpis* and *P. papatasi*. Asterisk indicates olfactory genes identified as primary candidates to induce behavioural responses via CRISPR-Cas9 knockout.

Gene	Organism	Function/Kairomone response	*L. longipalpis* orthologue(s)	*P. papatasi* orthologue(s)
OBP1	*An. gambiae*	Response to Indole	n/a	PPAI002561
Orco*†	*An. gambiae*	Co-receptor	LLOJ003114	PPAI006929
Or1	*An. gambiae*	4 methylphenol in human sweat	n/a	PPAI013272 PPAI013274 PPAI013279 PPAI013280 PPAI013299
Or2	*An. gambiae*	Aromatics e.g Indole	LLOJ010522 (OR3)	PPAI009955

			LLOJ010523 (OR2)	PPAI009956
			LLOJ010521 (OR5)	
			LLOJ010520 (OR6)	
Or5	*An. gambiae*	2,3 butandione a skin microflora product	n/a	n/a
Or8*	*An. gambiae*	1-octen-3-ol in human breath/sweat	LLOJ004500 (OR4)	PPAI013234
Or28	*An. gambiae*	broader response than OR8	LLOJ010880 (OR6)	PPAI011153
				PPAI013272
				PPAI013274
Or65	*An. gambiae*	2 ethylphenol in urine	n/a	PPAI013279
				PPAI013280
				PPAI013299
Gr1	*An. gambiae*	CO_2	n/a	n/a
				PPAI003611
			LLOJ003494 (Gr4)	PPAI013104
Gr2*†	*An. gambiae*	CO_2	LLOJ005747 (Gr11)	PPAI013138
			LLOJ005748 (Gr10)	PPAI013139
Gr3	*An. gambiae*	CO_2	n/a	n/a
Gr22*	*An. gambiae*	Neuron responding to CO_2	LLOJ006888 (Gr1)	n/a
Gr23*	*An. gambiae*	Neuron responding to CO_2	LLOJ010835 (Gr2)	PPAI013102
				PPAI013119
Gr24*	*An. gambiae*	Neuron responding to CO_2	LLOJ001875	PPAI000379
Ir8a*†	*An. gambiae*	Acids and amine (co-receptor)	LLOJ003646 (Ir8a)	PPAI001607
Ir25a*	*An. gambiae*	Acids and amine (co-receptor)	LLOJ001989 (Ir25a)	PPAI005280
				PPAI010236
Ir41a	*An. gambiae*	Ammonia and lactic acid via grooved peg sensilla	LLOJ003016	n/a
Ir41a.2	*An. gambiae*	Ammonia and lactic acid via grooved peg sensilla	n/a	n/a
Ir41c	An. gambiae	Tuning receptor highly expressed in antennae	n/a	n/a
Ir64a	An. gambiae	Response to acids	LLOJ002191	n/a
Ir75d.2	An. gambiae	Ammonia and lactic acid via grooved peg sensilla	LLOJ001956 (Ir75d)	PPAI002891
Ir75k	An. gambiae	Carboxcylic acid in sweat	n/a	n/a
Ir76b*	An. gambiae	Butylamine in vertebrate odour (acids and amine co-receptor)	LLOJ005467 (Ir76b)	PPAI007495

				PPAI007496
Ir87a.3	Ae. albopictus	Ammonia and lactic acid via grooved peg sensilla	LLOJ009667 (Ir87a)	PPAI007671 PPAI007672
Ir92a	D. melanogaster	Ac1 –sensilla - KO might limit response to ammonia	LLOJ010964 (Ir92a-1)	PPAI013184

** indicates shortlisted gene targets based on putative conservation of function across species.*

† indicates gene targets agreed by mosquito olfaction experts: Prof. Leslie Vosshall (Rockerfeller University), Ass.Prof. Conor McMeniman (Johns Hopkins), Dr Matthew DeGennaro (FIU).

Review of olfactory candidate genes

Orco gene

As previously described (see Chapter 2), odorant receptors (ORs) are transmembrane proteins located on olfactory receptor neurons (ORNs) that allow odorants to cross the lymph of the olfactory sensilla, into the neurons. They are involved in general response to attractive odorants, and consist of a ligand-selective OR, and a co-receptor. The co-receptor (*Orco*) is fundamental to the functioning of the OR-*Orco* complexes for odour responses. *Orco* genes have been shown to be important in host seeking behaviour in several disease vectors through knockout experiments. *Ae. aeygpti* lose attraction to skin odours without CO_2 present (DeGennaro et al., 2013), *Ae. albopictus* have lower feeding rate and less response to human odours (H. Liu et al., 2016), *An. coluzzi* have reduced response to human odours (Sun, Liu, Ye, Baker, & Zwiebel, 2020), and *An. sinensis* had reduced preference for humans (Y. Wang et al., 2022).

Gr2 gene

Gustatory receptors (GRs) are ligand-gated ion channels are involved in the detection of sugars, bitter compounds and pheromones, and importantly for the detection of CO_2. *Gr2* is one of a triumvirate of Grs acutely linked to CO_2 reception, which is a key compound for host-seeking behaviour in vector species. Recent knockout experiments of *Gr23* (*Gr2* orthologue) in *An. coluzzii* reduced responses to CO_2 (F. Liu, Ye, Baker, Sun, & Zwiebel, 2020).

IR8a gene

Ionotropic receptors (IRs) are expressed in olfactory sensory neurons, forming ligand-gated ion channels with co-receptors, *Ir8a*, *Ir25a* and *Ir76b* (see Chapter 2). These IR-co-receptor channels detect a range of volatile organic compounds such as amines and acids which are attractive odours for host seeking. Ir8 channels in particular are tuned to detect volatile acids such as lactic acid which is a large component of human odour when compared to other animals (Raji et al., 2019). Ir8a knockouts studies *Ae. aegypti* do not detect lactic acid, and have reduced attraction to humans (Raji et al., 2019).

Rationalisation of candidate Olfactory genes

The *Orco* gene (FBgn0037324) from *D. melanogaster* was used as a reference to identify orthologues via Clustal Omega protein sequence alignments in VectorBase. The gene identified in *L. longipalpis* (LLOJ003114), consisting of six exons, located at on scaffold 193: 192,178-196,148 in the forward strand, and consisting of 474 amino acid residues. A *P. papatasi* orthologue was also identified (PPAI006929) consisting of 205 amino acids and two exons located on the forward strand at AJVK01059062.1: 6,716-7,573.

Phylogenetic analysis revealed sand fly *Orco* genes have close identity (Figure 8A). *Orco* genes from the mosquito species are conserved and also share high identity to *D. melanogaster Orco*. These genes (*Aedes*, *Anopheles*, *Culex*, and *Drosophila* orthologues) diverged from the sand fly orthologues prior to the occurrence of the *L. longipalpis* / *P. papatasi* split, yet are more closely related to the sand fly *Orco* genes than to the *B. mori Orco* outgroup.

Similarly, the sand fly *Ir8a* genes are conserved, and more distantly related to *D. melanogaster* and mosquito *Ir8a* genes. All Dipteran *Ir8a* genes are more phylogenetically related than the *Ir8a* in Lepidopteran species *B. mori* (Figure 8B).

The *Gr2* gene in *L. longipalpis* is highly conserved with respect to a *P. papatasi* orthologue, and also related to mosquito GRs (*Gr23* in *Anopheles* and more distantly *Gr2* in *Aedes* and *Culex*). The sand fly and mosquito *Gr2* orthologues are more divergent with respect to the *Drosophila Gr21a* orthologue, and *B. mori Gr22* orthologue, likely due to the importance of CO_2 detection in vectors compared to non-vector species (Figure 8C).

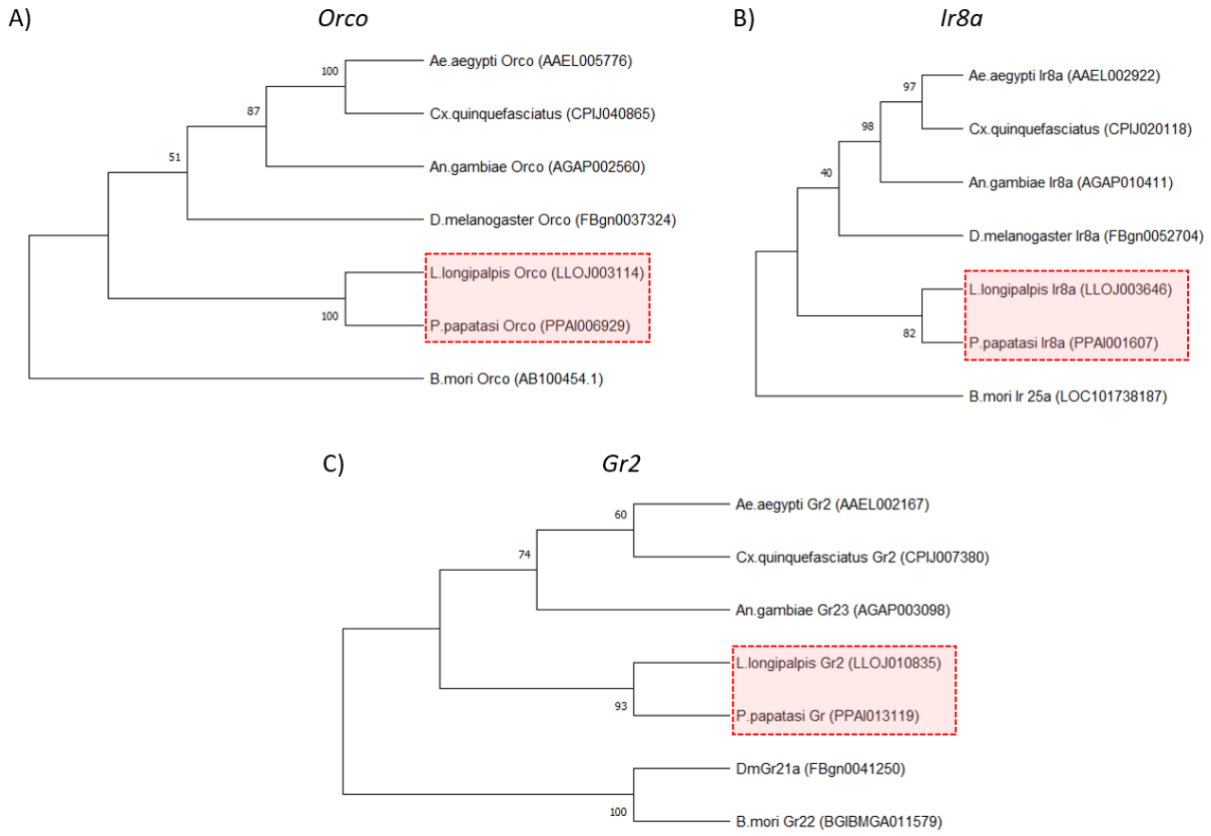


Figure 8. Phylogenetic analysis of sand fly olfactory genes. Orco (A), Ir8a (B) and Gr2 (C) orthologue phylogenies were estimated using *L. longipalpis*, *P. papatasi*, *Ae. aegypti*, *An. gambiae*, *Cx. quinquefasciatus*, *D. melanogaster* and *B. mori* protein sequences aligned with Muscle algorithm. Resulting phylogenies were visualised with MEGA X. Maximum likilhood (ML) and JTT matrix-based model for tree estimation. Tree is rooted at the node leading to *Orco*, *Ir8a*, and *Gr2* (A, B, and C respectively). Branch support based on 500 bootstrap replications. Sequences were accessed from VectorBase, FlyBase and NCBI.

Design of gRNAs to target phenotypic and olfactory genes

gRNAs were designed for insertion into pDCC6 plasmid backbones following the methods described above (see also Chapter 4, *Generation of Plasmid constructs via Gibson Assembly*). For olfactory genes two gRNAs were designed for each of the three genes (*Orco, Gr2,* and *IR8a*). All six gRNAs had high predicted efficiency scores (65.89 – 72.48), with zero self-complementarity within the genome (Table 3). *Gr2* gRNAs target the beginning of exon 3, with a distance of 426bp between the sequences. *Orco* gRNAs target exon 2 and exon 4, with a distance of 1,455bp between the sequences, and *Ir8a* gRNAs target exon 1 and 2, with 855bp between the sequences.

For wing development gene targets three plasmid gRNAs were designed for each of the *Vestigial* and *Rudimentary* genes. The six gRNA sequences had efficiency scores predicted between 46.05 – 73.70, with some self-complementarity predicted for two *Rudimentary* gRNAs and one *Vestigial* gRNA. The *Rudimentary* gRNAs target the start of exon 1, exon 4, and the end of exon 11, and the *Vestigial* gRNAs target exons 3, 6 and 7.

gRNAs for *in vitro* transcription were designed as described in Chapter 4 (see Methods, *In vitro transcription of gRNAs*) and methods above. For *Gr2* the gRNAs target the 3' end of exon 3, with 184bp between the two gRNA sequences. *Ir8a* gRNAs target the 5' end of exon 1 with 225bp between the sequences, and *Orco* gRNAs target exon 1 and 2 with 554bp between the sequences. *Rudimentary* gRNAs target the 3' end of exon 6 with 247bp between the sequences, and the *Vestigial* gRNA target the exons 3 and 7 with 17,645bp between the sequences (Table 4).

Table 3. Phenotypic gRNA sequences for downstream CRISPR construct (pDCC6) mediated gene editing. Gene IDs were input into ChopChop and gRNAs sequences identified, including self-complementarity and efficiency scores determined via the ChopChop algorithm.

Gene	*Rank*	*Target sequence*	*Genomic location*	*Strand*	*GC content (%)*	*Self-complementarity*	*Efficiency score*
GR2	2	AAACATAAAGAGGCAATACGTGG	Scaffold895:59205	+	35	0	72.48
GR2	4	GCTGTGTACAAGACAATGTGGGG	Scaffold895:59608	-	45	0	71.67
Orco	3	TACAGCAATCAAGTATTGGGTGG	Scaffold193:194147	+	40	0	65.89
Orco	1	TGTGAGATACATGACCAACAAGG	n/a	+	40	0	76.01
Ir8a	2	AGCAATAGCGAGACCTGCGAGGG	Scaffold216:65688	-	55	0	71.16
Ir8a	3	TTGAGTGTAAAATCCCGACAGGG	Scaffold216:64850	-	40	0	71.13
Rudimentary	27	AGCATTGGAAGACACAGGGT	18084	n/a	50	0	66.21
Rudimentary	4	TTGTAGCCCATCTTCACGA	36555	n/a	40	2	73.70
Rudimentary	1	GGAACTATGGCATTCCGTG	40520	n/a	60	2	69.15
Vestigial	2	TCATCATTACGGTTCCTACGCGG	39731	-	45	0	70.74
Vestigial	16	GGAAAATTTCTCGCCGACAT	41612	n/a	45	0	60.21
Vestigial	38	TCGCGGACACGTATTGTGCT	57320	n/a	55	2	46.05
Ebony	2	TCGCATTCAGCACATCCTTG	Scaffold369:34007	-	50	0	67.76
Ebony	3	AAAGTGCATGGTAATCAGGA	Scaffold369:34144	-	40	0	64.69
Ebony	4	CACATTTCCATATGGCCAGG	Scaffold369:34070	+	50	0	64.08
Ebony	5	GTGGCTATCAAGTTGTCCGA	Scaffold369:34093	+	50	0	63.20

Caspar	1	GGATTCTGAGAGTTCC ATGG	Scaffold31 43:13731	+	50	0	74.80
Caspar	2	TATCCTCAAGAATCTC AATG	Scaffold31 43:13123	+	35	0	73.73
Caspar	5	GATGAAGCAACGCTGT AGGG	Scaffold31 43:12725	-	55	0	69.26
Caspar	6	TGAGGCTATCTATTTG CTAG	Scaffold31 43:12532	+	40	0	68.11
KMO	3	CCCCCATGAATATCGG CCCTTCC	n/a	-	50	1	54.85
KMO	n/a	CCTTTGTGAAGATAGA CGATGAC	n/a	n/a	n/a	n/a	n/a
Yellow	5	CCCCACTGCATCGACC ATTGCT	n/a	n/a	n/a	n/a	n/a
Yellow	7	CCCACATCATCTCTCA GCCTGAA	n/a	n/a	n/a	n/a	n/a

Table 4. Olfactory and Phenotypic gene targeting gRNA sequences for *in vitro* transcription. Gene IDs were input into ChopChop and gRNAs sequences identified, including self-complementarity and efficiency scores determined via the ChopChop algorithm.

Gene sgRNA	*Rank*	*Target sequence*	*Genomic location*	*Strand*	*GC content (%)*	*Self-complementarity*	*Efficiency score*
Gr2	33	GTTAACAGCCACTGTGGATC	Scaffold89 5:60008	-	50	3	47.01
Gr2	8	TTACCCGATCACAACCCGGA	Scaffold89 5:59844	-	55	0	62.83
Ir8a	75	TGAGATTGTACGTCCGAAAA	Scaffold21 6:64492	+	40	0	39.55
Ir8a	59	GTGAGACAGACATCCCTTTT	Scaffold21 6:64691	-	45	0	45.89
Orco	36	GTGCAACCGCAAAAATATCA AGG	Scaffold19 3:192184	+	40	0	47.80
Orco	1	TGTGAGATACATGACCAACA AGG	Scaffold19 3:192715	+	40	0	76.01
LLRud gRNA#1	32	GGATTCAACAAGTGGCCATG GGG	Scaffold84 1:24324	-	50	0	65.65
LLRud gRNA#2	172	GGGGAGGTTGCGTATGTTGA TGG	Scaffold84 1:24548	-	55	0	49.04
LLVest gRNA#1	38	TCGCGGACACGTATTGTGCT	Scaffold91: 57320	+	55	2	46.05
LLVest gRNA#2	22	CCATGCGGCTCACGCTTACC ACC	Scaffold91: 39695	+	65	1	56.52
LLYellow gRNA#3	49	CCGTAGTCAACTTTACAATG AGG	Scaffold61 4:30342	+	40	0	60.29

LLYellow gRNA#4	38	TCAGTGCCACGATGGAAGGACGG	Scaffold614	n/a	55	0	62.27
Kmo#1	-	CCTCAGACAAAGTATCCGGAAAC	Scaffold93	-	48	1	46.54
Kmo#2	7	TACCACGGAGCTCATTGGAGGG	Scaffold93:170162	-	45	0	72.45
Scarlet	35	GCGCATCATCAATAACGTGA	Scaffold130:54837	-	45	0	62.27
Cinnabar	104	GATTTGATCATCGGCTGCGA	Scaffold93:168174	-	50	1	60.59
Martin-Martin gRNA5	n/a	CTGATGTCTGCGGTAGGCTT	Scaffold614:35574	+	57	0	49.00
Martin-Martin gRNA6	n/a	CGCATCAAGGCTGATGTCTG	Scaffold614:35564	+	61	1	46.00

Discussion

In this Chapter, we initially identify, screen and rationalise a panel of non-lethal gene targets (targeting phenotypic characteristics and olfaction) and generate associated gRNA sequences for CRISPR based gene editing. Targets inducing a phenotypic effect include cuticle pigmentation genes (*y* and *e*), eye pigmentation genes (*sc, cin,* and *W*), and wing development genes (*Vg* and *r*). Putative *L. longipalpis* and *P. papatasi* gene orthologue were identified within the available genomes of these species using a bioinformatics approach via genome databases, and identification of orthologues (VectorBase) with phylogenetic similarity to help inform those likely to have conserved function in sand flies.

The phylogenetic analysis of these phenotypic marker genes demonstrated a high degree of genetic identity between genes in *D. melanogaster* and major mosquito vectors, and sand flies. Sand fly orthologues of *Yellow*, *Ebony*, *Scarlet* and *Cinnabar* were more closely related to the *D. melanogaster* genes, and *Rudimentary* and *Vestigial* more closely related to *An. gambiae* genes. The phylogenetic relatedness of these genes suggests conserved function across species, and knockout studies targeting these genes in diptera, and more distantly related coleoptera (*H. vigintioctopunctata H. axyridis* and *T. castaneum*) (Grubbs et al., 2015; Ohde et al., 2009) and lepidoptera (*B. mori*, *V. cardui*, *J. coenia* and *B. anynana*) (Futahashi et al., 2008; L. Zhang et al., 2017) confirm conserved function. This contributes some level of confidence that genes are suitable knockout targets for CRISPR guided by appropriate sand fly specific sgRNAs.

Phenotypic gene targets were taken forward to for *in vitro* assessment as described in Chapter 4. A multiplex gRNA knockout strategy was pursued whereby more than one gRNA was delivered following

gRNA synthesis. The synchronous introduction of multiple gRNAs targeting the same gene increases potential for mutagenesis and assists screening (M. Li et al., 2020).

We have described a pipeline for identifying phenotypic targets to modify in sand flies; by initial literature reviews, identification of genes of interest, consensus from academic field leaders, phylogenetic analysis of orthologues, and bioinformatics webtools against model organism databases. The process was an important initial step prior to synthesising gRNAs to further derive an *in vitro* platform (Chapter 4). The pipeline was successfully applied to olfactory gene targets that may have also impact sand fly host-seeking behaviour.

Here initial literature reviews and subsequent rationalisation identified eight key genes (*Gr1, Gr2, Gr3*, *Orco*, *Ir8a*, *Ir25a* and *Ir76b*) involved in host seeking behaviour within mosquito vectors, with many knockout studies demonstrating behavioural change particularly in genes associated with human odours (CO_2, L-(+) lactic acid, hexanoic acid, nonanal, 1-octen-3-ol, Ammonia), or human hosts as a result (see Chapter 2). These studies along with extensive discussion with academic researchers provides the rationale for determining the presence of these genes within sand flies.

Sand fly olfactory genetics in relation to identifying hosts for blood meals is understudied, however recent annotation of *L. longipalpis* and *P. papatasi* genomes has helped identify chemoreceptors related to important olfactory host-seeking genes (Hickner et al., 2020). Orthologues of these genes were identified in *this* study, and phylogenetic analysis confirmed the relatedness to the *Ae. aegypti*, *An. gambiae*, and *Cx. quinquefasciatus* orthologues. The same degree of relatedness was observed with the three genes analysed (*Gr2*, *Orco*, and *IR8a*), with the sand fly orthologues being more closely related to *D. melanogaster* and the mosquito vectors, than to the *B. mori* outgroup.

As with the phenotypic genes, previously identified, the olfactory genes selected show a high degree of relatedness to other vector species suggesting the potential for conserved function relating to host-seeking behaviour, observed in gene knockout studies. In addition, gene knockouts in more distantly related species demonstrate similar response or lack of response to key components of human odour (Q. Liu et al., 2017).

GRs such as *Gr1/Gr22* and *Gr3/Gr24* are highly conserved in mosquitoes, and amino acid sequences are highly conserved in *D. melanogaster* with up to 70% identity. This degree of conservation indicates that these genes and associated receptor structures, are vital for CO_2 binding and detection (Kent, Walden, & Robertson, 2008). Interestingly, only two (*Gr1* and *Gr2*) of the three CO_2 receptors are found in sand flies (Hickner et al., 2020). This reduced number of known CO_2 receptors provides notable targets for affecting host-seeking behaviour, as these remaining genes may be compensating for the loss of the *Gr3* CO_2 receptor. In particular, *Gr2* (or its orthologues *AaGr2* and *AgGr23*), may lead to reduced responses to CO_2 when knocked out, making this a good target in sand flies.

Orco provides another primary target for attempting to interrupt host-seeking behaviour in sand flies due to its role as a fundamental co-receptor to ORs in several mosquito vectors. This gene has been identified as key in the detection of human skin odours via knockout studies in *Ae. aeygpti* (DeGennaro et al., 2013), *Ae. albopictus* (H. Liu et al., 2016), *An. coluzzi* (Sun et al., 2020), and *An. sinensis* (Y. Wang et al., 2022). Importantly several sand fly species have demonstrated attraction to a key OR-Orco complex ligand, 1-octen-3-ol, including *L. longipalpis* (Andrade, Andrade, Dias, Pinto, & Eiras, 2008) and *P. papatasi* (Beavers, Hanafi, & Dykstra, 2004). Disrupting the *Orco* gene in sand flies may therefore result in similar reduction in odour responses.

Finally, *Ir8a* has been identified as a non-redundant co-receptor to IRs involved in vector detection of human volatile acids such as lactic acid (Raji et al., 2019). Phylogenetic analysis of sand fly orthologues in this study demonstrate conservation of protein sequences, suggesting conserved function. Additionally attraction to lactic acid amongst other acids has been demonstrated in *L. longipalpis* (Andrade et al., 2008), marking this co-receptor as an essential target for attempting to interfere with host-seeking in sand flies.

Overall, phenotypic marker orthologues were rationalised within *L. longipalpis* to target pigmentation of the eyes and cuticle, and to interfere with development of the wings. Genomes assessed through webtools and alignment software (VectorBase and MEGA X) were used to design gRNAs to be taken forward to *in vitro* assessment (see Chapter 4) including incorporation within CRISPR plasmids . This pipeline was utilised to rationalise olfactory gene targets that are known to influence host-seeking in mosquitoes. Together the results form an important contribution to begin sand fly specific CRISPR mediated knockouts pursued in following chapters.

References

Anderson, M. E., Mavica, J., Shackleford, L., Flis, I., Fochler, S., Basu, S., & Alphey, L. (2019). CRISPR/Cas9 gene editing in the West Nile Virus vector, Culex quinquefasciatus Say. *PLoS ONE*, *14*(11), 1–10. https://doi.org/10.1371/journal.pone.0224857

Andrade, A. J., Andrade, M. R., Dias, E. S., Pinto, M. C., & Eiras, Á. E. (2008). Are light traps baited with kairomones effective in the capture of Lutzomyia longipalpis and Lutzomyia intermedia? An evaluation of synthetic human odor as an attractant for phlebotomine sand flies (Diptera: Psychodidae: Phlebotominae). *Memorias Do Instituto Oswaldo Cruz*, *103*(4), 337–340. https://doi.org/10.1590/s0074-02762008000400004

Bae, S., Park, J., & Kim, J. S. (2014). Cas-OFFinder: A fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics*, *30*(10), 1473–1475. https://doi.org/10.1093/bioinformatics/btu048

Basu, S., Aryan, A., Overcash, J. M., Samuel, G. H., Anderson, M. A. E., Dahlem, T. J., … 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 of the United States of America*, *112*(13), 4038–4043. https://doi.org/10.1073/pnas.1502370112

Beavers, G. M., Hanafi, H. A., & Dykstra, E. A. (2004). Evaluation of 1-octen-3-ol and carbon dioxide as attractants for Phlebotomus papatasi (Diptera: Psychodidae) in southern Egypt. *Journal of the American Mosquito Control Association*, *20*(2), 130–133.

Benedict, M. Q., Besansky, N. J., Chang, H., Mukabayire, O., & Collins, F. H. (1996). Mutations in the Anopheles gambiae pink-eye and white genes define distinct, tightly linked eye-color loci. *Journal of Heredity*, *87*(1), 48–53. https://doi.org/10.1093/oxfordjournals.jhered.a022952

Bernard, F., Lalouette, A., Gullaud, M., Jeantet, A. Y., Cossard, R., Zider, A., … Silber, J. (2003). Control of apterous by vestigial drives indirect flight muscle development in Drosophila. *Developmental Biology*, *260*(2), 391–403. https://doi.org/10.1016/S0012-1606(03)00255-0

Bernard, Frédéric, Kasherov, P., Grenetier, S., Dutriaux, A., Zider, A., Silber, J., & Lalouette, A. (2009). Integration of differentiation signals during indirect flight muscle formation by a novel enhancer of Drosophila vestigial gene. *Developmental Biology*, *332*(2), 258–272. https://doi.org/10.1016/j.ydbio.2009.05.573

Burnet, B., & Wilson, R. (1980). Pattern mosaicism for behaviour controlled by the yellow locus in Drosophila melanogaster. *Genetical Research*, *36*(3), 235–247. https://doi.org/10.1017/S0016672300019868

Chaverra-Rodriguez, D., Benetta, E. D., Heu, C. C., Rasgon, J. L., Ferree, P. M., & Akbari, O. S. (2020). Germline mutagenesis of Nasonia vitripennis through ovarian delivery of CRISPR-Cas9 ribonucleoprotein. *Insect Molecular Biology*, 569–577.

Clark-Hachtel, C. M., Linz, D. M., & Tomoyasu, Y. (2013). Insights into insect wing origin provided by functional analysis of vestigial in the red flour beetle, Tribolium castaneum. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(42), 16951–16956. https://doi.org/10.1073/pnas.1304332110

Cornel, A. J., Q. Benedict, M., Salazar Rafferty, C., Howells, A. J., & Collins, F. H. (1997). Transient expression of the Drosophila melanogaster cinnabar gene rescues eye color in the white eye (WE) strain of Aedes aegypti. *Insect Biochemistry and Molecular Biology*, *27*(12), 993–997.

https://doi.org/10.1016/S0965-1748(97)00084-2

DeGennaro, M., McBride, C. S., Seeholzer, L., Vosshall, L. B., Jasinskiene, N., James, A. A., … Dennis, E. J. (2013). orco mutant mosquitoes lose strong preference for humans and are not repelled by volatile DEET. *Nature*, *498*(7455), 487–491. https://doi.org/10.1038/nature12206

Delanoue, R., Legent, K., Godefroy, N., Flagiello, D., Dutriaux, A., Vaudin, P., … Silber, J. (2004, January). The Drosophila wing differentiation factor vestigial-scalloped is required for cell proliferation and cell survival at the dorso-ventral boundary of the wing imaginal disc. *Cell Death and Differentiation*, Vol. 11, pp. 110–122. https://doi.org/10.1038/sj.cdd.4401321

Doench, J. G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E. W., Donovan, K. F., … Root, D. E. (2016). Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nature Biotechnology*, *34*(2), 184–191. https://doi.org/10.1038/nbt.3437

Dreesen, T. D., Johnson, D. H., & Henikoff, S. (1988). The Brown Protein of Drosophila melanogaster Is Similar to the White Protein and to Components of Active Transport Complexes. *Molecular and Cellular Biology*, *8*(12), 5206–5215. https://doi.org/10.1016/b978-0-12-429350-2.50019-x

Fausto-Sterling, A., & Hsieh, L. (1976). Studies on the female-sterile mutant rudimentary of Drosophila melanogaster: An analysis of the rudimentary wing phenotype. *Developmental Biology*, *51*(2), 269–281. https://doi.org/10.1016/0012-1606(76)90143-3

Felsenstein, J. (1985). Confidence limits on phylogenies: An approach using the bootstrap. *Evolution*, *39*(2), 783-791.

Futahashi, R., Sato, J., Meng, Y., Okamoto, S., Daimon, T., Yamamoto, K., … Fujiwara, H. (2008). yellow and ebony are the responsible genes for the larval color mutants of the silkworm Bombyx mori. *Genetics*, *180*(4), 1995–2005. https://doi.org/10.1534/genetics.108.096388

Gratz, S. J., Cummings, A. M., Nguyen, J. N., Hamm, D. C., Donohue, L. K., Harrison, M. M., … O'connor-Giles, K. M. (2013). Genome engineering of Drosophila with the CRISPR RNA-guided Cas9 nuclease. *Genetics*, *194*(4), 1029–1035. https://doi.org/10.1534/genetics.113.152710

Grubbs, N., Haas, S., Beeman, R. W., & Lorenzen, M. D. (2015). The ABCs of eye color in Tribolium castaneum: Orthologs of the Drosophila white, scarlet, and brown genes. *Genetics*, *199*(3), 749–759. https://doi.org/10.1534/genetics.114.173971

Hickner, P. V., Timoshevskaya, N., Nowling, R. J., Labbé, F., Nguyen, A. D., McDowell, M. A., … Syed, Z. (2020). Molecular signatures of sexual communication in the phlebotomine sand flies. *PLoS Neglected Tropical Diseases*, *14*(12), 1–18. https://doi.org/10.1371/journal.pntd.0008967

Ismail, N. I. B., Kato, Y., Matsuura, T., & Watanabe, H. (2018). Generation of white-eyed Daphnia magna mutants lacking scarlet function. *PLoS ONE*, *13*(11), 1–11. https://doi.org/10.1371/journal.pone.0205609

Jones, D. T., Taylor, W. R., & Thornton, J. M. (1992). The rapid generation of mutation data matrices from protein sequences. *Bioinformatics*, *8*(3), 275–282. https://doi.org/10.1093/bioinformatics/8.3.275

Kent, L. B., Walden, K. K. O., & Robertson, H. M. (2008). The Gr family of candidate gustatory and olfactory receptors in the yellow-fever mosquito Aedes aegypti. *Chemical Senses*, *33*(1), 79–93. https://doi.org/10.1093/chemse/bjm067

Khan, S. A., Reichelt, M., & Heckel, D. G. (2017). Functional analysis of the ABCs of eye color in Helicoverpa armigera with CRISPR/Cas9-induced mutations. *Scientific Reports*, *7*(January), 1–14. https://doi.org/10.1038/srep40025

Kim, J., Sebring, A., Esch, J. J., Kraus, M. E., Vorwerk, K., Magee, J., & Carroll, S. B. (1996). Integration of positional signals and regulation of wing formation and identity by Drosophila vestigial gene. *Nature*, *382*, 133–138. Retrieved from https://www.nature.com/articles/382133a0

Klein, T., & Arias, A. M. (1999). The Vestigial gene product provides a molecular context for the interpretation of signals during the development of the wing in Drosophila. *Development*, *126*(5), 913–925. https://doi.org/10.1242/dev.126.5.913

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

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution*, *35*(6), 1547–1549. https://doi.org/10.1093/molbev/msy096

Labun, K., Montague, T. G., Krause, M., Torres Cleuren, Y. N., Tjeldnes, H., & Valen, E. (2019). CHOPCHOP v3: Expanding the CRISPR web toolbox beyond genome editing. *Nucleic Acids Research*, *47*(W1), W171–W174. https://doi.org/10.1093/nar/gkz365

Li, M., Li, T., Liu, N., Raban, R. R., Wang, X., & Akbari, O. S. (2020). Methods for the generation of heritable germline mutations in the disease vector Culex quinquefasciatus using clustered regularly interspaced short palindrome repeats-associated protein 9. *Insect Molecular Biology*, *29*(2), 214–220. https://doi.org/10.1111/imb.12626

Li, Ming, Bui, M., Yang, T., Bowman, C. S., White, B. J., & Akbari, O. S. (2017). Germline Cas9 expression yields highly efficient genome engineering in a major worldwide disease vector, Aedes aegypti. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(49), E10540–E10549. https://doi.org/10.1073/pnas.1711538114

Liu, F., Ye, Z., Baker, A., Sun, H., & Zwiebel, L. J. (2020). Gene editing reveals obligate and modulatory components of the CO2 receptor complex in the malaria vector mosquito, Anopheles coluzzii. *Insect Biochemistry and Molecular Biology*, *127*(August), 1–9. https://doi.org/10.1016/j.ibmb.2020.103470

Liu, H., Liu, T., Xie, L., Wang, X., Deng, Y., Chen, C. H., … Chen, X. G. (2016). Functional analysis of Orco and odorant receptors in odor recognition in Aedes albopictus. *Parasites and Vectors*, *9*(1), 1–10. https://doi.org/10.1186/s13071-016-1644-9

Liu, Q., Liu, W., Zeng, B., Wang, G., Hao, D., & Huang, Y. (2017). Deletion of the Bombyx mori odorant receptor co-receptor (BmOrco) impairs olfactory sensitivity in silkworms. *Insect Biochemistry and Molecular Biology*. https://doi.org/10.1016/j.ibmb.2017.05.007

Liu, T., Yang, W. Q., Xie, Y. G., Liu, P. W., Xie, L. H., Lin, F., … Chen, X. G. (2019). Construction of an efficient genomic editing system with CRISPR/Cas9 in the vector mosquito Aedes albopictus. *Insect Science*, *26*(6), 1045–1054. https://doi.org/10.1111/1744-7917.12645

Mackenzie, S. M., Brooker, M. R., Gill, T. R., Cox, G. B., Howells, A. J., & Ewart, G. D. (1999). Mutations in the white gene of Drosophila melanogaster affecting ABC transporters that determine eye colouration. *Biochimica et Biophysica Acta - Biomembranes*, *1419*(2), 173–185. https://doi.org/10.1016/S0005-2736(99)00064-4

Mali, P., Aach, J., Stranges, P. B., Esvelt, K. M., Moosburner, M., Kosuri, S., … Church, G. M. (2013). CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nature Biotechnology*, *31*(9), 833–838. https://doi.org/10.1038/nbt.2675

Martin-Martin, I., Aryan, A., Meneses, C., Adelman, Z. N., & Calvo, E. (2018). Optimization of sand fly embryo microinjection for gene editing by CRISPR/Cas9. *PLoS Neglected Tropical Diseases*,

12(9), 1–18. https://doi.org/10.1371/journal.pntd.0006769

Mount, S. M. (1987). Sequence similarity. *Nature*, *325*, 487.

Ohde, T., Masumoto, M., Morita-Miwa, M., Matsuura, H., Yoshioka, H., Yaginuma, T., & Niimi, T. (2009). Vestigial and scalloped in the ladybird beetle: A conserved function in wing development and a novel function in pupal ecdysis. *Insect Molecular Biology*, *18*(5), 571–581. https://doi.org/10.1111/j.1365-2583.2009.00898.x

Pimmett, V. L., Deng, H., Haskins, J. A., Mercier, R. J., LaPointe, P., & Simmonds, A. J. (2017). The activity of the Drosophila Vestigial protein is modified by Scalloped-dependent phosphorylation. *Developmental Biology*, *425*(1), 58–69. https://doi.org/10.1016/j.ydbio.2017.03.013

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

Raji, J. I., Melo, N., Castillo, J. S., Gonzalez, S., Saldana, V., Stensmyr, M. C., & DeGennaro, M. (2019). Aedes aegypti Mosquitoes Detect Acidic Volatiles Found in Human Odor Using the IR8a Pathway. *Current Biology*, *29*(8), 1253-1262.e7. https://doi.org/10.1016/j.cub.2019.02.045

Ran, F. A., Hsu, P. D., Lin, C.-Y., Gootenberg, J. S., Konermann, S., Trevino, A. E., … Zhang, F. (2013). Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell*, *154*(6), 1380–1389. https://doi.org/10.1016/j.cell.2013.08.021

Sethuraman, N., & O'Brochta, D. A. (2005). The Drosophila melanogaster cinnabar gene is a cell autonomous genetic marker in Aedes aegypti (Diptera: Culicidae). *Journal of Medical Entomology*, *42*(4), 716–718. https://doi.org/10.1093/jmedent/42.4.716

Sun, H., Liu, F., Ye, Z., Baker, A., & Zwiebel, L. J. (2020). Mutagenesis of the orco odorant receptor co-receptor impairs olfactory function in the malaria vector Anopheles coluzzii. *Insect Biochemistry and Molecular Biology*, *127*(October). https://doi.org/10.1016/j.ibmb.2020.103497

Tatematsu, K. ichiro, Yamamoto, K., Uchino, K., Narukawa, J., Iizuka, T., Banno, Y., … Daimon, T. (2011). Positional cloning of silkworm white egg 2 (w-2) locus shows functional conservation and diversification of ABC transporters for pigmentation in insects. *Genes to Cells*, *16*(4), 331–342. https://doi.org/10.1111/j.1365-2443.2011.01490.x

Tearle, R. G., Belote, J. M., McKeown, M., Baker, B. S., & Howells, A. J. (1989). Cloning and characterization of the scarlet gene of Drosophila melanogaster. *Genetics*, *122*(3), 595–606. https://doi.org/10.1093/genetics/122.3.595

Thyme, S. B., Akhmetova, L., Montague, T. G., Valen, E., & Schier, A. F. (2016). Internal guide RNA interactions interfere with Cas9-mediated cleavage. *Nature Communications*, *7*, 1–7. https://doi.org/10.1038/ncomms11750

Tsubota, S., Ashburner, M., & Schedl, P. (1985). P-element-induced control mutations at the r gene of Drosophila melanogaster. *Molecular and Cellular Biology*, *5*(10), 2567–2574. https://doi.org/10.1128/mcb.5.10.2567-2574.1985

Wang, H., Yang, H., Shivalila, C. S., Dawlaty, M. M., Cheng, A. W., Zhang, F., & Jaenisch, R. (2013). One-step generation of mice carrying mutations in multiple genes by CRISPR/cas-mediated genome engineering. *Cell*, *153*(4), 910–918. https://doi.org/10.1016/j.cell.2013.04.025

Wang, Y., He, X., Qiao, L., Yu, Z., Chen, B., & He, Z. (2022). CRISPR/Cas9 mediates efficient site-specific mutagenesis of the odorant receptor co-receptor (Orco) in the malaria vector Anopheles sinensis. *Pest Management Science*, (April). https://doi.org/10.1002/ps.6954

Wilson, R., Burnet, B., Eastwood, L., & Connolly, K. (1976). Behavioural pleiotropy of the yellow gene in Drosophila melanogaster. *Genetical Research*, *28*(1), 75–88. https://doi.org/10.1017/S0016672300016748

Wittkopp, P. J., True, J. R., & Carroll, S. B. (2002). Reciprocal functions of the Drosophila Yellow and Ebony proteins in the development and evolution of pigment patterns. *Development*, *129*(8), 1849–1858. https://doi.org/10.1242/dev.129.8.1849

Wu, X., Kriz, A. J., & Sharp, P. A. (2014). Target specificity of the CRISPR-Cas9 system. *Quantitative Biology*, *2*(2), 59–70. https://doi.org/10.1007/s40484-014-0030-x

Zhang, L., Martin, A., Perry, M. W., van der Burg, K. R. L., Matsuoka, Y., Monteiro, A., & Reed, R. D. (2017). Genetic basis of melanin pigmentation in butterfly wings. *Genetics*, *205*(4), 1537–1550. https://doi.org/10.1534/genetics.116.196451

Zhang, L., & Reed, R. D. (2017). A praactical guide to genome editing using CRISPR/Cas9 in Lepidoptera. *Springer Press*, 155–172. https://doi.org/10.1007/978-981-10-4956-9

Zhang, X., Koolhaas, W. H., & Schnorrer, F. (2014). A versatile two-step CRISPR- and RMCE-based strategy for efficient genome engineering in Drosophila. *G3: Genes, Genomes, Genetics*, *4*(12), 2409–2418. https://doi.org/10.1534/g3.114.013979

Chapter 4

Utilisation and optimisation of an *in vitro* platform for assessment of PiggyBac and CRISPR components

Introduction

Application and validation of gene editing approaches, including CRISPR and PiggyBac based methods, applied *in vivo* to insect vectors is still an inefficient, time consuming and logistically difficult process. Therefore, validation of CRISPR components prior to *in vivo* application can greatly facilitate successful downstream outcomes. In particular, *in vitro* systems based on cell lines can enable prior screening and optimisation of system components (Cas9, gRNAs, exogenous sequences, and associated expression promotors), gene targets to be affected, assessment of knockout and knockin plasmid-based constructs, along with delivery systems and optimising gene editing detection assays. In this Chapter, we describe an *in vitro* pipeline applied to sand fly cell lines to validate gene editing components and plasmid constructs to take forward to *in vivo* transfection.

Insect cell lines are invaluable tools for the study of virology, insect immunity, gene expression, host parasite interactions, and for the production of recombinant proteins. Recently cell lines have been utilised for CRISPR and other gene editing methodologies applied to insect vectors of disease. Development of molecular tools has been conducted in different insect cell lines (Sf9, High Five, S2R+ and BmN), and a range of mosquito cell lines have been used in CRISPR experiments, across the three major genera (*Anopheles*, *Aedes*, and *Culex*).

With respect to vector-borne diseases, mosquito cell lines have been studied extensively to investigate anti-viral immunity. Here, knockout studies of the *Dcr2* gene in *Aedes aegypti* cell line Aag2 demonstrated increased viral replication of Semliki Forest virus (amongst other viruses), but not Zika virus, implicating this gene in virus repression within the vector. However, knockout studies of an alternative gene, *Ago2*, show no reduction in viral replication suggesting it plays no direct role in viral suppression within *Ae. aegypti* (Scherer et al., 2021; Varjak, Kean, Vazeille, Failloux, & Kohl, 2017). These cell studies have helped elucidate immunological pathways that may lead to potential innovations in reduction of disease transmission.

More recently, use of cell lines have been used to investigate sex determination within insect vectors, with acute relevance to vector control (see Chapter 1, *Sex ratio distortion for population suppression*). *Aedes albopictus* cell line C6/36 with CRISPR based knockouts of the *Nix* gene demonstrated interference with male sex determination genes (*dsx* and *fru*), skewing towards female isoform variants (Liu et al., 2020a). When such knockouts were applied *in vivo*, the result was feminisation of reproductive organs in adult males. Furthermore, cell lines have facilitated optimisation of CRISPR reagents, including promoter viability and assessing plasmid constructs. For example, Anderson *et al.,* (2020) used *Ae*. *aegypti* cell lines (Aag2 and U4.4), and *Cx*. *quinquefacsiatus* cell lines (Hsu) to assess Pol III promoters (U6) for the expression of sgRNAs. It was shown that in some cell lines it was important to have promoters that were specific to the original species (Aag2 and Hsu), but non-specific

promoters (from alternative mosquito species) functioned well in others (U4.4 cells). In particular, none of the non-mosquito U6 promoters (from alternative insect species) used for sgRNA expression functioned above background levels of expression (Anderson et al., 2020). Further work by Rozen-Gagnon *et al.,* (2021) generated and characterised a CRISPR-Cas9 plasmid-based system for editing U4.4 and Aag2 cell lines, using modified plasmids with species-specific promoters. They demonstrated efficient editing and reduced expression in multiple loci induced by targeted knockouts, and also demonstrated homology directed repair (HDR) knockin of an RFP fluorescence marker. Ovarian Hsu cell lines were used to validate CRISPR plasmid reagents (Feng, Kambic, et al., 2021), and Viswanatha *et al.,* (2021) recently developed a Cas9 expressing Sua-5B (*Anopheles coluzzi*) cell line by optimising recombination-mediate cassette exchange (RMCE), to deliver CRISPR sgRNAs into the cells for screening.

Cell lines have been generated in many insect species (Bairoch, 2018) , including two sand fly cell lines. Tesh and Modi (1983) produced a *L. longipalpis* cell line, LL-5, which consisted of two cell types, (epithelial and fibroblastoid cells with elongated protrusions). These cells were initially used to determine susceptibility to arboviruses (both sand fly and mosquito viruses multiplied within the cells (Tesh & Modi, 1983)). A second *L. longiplapis* cell line, Lulo, was developed from embryonic tissue (epithelioid cells) and viral susceptibility was assessed with phleboviruses able to replicate within this line (Rey, Ferro, & Bello, 2000). Subsequently Lulo has been used to study insect-parasite interactions, focused on promastigote adhesion, of *Leishmania* species (Bello et al., 2005; L. M. M. Côrtes et al., 2012; L. M. Côrtes et al., 2011), and elucidation of immunological pathways activated by *Leishmania* (Tinoco-Nunes et al., 2016).

Although LL-5 and Lulo sand fly cell lines still exist they are not easily available to the international research community despite exhaustive efforts However, two new L. longipalpis cell lines have been generated, LLE/LULS40 and LLE/LULS45 (Bell-Sakyi et al., 2021; Bell-Sakyi, Darby, Baylis, & Makepeace, 2018), and are available to the research community (Tick Cell Biobank, University of Liverpool, UK).We have used these cell lines towards the development of the *in vitro* platform for CRISPR .

In vitro transcription of gRNAs

Only limited reports exist of gene editing being performed in sand flies at the time of writing (see, Chapter 5). For example, a single publication has described successful application of CRISPR applied to *P. papatasi* (Louradour, Ghosh, Inbar, & Sacks, 2019), and PiggyBac, a widely used gene editing approach has yet to be successfully applied to sand flies.

To achieve an effective gene editing strategy (including PiggyBac and CRISPR based approaches) a number of parameters need to be in place. Briefly these are, established cultured cell lines (described

above), a rationalised panel of gene targets to affect (see Chapter 2), for example those associated with non-lethal phenotypic effects (as proof-of-principle), olfaction or those implicated in establishing infection in insect vectors. Following this, generation of associated gRNAs to target genes of interest (CRISPR), described in Chapter 3. In plasmid based approaches suitable expression promotors must be validated to minimally drive gRNA and Cas9 and incorporated into backbone plasmids for knockout studies. For more refined HDR approaches, for insertion and expression of exogenous DNA, companion plasmids containing sequence templates are also required. Finally, there must be in place a method for effective cell transfection and detection of gene edits. The rationale here is that outputs will contribute working gRNAs, promotors and validated constructs (PiggyBac and CRISPR) that can be applied to downstream *in vivo* experiments (Chapter 5). Each are discussed in more detail below.

Inducing double stranded breaks (DSBs) in genomic DNA using CRISPR-Cas9 requires different components to be delivered to the target organisms or cells. One method is direct injection of the CRISPR reagents (Cas9 protein, sgRNAs, and donor DNA). This method is relatively simple as Cas9 protein is commercially available for direct injection, sgRNAs can be ordered pre-transcribed or in a form requiring in-house kit-based transcription (gRNA to sgRNA) prior to injection. Overall, the direct injection pathway initially requires a selection of rationalised gene targets, design and sequence verification of gRNAs, and production of usable sgRNAs through *in vitro* protocols, incorporating a validatory cleavage assay. Direct injection of CRISPR components will provide a rapid demonstration that components are functioning and precise targeting and knockouts can be achieved in sand flies.

An alternative approach (primarily pursued in this work) is to deliver the CRISPR components described in plasmids for CRISPR knockouts, an approach which can be adapted to knockins with an additional plasmid containing a HDR template for insertion and expression of exogenous sequences. The latter approach has been important in development of gene drives and insertion of anti-parasite peptides approaches.

The use of a single sgRNA to disrupt a gene is commonly used to induce indels in DNA. However, indels are typically small (<100bp) and can be difficult to screen by gel electrophoresis, but can be easily and effectively detected by heteroduplex cleavage assay. To facilitate screening by PCR electrophoresis, multiple sgRNAs can be delivered simultaneously targeting different loci (Ran et al., 2013) inducing larger knockouts (~200-2000bp) detectable directly via gel electrophoresis without the need for heteroduplex assays, or sequencing based approaches.

In vitro assessment of expression promoters

Identification of effective promoters that drive expression in sand flies is an important step towards developing CRISPR tools for modification of genes of interest, and to future development of gene drive systems. An initial step in transformation of insects often involves inserting markers, such as

fluorescence genes, and delivering these markers using transformation systems including PiggyBac and CRISPR plasmids.

For effective gene editing mediated by CRISPR-Cas9 endonuclease needs to be expressed within the insect, as does the gRNA sequence. Many studies have been successfully demonstrated expression of Cas9, driven by a variety of promoters (*hsp70*, *nanos*, *vasa*, and *zpg*) showing expression in both somatic and germline specific tissues. Germline specific promoters are key for gene drives and inheritance.

Hsp70Bb has been used to promote the Cas9 gene from *S. pyogenes* in *D. melanogaster* with somatic cell and germline cell expression (Gokcezade, Sienski, & Duchek, 2014). An endogenous *vasa* promoter was identified in *An. gambiae*, with expression localization to male and female gonads (Papathanos, Windbichler, Menichelli, Burt, & Crisanti, 2009). Utilising this promoter allowed mutagenesis to occur in early embryos. Port *et al*., (2014) compared *nanos* and *vasa* promoters to restrict Cas9 expression to the germline in *Drosophila,* with *nanos*-driven gene expression occurring in somatic and germline cells, and *vasa*-driven expression being restricted to the germline. Restriction of transgene expression to germline cells has also been observed in *Ae. aegypti* using a *vasa* promoter (Akbari, Papathanos, Sandler, Kennedy, & Hay, 2014), and in *An. stephensi* using a *nanos* promoter (Biedler, Qi, Pledger, James, & Tu, 2015). More recently, the *zero population growth* (*zpg*) promoter was used to restrict expression of Cas9 to the germline in *An. gambiae*, more effectively than *vasa* promoters (Kyrou et al., 2018; Simoni et al., 2020).

For gRNAs (and short RNA sequences), the RNA Polymerase III promoter *U6* has been used for expression. gRNAs must be transcribed by a Pol III promoter to avoid modifications made to mRNA (Ma et al., 2014). The *U6* promoter sequence was initially characterised and used successfully in *Drosophila* S2 cells for the expression of short-hairpin RNA (Wakiyama, Matsumoto, & Yokoyama, 2005). Three *U6* promoters exist in *Drosophila* (*U6:1*, *U6:2*, and *U6:3*), demonstrating differing degrees of expression. Port *et al.*, (2014) demonstrated these differences by generating plasmids with one of each of the promoters and an identical gRNA. *U6:3* was shown to promote the strongest gRNA activity, and *U6:2* the weakest, indicating the importance of gRNA promoter selection.

U6 promoters have been identified in both *Ae. aegypti* and *An. gambiae,* and utilised to knockout luciferase expression in genetically modified cell lines. New Pol III promoters (*7SK*) have recently been identified by Anderson *et al.,* (2020). A comparison of promoter activity in Culicine cell lines (Aag2, Hsu, and U4.4) demonstrated strongest promoter activity when *U6* was phylogenetically similar to the cell line of origin, for *Ae. aegypti*, and *Cx. quinquefasciatus*, but not *Ae. albopictus*. The *7SK* promoters had greater activity than *U6* promoters, and *U6* promoters from *Drosophila* and *Plutella* do not function in mosquito cells.

Currently sand fly-specific (homologous) promoters have not been identified due to the limited genetic modification research conducted in sand flies, as a result, promoters that function in other species are important to study, and determine whether they could function within sand flies. These heterologous promoters successfully function within Dipterans such as *D. melanogaster* and mosquito species (see above), making them prime candidates for translation in to sand flies.

Saraiva *et al*., (2009) used two sand fly cell lines (LL-5 and PP-9) to express a luciferase reporter gene under the control of several heterologous promoters from other species. Specifically, the Heat shock protein 70 (*hsp70*) from *D. melanogaster*, human cytomegalovirus (*CMV*) and simian virus 40 (*SV40*). For the expression of the transient luciferase by these promoters, the transcriptional machinery of the sand fly cells must recognise the promoter sequences. *Hsp70* functioned within the sand fly cell lines, *SV40* was more active in the LL-5 cell lines than the *CMV* after one day of transfection, however expression from both plasmids was similar after 3 days. With the PP-9 cell line, the expression was greater with the *CMV* promoter. This provides evidence of effective gene expression in a sand fly cell line driven by multiple non- homologous promoters of relevance, with implications for the development and optimisation of genetic modification components *in vitro* and *in vivo*.

Cell transfection

Delivery of plasmid constructs to cells can be conducted via multiple approaches including physical methods such as electroporation or chemical transfection. Electroporation increases the permeability of cell membranes by running an electrical current through the cell culture, and is often used for cell lines that are hard to transfect (Du et al., 2018; Jordan, Collins, Terefe, Ugozzoli, & Rubio, 2008). However, the electroporation process can cause apoptosis, destroying the cells (Piñero, López-Baena, Ortiz, & Cortés, 1997). Chemical transfection can be achieved by the use of commercially available liposomal-based transfection reagents. Positively charged lipid molecules, containing the plasmids to be transfected, merge with the phospholipid bilayer of host cells, allowing the plasmid to be delivered without major damage to the cells (Kim & Eberwine, 2010). Chemo-transfection reagents are commonly used for insect cell transfections (González et al., 2011; Vidigal et al., 2018; Viswanatha et al., 2021; Zhang et al., 2014), and liposomal transfection reagents have been used to successfully transfect sand fly cell lines (Saraiva et al., 2009), suggesting this approach is applicable to novel sand fly cell lines.

Mutation detection

Once CRISPR plasmid transfections have been affected, analysis is required to determine whether modifications have occurred. Methods are available for evaluation including mutation detection assays, Sanger sequencing (Brinkman, Chen, Amendola, & Van Steensel, 2014) followed by algorithmic

deconvolution of sequence data (described in Chapter 5), and next generation sequencing (NGS) (Gagnon et al., 2014).

NGS is a high-throughput DNA sequencing technology that allows for analysis of large amounts of genetic material. It can be used to validate CRISPR-mediated genomic modifications by providing a detailed analysis of the targeted regions of the genome. This method has been widely used to detect CRISPR modifications in insect vectors (Hall et al., 2015; Kyrou et al., 2018) and cell lines (Viswanatha et al., 2021), however preparation for sequencing can be time-consuming and costly.

An alternative to NGS methods is Heteroduplex assays, also known as T7 endonuclease assays. The assay is based on the ability of T7 endonuclease to selectively cleave DNA strands that contain mismatches, which has been optimised for the detection of CRISPR modification. The assay typically includes the following steps: PCR amplification of the targeted region of interest using the wildtype and edited DNA as templates, annealing of the wildtype and edited PCR products - forming heteroduplexes, incubation of the heteroduplexes with T7 endonuclease, which will cleave at the mismatch points, and analysis of the cleaved products by gel electrophoresis. This method has been used for detection of editing in *Ae albopictus* C6/36 cells (Liu et al., 2020a) and *Cx. quinquefasciatus* Hsu cells (Torres, Prince, Robison, & Rückert, 2022), and in multiple insect species (Kotwica-Rolinska et al., 2019; Koutroumpa et al., 2016; Zhang et al., 2014). This assay is relatively simple and inexpensive compared to NGS and can be optimised for use in sand fly cell lines for validation of CRISPR components.

Aims

Building a gene editing platform through the development of protocols to assess components for genetic modification of sand flies, is an important step towards novel vector control techniques and interruption of transmission of leishmania. The initial use of cell lines can provide a rapid approach for efficiently screening gene editing components for CRISPR and PiggyBac based approaches, to identify those to take forward to *in vivo* studies. Here we describe the development of *in vitro* protocols for a sand fly genome editing platform (using PiggyBac and CRISPR based approaches) in newly acquired sand fly cell lines. More specifically, we identify and rationalise gene targets to be affected, describe protocols for *in vitro* screening of system components (Cas9, gRNAs, exogenous sequences, transfection components), validation of associated expression promotors, assessment of knockout and knockin plasmid-based constructs along with a comparison of cell line transfection systems applied to sand fly cell lines and gene editing detection methods. Together these form the foundation for researchers interested in gene editing applied to sand flies and disease control.

Methods

Insect cell cultures

Culture media

Different cell culture media and conditions were required to maintain the insect cell lines (Table 1). The sand fly cell lines (LLE45 and LLE40) require three types of media; Minimal L-15, complete L15, and H-Lac (Table 2). These three media were prepared individually and subsequently combined to make complete media. For insect cell lines used a comparative control (An4a3B, DS2, and SF21, see Table 1), each require one type of commercially available media (Schneider's Insect Medium and Insect Xpress, Table 1) supplemented with FCS, and antibiotics (Penicillin/Streptomycin). All media were syringe filtered (0.22μm), before combining for the complete sand fly cell growth media. Aliquots were incubated at 28°C overnight to test for contamination.

Table 1. Insect cell culture media and growth conditions.

Insect species	*Cell line*	*Culture media*	*Conditions*
Lutzomyia longipalpis	LLE45 (Jacobina strain)	L15B	T25 flasks with the cap sealed cap. Incubated at 28°C.
Lutzomyia longipalpis	LLE40 (Campo Grande strain)	Minimal L-15, complete L15, H-Lac.	T25 flasks with sealed cap. Incubated at 28°C.
Anopheles gambiae	An 4a3B	89% Schneider's Insect Medium, 10% FCS (heat inactivated), 1% Penicillin (10,000units)/ streptomycin (10,000μg).	T25 flasks with sealed cap. On initial seeding, media changed after 12-16 hours, topped up every 48 hours if necessary. Incubated at 28°C.
Drosophila melanogaster	DS2	89% Schneider's Insect Medium, 10% FCS (heat inactivated), 1% Penicillin (10,000units)/ streptomycin (10,000μg).	Flasks with vented caps. Maintained in a humidified box, incubated at 28°C
Spodoptera frugiperda	Sf21	97% Insect-Express protein free growth media, 2% FCS	Flasks with vented caps. Maintained in a

		(heat inactivated), 1% Penicillin (10,000units)/ streptomycin (10,000µg).	humidified box, incubated at 28ºC.

Table 2. Culture media components for LLE40 and LLE45 cell lines.

Culture media	*Components*
Minimal L-15	L-15 (Leibovitz) medium, Tryptose phosphate broth, L-glutamine 200mM, Penicillin (10,000units)/ streptomycin (10,000µg), FCS (heat-inactivated).
Complete L-15B	Minimal L-15 media, Tryptose phosphate broth, Bovine lipoprotein (ICN) 10% solution in L-15B, L-glutamine 200mM, Penicillin (10,000units)/ streptomycin (10,000µg), FCS (heat-inactivated). pH of the mixture is adjusted to 6.8 using 1N Sodium Hydroxide (0.8g of pellets in 20mL miliQ water).
H-Lac	Hanks balanced salt solution (HBSS), Lactalbumin hydrolysate 10% solution, L-glutamine 200mM, Penicillin (10,000units)/ streptomycin (10,000µg), FCS (heat-inactivated).

To passage cells, T25 flasks were positioned upright for 10 minutes. 2.2mL of media was removed from the flask, and 4.4mL of fresh media was added. Media was homogenised rapidly via pipetting, making sure to rinse the surface of the flask to remove adherent cells. A 2.2mL aliquot was then removed to seed a new flask, and incubated at 28ºC. Flasks were monitored for cell proliferation and contamination, and were re-passaged after 1-2 weeks.

Sand fly cell lines

LLE/LULS40 and LLE/LULS45 *Lutzomyia longipalpis* cell lines were acquired from the Tick Cell Biobank (TCB), referred to as LLE40 and LLE45 from here on. TCB derived cell lines are described by Rey *et al.,* (2000) and Tesh and Modi*,* (1983). Briefly, embryos from *L. longipalpis* were disinfected, homogenised, and retained in culture flasks with media (from Table 1 above). Cell passages were conducted 45 days after initiation, with multiple passages resulting in greater homology of cell morphology (Bell-Sakyi et al., 2021, 2018)

LLE40 and LLE45 cell lines were used in this thesis as the basis of the *in vitro* platform for the development and validation of CRISPR and PiggyBac components for genetic modification in *L. longipalpis*. The cell lines differ in morphology (Figure 1) with LLE40 cells growing in clusters, extending upwards from the surface of the flask, with the majority of the cells displaying epithelioid features. The cell population is heterogeneous, with the presence of fibroblastoid cells with cytoplasmic

protrusions growing on the flask surface. LLE45 cells are fibroblastoid cells covering the flask surface, forming fewer cell clusters.

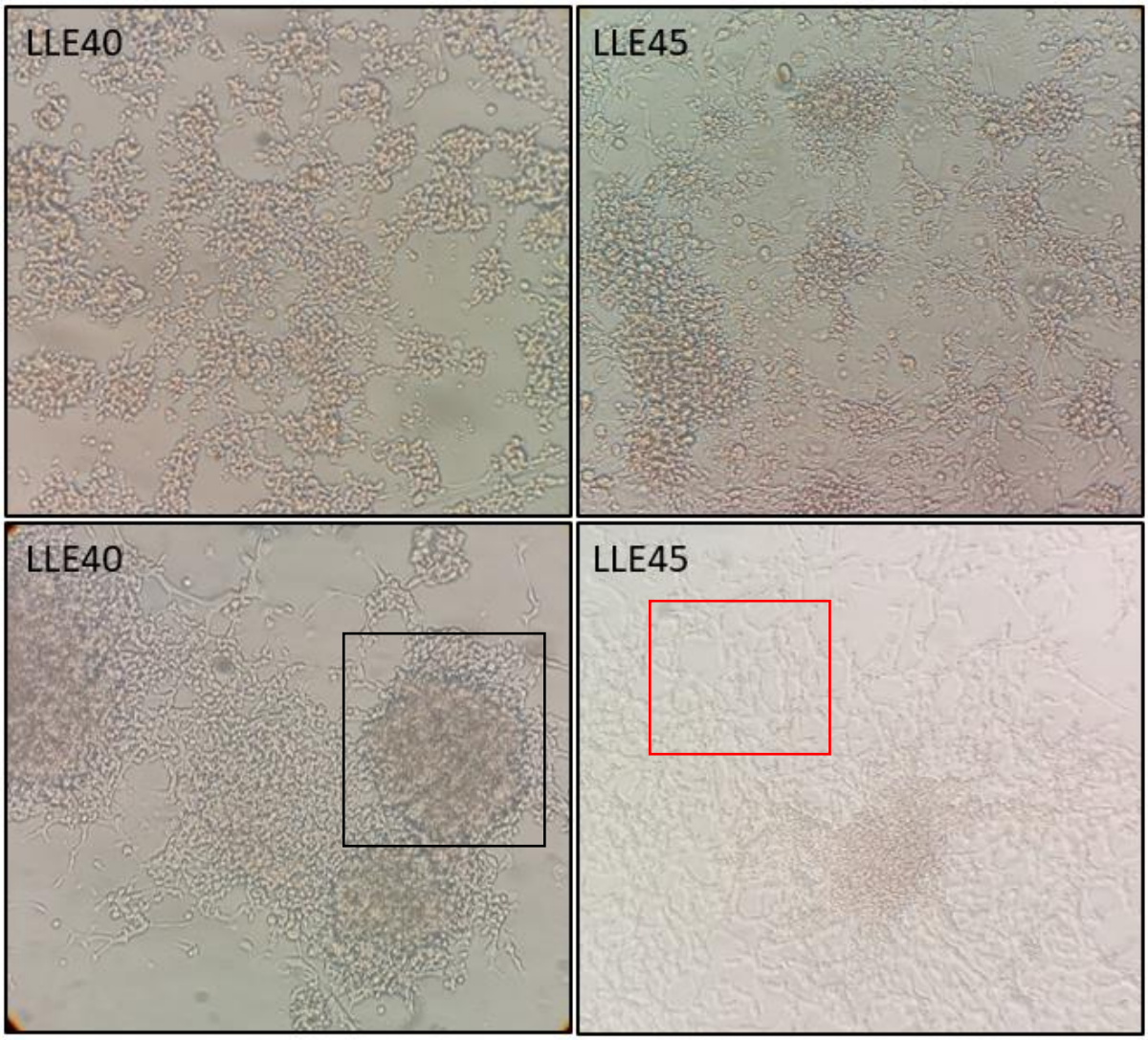


Figure 1. Novel *L. longipalpis* cell lines (LLE40 and LLE45) from the Tick Cell Biobank. Images show the formation of epithelioid cell clusters in LLE40 (black box), with few fibroblastoid cells in between clusters. The majority of LLE45 cells are fibroblastoid (red box), and grow along the culture flask surface. Images taken using a Motic B3 series light microscope (Motic, China) at 60x magnification.

Cryopreservation of cell lines

For cryopreservation, 900µL of cell culture was added to 100µL of DMSO (10%) in a cryogenic vial (cryovial), and homogenised. Cryovials were transferred to a Mr Frosty freezing container at room temperature, before being stored at -70°C for 24-48 hours. Vials were then transferred to Liquid Nitrogen.

To revive cells, cryovials containing frozen insect cells were removed from liquid nitrogen storage and immediately placed it into a 37°C water bath. The media and cells were thawed (< 1 minute) by gently swirling the vial in the water bath until thawed. The vials were transferred to a sterile laminar flow hood. Before opening, the outside of the vials were sterilised with 70% ethanol. 1 mL of pre-warmed complete growth media was added dropwise to the cells, and mixed before transfer to a 15mL centrifuge tube. An additional 8 mL of pre-warmed complete growth media was added to the tube and mixed. The cells were centrifuged at 200 × g for 10 minutes (room temperature). After centrifugation, supernatant was removed without disturbing the cell pellet and cells were re-suspended in 10 mL complete growth medium and transferred to a T25 flask. Flasks were incubated at 28°C, without CO_2.

Cell transfections

Plasmids

A selection of plasmids were assessed in the development of the *in vitro* sand fly platform (see Table 3). The purpose here was twofold. Firstly, to ascertain whether stock plasmids containing distinct expression promotors known to function in other insect species were also functional in sand fly cell lines, and therefore suitable for expression in CRISPR based plasmids (Cas9 and gRNAs). Secondly PiggyBac plasmids known to function in *D. melanogaster* (Kandul et al., 2019) were assessed to infer if promotors were functional in sand fly cell lines (including drivers of exogenous fluorescent markers), and were therefore candidates for downstream *in vivo* studies (Chapter 5). Promotors contained in each of the plasmids are shown in Table 3.

Four plasmids were assessed in cell transfections: pMaxGFP, Ac5-STABLE1-Neo, UbiqCas9.874W and pHome-T (Table 3). PiggyBac transfections required helper plasmids to provide a source of transposase. Three additional different helper plasmids containing *hsp70* promoters were used, with two hyperactive transposases **(**IhyPBase or MhyPBase) reported to increase transfection efficiency (Eckermann et al., 2018). The transposase coding sequence for these plasmids is downstream of the *Drosophila melanogaster* promoter (heat shock protein 70 gene (*hsp70*)).

Table 3. Plasmids used in sand fly cell transfections.

Plasmid	*Promoters (gene::promoter)*	*Size*	*Notes*
pMaxGFP	GFP::CMV(Cytomegalovirus)	3486bp	From Lonza pmaxCloning Vector Catalog #: VDC-1040
Ac5-STABLE1-Neo	GFP-T2A-Neo::Actin5C	6762bp	Gift from Rosa Barrio & James Sutherland (Addgene plasmid # 32425; http://n2t.net/addgene:32425; RRID:Addgene_32425)
UbiqCas9.874W (PiggyBac)	hSpCas9::Ubiquitin-63E EGFP::Ubiquitin-63E DsRed1::Opie2	14,636bp	Gift from Omar Akbari (Addgene plasmid # 112686; http://n2t.net/addgene:112686; RRID:Addgene_112686).
pHome-T (PiggyBac)	3xP3::GFP Actin5C::RFP	9,545bp	Gift from Tony Nolan, Imperial College, London, UK.
IhyPBase (transposase insect codon optimized)	transposase::Hsp70	-	Gift from Ernst Wimmer, Univeristy of Gottingen, Germany
MhyPBase (transposase mammalian codon optimized)	transposase::Hsp70	-	Gift from Ernst Wimmer, Univeristy of Gottingen, Germany
pDCC6 (CRISPR)	Cas9::hsp70Bb sgRNA::U6-96Ab (U6-2)	8,154bp	Gift from Peter Duchek (Addgene plasmid # 59985 ; http://n2t.net/addgene:59985 ; RRID:Addgene_59985)

Transfection reagents and protocols

Transfection protocols applied to sand fly cell lines do not exist or are not formalised in the published literature. There are considerable differences in transfection efficiencies in different cell lines depending on the approach used. To address this omission we undertook a comparison of five different commercially available transfection reagents and associated transfection protocols in order to identify those that would be most appropriate for an *in vitro* platform in sand fly cells lines. Each are briefly described below.

Lipofectamine 3000

48-well-plates were seeded and maintained to 70-90% confluence for transfection. Culture media was removed, and wells topped up with 225µL of Opti-MEM media. Transfection solutions were assembled according to the manufacturer's protocol (Invitrogen, USA). Transfection consisted of 25µl Opti-Mem media, 250ng of Plasmid DNA, 0.5µL P3000, and 0.45µL Lipofectamine 3000 reagent. Mixture was incubated for 10-15 minutes at room temperature, before 25µL of mixture was added to the wells. The plate was sealed and incubated at 28°C for 12-18 hours. Opti-MEM media was removed and 250µL of culture media was added to the wells. The plate was again sealed and incubated for 48 hours, downstream processing.

FlyFectin optimisation

Initial testing showed that the FlyFectin (OZ Biosciences, France) transfection protocol required optimisation beyond the manufacturer's suggestions. Optimisation of the FlyFectin transfection protocol was required, using three different DNA concentrations (0.25, 0.50, and 1.0 µg DNA) (Table 4), and three different ratios of FlyFectin to DNA (2:1, 3:1, and 4:1 of Flyfectin:DNA) (Table 5). The protocol was applied, and was repeated in all cell lines (LLE40, LLE45, An4a3B, and DS2 or SF21) to determine transfection efficacies in cell lines derived from different insect species. A 48-well-plate was seeded to achieve >50% (60-80% ideally) confluence before transfection. Different reaction mixtures were prepared following a five step procedure.

Step 1: DNA solutions were prepared by diluting DNA in 100µL of culture medium without serum and without antibiotics (Table 4). For each of the three different ratios of FlyFectin tested, 3x master mixes of the DNA solutions were prepared. *E.g., DNA at a concentration of 0.25µg/100µL of media for use in combination with three different FlyFectin solution ratios (2:1, 3:1, 4:1).*

Step 2: FlyFectin solutions were prepared (tubes 1 – 9) by mixing FlyFectin and 100µL of serum free media (Table 5).

Step 3: DNA solutions were mixed with the relevant FlyFectin solutions, and incubated at room temperature for 15-20 minutes (Table 6).

Step 4: Whilst the complex (DNA and FlyFectin solutions combined) was incubated, media was removed from the wells, cells were rinsed with culture media (without antibiotics). 50µL of antibiotic free media was added to each well, followed by 200 µL of the *FlyFectin:DNA* complexes. Cells were incubated for 4 to 6 hours.

Step 5: After incubation 50 µL of complete media (containing FCS and Pen/Strep) was added to each well. Cells were incubated for a further 72 hours at 27°C, before undergoing fluorescent microscopy (directly in plate or on slides).

Table 4. DNA solution mastermixes for optimisation of the FlyFectin protocol. Tubes labelled A, B, and C have different concentrations of DNA.

Tube	*Plasmid DNA (µg)*	*Serum free media (µL)*	*Solution concentration (µg/100µL)*
A (1) 0.25	0.75	300	0.25
B (2) 0.50	1.50	300	0.50
C (3) 1.00	1.00	300	1.00

Table 5. FlyFectin solutions prepared for optimisation of transfection reactions.

Tube	*FlyFectin volume (µL)*	*Serum free media volume (µL)*	*Ratio*
	For 0.25 µg DNA		
(1)	0.5	100	*2:1*
(2)	0.75	100	*3:1*
(3)	1.0	100	*4:1*
	For 0.5 µg DNA		
(4)	1.0	100	*2:1*
(5)	1.5	100	*3:1*
(6)	2.0	100	*4:1*
	For 1 µg DNA		
(7)	2.0	100	*2:1*
(8)	3.0	100	*3:1*
(9)	4.0	100	*4:1*

Table 6. FlyFectin:DNA complex reactions for cell transfections.

Ratio	*0.25 µg DNA*	*0.5 µg DNA*	*1 µg DNA*
2:1	100µL tube (1) + 100µL tube A	100µL tube (4) + 100µL tube B	100µL tube (7) + 100µL tube C
3:1	100µL tube (2) + 100µL tube A	100µL tube (5) + 100µL tube B	100µL tube (8) + 100µL tube C
4:1	100µL tube (3) + 100µL tube A	100µL tube (6) + 100µL tube B	100µL tube (9) + 100µL tube C

FuGene HD optimisation

The FuGene HD (Promega, USA) transfection reagent required initial optimisation prior to use in sand fly cell lines (suggested by Promega, USA)Three different ratios of FuGene HD to DNA concentration were prepared (1.5:1, 2:1, and 3:1), and repeated in LLE40, LLE45, An4a3B, and DS2 or SF21 cell lines. Cells were plated to achieve 80% confluence on the day of transfection. Reaction mixtures were prepared with reagents at room temperature (Table 7), and mixed by inverting the reaction tubes. The FuGene HD:DNA was incubated for 5–15 minutes at room temperature, before being added to wells containing 100µL of growth medium. Plates were incubated at 28°C for 24–48 hours.

Table 7. FuGene HD pre-optimisation transfection reactions for 96-well and 48-well culture plate formats.

	Ratio of FuGene HD:DNA								
Ratio	3:1 20x 96well	3:1 1x96well	3:1 1x48well	2:1 20x 96well	2:1 1x96well	2:1 1x48well	1.5:1 20x 96well	1.5:1 1x96well	1.5:1 1x48well
Medium to a final volume	100 µL	5	17.2 µL	100 µL	5	17.2 µL	100 µL	5	17.2 µL
DNA amount	2 µg	0.1	0.344 µg	2 µg	0.1	0.344 µg	2 µg	0.1	0.344 µg
FuGene HD reagent voume	6 µL	0.3	1.032 µL	4 µL	0.2	0.688 µL	3 µL	0.15	0.516 µL

Cellfectin

Cells were seeded to achieve >70% confluence for the transfection. Media containing serum and antibiotics was removed from the wells, and cells were washed twice with Opti-MEM media. 200µL of Opti-MEM was added to each well while the transfection mixture was prepared. The transfection complexes were prepared by diluting 100ng of plasmid DNA in 10µl of Opti-MEM, followed by 15–30 minutes incubation at room temperature. Cellfectin II (Invitrogen, USA) was diluted with Opti-MEM (0.8µl + 10µl) and incubated for 15–30 minutes. The diluted DNA and diluted Cellfectin II (total volume ~21µl) were combined and incubated for 5–15 minutes. The DNA-lipid mixture was added dropwise to the cells, and incubated for 3–5 hours at 27°C. The transfection mixture was removed, and replaced with 250µL of complete growth medium (L-15, 10% FCS, P/S, Table 1). Cells were incubated at 27°C for 24–48 hours, after which they were prepared for downstream processing (Imaging using Nikon ECLIPSE Ti2 inverted microscope, and flow cytometry using Becton Dickinson LSR-II flow cytometer, results sections *Identification of Expression Promoters for use in Sand fly mutagenesis* and *Comparison of Transfection Protocols applied to cells lines*, respectively).

Effectene

Cells were seeded to achieve >70% confluence for the transfection. The transfection mixture was prepared by diluting 0.15µg of DNA (dissolved in TE buffer, minimum DNA concentration: 0.1µg/µL) with Buffer EC, to a total volume of 50µL. 1.2µL of Enhancer was added, and the mixture briefly vortexed (Table 8). The mixture was incubated at room temperature for 2–5 minutes. 4µL of the Effectene (Qiagen, Germany) transfection reagent was combined with the DNA-Enhancer mixture, and incubated for 5–10 min at room temperature to allow transfection-complex formation. While complex formation was taking place, media was removed from the wells, and cells were washed with PBS. 150µL of growth media was added to the tube containing the transfection complexes, which was then added dropwise to the cells. Cells were incubated under normal growth conditions for an appropriate time for expression of the transfected gene.

Table 8. Effectene transfection reactions in 24-96-well plate formats.

Culture format	*DNA (µg)*	*Enhancer (µL)*	*Final volume of DNA in Buffer EC (µL)*	*Volume of Effectene Reagent (µL)*	*Volume of medium added to cells (µL) †*	*Volume of medium added to complexes (µL) †*
Protocol step	3	3	3	5	7	8
96-well	0.1	0.8	30	2.5*	100	0
48-well	0.15	1.2	50	4*	150	200
24-well	0.2	1.6	60	5	350	350

Asterisk (*) if transfections are performed in 96- or 48-well plates, dilute the Effectene Reagent with Buffer EC to a total volume of 20 µL or 50 µL, respectively, before addition to the diluted DNA–Enhancer mixture prepared in step 3.

† Medium should contain the same percentage of serum as routinely used for culturing cells.

Bioinformatics approach for rationalisation of gene targets to affect

Target gene identification requires a bioinformatics approach utilising genome databases (FlyBase and VectorBase), to select a suite of non-lethal genes that can be knocked out, eliciting a phenotypic effect. Seven genes were selected and rationalised via phylogenetic analysis of orthologous peptide sequences

(conducted in MEGA X software) as putative phenotypic marker genes, including *Scarlet*, *Cinnabar Yellow*, *Ebony*, *Vestigial* and *Rudimentary*. The same approach was used to identify three genes associated with olfaction in insects including *Gr2*, *Orco*, and *Ir8a*. gRNAs were designed using the ChopChop webtool (https://chopchop.cbu.uib.no/) (Labun et al., 2019) prior to *in vitro* transcription. The bioinformatics pipeline and design of gRNAs is comprehensively described in Chapter 2.

In vitro transcription of gRNAs

Following rationalisation of genes to affect, useable gRNAs were generated for CRISPR-Cas approaches. A multi-stage process was followed: design of gRNAs and confirmation of presence within the genome, template generation, and kit-based *in vitro* transcription followed by *in vitro* testing. Individual stages of the approach are described below.

gRNA Design

Sand fly gene IDs were used to identify and design gRNAs in ChopChop software using the *L. longipalpis* (LlonJ1.5) genome (described comprehensively in Chapter 2). The optimum sequence length for gRNAs is 20 nucleotides, immediately followed by a PAM sequence (NGG).

For gRNAs to undergo kit-based transcription, two features were added to the gRNA sequences (manufactured by Integrated DNA Technologies, USA) (Table 9). A T7 promoter was added to the 5' end of the gRNA sequence as the recognition site for the T7 RNA polymerase to initiate transcription. A gRNA backbone was added to the 3' end of the gRNA sequence for interaction with the Cas9 protein. A universal oligo was also manufactured to provide the reverse compliment of the gRNA backbone, required for gRNA template generation.

Table 9. Features required for transcription of gRNAs using kit-based methods. T7 promoter (green), additional guanine bases (to increase transcription yields, red), gRNA backbone (blue). The final row shows an example of the complete sequence incorporating features to be transcribed using a T7 kit-based reaction.

Feature	*Sequence*
T7 promoter	GAAATTAATACGACTCACTATA
Target region including GG for T7 transcription	GGNNNNNNNNNNNNNNNNNGG
gRNA backbone	GTTTTAGAGCTAGAAATAGC
Universal oligo (reverse)	AAAAGCACCGACTCGGTGCGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTT GCTATTTCTAGCTCTAAAAC
Example oligo	GAAATTAATACGACTCACTATAGGCCGTAGTCAACTTTACAATGAGG GTTTTAGAGCTAGAAATAGC

A double sgRNAs knockout strategy was pursued by designing gRNAs targeting two locations in the same gene, with approximately 200bp – 1,500bp between (see Chapter 2). A double knockout approach facilitates screening of transgenic individuals by inducing a larger sequence knockout. For example, if both gRNAs function, indels will result in a ~200bp knockout giving a visibly lower gel band compared to wildtype sand fly DNA.

Confirmation of gRNA sequences

Primers were designed to amplify gRNA sequences in colony (wildtype) sand fly DNA to confirm gRNA sequences designed were as expected by comparison to the available genome. The primers were designed using ChopChop software (https://chopchop.cbu.uib.no/) with a primer size of 18-25bp, a product size of approximately 500-550bp (in some cases product size was larger), melting temperature of 57-60°C, and a flanking distance from the gRNA to the primers of 200bp. The most suitable gRNAs were selected using a combination of efficiency scores and self-complementarity scores suggested by the software generating a ranking position. Primers amplified the target region, and Sanger sequencing was performed to confirm the presence of the gRNA sequences within the primer-amplified region.

gRNA Template generation

gRNAs and universal oligos were annealed via PCR to produce a double stranded template sequence incorporating the T7 promoter, gRNA sequence, and scaffold sequence. The annealing reaction comprised of a standard PCR mix , in addition to sgRNA oligos (10µM), and the universal oligo (10µM). Reactions were thermocycled and visualised via gel electrophoresis, followed by gel extraction (Qiaquick gel extraction kit, Qiagen, Germany). Template concentrations were determined by nanodrop sectrophotometer (see Appendices 3 and 4).

The universal oligo is 80bp in length and in tandem with the sgRNA oligo, produced a total template gRNA length of 118bp.

Kit-based in vitro transcription of sgRNAs

Different commercial kits are available to transcribe gRNAs into sgRNAs, the useable form for transfections both *in vitro* and *in vivo*, including HiScribe T7 High Yield RNA Synthesis Kit (New England Biolabs, NEB, USA), EnGen sgRNA Synthesis Kit (NEB, USA), and MEGAscript T7 Transcription Kit (Invitrogen, USA). HiScribe and Engen kits were investigated to identify which resulted in highest *in vitro* transcription efficiencies in our hands.

Transcription of sgRNAs using the HiScribe kit (NEB, USA) was conducted following the manufacturer's protocol (NEB, USA). Briefly, the reaction mixture was assembled, containing the template gRNA, T7 RNA polymerase mix, DNTPs and buffer. Reactions were incubated at 37°C for two hours, before gel electrophoresis. sgRNAs were extracted using Qiaquick gel extraction kit

(Qiagen, Germany) and eluted in H_2O. Concentrations were determined using a Nanodrop Spectrophotometer.

Transcription of sgRNAs was also conducted using the Engen kit (NEB, USA), following the manufacturer's protocol with some modifications. Briefly, the EnGen reaction mixture, target-specific DNA oligo, DTT ,and the EnGen sgRNA enzyme mix were combined and incubated for 45 minutes to 2 hours (to increase the yield, compared to the original protocol). DNase I was incubated with the mixture for 15 minutes at 37°C with subsequent visualisation by gel electrophoresis. The transcribed sgRNA product is ~100bp.The T7 promoter region is not transcribed, however the gRNA sequence and the overlap with the universal oligo are transcribed.

DNA Cleavage assay

A cleavage assay was used to determine whether transcribed sgRNAs successfully introduce double stranded breaks in genomic DNA (see results section *Cleavage Assay*). Two cleavage assay protocols were used: New England Biolabs (NEB) 'in vitro digestion of DNA', provided with the Cas9 Nuclease (product M0386) (NEB), and a protocol described by Martin-Martin *et al.,* (2018).

For the NEB protocol, reaction mixtures comprised 1µM Cas9 nuclease, 300nM *in vitro* transcribed sgRNA, NEB Buffer r3.1, and water. This mixture was incubated for 10 minutes at 25°C. Subsequently 30nM of amplified target DNA was added and the mixture incubated at 37°C for 15 minutes. 1µL Proteinase K was added and at room temperature for 10 minutes. Products were visualised on a 1.2% agarose gel at 100V.

The Martin-Martin protocol was followed by incubation of 2.5µL Cas9 nuclease at 37°C for 10 minutes with 45ng sgRNA, 2µL Cas9 buffer (NEB), DEPC- H_20. 300ng of amplified target DNA was added to the mixture for 1 hour 15 minutes at 37°C. 1µL 5µg/mL proteinase K was added for 10 minutes at room temperature to inactivate Cas9. Products were visualised on a 1.2% agarose as above.

The cleavage assays described help identification of sgRNA function in principle, prior to direct delivery both in vitro and in vivo, or for incorporation of sequence into plasmids constructs.

Generation of Plasmid constructs via Gibson Assembly

Gibson Assembly was used to construct phenotypic and olfactory gene targeting plasmids, for validation in cell lines and for *in vivo* studies. gRNAs, promotors and other exogenous DNA are inserted into backbone constructs (pDCC6) to produce CRISPR plasmids (see results section *Generation of transgenic Cas9-expressing sand fly cell line* and *Building CRISPR plasmids construction to target phenotypic and olfactory genes*). The protocol requires three key steps described below.

Step one requires amplification of the fragment to be inserted into the backbone plasmid. A Phusion High Fidelity PCR reaction was assembled including primers designed to amplify the fragments from

DNA or a plasmid and consists of the DNA, Phusion HF buffer, DNTPS, DMSO, and Phusion high Fidelity DNA Polymerase. The reaction was thermocyled in a 31-cycle program [30s at 98°C (10s at 98°C, 30s at 54°C, 30s at 72°C) x31, 5min at 72°C]. Products were visualised on a 1% gel (1hr, 80V, and 140mA). were excised and purified using Qiaex II gel extraction kit (Qiagen, Germany) following manufacturers protocols.

The second step requires amplification of the plasmid vector that will be incorporating the fragment amplified previously. The plasmid also requires overhangs for scarless insertion of the fragment. A Phusion high Fidelity PCR reaction was assembled as before, using primers for the vector plasmid, and the vector plasmid itself. The thermocycler program for this reaction, gel electrophoresis and purification were as previously described.

Step three was the Gibson assembly reaction to insert the fragment into the vector plasmid. The reaction mixtures consisted of 35ng of DNA fragment to be inserted, 200ng of recipient plasmid vector, Gibson Assembly mastermix, and H_2O. The reaction was thermocycled for 1 hour at 50°C, followed by gel electrophoresis on a 0.8% gel (1 hour, 80V, and 140mA).

The next steps were bacterial transformation, and plasmid purification described below.

Bacterial Transformation for Plasmid amplification

LB agar plates supplemented with 100µg/mL of ampicillin were prepared in a sterile lateral flow hood, before being incubated at 37°C. X10 gold competent *E. coli* cells were thawed on ice for 20 minutes. Cells were mixed and 4µL of β-ME mix was added. The mixture was placed on ice for 10 minutes (mixed every 2 minutes). 5µL DNA (10pg to 100ng) was added to 20µL of competent cells in a microcentrifuge tube and gently mixed. The competent cell/DNA mixture was kept on ice for 30 minutes. Transformation tubes were heat-shocked by placing the bottom 1/2 to 2/3 of the tube into a 42°C water bath for 30 seconds and tubes returned to ice for 2 minutes. 200µL of pre-warmed SOC media (without antibiotic) was added, before incubation in a 37°C shaking incubator for 1 hour (225-250rpm). Approximately 200µL of the transformation mixture was plated per LB agar plate, and incubated at 37°C overnight.

A pUC18 control was used to confirm transformation. pUC18 stock was diluted before the transformation as described above, replacing the DNA with 1 µL of diluted pUC18. 5µL of the control transformation was plated on LB agar before incubation. Approximately 50 colonies indicates a successful transformation.

Colonies were picked off the agar into 5mL LB broth containing 100µg/mL ampicillin in a 50mL centrifuge tubes. Tubes were placed in a shaking incubator at 37°C at 200rpm for 12-18 hours.

Miniprep Plasmid purification

Overnight bacterial cultures (X10 gold competent *E. coli*) were centrifuged at 30,000 x g for 5 min at room temperature. Bacteria were re-suspended in 250µL of buffer P1 and transferred to a microcentrifuge tube. 250µL of a second buffer P2 was added, and mixed by inversion of the tube. 350µL of buffer N3, was added and mixed by inversion, before centrifugation for 10 min at 13,000 rpm (~17,900 x g) in a table-top microcentrifuge. Supernatant, containing the plasmid, was removed and added to a QIAprep spin column (Qiagen, Germany), and centrifuged for 30–60 seconds. Two consecutive wash steps were conducted using buffer PB and PE. The QIAprep columns placed into sterile 1.5 mL microcentrifuge tubes, and DNA was eluted into 50µL of MiliQ via centrifugation for 1 minute. Plasmid concentrations were measured by nanodrop spectrophotometer.

Slide preparation for cell imaging

Transfected cell populations containing fluorescent markers were initially inspected microscopically. Cells were removed from the wells of a plate and centrifuged for six minutes at 1,000 rpm in 1.5mL tubes. The supernatant was removed and cells re-suspended in 100 µL of PBS before a second centrifugation step. The supernatant was removed and 4% paraformaldehyde was added and incubated for 30 minutes at room temperature. This was followed by additional washes with PBS. After the final wash, the pellet was re-suspended in approximately 5µL of PBS, applied to the centre of a polylysine slide and left to air dry. Subsequently 7µL of VectaShield mounting medium containing DAPI stain applied. Coverslips were added and the edge sealed with clear nail varnish. Once dry, the slides were stored at -4°C until imaging using a Nikon ECLIPSE Ti2 inverted microscope. Imaging was qualitative, systematically scanning slides for evidence of fluorescence in GFP and RFP channels, in addition to the DAPI channel to confirm fluorescence was from cells and not artefacts.

Flow cytometry

Flow cytometry was used as a quantitative method to assess transfection efficiency by identification of fluorescent cell events. Populations of plasmid transfected sand fly cells incorporating fluorescent cargo were prepared for flow cytometry. Briefly, media was removed from the wells and 250µL of PBS was added. The solution was agitated to remove adherent cells and transferred to 1.5mL Eppendorf tubes and centrifuged for six minutes at 1,000 rpm. The supernatant was removed without disturbing the pellet. Cells were re-suspended in 100µL of PBS, and centrifuged again. 4% paraformaldehyde (PFA) was added to the cells, mixed, and incubated for 30 minutes at room temperature. Two further washes with PBS were completed before re-suspending in a final volume of 1mL of PBS in Flow cytometry tubes and stored at 4°C.

Flow cytometry was performed on a Becton Dickinson LSR-II flow cytometer. FACSDiva 6.1.3 software was used for quality control of the flow cytometer by conducting a Setup and tracking (CST) system screen. FACSDiva 6.1.3 was also used to acquire the data (analysis described in the results

section). Flow cytometric data analysis was conducted using FlowJo v10.8.1 software (see results section *Comparison of Transfection Protocols applied to cells lines*)

G418 selection assays

The plasmids UbiqCas9-NeoR plasmid containing neomycin resistance selection markers was used in cell transfection assays, towards the development of a Cas9-expressing cell line. Cells were selected for by the addition of antibiotic (G418) using the following protocol. Cells were seeded in a 48-well plate with $0.5x10^6$ cells, and media topped up to 250µL per well. Plates were sealed and incubated for 16-24 hours at 28°C. After incubation cells were transfected with the Ubiq-Cas9-Neo plasmid, controls were transfected with the UbiqCas9.874W plasmid. The cells were again incubated, for 72 hours. G418 was added to the wells at the desired concentration. At 5-7 day intervals the wells were checked for cell death, before G418 was removed and replaced. After 30 days, cells were imaged using a fluorescent microscope. G418 concentrations calculations can be found in Appendix 7.

Alamar blue

The Alamar blue reagent was used to determine cell viability after selection with antibiotic (G418). The non-toxic reagent is converted by living cells to a fluorescent product, resorufin, which is analysed by a plate reader. A 48-well plate was seeded with LLE40 and LLE45 cells as per the protocol described for G418 selection assays (above), and allowed to reach 70-80% confluence. G418 was added to the wells at a range of concentrations, and the plate was incubated for 48 hours at 28°C. Alamar blue (25µL of 440µM stock) was added to the wells and viewed at 555-585nm absorbance (SpectraMax iD5). Concentration-Response nonlinear regression analysis was conducted in Graphpad Prism 9.0.0.

Results

In vitro transcription: Confirmation of gRNA sequences

Shortlisted olfactory and phenotypic gene targets of interest (see Chapter 2) were initially PCR amplified from DNA extracted from colony reared *L. Longipalpis* sand flies, to confirm expected fragment sizes via gel electrophoresis. Sequencing of gene targets was subsequently performed, and gRNA sequence sites confirmed within the fragment (Table 10). Sequencing identified the presence of multiple gRNAs within the amplified regions of interest: *Gr2*, *Orco* and *Ir8a* (six, three, and five gRNAs per locus respectively). Two gRNAs were unable to be identified within the Orco amplified region, and one gRNA had no identity within the Ir8a region. Sequencing also confirmed multiple gRNAs of phenotypic gene targets *Kmo*, *Rudimentary*, *vestigial* and *Yellow* (four, five, five, and four gRNAs respectively). For each of the targets, two gRNAs were taken forward to template generation and *in vitro* transcription.

Table 10. Sequencing primers for amplification of olfactory and phenotypic target genes within wildtype DNA. Primers, flank gRNA sequences. In the gRNA sequences column, black text represents the guide sequences found within the amplified region, red text indicates a nucleotide change in the guide sequence compared to the available genome sequence (VectorBase), purple text indicates that the guide sequence was not found within the amplified region. Asterisk (*) indicates gRNAs taken forward to *in vitro* transcription.

L. longipalpis gene	*Primers*	*gRNA sequences*	*Fragment size (bp)*
GR2 (LLOJ010835)	TGGCTTGGCATTCTAAGAAGA CCAGGACAATAACCGTGGTG	*GTTAACAGCCACTGTGGATCCGG GTATGGGATATTTCGCAGAAGGG GCTTCATGGCGACATGAACAGGG AAACATAAAGAGGCAATACGTGG GCTGTGTACAAGACAATGTGGGG *TCCGGGTTGTGATCGGGTAA	549
Orco (LLOJ003114)	CTGAAACATCCAGATTGTCCAT TTTCGTGATGCTGTGTGTGA	TGGGCAAAAGATCGGCCACAAGG GCAAAAATATCAAGGCCTTGTGG TACAGCAATCAAGTATTGGGTGG *GTGCAACCGCAAAAATATCAAGG *TGTGAGATACATGACCAACAAGG	539
Ir8a (LLOJ003646)	GATTCCCGAAAATGACGAGA GAGCCATACTTCCAACCGAA	GAGATTGTACGTCCGAAAACGGG TGTAGGAAGGTACGTTGATGAGG AGCAATAGCGAGACCTGCGAGGG TTGAGTGTAAAATCCCGACAGGG *TGAGATTGTACGTCCGAAAACGG *GTGAGACAGACATCCCTTTT	550
Ir25a (LLOJ001989)	GCAGCTACCAAGTACCCCAA AATCTCCCCTCAATGTCTGC	GCCTAGTGGCAGCAACGTGGTGG GAATCTCTCAGGACGCCTAGTGG GACGCCTAGTGGCAGCAACGTGG	550
Ir76b (LLOJ005467)	TCTCCCCTTTATTTCCGTCC AAAGTCCGCTTTCCCATTTT	AGATGGTGCTGGATCAGCTGAGG GAGTAATATGACATCGGGTGAGG GGCATGAGTAATATGACATCGGG	524
Gr1 (LLOJ006888)	TCTATGCAGTCAGCTGTGGG TAGAGATCATTCAAATGCCGC	GAGCAACATGATTCATCCGGAGG GGATCTGTTTCTCGCTCCTGTGG GTCCAATGAAATGTCGTACGAGG	550
Or4 (LLOJ004500)	TTTAAAGCCCCAAAGCAGTC AATGAGGTAGGACTGCCGAC	TGAGCTCCCCCACAACACAGAGG GTCAAAGGGCACGAGCCCAATGG GGTATCCGATAGCCAGGATGTGG	511
Or2 (LLOJ010523)	AAAAATGGCAGGTAAAAACGAA GAGTTGGGTCATTATCAGCACA	TGACAACATAGTGCAACATGAGG GGTATATAGAGGCTAGCTCCTGG GGCATAGTAAGTAGATCTGTCGG	516
Kmo (LLOJ009814)	AGCTCAACAAGGGCATAAGAAA ATTTGGAGATAAAAGTTTGGCG	CCCCCATGAATATCGGCCCTTCC CCTTTGTGAAGATAGACGATGAC *CCTCAGACAAAGTATCCGGAAAC *TACCACGGAGCTCATTGGAGGG	702
Rudimentary (LLOJ009278)	CCATTCTCGAGACAATCCTCC	AGCATTGGAAGACACAGGGTTGG TTGTAGCCCATCTTCACGAGGG	725

	CAAGTTGTTTCACTCTTCCCG	GGAACTATGGCATTCCGTGCGG *GGATTCAACAAGTGGCCATGGGG *GGGGAGGTTGCGTATGTTGATGG	
Vestigial (LLOJ009695)	TGGTATGCAATGAAATTGAGAA TTGAAATGATTAAAATCGGACA	TCATCATTACGGTTCCTACGCGG GGAAAATTTCTCGCCGACATCGG TCGCGGACACGTATTGTGCTCGG CCGCCGCGTACTCTAGCTACCCC CCATGCGGCTCACGCTTACCACC	834
Yellow (LLOJ007802)	GAGACTTTGGTGGCAATCAAT GCAACACGTCAAATGTGAAAAT	CCCCACTGCATCGACCATTGCT CCCACATCATCTCTCAGCCTGAA *CCGTAGTCAACTTTACAATGAGG *TCAGTGCCACGATGGAAGGACGG	658
Scarlet (LLOJ001495)	AACTCCATCGGTGAAAGTGC GGTTAAGGTGCCCACAAAGA	*GCGCATCATCAATAACGTGACGG	526
Cinnabar (LLOJ009814)	ATGTGGACGGTTCATTGACAGT ACTGCCCGAAGAAACCAA	*GATTTGATCATCGGCTGCGA	597
Martin-Martin gRNA5 *Martin-Martin gRNA6*	n/a	*CTGATGTCTGCGGTAGGCTTTGG *CGCATCAAGGCTGATGTCTGCGG	372

In vitro transcription: Template generation

Transcription of sgRNA oligos was initially attempted using the HiScribe kit (NEB, USA) targeting *Scarlet*, *Cinnabar*, and the *Yellow* genes (Table 11) from Martin-Martin *et al.,* (2018). Templates were successfully generated for *Scarlet* and *Cinnabar* with clear bands observed at the expected size (~118bp) (Figure 2A). Controls (no oligo, universal oligo alone, scarlet oligo alone, and cinnabar oligo alone) showed an absence of bands above 100bp.

The sgRNA template generation was successful with two gRNAs for each of the following genes: *Kmo*, *Rudimentary*, *Vestigial*, and *Yellow* (Table 11). Briefly, a reaction mixture containing standard PCR components in addition to the CRISPR universal oligo and the forward oligos for each of the sgRNAs, along with three controls (universal oligo alone, sgRNA oligo (Kmo1), and no oligos) (Table 11). sgRNA templates were visible at ~118bp, as expected (Figure 2B), in addition to some faint bands visible ~80bp, which was likely to be remaining universal oligo. The three controls showed no band between 200-100bp, confirming that the products observed were genuine template (Figure 2B, red box). The sgRNAs templates were gel extracted (Qiaquick), and concentrations determined by nonodrop sectrophotometer (see Chapter 3 Appendix 3).

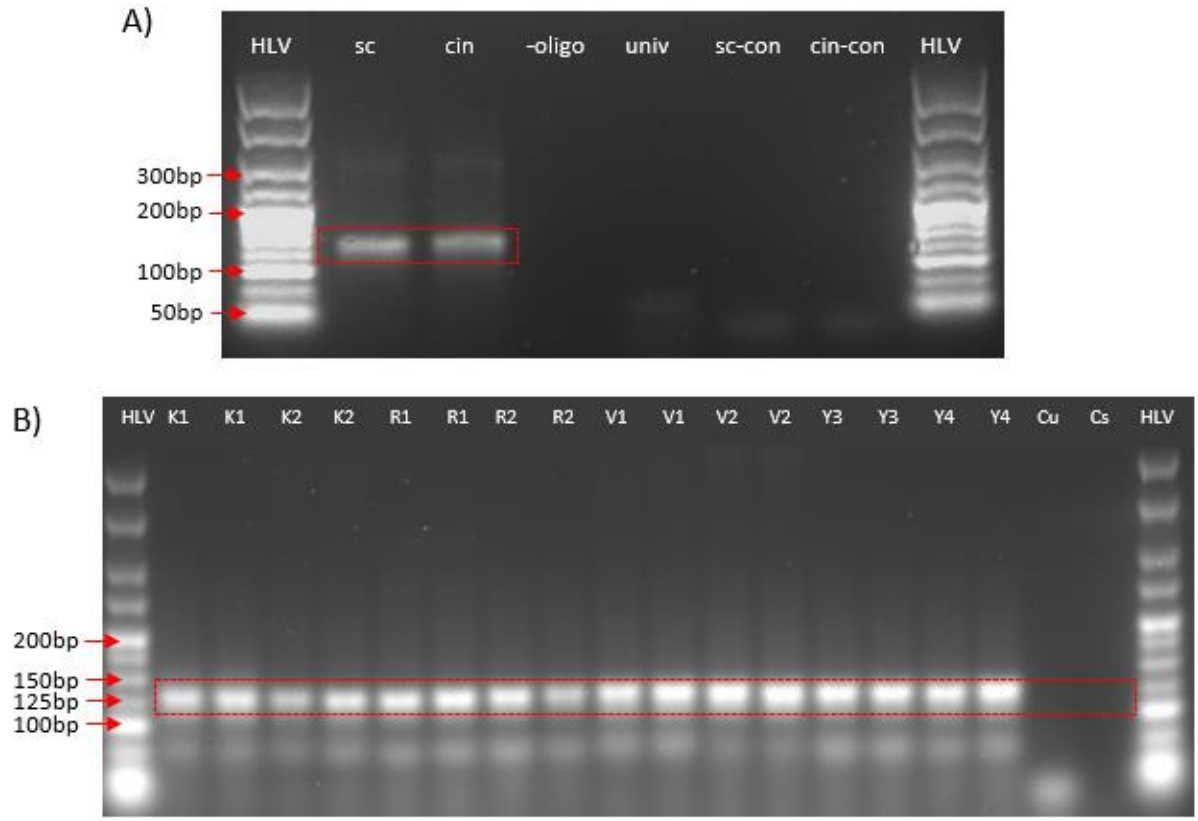


Figure 2. sgRNA template generation for HiScribe (NEB) *in vitro* transcription. The universal oligo was annealed to phenotypic gRNA oligos for *Scarlet* and *Cinnabar* genes. Expected template size is 118bp. Red box indicates bands at the expected size. A) Annealed oligos formed templates Scarlet (sc) and Cinnabar (cin), visualised on a 1.5% agarose gel with Hyperladder V (Bioline). Controls contained no oligo (-oligo), the universal oligo alone (univ), and scarlet (sc-con) and cinnabar (cin-con) oligos alone (*lanes 4-7*). B) sgRNA template generation for *Kmo (lanes 2-5)*, *Rudimentary (lanes 6-8)*, *Vestigial*, and *Yellow* phenotypic gRNA sequences; *Kmo* gRNA#1 and #2 (K1, K2); *Rudimentary* gRNA#1 and #2 (R1, R2), *Vestigial* gRNA#1 and #2 (V1, V2), *Yellow* gRNA#3 and #4 (Y3,Y4). Controls contained universal oligo (Cu), and sgRNA oligo (Cs) (*lanes 18-19*).

For T7 Transcription using the Engen kit, 1µM concentration of sgRNA template was required. The concentrations generated were below the requirement (with only two above, V1 and Y3, with concentrations of 1.06 µM and 1.2 µM respectively), therfore template generation was repeated until the transcription reaction could be performed (molar concentrations between 2.47 and 3.26 µM, see Chapter 3 Appendix 4). The sgRNA templates were taken forward to the Engen transcription kit reaction using the manufacturer's protocol, with modification (see Methods, DNA Cleavage assay).

Table 11. Phenotypic and olfactory gRNA oligos for synthesis of sgRNA templates. Red text represents 5' T7 promoter, green text represents the gene-specific target (gRNA sequence), and blue text represents the scaffold region complementary to the Universal oligo. gRNA5 and gRNA6 are controls derived from Martin-Martin *et al.,* (2018), targeting the *Yellow* gene.

Name	*Oligonucleotide sequence*	*Transcribed sgRNA size (bp)*
CRISPR Universal oligo	AAAAGCACCGACTCGGTGCCACTTTTTCAA GTTGATAACGGACTAGCCTTATTTTAACTT GCTATTTCTAGCTCTAAAAC	
LLKmo gRNA#1	TAATACGACTCACTATAGGCCTCAGACAAAGTATCCGGAAAC GTTTTAGAGCTAGAAATAGC	104
LLKmo gRNA#2	TAATACGACTCACTATAGGTACCACGGAGCTCATTGGAGGG GTTTTAGAGCTAGAAATAGC	103
LLRud gRNA#1	TAATACGACTCACTATAGGGGATTCAACAAGTGGCCATGGGG GTTTTAGAGCTAGAAATAGC	104
LLRud gRNA#2	TAATACGACTCACTATAGGGGGAGGTTGCGTATGTTGATGG GTTTTAGAGCTAGAAATAGC	104
LLVest gRNA#1	TAATACGACTCACTATAGGCCGCCGCGTACTCTAGCTACCCC GTTTTAGAGCTAGAAATAGC	104
LLVest gRNA#2	TAATACGACTCACTATAGGCCATGCGGCTCACGCTTACCACC GTTTTAGAGCTAGAAATAGC	104
LLYellow gRNA#3	TAATACGACTCACTATAGGCCGTAGTCAACTTTACAATGAGG GTTTTAGAGCTAGAAATAGC	104
LLYellow gRNA#4	TAATACGACTCACTATAGGTCAGTGCCACGATGGAAGGACGG GTTTTAGAGCTAGAAATAGC	104
Scarlet	TAATACGACTCACTATAGGCGCATCATCAATAACGTGACGG GTTTTAGAGCTAGAAATAGC	104
Cinnabar	TAATACGACTCACTATAGGATTTGATCATCGGCTGCGA GTTTTAGAGCTAGAAATAGC	101
Gr2-1	TAATACGACTCACTATAGGTTAACAGCCACTGTGGATCCGG GTTTTAGAGCTAGAAATAGC	103
Gr2-2	TAATACGACTCACTATAGGTCCGGGTTGTGATCGGGTAA GTTTTAGAGCTAGAAATAGC	101
Orco-1	TAATACGACTCACTATAGGTGCAACCGCAAAAATATCAAGG GTTTTAGAGCTAGAAATAGC	103
Orco-2	TAATACGACTCACTATAGGTGTGAGATACATGACCAACAAGG GTTTTAGAGCTAGAAATAGC	104
Ir8a-1	TAATACGACTCACTATAGGTGAGATTGTACGTCCGAAAACGG GTTTTAGAGCTAGAAATAGC	104
Ir8a-2	TAATACGACTCACTATAGGTGAGACAGACATCCCTTTT GTTTTAGAGCTAGAAATAGC	101
gRNA5	TAATACGACTCACTATAGGCTGATGTCTGCGGTAGGCTTTGG GTTTTAGAGCTAGAAATAGC	104
gRNA6	TAATACGACTCACTATAGGCGCATCAAGGCTGATGTCTGCGG GTTTTAGAGCTAGAAATAGC	104

In vitro transcription: sgRNA transcription

Transcription of *Scarlet* and *Cinnabar* sgRNAs using the HiScribe kit (NEB, USA) was conducted following the manufacturer's protocol (NEB, USA). Using this method, no products of the expected size (~100bp) for *scarlet* or *cinnabar* sgRNAs were obtained after multiple attempts. As transcription of sgRNAs was unsuccessful using the HiScribe kit an alternative transcription kit, Engen (NEB, USA), was attempted. Engen was selected because the annealing and transcription could be carried out in one reaction, taking under two hours, compared to transcription using the HiScribe which required two days.

The Engen transcription protocol functioned for eight sgRNAs (Scarlet, cinnabar, gRNA5 and gRNA6 (both targeting *Yellow*) shown in Figure 3, and K2, R1, V1, Y3 Chapter 3 appendix), however for three of these sgRNAs the products were not the expected size (K1=125-150bp, K2=150-175bp, and V2=175bp). Additionally, no transcription occurred with R2 and Y4 gRNAs (Chapter 3 Appendix). The larger than expected products may be due to dimerization or the type of gel being used. The NEB protocol suggests running the RNAs on a TBE-Urea denaturing gel as the sgRNAs may run as multiple bands if not fully denatured. Smearing may also suggest contamination with RNase.

Although not every gRNA appeared to be transcribed as expected, the transcription products were taken forwards to cleanup. Briefly, a Monarch RNA cleanup kit (NEB, USA) was used following the manufacturer's protocol. Binding buffer and ethanol (≥ 95%) were added to the sgRNAs and mixed to enable binding of the RNA. The mixture was centrifuged in a column followed by two repeats of washing with RNA wash buffer provided with the kit. The sgRNAs were eluted into H_2O, concentration determined, and the samples stored at -70°C.

The Monarch cleanup protocol was followed for all 12 sgRNAs transcribed using the Engen kit, including R2 and Y4 where no product was visible (for concentrations see Appendix 5). After cleanup of extracted sgRNA concentrations of between 100.58 ng/µL and 353.574ng/µL were achieved.

In summary, products obtained using the Engen kit were clearer (strong bands and no smearing) than those from the HiScribe kit, although most were observed lower than the expected ~100bp size. This may be a result of single stranded RNA migrating comparatively faster in gel electrophoresis than DNA. Controls confirmed this, with products observed at ~75bp rather than 102bp (as expected). When using the Monarch RNA clean up kit (NEB, USA), a much higher concentration of RNA was achieved than via gel extraction. The EnGen kit was used in future annealing and transcription reactions beacause the transcribed sgRNAs genereated were easier to visualise via gel electrophoresis and had higher concentrations. The Monarch RNA cleanup kit was selected to purify the products.

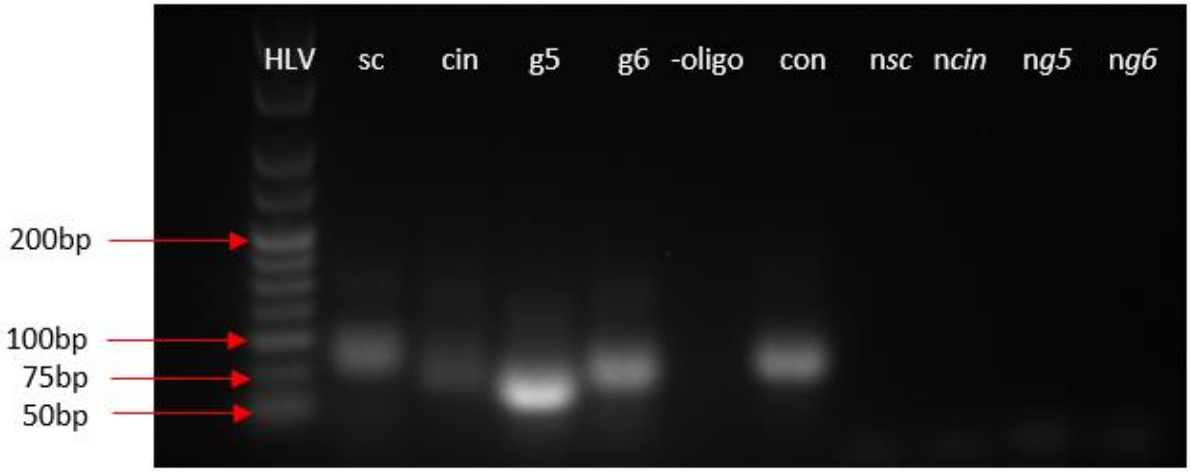


Figure 3. Transcription of sgRNAs using Engen sgRNA synthesis *S. pyogenes* kit (NEB, USA). Transcription of scarlet (sc), cinnabar (cin), gRNA5 (g5) and gRNA6 (g6). Expected sizes 104bp, 101bp, 104bp, 104bp for sc, cin, g5 and g6 respectively. Several controls were used including control with no sgRNA(-oligo), transcribed control oligonucleotide provided with the kit (con), non-transcribed single stranded sgRNA oligos for scarlet cinnabar, gRNA5 and gRNA6 (nsc, ncin, ng5, and ng6 respectively).

Cleavage Assay Optimisation

For the development of CRISPR-Cas9 in sand flies, transcription of gRNAs to sgRNAs must be achieved. Following transcription it is important to determine whether the transcribed sgRNAs are functional. A cleavage assay was developed against wildtype DNA amplified from the target gene. Here primers flanking each region were designed (Table 10), and amplification was performed using a standard PCR and Phusion High-Fidelity PCR. This provides the DNA template to assess cleavage facilitated by gRNA (below). Target region DNA cleavage templates were generated for all phenotypic and olfactory target genes (Table 11). These were taken forwards to optimisation of the DNA cleavage assay.

Cleavage Assay

Optimisation of the *in vitro* cleavage assay using transcribed gRNAs involved two protocols; An NEB protocol for *in vitro* digestion of DNA protocol provided with Cas9, and a published protocol in Martin-Martin *et al.,* (2018) (see Methods, DNA Cleavage assay). Briefly, target DNA is amplified, followed by incubation with *in vitro* transcribed gRNAs and Cas9 nuclease to determine whether gRNAs cleave target DNA as expected.

The NEB protocol for cleavage of DNA resulted in partial digestion with the *Cinnabar* sgRNA, producing a faint fragment visible above 200bp (Figure 4A). The expected digestion fragments for *Cinnabar* DNA (597bp) are 263bp and 274bp. Digestion of the *Scarlet* DNA template (526bp) demonstrated partial cleavage, with fragments visible at ~300bp, and ~200bp, when expected cleavage with the sgRNA would result in 277bp and 189bp fragments. *Yellow* DNA (372bp) appears to have had incomplete digestion with the sgRNA5, producing fragments at 350bp and 200bp compared to the expected 150bp and 214bp fragments. There was also incomplete digestion with sgRNA6. Fragments were visible at 350bp and 200bp, compared to 148bp and 224bp as expected.

Overall, although the NEB protocol appeared to demonstrate a degree of cleavage of target DNA using *in vitro* transcribed sgRNAs, the results were inconsistent; therefore, we attempted a second protocol published by Martin-Martin *et al.,* (2018). Using this cleavage protocol (Figure 4B), *Scarlet* sgRNA cleaved the *Scarlet* DNA template (526bp, lane 3), producing fragments sizes of ~200bp and ~300bp (lane 2). These align with the expected fragment sizes of 277bp and 189bp, suggesting complete digestion occurred. The *Yellow* sgRNA used as a control (sgRNA5), cleaved the *Yellow* template DNA (372bp), producing fragments of ~150bp and ~200bp (lane 4). These also matched the expected fragment sizes for this cleavage (158bp and 214bp), although incomplete digestion occurred. The expected fragment from digestion of *Yellow* DNA with the second control sgRNA (sgRNA6, from Martin-Martin *et al.,* 2018) are 148bp and 224bp, the faint fragment visible (lane 5) are ~200bp and ~350bp suggesting incomplete digestion may have occurred. Overall, the protocol from Martin-Martin *et al.,* (2018) demonstrated more complete cleavage, judged by gel electrophoresis visualisation, than the NEB protocol, therefore this assay was taken forward to analyse *in vitro* transcribed sgRNA function.

The function of five olfactory sgRNAs was determined using the Martin-Martin protocol (Martin-Martin, Aryan, Meneses, Adelman, & Calvo, 2018), with four resulting in successful cleavage (Figure 4C). *Gr2* DNA (549bp, lane 2) was cleaved with the GR2-1 sgRNA, giving fragments of ~250bp and ~350bp (lane 3), as expected (240bp and 309bp). Gr2-2 sgRNA cleaved DNA giving a 450bp fragment, and a faint 100bp fragment. Expected fragments for this cleavage are 76bp and 473bp. *Ir8a* DNA (689bp, lane 5) was cleaved with Ir8a-1 sgRNA producing fragments at 600bp, and a faint fragment at ~100bp (lane 6), compared to expected 85bp and 604bp bands. Ir8a-2 sgRNA did not cleave the DNA (expected bands at 261bp and 428bp), demonstrating Ir8a-2 sgRNA was not functional. *Orco* DNA (635bp, lane 8) cleaved with Orco-2 sgRNA (lane 9), giving fragments at ~300-400bp compared to expected 317bp and 318bp fragments. High concentrations of un-cleaved DNA were visible in all lanes, suggesting incomplete digestion.

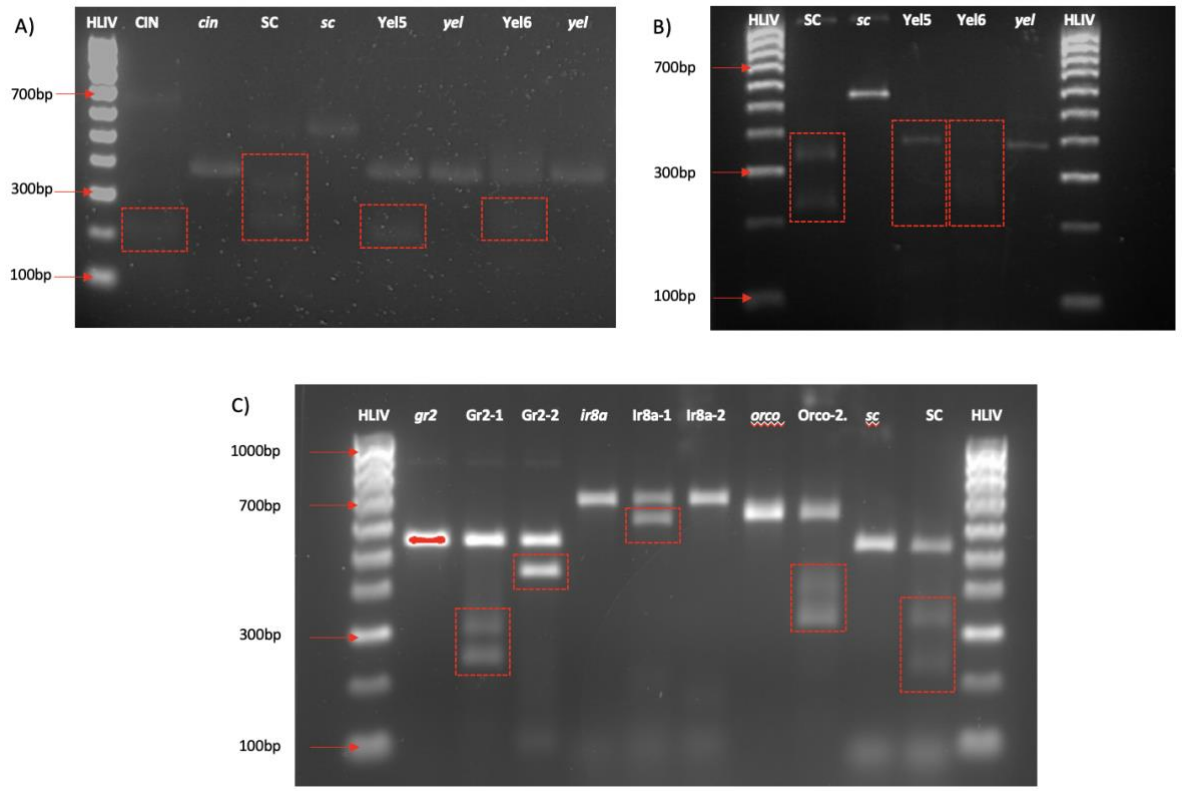


Figure 4. *In vitro* cleavage assay optimisation. A) Cleavage assay based on an NEB protocol. Digested samples with sgRNAs were loaded next to the equivalent control containing no sgRNA: CIN is digested Cinnabar DNA digested (CIN), control (cin); Scarlet DNA digested (SC), control (sc); Yellow DNA digested with sgRNA5 (Yel5) and sgRNA6 (Yel6), control (yel). B) Cleavage assay based on the Martin-Martin protocol for *Yellow* and *Scarlet* sgRNAs. C) Cleavage assay based on Martin-Martin protocol for olfactory sgRNAs Gr2-1, GR2-2, Ir8a-1, IR8-2 and Orco, Scarlet gRNA (*lanes 3, 4, 6, 7, 9, and 11*). Controls with no sgRNA were loaded next to samples digested with sgRNAs: Gr2 control (gr2) (*lane 2*), Ir8a control (ir8a) (*lane 5*); contains Orco control (orco) (*lane 8*) Scarlet control (l*ane 10*).

Discussion of sgRNA *In vitro* transcription and cleavage assays

Production of CRISPR Knock-out plasmids can be a time-consuming process, therefore it is important to determine the functionality of gRNA sequences that will be used to target specific genes prior to both construction of plasmids and microinjection into sand fly embryos.

An *in vitro* transcription assay was optimised from stock protocols to produce functional sgRNAs for validation in sand fly cell lines prior to use in a direct microinjection strategy, or prior to incorporation into plasmid constructs. Testing the function of gRNAs prior to construction of knock-out plasmids can help identify gRNAs likely to work *in vivo*. The cleavage assay allows the sgRNAs to be tested for functionality by incubating them with target DNA and Cas9 endonuclease to determine if the sgRNAs cause specific cleavage of the DNA.

During the development of the protocol, sgRNAs targeting *Scarlet* and *Cinnabar* genes were selected to aid optimisation as these were both promising and have been proven to cause phenotypic changes in other insects (see Chapter 2). Two sgRNAs were selected for use as controls during optimisation, gRNA5 and gRNA6 (Table 11), sgRNA which had previously demonstrated functionality *in vitro* by Martin-Martin *et al.,* (2018). These controls enabled optimisation of the protocol for use with un-tested gRNAs, via comparison of transcription products generated. Once optimised, the protocol was used to validated five sgRNAs targeting three olfactory genes (*Gr2*, *Ir8a* and *Orco*).

HiScribe kit (NEB, USA) was initially used as this kit had successfully transcribed gRNAs in previous studies (Koutroumpa et al., 2016). *Scarlet* and *Cinnabar* sgRNAs were transcribed following the manufacturer's protocols, however the results were inconsistent and cleavage products of poor quality as judged by electrophoresis. After multiple attempts with the HiScribe kit achieving varied results, the Engen kit (NEB, USA) was selected for optimisation. Benefits of the Engen kit include the annealing and transcription occurring in a single step reaction, taking under two hours. Fragments obtained from the Engen protocol were clearer when visualised via gel electrophoresis than those from the HiScribe protocol. Overall, the Engen transcription kit provided the most efficient *in vitro* sgRNA transcription in our hands.

Initial cleavage attempts using two different protocols (NEB protocol, and Martin-Martin *et al.,* 2018 protocol) were unclear, with the presence of several un-expected fragments, particularly with the *Cinnabar* DNA and when digested with the sgRNA. However, *Scarlet* and *Yellow* control fragments appeared at the expected size, and when digested with sgRNAs fragments were the expected size. In our hands, higher DNA concentrations and longer incubation times produced superior cleavage products via electrophoresis (Figure 4). Complete digestion was achieved with *Scarlet* sgRNA and incomplete digestion was achieved with the *Yellow* sgRNAs.

Finally, using the optimised annealing, transcription, and *in vitro* cleavage assay protocols, five olfactory gene targeted sgRNAs were produced and tested. Two sgRNAs (Gr2-1 and Gr2-2) for the gene *Gr2* and two sgRNAs (Ir8a-1 and Ir8a-2) for the *Ir8a* gene were analysed. The two sgRNAs for each gene were designed to target sequences 200bp apart in order to give a large deletion. One sgRNA (Orco-2) targeting the *Orco* gene was analysed. There was successful, yet incomplete cleavage of *Gr2* DNA with both sgRNAs (Figure 4).

Identification of Expression Promoters for use in Sand fly mutagenesis

Identification of expression promoters that function in sand flies is key for development of CRISPR approaches for modification of genes of interest, and further development of gene drives. Initial steps in transformation of insects often involves using fluorescent markers, and delivering these markers through transformation systems including PiggyBac and CRISPR based methods. To date few

promoters have been tested within sand fly cell lines via transfection experiments. Cell transfections were performed using LLE40 and LLE45 sand fly cell lines using a range of plasmids containing different promoters which had been shown to function in other insect species (Table 3).

Both LLE40 and LLE45 cell lines were successfully transfected using the pMaxGFP plasmid (Figure 5) The cell lines demonstrated numerous GFP fluorescent cells, with GFP expression observed in the cytoplasmic protrusions of fibroblastoid-like cells in LLE45 (Figure 5A). The GFP expression appeared more localised to epithelioid-like cells in LLE40 (Figure 23B). This result illustrates that the cytomegalovirus (*CMV*) promoter, from plasmid pMaxGFP, functions within sand fly cell lines.

Transfection with Ac5-STABLE1-Neo plasmid in LLE40 cells also demonstrated GFP expression. The majority of fluorescence was observed in cell clusters (Figure 6A and B). Positive GFP expressing cells were also observed in LLE45 (not shown).

A comparative transfection of LLE40 cells with the Ubiq-Cas9.874W (*D.melanogaster* Ubiquitin63E promoter for Cas9 and EGFP; baculovirus Opie2 promoter for DsRed1) PiggyBac plasmid and the helper transposase plasmids Hsp70, IhyPBase, and MhyPBase showed successful transfections, demonstrated by GFP expression, when the two hyperactive forms of the transposase were used (Figure *6*D and E), but not the helper Hsp70 plasmid (Figure *6*C). GFP was most evident in cell clusters. With the MhyPBase (mammal codon specific Hsp70) hyperactive transposase plasmid, the GFP expression was comparatively more intense than the insect codon specific Hsp70, when compared qualitatively by fluorescence microscopy. UbiqCas9.8774W transfections in LLE45 cells demonstrated very low levels of GFP fluorescence in individual cells (not shown) with the IhyPBase helper. Transfection with a second PiggyBac plasmid, pHome-T and IhyPBase helper transposase plasmid, also demonstrated positive fluorescence in LLE45 cells (Figure 7).

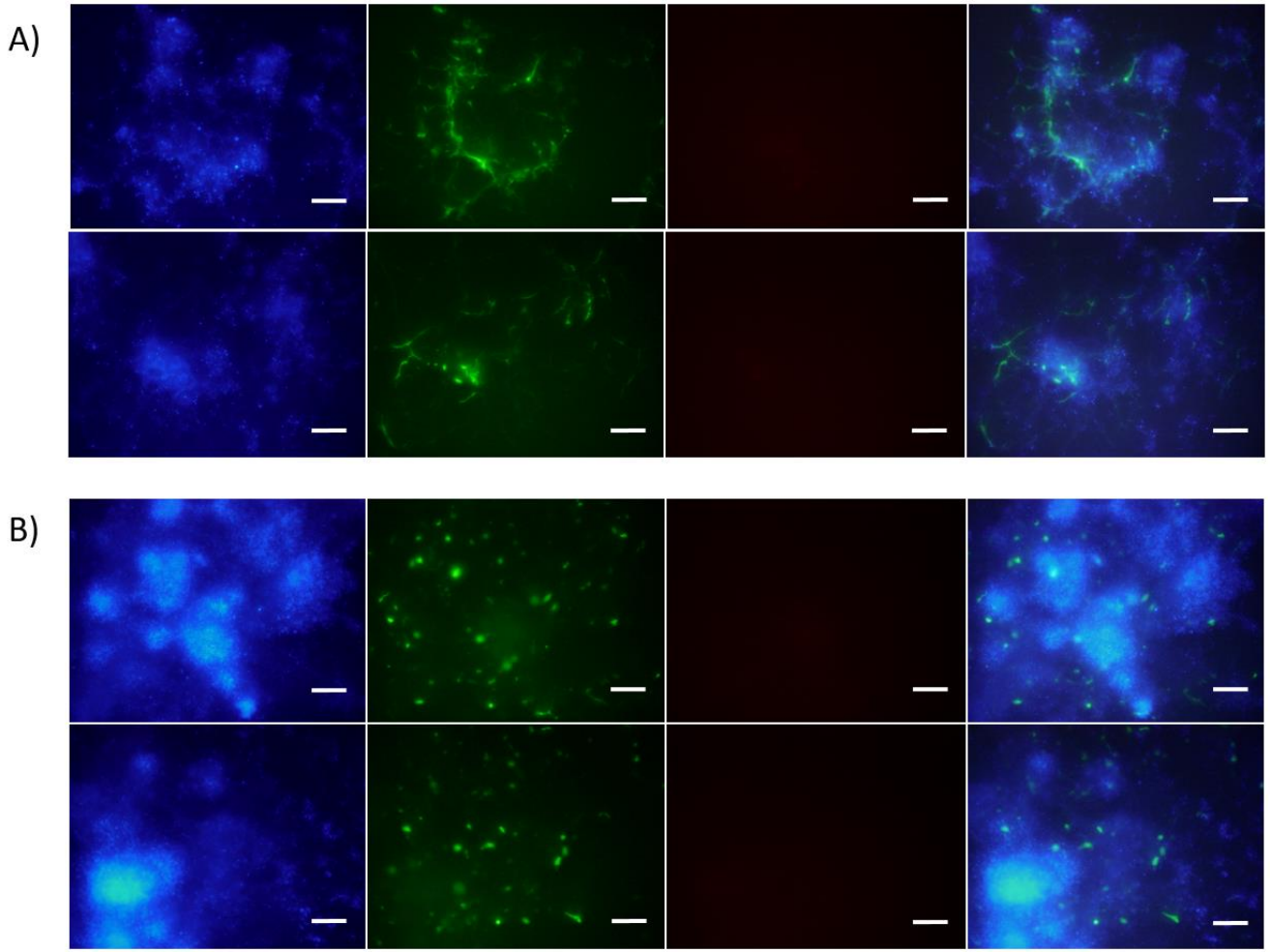


Figure 5. Sand fly cell lines transfected with pMaxGFP plasmid and Lipofectamine 3000. A) LLE40 cells, and (B) LLE45 cells. GFP expression is driven by the cytomegalovirus (CMV) promoter. Controls performed but not shown. Images were taken of live cells. From left to right, DAPI stain, GFP, RFP, and composite images. Images taken using Nikon ECLIPSE Ti2 inverted microscope. Scale bars 100µm

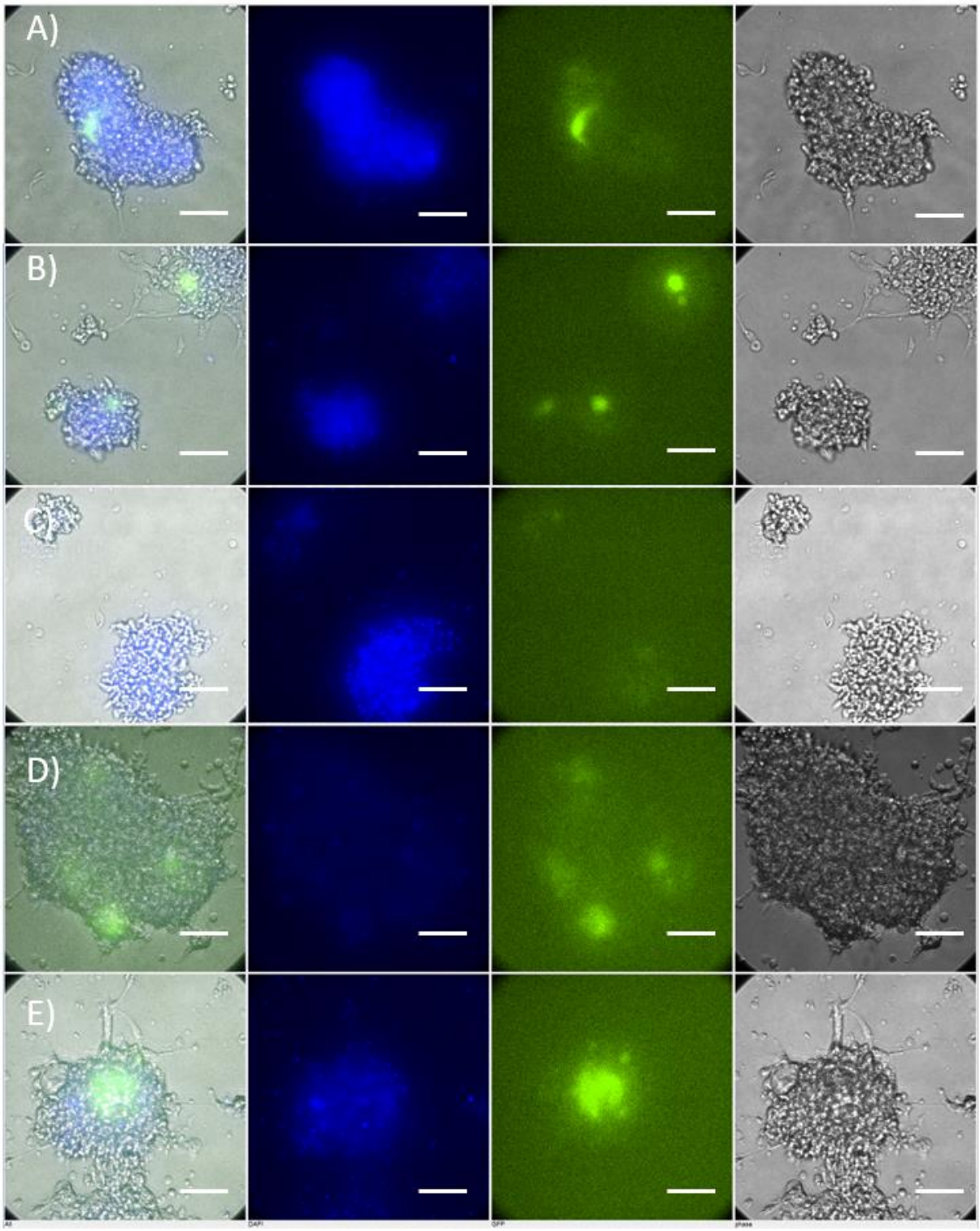


Figure 6. LLE40 sand fly cell transfections using Lipofectamine 3000. A) LLE40 cells transfected with Ac5-STABLE1-Neo, repeated in (B). C) Transfections with UbiqCas9.874w plasmid and Hsp70 transpose helpers, IhyPBase transposase (D), and MhyPBase transposase (E). Controls performed but not shown. Images were taken of live cells before flow cytometry. From left to right, composite images, DAPI stain, GFP, and bright-field. Images taken using Nikon ECLIPSE Ti2 inverted microscope. Scale bars 50µm.

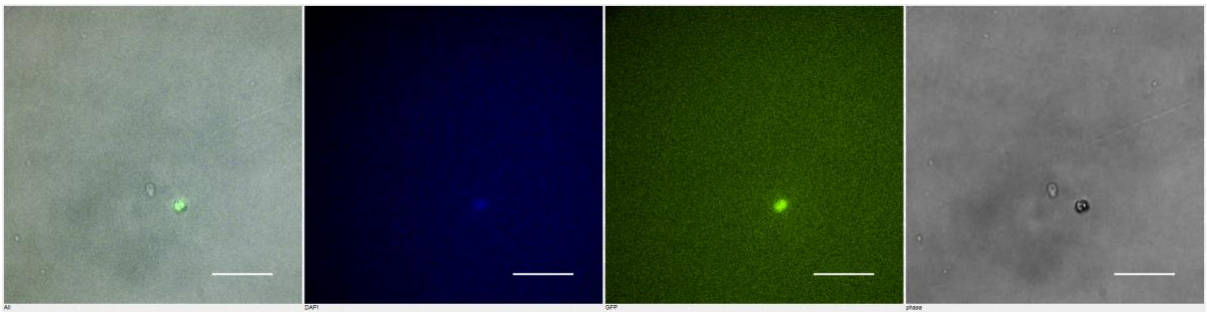


Figure 7. LLE45 cells transfected using Lipofectamine 3000. Transfections with pHome-T PiggyBac plasmid and IhyPBase transpose. Images were taken of live cells before flow cytometry. From left to right, composite images, DAPI stain, GFP, and bright-field. Images taken using Nikon ECLIPSE Ti2 inverted microscope. Scale bars 50µm.

Transfections with the pMaxGFP appeared to generate high levels of gene expression, when qualitatively visualised by microscopy, clearly demonstrating gene expression promoted by *CMV* across both cell lines. The Ac5-STABLE1-Neo also resulted in fluorescence, demonstrating that the *Actin5c D. melanogaster* promoter expresses within the sand fly cell lines. Expression was comparatively less in LLE45 cells by qualitative imaging (see discussion).

PiggyBac transfections conducted with UbiqCas9.874W and helper transposase plasmids demonstrate the *Ubiquitin 63e* promoter of Cas9 and GFP are able to express in sand fly cell lines, as well as the *hsp70* promoter of transposase, which is required for the intragenic insertion of PiggyBac plasmid. pHome-T transfection demonstrated positive expression of GFP driven by the synthetic *3xP3* promoter, at low frequency (low number of fluorescent cells, (Figure 7)). Importantly, this demonstrates that the *3xP3* promoter is possible candidate for use in sand flies.

In summary, Lipofectamine 3000 (Invitrogen) transfections resulted in successful expression of fluorescent using the manufacturer's protocol for transfection in two novel *L. longipalpis* cell lines. However transfection efficiencies appeared to be low (<1%) assessed visually, and protocols require further optimisation. Increasing transfection efficiencies using different approaches was subsequently undertaken (see results section *Comparison of Transfection Protocols applied to cells lines*). Importantly, function of multiple promoters (*Ac5*, *CMV*, *Ubiquitin 63e*, *hsp70*, and *3xP3*) was confirmed within sand fly cell lines (LLE40 and LLE45), which provides an important step towards demonstrating the utilization of these cells as part of a gene editing platform. We demonstrate that the *hsp70* promoter functions within the sand fly cell lines, which is important for the development of functioning PiggyBac and CRISPR tools. The CRISPR backbone plasmid used for the delivery of gRNAs targeting phenotypic and olfactory genes in this thesis (pDCC6) contains *hsp70* to promote Cas9, in addition to a *U6* promoter for the gRNAs within the constructs.

Quantitative comparison of transfection efficiency protocols in sand fly cells lines against insect controls

Transfections efficiencies in LLE40 and LLE45 cell lines using manufacturer's protocols for Lipofectamine 3000 was qualitatively low, typically less than 1 cell per field (40x magnification, Nikon ECLIPSE Ti2 inverted microscope). A range of transfection reagent were identified from the literature that had previously been used for transfections of insect cell lines (Cellfectin II, Effectene, Flyfectin, and FuGene HD). A comparison of transfections and optimisation of protocols were performed, to identify the reagents with high transfection efficiency in the context of CRISPR/PiggyBac components for *in vitro* assessment. Prior to comparison of the transfection agents selected, initial optimisation of stock protocols were required for Flyfectin and FuGene HD (see Methods, *Transfection reagents and protocols)*.

After optimisation, LLE45 cells were transfected with the Ac5-STABLE1-Neo plasmid following the optimised protocols for Flyfectin and FuGene HD. In general, all transfections had a low number of cells when visualised qualitatively by GFP fluorescent microscopy (<1 cell per field of view) (Figure 8).

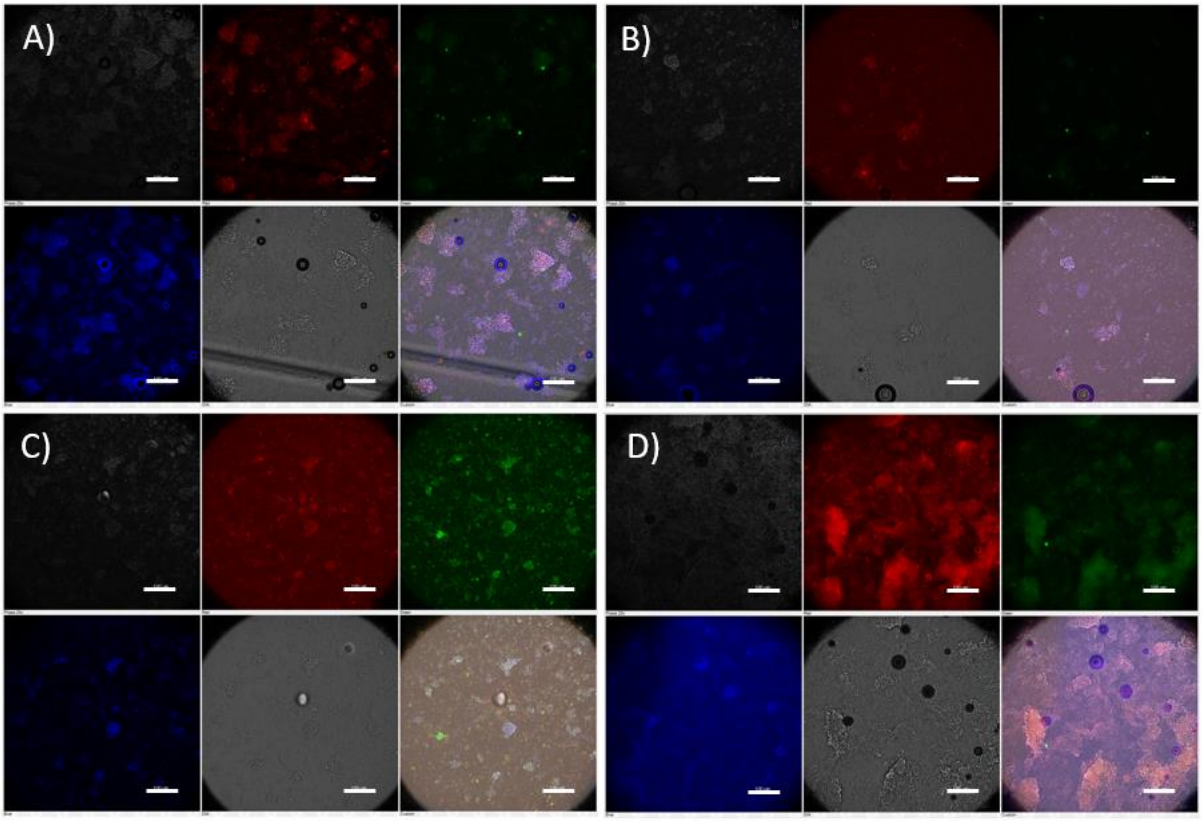


Figure 8. LLE45 cells transfected with Ac5-STABLE1-Neo plasmid using different transfection reagents. A) Cellfectin, B) Effectene, C) FlyFectin, and D) FuGene HD. Each panel of six images includes from left to right, phase 20x, RFP, and GFP (top), and DAPI DNA stain, diascopic, and overlay

composite image (bottom). Images taken using Nikon ECLIPSE Ti2 inverted microscope. Scale bars 100μm.

To assess transfection efficiencies quantitatively, flow cytometry was conducted (see methods, *Flow Cytometry*). To detect cell counts gates were not applied to the events detected by the flow cytometer, to ensure all cells were included within the analysis. Using a singlet gating strategy would have prevented recognition of fluorescent cells that often appeared in clumps in LLE45 cell lines.

For LLE45 cells, 102,797 - 103,558 events were recorded for each of the different transfection reagents. Approximately 50% of all event recorded occurred in the treatment (plasmid and reagents), and ~50% occurred in the controls (reagent only) (49.77 – 49.89%) (Table 12). For insect cell lines controls (DS2, Sf21, and An4a3B), fewer total events were recorded with each of the transfection reagents, compared to the sand fly cell lines (17,915 – 30,077; 21,895 – 31,087; 34,794 – 309,363, respectively). For these cell lines, the ratio of events recorded with plasmid and transfection reagent (treatments) and transfection reagent alone (controls), were often greater than 50.00% (49.00% for DS2 FuGene HD treatment, to 96.56% for An4a3B Cellfectin and FuGene HD treatments). This disparity in the total number of events between the cell lines was a result of the availability of cells. Importantly, the number of events recorded was robust enough for analysis.

All events were initially plotted on a forward-scatter-area (FSC-A) by side scatter area (SSC-A) plot, showing the size and internal complexity of the cells. Data was also plotted on a GFP/count histogram, with GFP- and GFP+ gates determined by control data, and applied to all treatment data. The histograms for all cell line reagent combinations show the majority of events fell below the fluorescence threshold value used to gate for GFP+ events (~10^3 on X-axis) (Figure 9, and Chapter 3 appendices 9-11). For LLE45, no GFP+ events were visible in the GFP+ gates when transfected with the Ac5-STABLE1-Neo plasmid, due to the count numbers being below the limit of detection (Table 12 and Table 13).

The FSC-A/SSC-A plots and GFP/count histograms do not clearly demonstrate the differences in transfection efficiencies between the different cell types and transfection reagents due to transfection event being rare, below the limit of detection. Statistical analysis of the data (Chi-squared test) were performed to overcome this with two objectives: (1) to determine if there is a statistically detectable difference in GFP expression between the control and the treatment for each type of transfection reagent, and (2) to compare the four different transfection reagents to determine which protocol has the highest transfection efficiency.

To address the first objective, a Chi-squared test was conducted with the total number of GFP+ events for each protocol, and GFP+ events in the control (where no plasmid was added to the transfections) (Table 13). P-values were statistically significant for all reagents used in transfections applied to DS2,

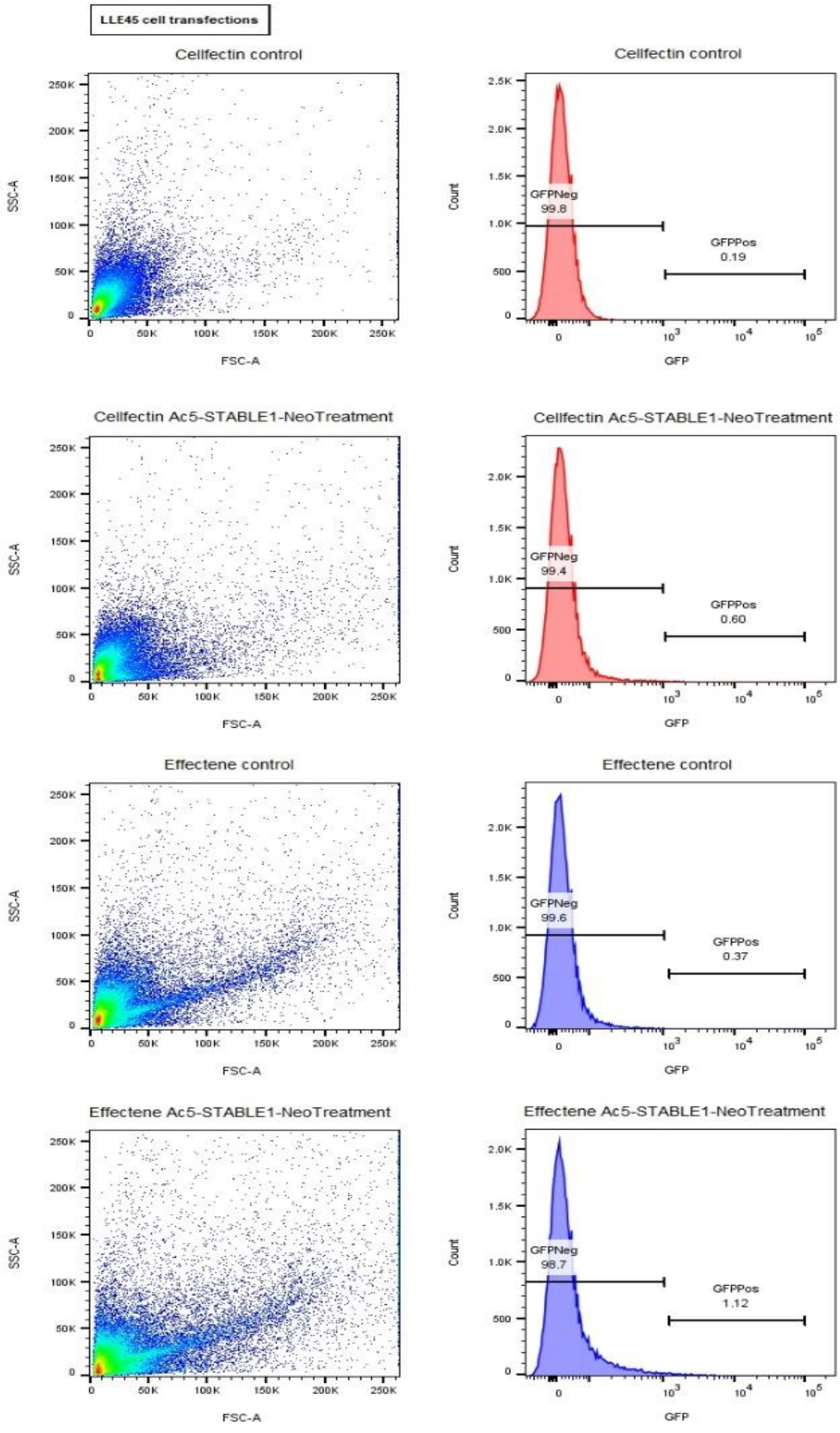
Sf21 and LLE45, evidenced by an increase in GFP expression in treatment samples. Effectene and Flyfectin did not show a significant difference in GFP expression in An4a3B cells (p=0.173 and 0.084, respectively), and surprisingly the Effectene control in Sf21 and FuGene HD control in An4a3B showed significance in GFP expression compared to the treatments. This is unexpected, however may be a result of insufficient cleaning of the flow cytometer.

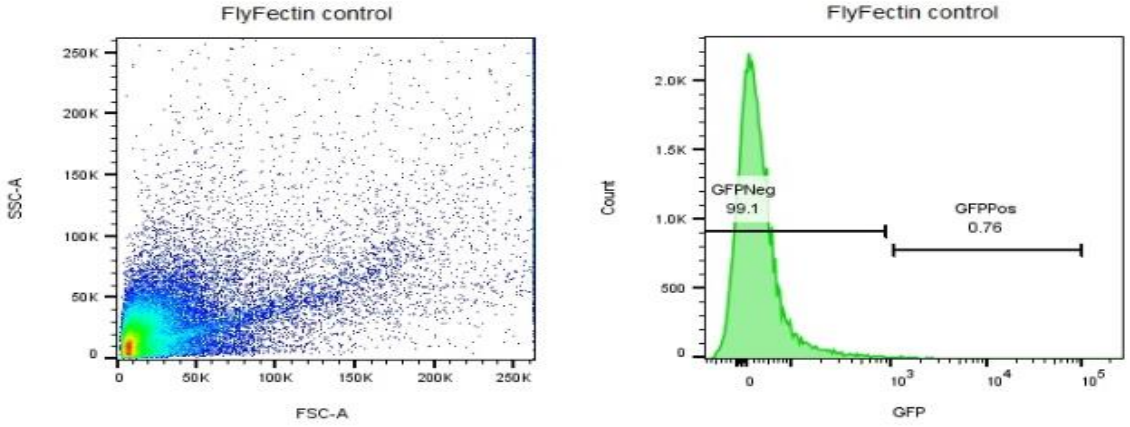
The overall comparison of transfection percentages (Chi-squared test) demonstrated that all of the transfection reagents had more GFP+ events when the transfection reagents were used, compared to controls without transfection reagents. However, this analysis is unable to show the magnitude of the difference in effectiveness between treatment and control, and cannot compare the effectiveness of the transfection reagents to one another (objective 2, above).

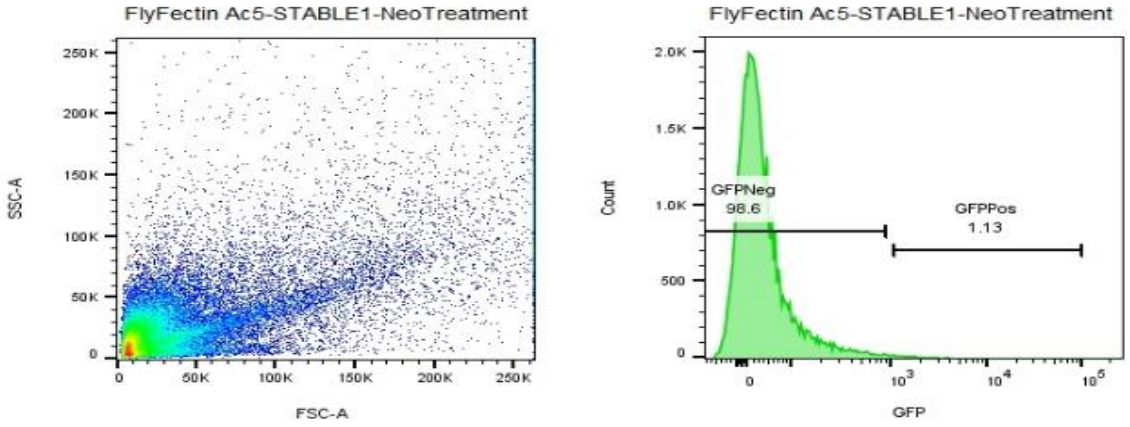
To address the second objective, the odds and odds ratio were calculated to determine the magnitude of the treatment response versus the control response. Odds are defined as *number of GFP+ / number of GFP-*, calculated independently for controls and treatments for each type of treatment. Odds values alone do not provide useful information in this context, however the ratio of the treatment odds over the control odds, can provide the magnitude metrics to answer objective two (above).

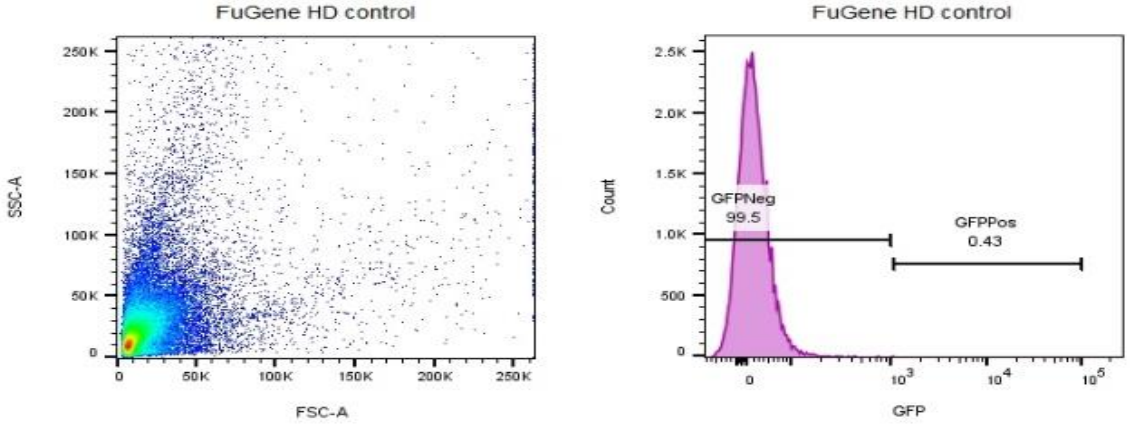
The odds ratios analysis demonstrated that different transfection reagents worked more successfully in different cell lines. The greatest odds of DS2 cells being GFP+ was 130.07 [48.49, 348.91] times higher when treated with Effectene compared to the control (

Table 14). In Sf21, the greatest odds was with Flyfectin (2.43 [2.03, 2.90]), in An4a3b the greatest odds was with Cellfectin (8.67 [1.21, 61.89]), and in LLE45 the greatest odds was with Cellfectin (3.08 [2.46, 3.86]) closely followed by Effectene (3.04 [2.59, 3.58]). Overall, these results indicates the importance of identifying optimal transfection reagents for the cell lines being studied. For the sand fly cell lines, Cellfectin and Effectene protocols result in the highest transfection efficiency, and should be taken forward for delivery of PiggyBac plasmids and CRISPR constructs.


LLE45 cell transfections
Cellfectin control
SSC-A
FSC-A
Count
GFP
GFPNeg
99.8
GFPPos
0.19
Cellfectin Ac5-STABLE1-NeoTreatment
GFPNeg
99.4
GFPPos
0.60
Effectene control
GFPNeg
99.6
GFPPos
0.37
Effectene Ac5-STABLE1-NeoTreatment
GFPNeg
98.7
GFPPos
1.12


FlyFectin control
SSC-A
FSC-A
FlyFectin control
Count
GFP
GFPNeg
99.1
GFPPos
0.76


FlyFectin Ac5-STABLE1-NeoTreatment
SSC-A
FSC-A
FlyFectin Ac5-STABLE1-NeoTreatment
Count
GFP
GFPNeg
98.6
GFPPos
1.13


FuGene HD control
SSC-A
FSC-A
FuGene HD control
Count
GFP
GFPNeg
99.5
GFPPos
0.43

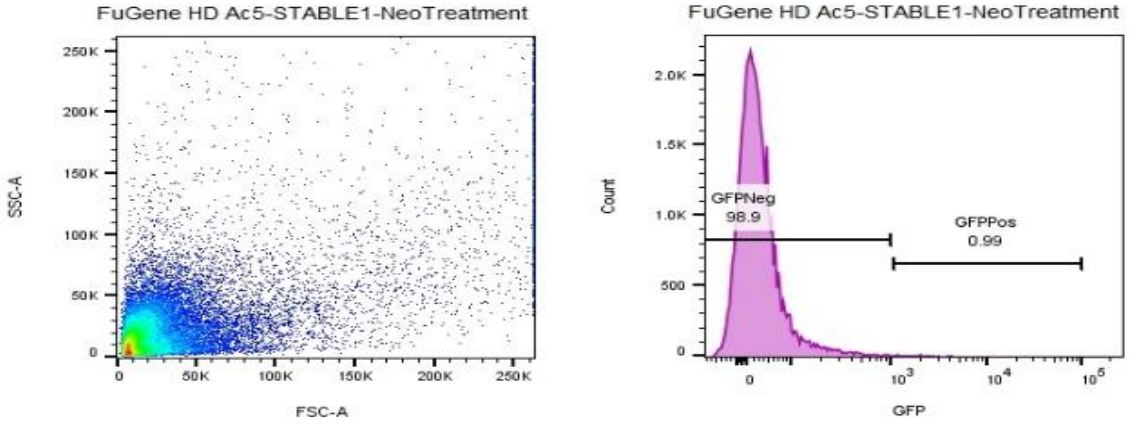

FuGene HD Ac5-STABLE1-NeoTreatment
SSC-A
FSC-A
FuGene HD Ac5-STABLE1-NeoTreatment
Count
GFP
GFPNeg
98.9
GFPPos
0.99

Figure 9. Flow cytometry of LLE45 cells transfected with Ac5-STABLE1-Neo. FSC-A/SSC-A plots are shown next to GFP/Count plots with GFP+ and GFP- gates applied. Controls plots are presented above treatment plots, in which the plasmid was used in transfection. Each histogram colour represents a different transfection reagent used. Cellfectin (red), Flyfectin (green), Effectene (blue), and FuGene HD (purple). Flow cytometry plots for DS2, Sf21, and An4a3B cell lines are presented in Chapter 3 appendices 9-11.

Table 12. Descriptive summary of Flow cytometry cell counts for transfection of insect cells with Ac5-STABLE1-Neo using Cellfectin, Effectene Flyfectin and FuGene HD transfection reagents. *Total samples* is the total number of cell event identified by flow cytometry, *Treatment* is the proportion of the cell events that were identified in treated samples (plasmid and transfection reagent), *Control* is the proportion of cell events identified (transfection reagent without plasmid).

Cell line	*Transfection Treatment*	*Total samples, N*	*Treatment, n (%)*	*Control, n (%)*
DS2	Cellfectin	30,077	20,050 (67.00)	10,027 (33.00)
	Effectene	19,965	12,013 (60.00)	7952 (40.00)
	Flyfectin	19,760	9661 (49.00)	10,099 (51.00)
	FuGene HD	17,915	9648 (54.00)	8267 (46.00)
Sf21	Cellfectin	31,087	15,270 (49.12)	15,817 (50.88)
	Effectene	28,213	15,024 (53.25)	13,189 (46.75)
	Flyfectin	21,895	14,375 (65.65)	7,520 (34.35)
	FuGene HD	28,285	17,118 (60.52)	11,167 (39.48)
An4a3B	Cellfectin	309,363	29663 (96.56)	1273 (3.44)
	Effectene	53,844	51596 (95.82)	2248 (4.18)
	Flyfectin	54,101	51662 (95.49)	2439 (4.51)
	FuGene HD	34,794	33598 (96.56)	1196 (3.44)
LLE45	Cellfectin	103,558	51662 (49.89)	51896 (50.11)
	Effectene	103,092	51306 (49.77)	51786 (50.23)
	Flyfectin	102,797	51265 (49.87)	51532 (50.13)
	FuGene HD	103,210	51453 (49.85)	51757 (50.15)

Table 13. Chi-squared analysis of GFP+ events from flow cytometric analysis between control and treatment samples using different transfection reagents.

Cell line	*Treatment type*	*Treatment GFP+, n (%)*	*Control GFP+, n (%)*	*Pearson's χ^2 test*	*p-value (p< 0.05)*
DS2	Cellfectin	462 (2.30)	3 (0.03)	221.84	<0.00001
	Effectene	302 (2.50)	4 (0.05)	193.66	<0.00001
	Flyfectin	10 (0.10)	2 (0.02)	5.70	0.016946
	FuGene HD	15 (0.20)	4 (0.05)	4.83	0.02821
Sf21	Cellfectin	389 (2.55)	256 (1.62)	33.71	<0.00001
	Effectene	126 (0.84)	170 (1.29)	13.74	0.00021
	Flyfectin	701 (4.88)	151 (2.01)	111.33	<0.00001
	FuGene HD	390 (2.28)	181 (1.62)	14.94	0.000111
An4a3B	Cellfectin	202 (0.68)	1 (0.078)	6.74	0.009407
	Effectene	298 (0.57)	8 (0.35)	1.86	0.173072
	Flyfectin	284 (0.55)	7 (0.29)	2.98	0.084307
	FuGene HD	28 (0.083)	8 (0.66)	38.03	<0.00001
LLE45	Cellfectin	310 (0.60)	101(0.19)	106.80	<0.00001
	Effectene	582 (1.12)	193 (0.37)	197.41	<0.00001
	Flyfectin	589 (1.13)	394 (0.76)	39.33	<0.00001
	FuGene HD	517 (0.99)	224 (0.43)	116.75	<0.00001

Table 14. Odds Ratio comparison of cell transfection reagents to determine which has highest transfection efficiency, determined by GFP expression in cells

Cell line	*Treatment type*	*Treatment GFP+, odds*	*Control GFP+, odds*	*Odds Ratio [95%CI]*	*p-value*
DS2	Cellfectin	462/20,050	3/10,024	76.99 [24.73, 239.66]	<0.0001
	Effectene	302/11,638	4/20,050	130.07 [48.49, 348.91]	<0.0001
	Flyfectin	10/9,646	2/10,094	5.23 [1.15, 23.89]	<0.0327
	FuGene HD	15/9,625	4/8,263	3.22 [1.07, 9.70]	<0.0378
Sf21	Cellfectin	389/15,270	256/15,817	1.57 [1.34, 1.85]	<0.0001
	Effectene	126/15,024	170/13,189	0.65 [0.51, 0.82]	0.0003
	Flyfectin	701/14,375	151/7,520	2.43 [2.03, 2.90]	<0.0001
	FuGene HD	390/17,118	181/11,167	1.41 [1.18, 1.68]	0.0002
An4a3B	Cellfectin	202/29,663	1/12,73	8.67 [1.21, 61.89]	0.0313
	Effectene	298/51,596	7/2,248	1.85 [0.88, 3.93]	0.1067
	Flyfectin	284/51,662	7/2,439	1.92 [0.90, 4.06]	0.0898
	FuGene HD	28/33,598	8/1,196	0.12 [0.06, 0.27]	<0.0001
LLE45	Cellfectin	310/51,662	101/51,896	3.08 [2.46, 3.86]	<0.0001
	Effectene	582/51,306	193/51,786	3.04 [2.59, 3.58]	<0.0001
	Flyfectin	589/51,265	394/51,532	1.50 [1.32, 1.71]	<0.0001
	FuGene HD	517/51,453	224/51,757	2.32 [1.98, 2.71]	<0.0001

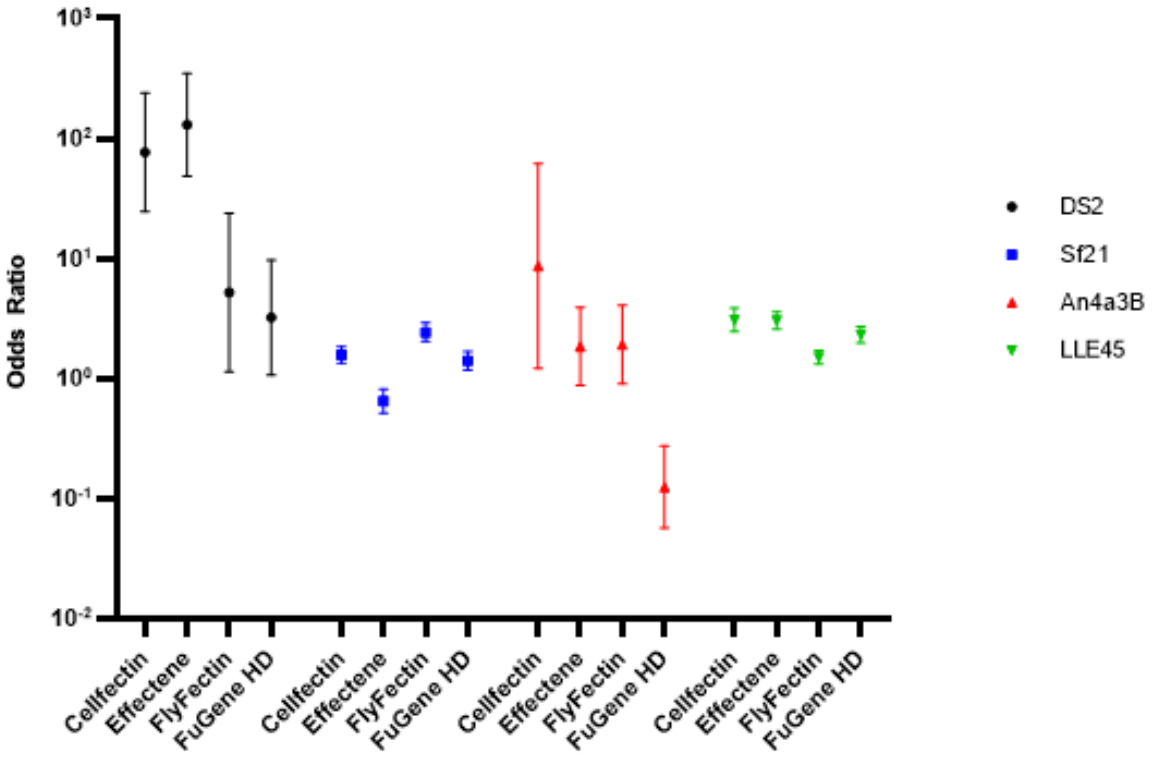


Figure 10. Odds ratio comparison of cell transfection reagents in four insect cell lines using the Ac5-STABLE1-Neo plasmid. Odds ratio value is represented by the shape symbol, with 95% Confidence Interval bars.

Discussion of comparison of transfection reagents and protocols

Initial attempts at cell transfections using Lipofectamine3000 (Invitrogen, USA) using stock manufacturer's protocols which demonstrated the ability to successfully transfect LLE40 and LLE45 cells, albeit at low transfection efficiencies. Transfections were attempted using four transfection reagents (Cellfectin, Flyfectin, Effectene, and FuGene HD) in two novel sand fly cell lines (LLE409 and LLE45) with optimised protocols, and in addition to commercially available insect cell lines (DS2, An4a3B, and Sf21). The protocols for transfections using these reagents were optimised.

Flow cytometry was used to quantitatively identify the presence of transfected (fluorescent) cells to compare transfection efficiencies. Overall, positive transfections were identified in all of the insect cell lines, and all of the reagents used successfully resulted in positive transfections with varying degrees of success. It was shown that the efficiency of transfections in the LLE45 cells were not dissimilar to the efficiency of transfections in *Anopheles* and *Drosophila* cell lines, in our hands. Additionally, Cellfectin and Effectene were identified as the optimal transfection reagents for the LLE45 cells, however, overall efficiencies were still relatively low (odds ratio of 3.08 and 3.04, respectively).

The *Lutzomyia* cell lines were obtained through the Tick Cell biobank, newly derived from embryos, containing differentiated mixtures of cell types (Bell-Sakyi et al., 2018). Time permitting fuller characterisation and sub-cloning of LLE45 and LLE40 cell lines would be beneficial to produce a less differentiated cell line. This would have been performed in house, however COVID had an impact on timelines, preventing this from being pursued further.

The use of alternative insect cell lines that have been fully characterised is beneficial when determining whether generic constructs and promoters function as expected. This is the case for all of the constructs tested in this chapter. This could provide evidence of functioning elements across multiple insect orders (including Diptera), increasing the likelihood that the constructs would function in sand fly cell line, or *in vivo*, if confirmed within alternative cell lines.

Given more time, further optimisation of the sand fly cell line transfections would be hugely beneficial (to increase transfection efficiencies), particularly regarding characterisation of sand fly cell lines for testing constructs with species-specific promoters (such as U6) and gRNAs.

Generation of transgenic Cas9-expressing sand fly cell line

Previously we demonstrated the identification and rationalisation of gene targets to affect genomic modification (Chapter 2), subsequent transcription and *in vitro* testing (via cleavage assays) of gRNAs, and identification of promotors that can drive expression of exogenous fluorescent cargo, Cas9 and gRNAs. A additional useful tool to rapidly assess CRISPR mediated gRNAs *in vitro* would be the derivation of a Cas9-expressing sand fly cell line. Here we attempt to create a Cas9-expressing cell line by application of a modified PiggyBac plasmid (UbiqCas9.874W backbone) using the protocols described below.

Briefly, the UbiqCas9.874W PiggyBac plasmid was modified to include a neomycin resistance marker gene excised from UbiqCas9-NeoR. Sand fly cell lines were transfected (as previously described) with this newly constructed plasmid with required helper transposase plasmids, followed by an antibiotic (G418) selection assay.

A protocol was developed based on *Drosophila* cells (González et al., 2011). Transfections were conducted using a newly constructed Cas9 plasmid containing the neomycin resistance gene excised from the Ac5-STABLE1-Neo plasmid. An Alamar Blue assay was also conducted in parallel to determine *in vitro* cytotoxicity, with the aim of selecting cells expressing the Cas9. Detailed methods are described in the following sections.

Cas9 plasmid construction containing NeoR marker

A Gibson Assembly methodology (as described previously) was followed to construct a Cas9 expressing plasmid containing a neomycin resistance marker (NeoR) from the Ac5-Stable-Neo plasmid (Addgene, Plasmid #32425) conferring resistance to the antibiotic G418.

The Neomycin resistance marker (including the T2A ribosome skipping element) was selected for insertion directly after the EGFP tag in the Ubiq-Cas9.874W plasmid. Successful integration of *Ubiquitin 63e* promoter would initiate transcription for and expression of the Cas9 protein, which is tagged with EGFP, in addition to the NeoR. GFP fluorescence would indicate the *Ubiquitin 63e* promoter functions, and that the Cas9 and NeoR are expressed.

The Gibson Assembly (see methods) initially involved two PCR reactions to linearise the vector plasmid (UbiqCas9.874W), and to amplify the Neomycin resistance fragment from the donor plasmid (Ac5-STABLE1-Neo) (Figure 11). Primers were designed for linearisation of the vector UbiqCa9.874w plasmid, and for excision of the T2A-NeoR fragment, from the Ac5-STABLE1-Neo (Table 15). The two PCRs to digest the vector and to amplify the insert were run in tandem as shown in Figure 11.

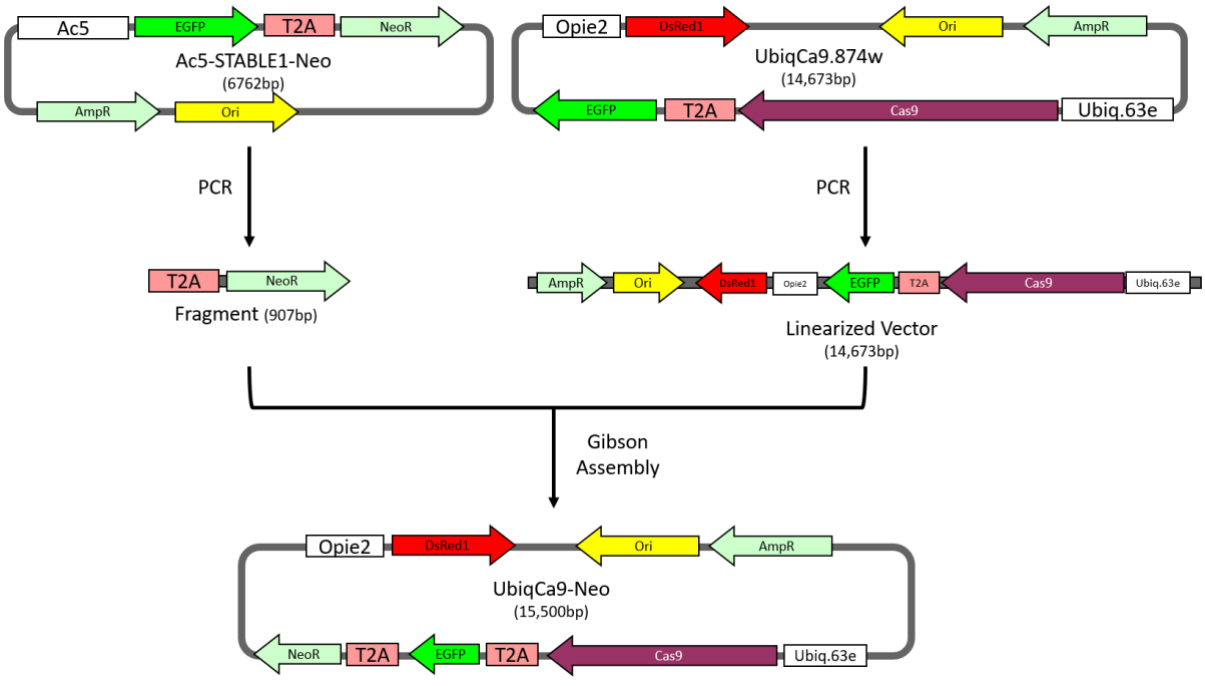


Figure 11. Gibson assembly of UbiqCas9-NeoR PiggyBac plasmid. The T2A-NeoR fragment is amplified from the Ac5-STABLE1-Neo plasmid, and the UbiqCas9.874W plasmid linearised. A Gibson assembly inserts the T2A-NeoR fragment in to the UbiqCas9.874W.

Table 15. Primer sequences for Gibson Assembly of UbiqCas9-NeoR PiggyBac plasmid. Primers were designed using SnapGene software (www.snapgene.com).

Primer	*Sequence*	*Notes*
Confirmation fwd	GCATGGACGAGCTGTACAAG	Primer to confirm the presence of the NeoR insert into the UbiqCas9-NeoR construct.
Confirmation rev	GATTCATTCTAGTTAATTAA	Primer to confirm the presence of the NeoR insert into the new construct.
NeoR_Frag_For_Pri	GCATGGACGAGCTGTACAAGCTTGAGGGCAGA	Primer to amplify Neomycin Fragment from the Ac5-STABLE-Neo plasmid.
NeoR_Frag_Rev_Pri	GATTCATTCTAGTTAATTAATCAGAAGAACTCGTCAAGAAGGC	Primer to amplify Neomycin Fragment from the Ac5-STABLE-Neo plasmid.
UbiqCas9_Vec_For_Pri	TTCTTGACGAGTTCTTCTGATTAATTAACTAGAATGAATCGTTTTTAAAATAACAAATCAATTGT	Primer to linearise UbiqCas9.974W vector to receive of fragment.
UbiqCas9_Vec_Rev_Pri	AGACTTCCTCTGCCCTCAAGCTTGTACAGCTCGTCCATGCCGAG	Primer to linearise UbiqCas9.974W vector to receive of fragment.

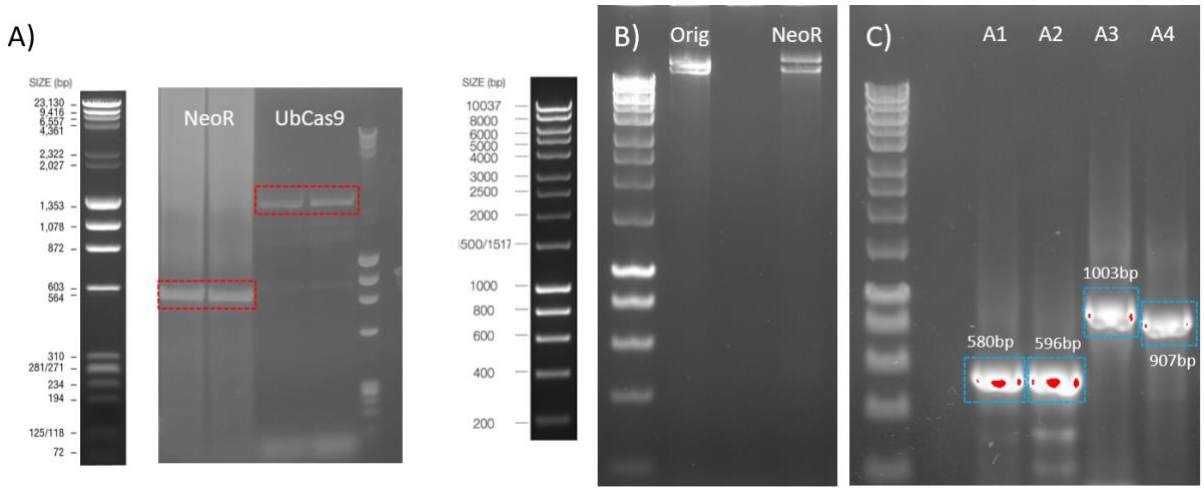


Figure 12. Construction of UbiqCas9-NeoR PiggyBac plasmid via Gibson Assembly. A) NeoR fragment from Ac5-STABLE-Neo plasmid, expected size 907bp. Vector plasmid UbiqCas9.84W, expected size 14,673bp. B) Comparison of Original UbiqCas9 plasmid (Orig, 14,636bp), and UbiqCas9-NeoR plasmid (NeoR , 15,500bp) constructed via Gibson Assembly. C) Plasmid construct confirmation PCR for UbiqCas9-NeoR construct. Confirmation primer pairs A1, A2, A3, and A4 (see Table 16) were used to determine the presence of the NeoR fragment within the newly constructed plasmid, UbiqCas9-NeoR.

The PCR products were visualised via gel electrophoresis. The expected size of the NeoR fragment was ~907bp, and 14,673bp for the linearised UbiqCas9.874w vector plasmid (Figure 12A). In initial attempts, the NeoR fragment was the correct size, however, artefacts were visible in the NeoR fragment

lanes. Two bands were visible between 2,322bp and 4,361bp for the linearised UbiqCas9.874W vector, raher than 14,673bp (Figure 12A). The result were inconclusive, therefore the PCR was optimised. After optimisation, the Gibson assembly derived the UbiqCas9-NeoR plasmid. Gel electrophoresis compared the vector UbiqCas9.874w plasmid (14,636bp) and the UbiqCas9-NeoR plasmid (15,500bp) (Figure 12B), confirming the presence of the Neomycin resistance marker (Table 16).

Table 16. Confirmation primers for NeoR insertion into the UbiqCas9.874W vector to generate the UbiqCas9-NeoR plasmid construct.

Primer name	*Primer sequence*	*Expected fragment size (bp)*	*Primer pair label*
Neo confirm 1	GTCCTGCTGGAGTTCGTGAC	580	A1
Neo confirm 3	CTTCCATCCGAGTACGTGCT		
Neo confirm 2	GCTATTGGGCGAAGTGCCGG	596	A2
Neo confirm 4	CATAATCAAAGAATCGTACG		
Neo confirm 1	GTCCTGCTGGAGTTCGTGAC	1003	A3
Neo confirm 4	CATAATCAAAGAATCGTACG		
Conf_NeoR_Fwd	GCATGGACGAGCTGTACAAG	907	A4
Conf_NeoR_Rev	GATTCATTCTAGTTAATTAA		

Bacterial transformation was conducted with the putative UbiqCas9-NeoR construct, and cells plated as described previously (see Bacterial Transformation for Plasmid amplification in Method section). Colonies were collected, grown, and purified via minprep before PCR confirmation was attempted via gel electrophoresis. Neomycin confirmation primers produced amplicons of expected size (580bp and 596bp, A1 and A2), with a smaller fragments visible in both lanes (~300bp and ~200bp) (Figure 12C). With Neomycin confirmation primer pairs (A3 and A4), amplicons produced were slightly lower than expected (1003bp and 907bp, respectively) (Figure 12C). These were promising results, however further confirmation was required via Sanger sequencing.

PCR products A1 and A2 were sequenced (Figure 12C) using the confirmation primer pairs (Table 16). Sequences demonstrated successful insertion of the NeoR, with some mutations (Figure 31). A base change occurred in the third amino acid of the NeoR gene, however, this was a silent mutation from *TCG* (Serine) to *TCA* (Serine). A base change also occurred six amino acids from the end of the NeoR gene from *CTT* (Leucine) to *ATT* (Isoleucine).

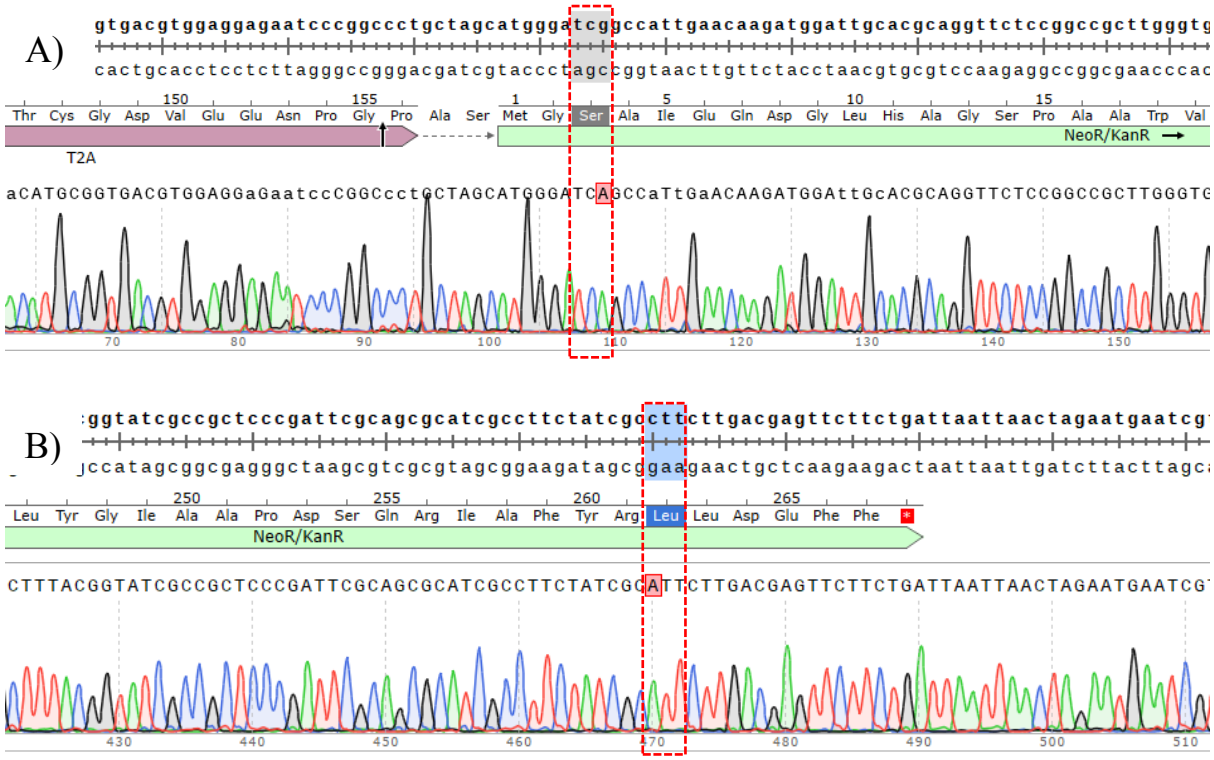


Figure 13. UbiqCas9-NeoR sequence confirmation (NeoR-confirm primer pairs). Mutations in the NeoR gene are highlighted (red boxes). A silent mutation from TCG (Ser) to TCA (Ser) occurs at the third amino acid in the NeoR gene (A), and a mutation from CTT (Leu) to ATT (Ile) at the 261st amino acid (B).

A final confirmation of the UbiqCas9-NeoR structure was performed by restriction digestion. SnapGene software (www.snapgene.com) was used to identify enzymes with a single cut site within the plasmid, and to predict the product sizes from the digestion. (15,500bp for UbiqCas9-NeoR). A simple digest protocol was followed and the product visualised on gel to determine the size of the fragments and size of the assembled plasmid construct. Briefly, a reaction mixture consisting of UbiqCas9-NeoR plasmid construct (1000ng), 5µL CutSmart buffer (NEB, USA), 1 µL *NcoI-HF* (20,000 units/mL,NEB), 1µL *NotI-HF* (20,000 units/mL,NEB), and H_2O upto a final volume of 50µL was incubated at 37°C for 15 minutes. The reaction product was then run on a 1% gel for 1 hour 45 at 90V and 140 mA.

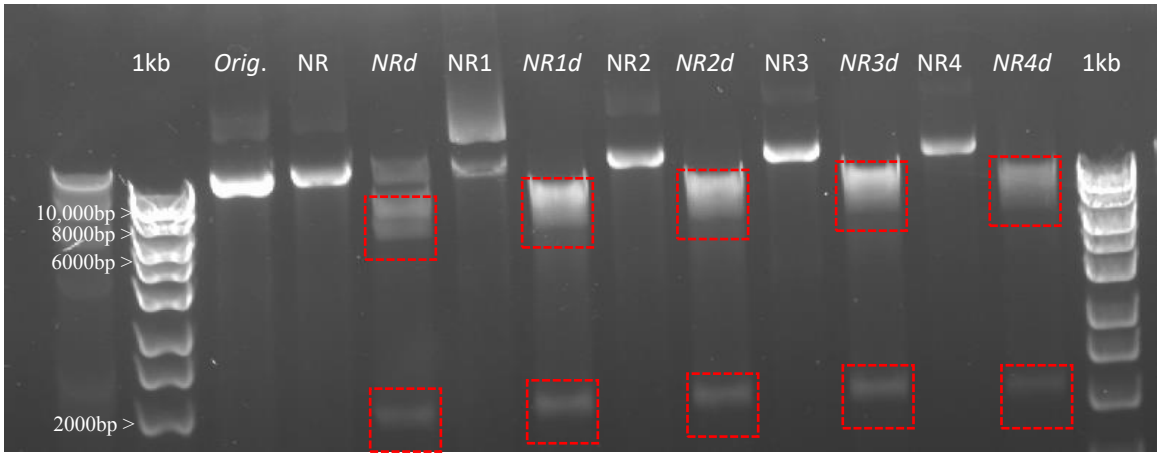


Figure 14. Enzyme digest confirmation of UbiqCas9-NeoR plasmids constructed using Gibson Assembly. Nco1 and NotI digest resulted in expected fragment sizes (7,552bp, 6,058bp, and 1890bp). The original UbiqCas9.874W plasmid (*Orig*) was run alongside the newly constructed UbiqCas9-NeoR plasmids (*NR, NR1, NR2, NR3, NR4*), *lanes 4, 6, 8, 10, 12*, and the digests of these newly constructed plasmids (*NRd*, *NR1d, NR2d, NR3d, NR4d)*, *lanes 5, 7, 9, 11, 13*.

The UbiqCas9.874w plasmid (14,636bp) was run as a control, alongside the undigested UbiqCas9-NeoR construct, and the enzyme-digested UbiqCas9-NeoR constructs (Figure 32). Plasmid constructs appeared to be larger in size than the original plasmid as expected due to the insert of the T2A-NeoR fragment. The undigested plasmid constructs appear the same size, and the digested versions of these constructs show three bands, which correspond to the expected sizes (7552bp, 6058bp, 1890bp). There is some evidence of incomplete plasmid digestion (Figure 32) remaining in the digested NRd sample, resulting in an extra band.

In summary a modified PiggyBac plamid (UbiqCas9-NeoR) was derived, containing a nemycin resistance marker (NeoR), which was inserted into the vector plasmid (UbiqCas9.874W. Structure was confirmed by restriction digestion, PCR, and sequencing.

G418 selection of transgenic Cas9-expressing cells

Transfections were subsequently conducted with the confirmed UbiqCas9-NeoR plasmid, with previous evidence of integration observed via red fluorescence from the DsRed1 PiggyBac integration marker. (not shown). Although transfection efficiency appeared to be low, selection for transfected cells expressing a neomycin resistance marker could allow the development of a Cas9 expressing cell line. Multiple methods are available for isolation of specific cells and generation of clonal cell populations, including serial dilution of cells (Nilsen & Castellino, 1999), and fluorescence-activated cell sorting (FACS) (Feng, López Del Amo, et al., 2021; Liu et al., 2020b; Viswanatha et al., 2021). Another method for isolating cells is via drug selection.(González et al., 2011).

The aminoglycoside antibiotic G418 was used as a selective agent. A range of G418 concentrations (0, 600, and 2000 µg/mL) were used to select for S2R+ and Kc167 *Drosophila* cells that had been transfected with the plasmid containing a NeoR resistance marker (González et al., 2011). Few control cells survived G418 selection over a 30 day period, however cells expressing GFP made up 68% of all survivors with the highest concentration of G418 (2000µg/mL G418).

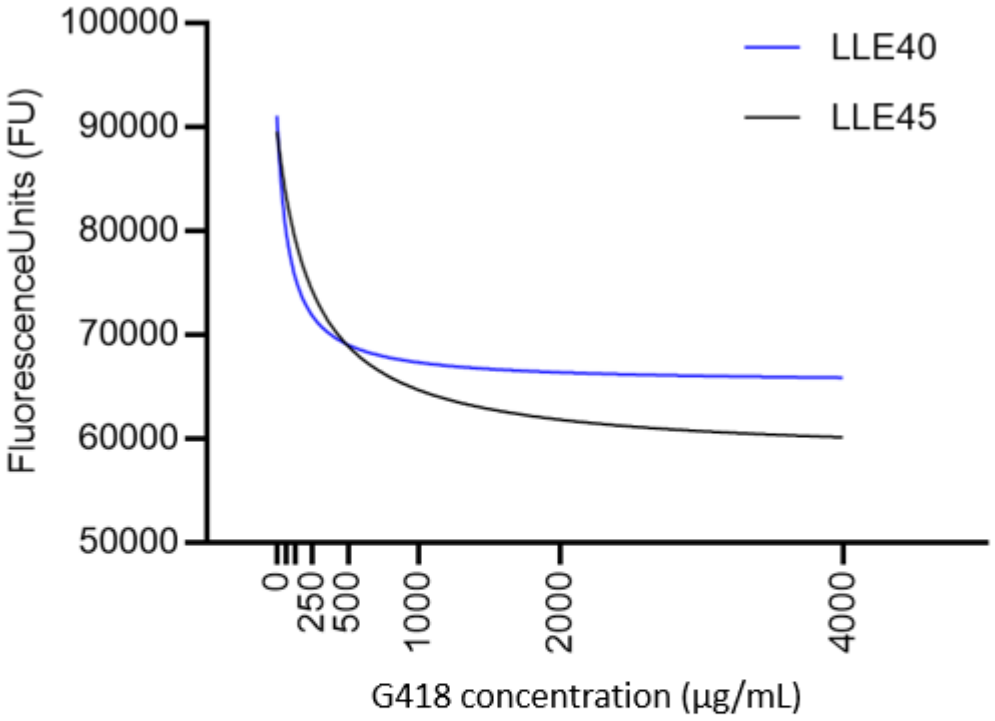
This study formed the basis of a protocol to determine the G418 concentrations required to select for the LLE40 and LLE45 cell successfully transfected with the UbiqCas9-NeoR plasmid. Initially, an Alamar Blue (AB) cytotoxicity assay was performed to identify drug concentrations that kill ~50% of un-transfected sand fly cells (IC50), as described below.

Seven G418 concentrations were selected (0, 62.5, 125, 250, 500, 1000, 2000, and 4000µg/mL) covering the range of concentrations used in previous *Drosophila* cell assays. The highest concentration (4000µg/mL) was selected as the previous study observed 2000µg/mL of G418 left ~32% un-transfected S2R+ and Kc167 cells to survive.

The AB assay was conducted as previously described (see Methods, *Alamar blue*), and data was collected at 48 hours via the SpectraMax iD5 plate reader. AB reagent contains a blue fluorescent dye, Resazurin, which is both non-toxic and able to penetrate cells. The dye is metabolised by cells, resulting in a pink product, Reofurin, which is highly fluorescent. The intensity of fluorescence in the assay is proportional to the number of cells respiring within the wells. A dose-response curve using non-linear regression models was used to analyse the data (see see Methods, *Alamar blue*).

As expected, the fluorescence intensity decreased with increasing concentration of antibiotic in both LLE40 and LLE45 cells (Figure 15A). The baseline fluorescence in LLE40 (65298 [62494, 67926] FU) was higher than the baseline in LLE45 (58179 [52437, 62487] FU). The IC50 (G418 concentration at the midpoint between the bottom and top fluorescence values) was comparatively lower for LLE40 (84.15 [46.02, 150.6] µg/mL) than to LLE45 (259.9[128.3, 593.8] µg/mL), indicating increased sensitivity to G418. The IC50 concentration for LLE45 was not dissimilar to G418 concentrations used to select *Ae*. *albopictus* C6/36 cells (400µg/mL) (Cano-Monreal, Williams, & Heidner, 2010) and Sf9 cells (300µg/mL) (Vidigal et al., 2018), however the concentration was significantly lower for the LLE40 cells. Both sand fly cell lines supported robust had large confidence intervals for the IC50 values (Figure 15).

A)



B)

Best-fit values	*LLE40 [95% CI]*	*LLE45 [95% CI]*
Bottom (FU)	65298 [62494.00, 67926.00]	58179 [52437.00,62487.00]
Top (FU)	91044 [86908.00, 95256.00]	89540 [85025.00, 94304.00]
IC50 (µg/ml)	84.15 [46.02, 150.60]	259.9 [128.30, 593.80]

Figure 15. G418 Dose-Response curve for LLE40 and LLE45 cells after 48 hours (A), with non-linear regression analysis results (B).

The AB assay conducted initially provided control data for the G418 assay. Subsequently, UbiqCas9-NeoR plasmid (with IhyPBase transposase) was transfected as previously described (see Methods, *G418 selection assay*, and *Alamar blue*). No viable cells were observed by the 30-day end point. Possible reasons are: 1) The AB assay was conducted after 48 hours treatment with G418, demonstrating a large knockdown observed with G418 concentrations below 500µg/mL. Therefore, the extended time (30 days) used in the selection assay may have been too long to allow any cells to survive. Additionally the transfection efficiency may have been low, as previously observed in these cell lines when using Lipofectamine 3000 with several plasmids (see above).

To optimise the development of a Cas9- expressing sand fly cell line, a more comprehensive assessment of the cytotoxicity of the cell lines is required, including a greater number of G418 concentrations between 0-500µg/mL to identify an accurate IC50. In addition, more time points for assessment of the cytotoxicity assay would indicate more accurately the optimum time for G418 selection pressure to be applied. Following this, a selection assay with cells transfected with the UbiqCas9-NeoR plasmid could be performed alongside FACS to isolate Cas9-expressing cells.

Overall, a Cas9-expressing plasmid (UbiqCas9-NeoR) was successfully constructed containing the NeoR marker for antibiotic selection. Plasmid structure was confirmed by sequencing, PCR and an restriction digestion assay, and confirmed by transfection experiments showing red fluorescence indicating PiggyBac integration. G418 antibiotic selection was attempted, however we was unable to successfully isolate cells that had been transfected with the plasmid, as the length of the selection pressure was inappropriate for the sand fly cell lines. The preliminary data generated provides a robust concentration assay of G418 antibiotic for use in further selection studies. One key limiting factor to the generation of this Cas9 cell line was the low transfection efficiency when using Lipofectamine 3000. Alternative transfection reagents may provide a solution to generate higher transfection efficiency, simplifying the process of cell isolation either via G418 drug selection or FACS.

Construction of CRISPR knockout plasmids targeting phenotypic and olfactory genes

Plasmids are widely used to introduce DNA into cells and to express genes of interest. Previous studies have demonstrated that using a single plasmid containing both the Cas9 and an sgRNA to target specific gene provides an efficient delivery method with increased modification efficiency compared to *in vitro* transcribed sgRNAs with recombinant Cas9 protein (Cong et al., 2013; Wang et al., 2013).

The pDCC6 plasmid (see Methods, Table 3) was selected as the basis for CRISPR-Cas9 knockout constructs due to its previous use in several insect orders (Gokcezade et al., 2014). The plasmid is bicistronic, with two promoters within the plasmid that have functioned effectively within insects; hsp70 drives expression of Cas9, and U6 expresses sgRNAs. In particular the hsp70 promoter has also been applied to sand fly cell lines to successfully drive luciferase expression (Saraiva et al., 2009), and was demonstrated in this study (see above) for LLE40 and LLE45 cell lines. This provides a degree of confidence that the pDCC6 plasmid will be functional in sand fly species, and could be an effective method to deliver Cas9 and gRNAs concurrently.

However, the U6-2 promoter (U6:96Ab) in the pDCC6 plasmid is *Drosophila*-specific. Sand fly specific U6 promoters were selected by identification of a TATA box and Pol III PSEA sequence (400bp), upstream of the transcription start codon for a U6 gene (LLOJ010426), and the pDCC6 plasmid was modified to include these to promote gRNAs (Figure 16). Species-specific primers were used to amplify the U6 promoter from the *L. longipalpis* and *P. papatasi* genome, and to linearise the vector

plasmid, followed by the Gibson assembly reaction (described in the methods) (Table 17). One *L. longipalpis*, and two *P. papatasi* pDCC6 plasmid backbones were successfully constructed (Llon1-U6-pDCC6, Ppap1-U6-pDCC6, and Ppap2-U6-pDCC6 respectively).

Table 17. Primers for construction of pDCC6 plasmids containing species-specific U6 promoter sequences.

Primer	*Sequence*
Llon1 U6 fwd	GGCAACTCGTGAAAGGTAGGCGGATCAGCGATTGAAGTGACAATTGAATATCCAACGGTT
Llon1 U6 rev	TAGCTCTAAAACAGGTCTTCTCGAAGACCCAATTCCATGGCAAATCTATTTTCCTTATAA
Ppap1 U6 fwd	GGCAACTCGTGAAAGGTAGGCGGATCAGCGTGTGATATCCCGTGGGCCAAATTTGAAATG
Ppap1 U6 rev	TAGCTCTAAAACAGGTCTTCTCGAAGACCCAAATGCATAGAAATCGAATTGATATATGAA
Ppap2 U6 fwd	GGCAACTCGTGAAAGGTAGGCGGATCAGCGCAAATCGGGTGAAAATCGGGTGAAAAATTT
Ppap2 U6 rev	TAGCTCTAAAACAGGTCTTCTCGAAGACCCAACTTTAACACATTTTAGGTGTAAAAATAT
Llon1 pDCC6 fwd	AACCGTTGGATATTCAATTGTCACTTCAATCGCTGATCCGCCTACCTTTCACGAGTTGCC
Llon1 pDCC6 rev	TTATAAGGAAAATAGATTTGCCATGGAATTGGGTCTTCGAGAAGACCTGTTTTAGAGCT
Ppap1 pDCC6 fwd	CATTTCAAATTTGGCCCACGGGATATCACACGCTGATCCGCCTACCTTTCACGAGTTGCC
Ppap1 pDCC6 rev	TTCATATATCAATTCGATTTCTATGCATTTGGGTCTTCGAGAAGACCTGTTTTAGAGCTA
Ppap2 pDCC6 fwd	AAATTTTTCACCCGATTTTCACCCGATTTGCGCTGATCCGCCTACCTTTCACGAGTTGCC
Ppap2 pDCC6 rev	ATATTTTTACACCTAAAATGTGTTAAAGTTGGGTCTTCGAGAAGACCTGTTTTAGAGCTA

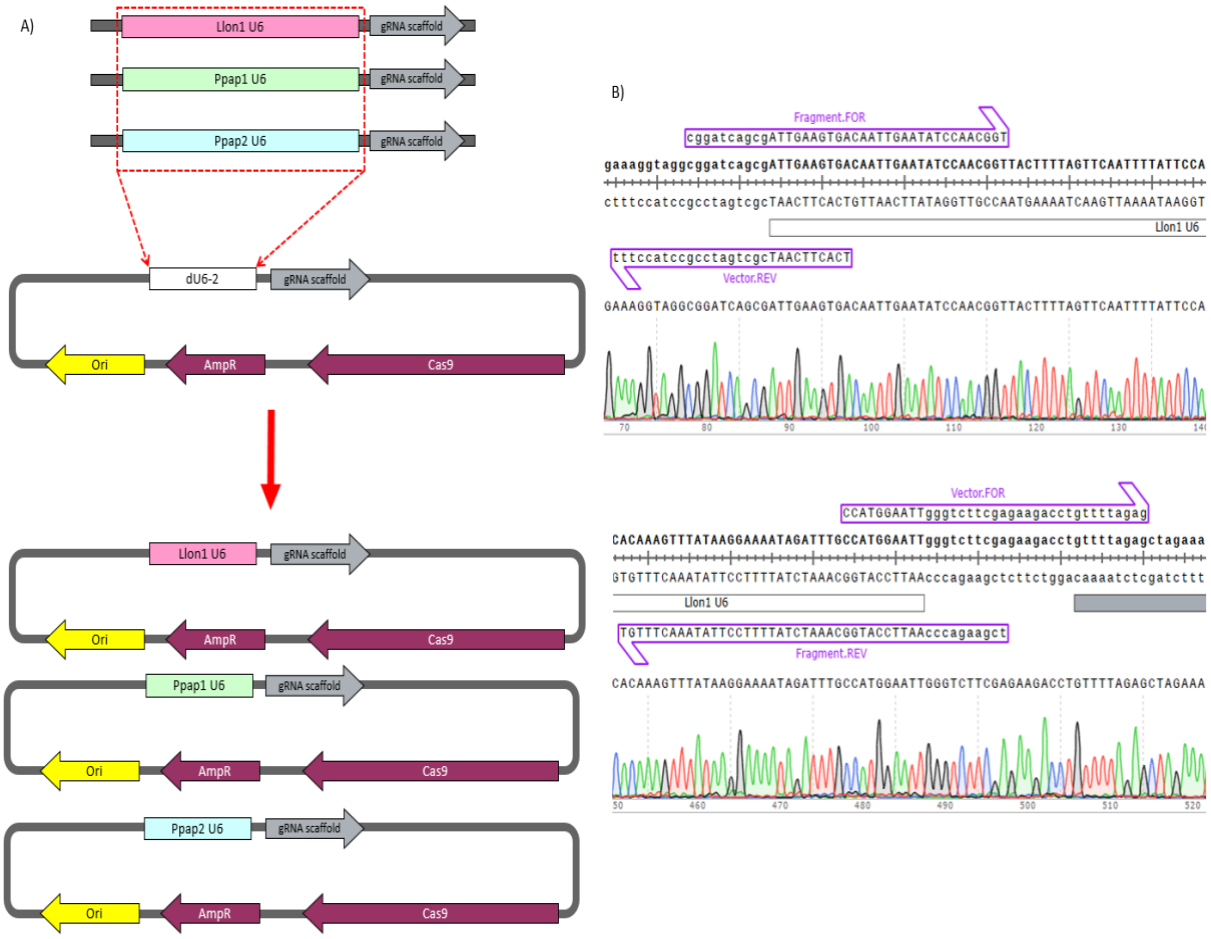


Figure 16. pDCC6 plasmid constructions with *L. longipalpis* and *P. papatasi* U6 promoters inserted to replace the D. melanogaster U6-2 promoter. A) Schematic of the promoter insertion. B) Sequence alignment of the Llon1 U6 promoter inserted into the pDCC6 plasmid at the 5' end (top), and 3' end (bottom).

Using the Llon1-U6-pDCC6 and Ppap-U6-pDCC6 as backbones, constructs were designed and built via Gibson assembly (see Method, *Generation of Plasmid constructs via Gibson Assembly*) to target phenotypic marker genes and genotypic targets in both *L. longipalpis* and *P. papatasi*. For *L. longipalpis* constructs were designed to target wing development genes, *Rudimentary* and *Vestigial*, and olfactory genes *Gr2*, *IR8a*, and *Orco*. For *P. papatasi* constructs were designed to target the *Ebony* gene involved in cuticle pigmentation, and previously used in other insect as a proof of concept (see Chapter 2). *Ebony* was selected as a target for *P. papatasi* because of the lighter colour of this species compared to *L. longipalpis*. In addition, four constructs were designed containing different gRNAs targeting the *Caspar* gene involved in the immune deficiency pathway due to *Caspar*'s involvement in vector-parasite interactions (research pursued by collaborators at Charles University, Czech Republic).

Briefly, plasmids were constructed as follows. The oligo pairs (gRNAs) were phosphorylated and annealed using T4 polynucleotide kinase. Annealed oligo concentrations were determined by Nanodrop

spectrophotometry. The species-specific pDCC6 plasmid backbone were digested using Fast digest BpiI enzyme, incubated and run on a gel before extracting the bands. A ligation reaction was performed using the digested pDCC6 plasmids and the annealed oligos with Quick Ligase. Competent *E. coli* cells were transformed and plated. Colonies were selected for Colony PCR to indicate whether the size determined by gel electrophoresis was as expected.

Three *L. longipalpis* constructs were built targeting each of the wing development genes (V1-pDCC6, V2-pDCC6, V3-pDCC6, R1-pDCC6, R2-pDCC6, and R3-pDCC6), with successful insertion of the gRNA sequences, and no modification observed at the gRNA insertion site within the pDCC6 backbones (Figure 35). Two constructs were targeted to each olfactory gene (Table 18) and confirmed by sequencing (Gr2-1-pDCC6, Gr2-2-pDCC6, Orco-1-pDCC6, Orco-2-pDCC6, Ir8a-1-pDCC6, and Ir8a-2-pDCC6). gRNAs were successfully inserted into all six olfactory constructs, and had no modifications at the insertion site (Figure 35). Additionally, Eight *P. papatasi* targeting plasmids *Ebony* and *Caspar* were successfully constructed, and confirmed by sequencing and alignment (Figure 35). Plasmid constructs were not tested in vitro due to time constraints, however they were taken forward to *in vivo* studies (see Chapter 5).

Table 18. pDCC6 Olfactory plasmid construction for *L. longipalpis*. gRNA sequence pairs (green) have additional overhangs (AATT for forward gRNAs and AAAC for reverse gRNAs) to facilitate insertion into the pDCC6 backbone via Gibson assembly. See Chapter 2 for *P. papatasi Ebony* and *Caspar* gRNA sequences inserted into pDCC6 backbone.

Gene	gRNA rank	gRNA sequence	PAM sequence	Direction	gRNA name	Construct name
Gr2	2	AAACATAAAGAGGCAATACG	TGG	Reverse	LL Gr2 gRNA #4	Gr2-1-pDCC6
		CGTATTGCCTCTTTATGTTT		Forward	LL Gr2 gRNA #3	
	4	GCTGTGTACAAGACAATGTG	GGG	Forward	LL Gr2 gRNA #5	Gr2-2-pDCC6
		CCCCACATTGTCTTGTACACAGC		Reverse	LL Gr2 gRNA #6	
Orco	1	TGTGAGATACATGACCAACA	AGG	Forward	LL Orco gRNA #1	Orco-1-pDCC6
		CCTTGTTGGTCATGTATCTCACA		Reverse	LL Orco gRNA #2	
	3	TACAGCAATCAAGTATTGGG	TGG	Forward	LL Orco gRNA #3	Orco-2-pDCC6
		CCACCCAATACTTGATTGCTGTA		Reverse	LL Orco gRNA #4	
Ir8a	2	AGCAATAGCGAGACCTGCGA	GGG	Reverse	LL Ir8a gRNA #2	Ir8a-1-pDCC6
		TCGCAGGTCTCGCTATTGCT		Forward	LL Ir8a gRNA #1	
	3	TTGAGTGTAAAATCCCGACA	GGG	Reverse	LL Ir8a gRNA #4	Ir8a-2-pDCC6
		TGTCGGGATTTTACACTCAA		Forward	LL Ir8a gRNA #3	

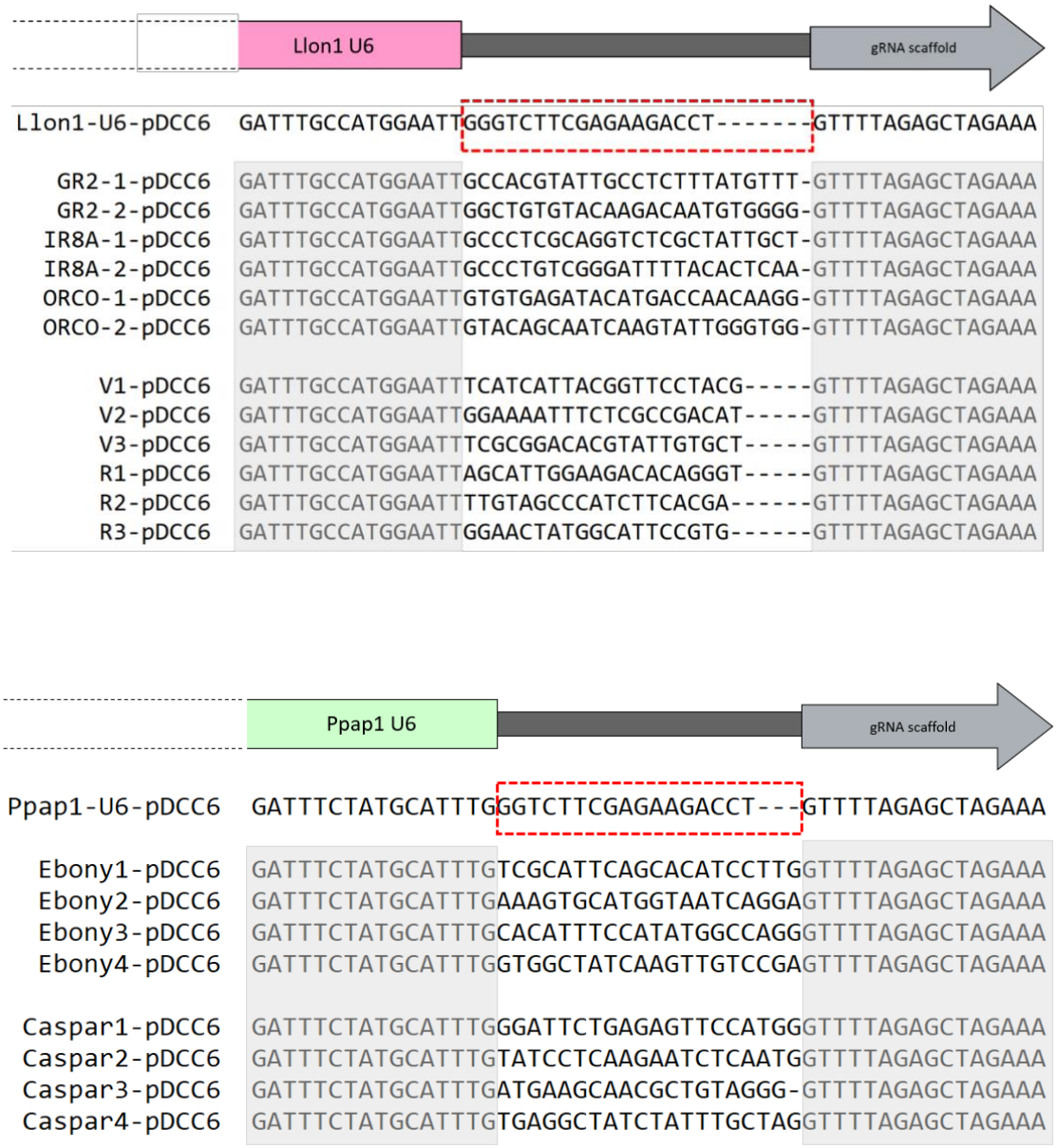


Figure 17. Sequence alignment of sand fly species-specific pDCC6 constructs incorporating gRNAs targeting olfactory genes, immunity genes, wing development and pigmentation genes. The gRNAs are inserted between the U6 promoter and the gRNA scaffold (red box). Llon1-U6-pDCC6 backbone was modified with olfactory and wing development gRNA sequences inserted via Gibson assembly. The sequences were as expected for Gr2-1, Gr2-2, Ir8a-1, Ir8a-2, Orco-1, and Orco-2. Ppap1-U6-pDCC6 backbone was modified to produce Ebony1, Ebony2, Ebony3, and Ebony4-pDCC6, and Caspar1, 2, 3, and 4-pDCC6.

Chapter Summary

In this chapter we developed tools and a pathway for an *in vitro* platform to assess gene editing components (PiggyBac and CRISPR), including validation of promoter function, *in vitro* transcription of gRNAs, and development of CRISPR knockout plasmids.

Insect cell lines have been used extensively for a range of purposes including virology, insect immunity and host-parasite interactions. However, only two *L. longipalpis* and two *P. papatasi* sand fly cell lines have been developed and maintained, despite the insights they could elucidate with respect to *Leishmania* parasite biology, cellular immune responses to infection, or gene expression. Here, two novel sand fly cell lines (LLE40 and LLE45) have been utilised towards the development of an *in vitro* platform for validating CRISPR and PiggyBac tools. Although cells were used to their current potential in terms of *in vitro* testing of gene editing components, further characterisation and sub cloning would be highly advantageous to produce a line of undifferentiated lines.

In vitro transcription of gRNAs targeting phenotypic genes and olfactory genes was achieved, demonstrating *in vitro* cleavage of sand fly DNA. Subsequent cell transfections confirmed promoters, including *Actin5C*, *CMV*, *3x3P, hsp70*, and *Ubiquitin 63e*, driving EGFP, GFP, Cas9, and hSpCas9 respectively, which are now validated candidates for CRISPR based approaches. This is an important step to efficiently screen effectors, prior to use in an *in vivo* system. Transfection optimisation of this process was attempted by comparing candidate transfection reagents in the context of quantitative transfection efficiencies. In the sand fly cell lines two transfection reagents were more efficient at plasmid transfection (Cellfectin and Effectene), determined by identification of GFP expression in flow cytometric analysis. However, we recommend that further optimisation should be considered particularly including a fully cloned and characterised undifferentiated sand fly cell lines.

In vitro transcription of gRNAs was optimised in this study, confirmed by sequence based approaches and through heteroduplex cleavage assays to facilitate selection of gRNAs to take forward to *in vivo* studies. Given more time (impacted by COVID) we would also continue to directly transfect transcribed gRNAs and synthetic Cas9 protein, in parallel to plasmid based approaches. Importantly a range of expression promoters were confirmed to function within the LLE40 and LLE45 cells for use in expression of CRISPR components. We have subsequently identified of alternative sand fly specific promoters (from the current genome, VectorBase) for the expression of Cas9, through *nanos*, *vasa* and *zpg*, which are reported here.

Generation of a Cas9-expressing cell line was attempted as a simplified tool for validating CRISPR components, before pursuing laborious *in vivo* studies. A modified plasmid (UbiqCas9-NeoR) was constructed to allow for antibiotic selection, and preliminary data was collected on approximate

concentrations of G418 use against the sand fly cells. The IC50 antibiotic concentration identified was not dissimilar to concentrations used for selection in other insect cells. However, we were not able to recover resistance cells within the timeframe available and repeat runs are being performed. Optimisation of both the transfection protocols, and G418 selection assays, would allow the isolation of transgenic Cas9 cells and facilitate rapid in vitro screening of gRNAs.

CRISPR knockout plasmids were successfully constructed for a range of targets (*Orco*, *Gr2*, *Ir8a*, *Vestigial*, *Rudimentary*, *Ebony*, *Caspar*) based around a pDCC6 backbone, both phenotypic and olfactory in *L. longipalpis* and phenotypic and immunity-based in *P. papatasi* sand flies. Species-specific RNA Pol III *U6* promoters were successfully incorporated into plasmid backbones to promote expression of gRNA sequences. Given the COVID effected timeframe we decided to assess selected plasmids immediately *in vivo* (Chapter 5). Therefore we would like to revisit and fully assess the full panel of plasmids within *in vitro* platform, to further determine gRNA expression activity, and to identify species-specific RNA Pol III promoters

Transgenic modification in sand fly vectors is difficult, and maintenance is complicated. Therefore, the development of an *in vitro* platform, as described here, is invaluable to rationalise gene editing components prior to *in vivo* approaches that were conducted in Chapter 5.

References

Akbari, O. S., Papathanos, P. A., Sandler, J. E., Kennedy, K., & Hay, B. A. (2014). Identification of germline transcriptional regulatory elements in Aedes aegypti. *Scientific Reports*, *4*, 35–39. https://doi.org/10.1038/srep03954

Anderson, M. A. E., Purcell, J., Verkuijl, S. A. N., Norman, V. C., Leftwich, P. T., Harvey-Samuel, T., & Alphey, L. S. (2020). Expanding the CRISPR Toolbox in Culicine Mosquitoes: In Vitro Validation of Pol III Promoters. *ACS Synthetic Biology*, *9*(3), 678–681. https://doi.org/10.1021/acssynbio.9b00436

Bairoch, A. (2018). The cellosaurus, a cell-line knowledge resource. *Journal of Biomolecular Techniques*, *29*(2), 25–38. https://doi.org/10.7171/jbt.18-2902-002

Bell-Sakyi, L., Beliavskaia, A., Hartley, C. S., Jones, L., Luu, L., Haines, L. R., … Makepeace, B. L. (2021). Isolation in natural host cell lines of wolbachia strains wpip from the mosquito culex pipiens and wpap from the sand fly phlebotomus papatasi. *Insects*, *12*(10). https://doi.org/10.3390/insects12100871

Bell-Sakyi, L., Darby, A., Baylis, M., & Makepeace, B. L. (2018). The Tick Cell Biobank: A global resource for in vitro research on ticks, other arthropods and the pathogens they transmit. *Ticks and Tick-Borne Diseases*, *9*(5), 1364–1371. https://doi.org/10.1016/j.ttbdis.2018.05.015

Bello, F. J., Mejía, A. J., Del Pilar Corena, M., Ayala, M., Sarmiento, L., Zuñiga, C., & Palau, M. T. (2005). Experimental infection of Leishmania (L.) chagasi in a cell line derived from Lutzomyia longipalpis (Diptera:Psychodidae). *Memorias Do Instituto Oswaldo Cruz*, *100*(6), 518–525. https://doi.org/10.1590/s0074-02762005000600004

Biedler, J. K., Qi, Y., Pledger, D., James, A. A., & Tu, Z. (2015). Maternal germline-specific genes in the Asian malaria mosquito Anopheles stephensi: Characterization and application for disease control. *G3: Genes, Genomes, Genetics*, *5*(2), 157–166. https://doi.org/10.1534/g3.114.015578

Brinkman, E. K., Chen, T., Amendola, M., & Van Steensel, B. (2014). Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Research*, *42*(22), 1–8. https://doi.org/10.1093/nar/gku936

Cano-Monreal, G. L., Williams, J. C., & Heidner, H. W. (2010). An arthropod enzyme, Dfurin1, and a vertebrate furin homolog display distinct cleavage site sequence preferences for a shared viral proprotein substrate. *Journal of Insect Science*, *10*(1), 1–16. https://doi.org/10.1673/031.010.2901

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

Côrtes, L. M. M., De Souza Pereira, M. C., Da Silva, F. S., Pereira, B. A. S., De Oliveira Junior, F. O., De Araújo Soares, R. O., … Alves, C. R. (2012). Participation of heparin binding proteins from the surface of Leishmania (Viannia) braziliensis promastigotes in the adhesion of parasites to Lutzomyia longipalpis cells (Lulo) in vitro. *Parasites and Vectors*, *5*(1), 1–10. https://doi.org/10.1186/1756-3305-5-142

Côrtes, L. M., Silva, R. M., Pereira, B. A., Guerra, C., Zapata, A. C., Bello, F. J., … Alves, C. R. (2011). Lulo cell line derived from Lutzomyia longipalpis (Diptera: Psychodidae): A novel model to assay Leishmania spp. and vector interaction. *Parasites and Vectors*, *4*(1), 3–7. https://doi.org/10.1186/1756-3305-4-216

Du, X., Wang, J., Zhou, Q., Zhang, L., Wang, S., Zhang, Z., & Yao, C. (2018). Advanced physical techniques for gene delivery based on membrane perforation. *Drug Delivery*, *25*(1), 1516–1525. https://doi.org/10.1080/10717544.2018.1480674

Eckermann, K. N., Ahmed, H. M. M., KaramiNejadRanjbar, M., Dippel, S., Ogaugwu, C. E., Kitzmann, P., … Wimmer, E. A. (2018). Hyperactive piggyBac transposase improves transformation efficiency in diverse insect species. *Insect Biochemistry and Molecular Biology*, *98*(May), 16–24. https://doi.org/10.1016/j.ibmb.2018.04.001

Feng, X., Kambic, L., Nishimoto, J. H. K., Reed, F. A., Denton, J. A., Sutton, J. T., & Gantz, V. M. (2021). Evaluation of Gene Knockouts by CRISPR as Potential Targets for the Genetic Engineering of the Mosquito Culex quinquefasciatus. *CRISPR Journal*, *4*(4), 595–608. https://doi.org/10.1089/crispr.2021.0028

Feng, X., López Del Amo, V., Mameli, E., Lee, M., Bishop, A. L., Perrimon, N., & Gantz, V. M. (2021). Optimized CRISPR tools and site-directed transgenesis towards gene drive development in Culex quinquefasciatus mosquitoes. *Nature Communications*, *12*(1). https://doi.org/10.1038/s41467-021-23239-0

Gagnon, J. A., Valen, E., Thyme, S. B., Huang, P., Ahkmetova, L., Pauli, A., … Schier, A. F. (2014). Efficient mutagenesis by Cas9 protein-mediated oligonucleotide insertion and large-scale assessment of single-guide RNAs. *PLoS ONE*, *9*(5), 5–12. https://doi.org/10.1371/journal.pone.0098186

Gokcezade, J., Sienski, G., & Duchek, P. (2014). Efficient CRISPR/Cas9 Plasmids for Rapid and Versatile Genome Editing in Drosophila . *G3: Genes|Genomes|Genetics*, *4*(11), 2279–2282. https://doi.org/10.1534/g3.114.014126

González, M., Martín-Ruíz, I., Jiménez, S., Pirone, L., Barrio, R., & Sutherland, J. D. (2011). Generation of stable Drosophila cell lines using multicistronic vectors. *Scientific Reports*, *1*. https://doi.org/10.1038/srep00075

Hall, A. B., Basu, S., Jiang, X., Qi, Y., Timoshevskiy, V. A., Biedler, J. K., … Tu, Z. (2015). A male-determining factor in the mosquito Aedes aegypti. *Science*, *348*(6240), 1268–1270. https://doi.org/10.1126/science.aaa2850

Jordan, E. T., Collins, M., Terefe, J., Ugozzoli, L., & Rubio, T. (2008). Optimizing electroporation conditions in primary and other difficult-to-transfect cells. *Journal of Biomolecular Techniques*, *19*(5), 328–334.

Kandul, N. P., Liu, J., Sanchez C, H. M., Wu, S. L., Marshall, J. M., & Akbari, O. S. (2019). Transforming insect population control with precision guided sterile males with demonstration in flies. *Nature Communications*, *10*(1), 1–12. https://doi.org/10.1038/s41467-018-07964-7

Kim, T. K., & Eberwine, J. H. (2010). Mammalian cell transfection: The present and the future. *Analytical and Bioanalytical Chemistry*, *397*(8), 3173–3178. https://doi.org/10.1007/s00216-010-3821-6

Kotwica-Rolinska, J., Chodakova, L., Chvalova, D., Kristofova, L., Fenclova, I., Provaznik, J., … Dolezel, D. (2019). CRISPR/Cas9 genome editing introduction and optimization in the non-model insect Pyrrhocoris apterus. *Frontiers in Physiology*, *10*(JUL), 1–15. https://doi.org/10.3389/fphys.2019.00891

Koutroumpa, F. A., Monsempes, C., François, M. C., De Cian, A., Royer, C., Concordet, J. P., & Jacquin-Joly, E. (2016). Heritable genome editing with CRISPR/Cas9 induces anosmia in a crop pest moth. *Scientific Reports*, *6*(February), 1–9. https://doi.org/10.1038/srep29620

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

Labun, K., Montague, T. G., Krause, M., Torres Cleuren, Y. N., Tjeldnes, H., & Valen, E. (2019). CHOPCHOP v3: Expanding the CRISPR web toolbox beyond genome editing. *Nucleic Acids Research*, *47*(W1), W171–W174. https://doi.org/10.1093/nar/gkz365

Liu, P., Jin, B., Li, X., Zhao, Y., Gu, J., Biedler, J. K., … Chen, X. G. (2020a). Nix is a male-determining factor in the Asian tiger mosquito Aedes albopictus. *Insect Biochemistry and Molecular Biology*, *118*(June 2019), 103311. https://doi.org/10.1016/j.ibmb.2019.103311

Liu, P., Jin, B., Li, X., Zhao, Y., Gu, J., Biedler, J. K., … Chen, X. G. (2020b). Nix is a male-determining factor in the Asian tiger mosquito Aedes albopictus. *Insect Biochemistry and Molecular Biology*, *118*(December 2019), 103311. https://doi.org/10.1016/j.ibmb.2019.103311

Louradour, I., Ghosh, K., Inbar, E., & Sacks, D. L. (2019). CRISPR/Cas9 Mutagenesis in Phlebotomus papatasi: the Immune Deficiency Pathway Impacts Vector Competence for Leishmania major. *MBio*, *10*(4), 1–14.

Ma, H., Wu, Y., Dang, Y., Choi, J. G., Zhang, J., & Wu, H. (2014). Pol III promoters to express small RNAs: Delineation of transcription initiation. *Molecular Therapy - Nucleic Acids*, *3*(March), e161. https://doi.org/10.1038/mtna.2014.12

Martin-Martin, I., Aryan, A., Meneses, C., Adelman, Z. N., & Calvo, E. (2018). Optimization of sand fly embryo microinjection for gene editing by CRISPR/Cas9. *PLoS Neglected Tropical Diseases*, *12*(9), 1–18. https://doi.org/10.1371/journal.pntd.0006769

Nilsen, S. L., & Castellino, F. J. (1999). Expression of human plasminogen in Drosophila Schneider S2 cells. *Protein Expression and Purification*, *16*(1), 136–143. https://doi.org/10.1006/prep.1999.1045

Papathanos, P. A., Windbichler, N., Menichelli, M., Burt, A., & Crisanti, A. (2009). The vasa regulatory region mediates germline expression and maternal transmission of proteins in the malaria mosquito Anopheles gambiae: A versatile tool for genetic control strategies. *BMC Molecular Biology*, *10*, 1–13. https://doi.org/10.1186/1471-2199-10-65

Piñero, J., López-Baena, M., Ortiz, T., & Cortés, F. (1997). Apoptotic and necrotic cell death are both induced by electroporation in HL60 human promyeloid leukaemia cells. *Apoptosis*, *2*(3), 330–336. https://doi.org/10.1023/A:1026497306006

Ran, F. A., Hsu, P. D., Lin, C.-Y., Gootenberg, J. S., Konermann, S., Trevino, A. E., … Zhang, F. (2013). Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell*, *154*(6), 1380–1389. https://doi.org/10.1016/j.cell.2013.08.021

Rey, G. J., Ferro, C., & Bello, F. J. (2000). Establishment and Characterization of a New Continuous Cell Line from Lutzomyia longipalpis (Diptera: Psychodidae) and its Susceptibility to Infections with Arboviruses and Leishmania chagasi. *Memorias Do Instituto Oswaldo Cruz*, *95*(1–2), 103–110. https://doi.org/10.1590/S0074-02762000000100017

Saraiva, E., Fampa, P., Cedeno, V., Bergoin, M., Mialhe, E., & Miller, L. H. (2009). Expression of Heterologous Promoters in <I>Lutzomyia longipalpis</I> and <I>Phlebotomus papatasi</I> (Diptera: Psychodidae) Cell Lines. *Journal of Medical Entomology*, *37*(6), 802–806. https://doi.org/10.1603/0022-2585-37.6.802

Scherer, C., Knowles, J., Sreenu, V. B., Fredericks, A. C., Fuss, J., Maringer, K., … Schnettler, E.

(2021). An aedes aegypti-derived ago2 knockout cell line to investigate arbovirus infections. *Viruses*, *13*(6), 1–19. https://doi.org/10.3390/v13061066

Simoni, A., Hammond, A. M., Beaghton, A. K., Galizi, R., Taxiarchi, C., Kyrou, K., … Crisanti, A. (2020). A male-biased sex-distorter gene drive for the human malaria vector Anopheles gambiae. *Nature Biotechnology*, *38*(9), 1054–1060. https://doi.org/10.1038/s41587-020-0508-1

Tesh, R. B., & Modi, G. B. (1983). Development of a continuous cell line from the sand fly Lutzomyia longipalpis (Diptera: Psychodidae), and its susceptibility to infection with arboviruses. *Journal of Medical Entomology*, *20*(2), 199–202. https://doi.org/10.1093/jmedent/20.2.199

Tinoco-Nunes, B., Telleria, E. L., Da Silva-Neves, M., Marques, C., Azevedo-Brito, D. A., Pitaluga, A. N., & Traub-Csekö, Y. M. (2016). The sandfly Lutzomyia longipalpis LL5 embryonic cell line has active Toll and Imd pathways and shows immune responses to bacteria, yeast and Leishmania. *Parasites and Vectors*, *9*(1), 1–11. https://doi.org/10.1186/s13071-016-1507-4

Torres, T. Z. B., Prince, B. C., Robison, A., & Rückert, C. (2022). *Optimized In Vitro CRISPR / Cas9 Gene Editing Tool in the West*. 1–13.

Varjak, M., Kean, J., Vazeille, M., Failloux, A., & Kohl, A. (2017). Aedes aegypti Piwi4 Is a Noncanonical. *MSphere*, *2*(3), e00144-17.

Vidigal, J., Fernandes, B., Dias, M. M., Patrone, M., Roldão, A., Carrondo, M. J. T., … Teixeira, A. P. (2018). RMCE-based insect cell platform to produce membrane proteins captured on HIV-1 Gag virus-like particles. *Applied Microbiology and Biotechnology*, *102*(2), 655–666. https://doi.org/10.1007/s00253-017-8628-3

Viswanatha, R., Mameli, E., Rodiger, J., Merckaert, P., Feitosa-Suntheimer, F., Colpitts, T. M., … Perrimon, N. (2021). Bioinformatic and cell-based tools for pooled CRISPR knockout screening in mosquitos. *Nature Communications*, *12*(1). https://doi.org/10.1038/s41467-021-27129-3

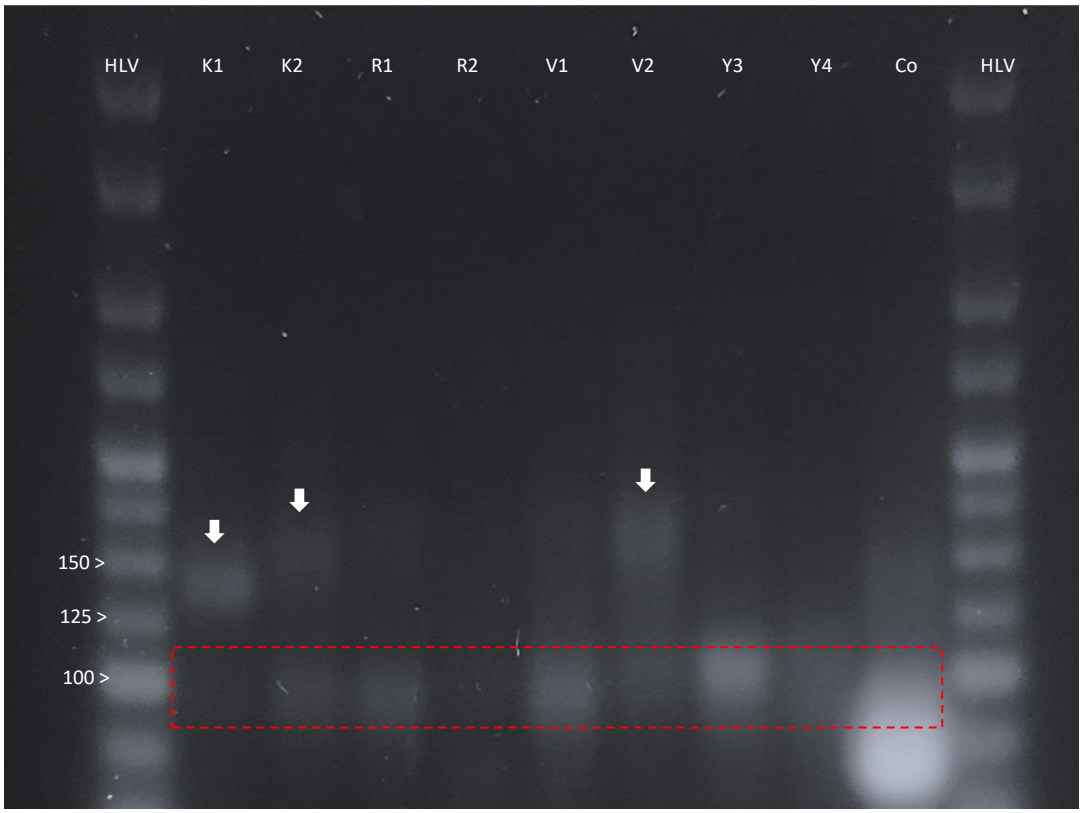
Wakiyama, M., Matsumoto, T., & Yokoyama, S. (2005). Drosophila U6 promoter-driven short hairpin RNAs effectively induce RNA interference in Schneider 2 cells. *Biochemical and Biophysical Research Communications*, *331*(4), 1163–1170. https://doi.org/10.1016/j.bbrc.2005.03.240

Wang, H., Yang, H., Shivalila, C. S., Dawlaty, M. M., Cheng, A. W., Zhang, F., & Jaenisch, R. (2013). One-step generation of mice carrying mutations in multiple genes by CRISPR/cas-mediated genome engineering. *Cell*, *153*(4), 910–918. https://doi.org/10.1016/j.cell.2013.04.025

Zhang, J., Liu, Y., Chang, J., Shi, R., Wang, X., Zhao, P., … Lu, W. (2014). CRISPR/Cas9 mediated multiplex genome editing and heritable mutagenesis of BmKu70 in Bombyx mori. *Scientific Reports*, *4*(1), 1–6. https://doi.org/10.1038/srep04489

Chapter 4 Appendices

Appendix 1. PCR amplification of target regions (wildtype *Lu. longipalpis* DNA) of olfactory gene targets prior to gel extraction/PCR cleanup. White box represents expected fragment sizes, between 500-550 bp. Expected fragment sizes (bp): OR2 (516), OR4 (511), GR1 (550), IR76b (524), IR25a (550), IR8a (550), Orco (539), GR2 (549).



Appendix 2. sgRNA for *Kmo, Rudimentary, Vestigial,* and *Yellow* genes transcribed using Engen T7 Transcription kit. Expected band size for the sgRNA is ~100bp (red box), as is the control oligo (Co) provided with the kit. Arrows indicate sgRNAs with bands larger than expected.

Appendix 3. Concentrations for gRNA template from gel electrophoresis determined by nanodrop spectrometry. The molar concentrations for each of the combined templates is presented. For Engen transcription (1uM) is required. An asterisk indicates templates above the required concentration (1µM).

gRNA	*Concentration (ng/µL)*	*Combined concentration of the two repeats (ng/µL),*	*Molar concentration (µM).*
K1	6.339	1.745	0.760
K1	4.751		
K2	5.328	2.142	0.931
K2	3.989		
R1	5.076	1.762	0.766
R1	4.232		
R2	4.355	1.540	0.670
R2	4.572		
V1	8.324	2.425	1.06*
V1	5.055		
V2	8.239	2.021	0.958
V2	3.086		
Y3	7.913	2.758	1.20*
Y3	2.487		
Y4	7.933	1.831	0.797
Y4	2.646		

Combined concntrations eluted into 30µL sterile H_2O. Molar concentration of 1µM required for Engen transcription kit. Asterisk (*) indicates concentrations above the threshold for transcription.

Appendix 4. Concentrations for phenotypic gRNA templates extracted from agarose gels. Concentrations are determined by Nanodrop spectrometry. The molar concentrations for each of the combined templates is presented. For Engen transcription (1uM) is required. An asterisk indicates templates above the required concentration (1μM).

sgRNA	*Concentration (ng/μL)*	*Molar concentration (μM)*
K1	6.73	2.93
K2	6.31	2.75
R1	5.91	2.57
R2	6.42	2.79
V1	6.18	2.69
V2	5.67	2.47
Y3	7.48	3.26
Y4	6.07	2.64

Appendix 5. Phenotypic sgRNAs transcribed using Engen transcription kit and protocol and purified using Monarch cleanup kit (NEB, USA). Concentrations calculated via Nanodrop spectrophotometer.

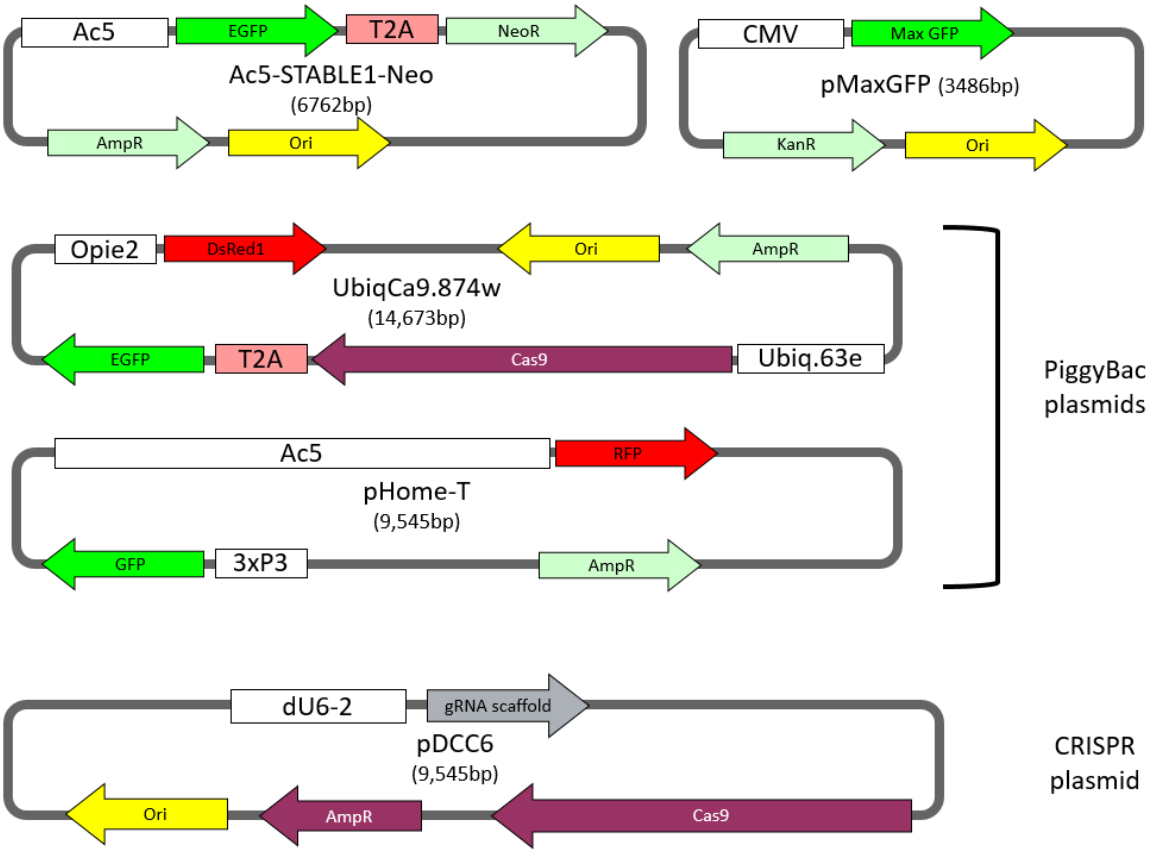
Transcribed sgRNA	*Concentration (ng/μL)*	*260/280*
K1	353.32	2.26
K2	296.53	2.26
R1	257.82	2.27
R2	100.58	2.29
V1	306.05	2.13
V2	267.05	2.20
Y3	353.57	2.23
Y4	206.06	2.20

Appendix 6. Primers for target DNA amplification. DNA used in cleavage assays to determine function of in vitro transcribed phenotypic sgRNAs.

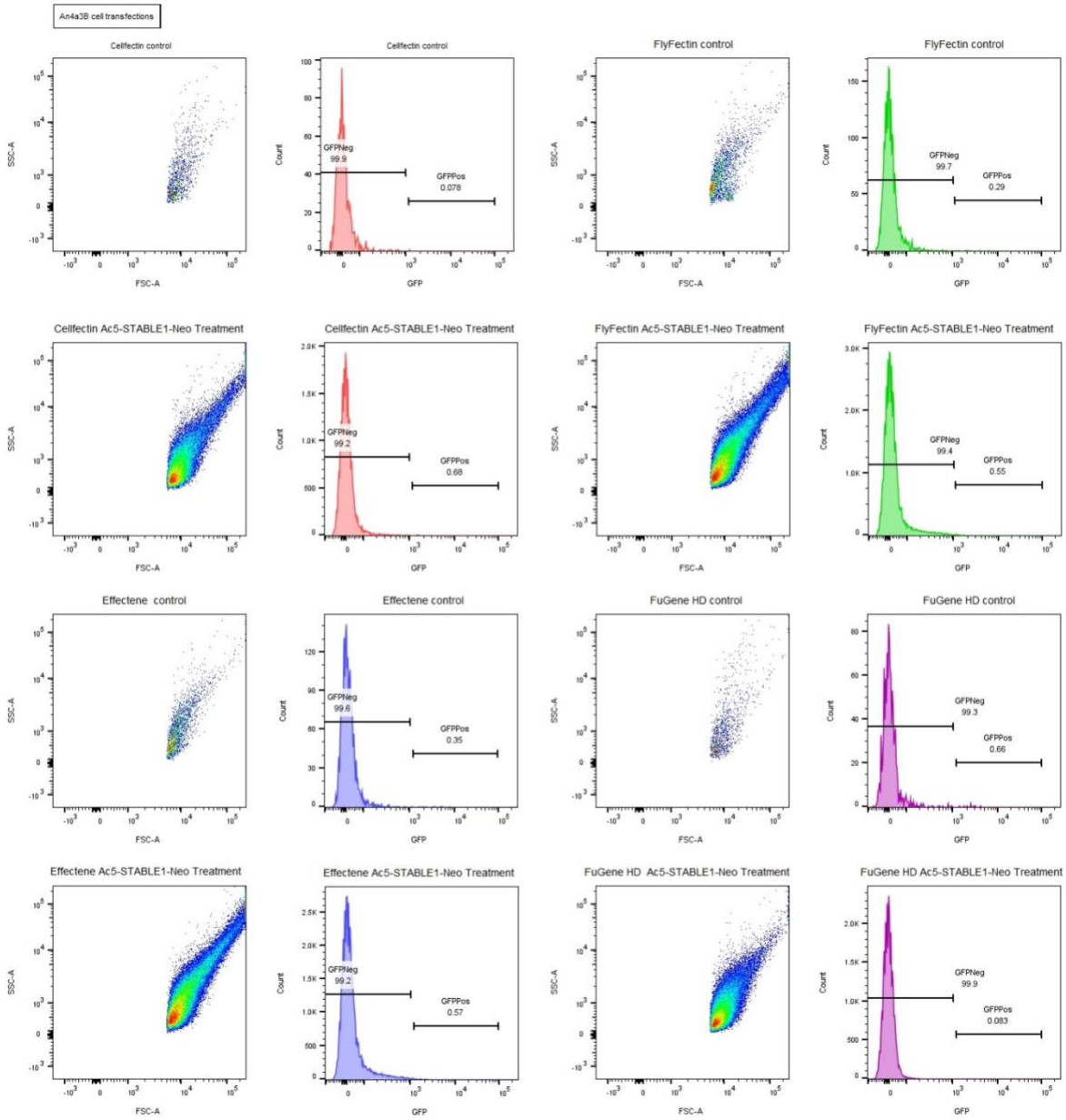
Gene	*Transcribed gRNA*	*Primer 1*	*Tm*	*Primer 2*	*Tm*	*Tm (Phusion MasterMix)*	*Size (bp)*
Kmo	K1 K2	LLKmo Prim#1 AGCTCAACAAGGGCA TAAGAAA	60	LLKmo Prim#2 ATTTGGAGATAAAA GTTTGGCG	57	61	702
Rudimentary	R1 R2	LLRudimentary Prim#1 CCATTCTCGAGACAAT CCTCC	60	LLRudimentary Prim#2 CAAGTTGTTTCACTC TTCCCG	59	62	725
Vestigial	V1 V2	LLVestigial Prim#1 TGGTATGCAATGAAAT TGAGAA	56	LLVestigial Prim#2 TTGAAATGATTAAA ATCGGACA	54	58	834
Yellow	Y3 Y4	LLYellow Prim#1 GAGACTTTGGTGGCAA TCAAT	59	LLYellow Prim#3 GCAACACGTCAAAT GTGAAAAT	58	61	459

Appendix 7. G418 and growth media volumes to reach final concentrations of G418 from 16mg/ml stock solution.

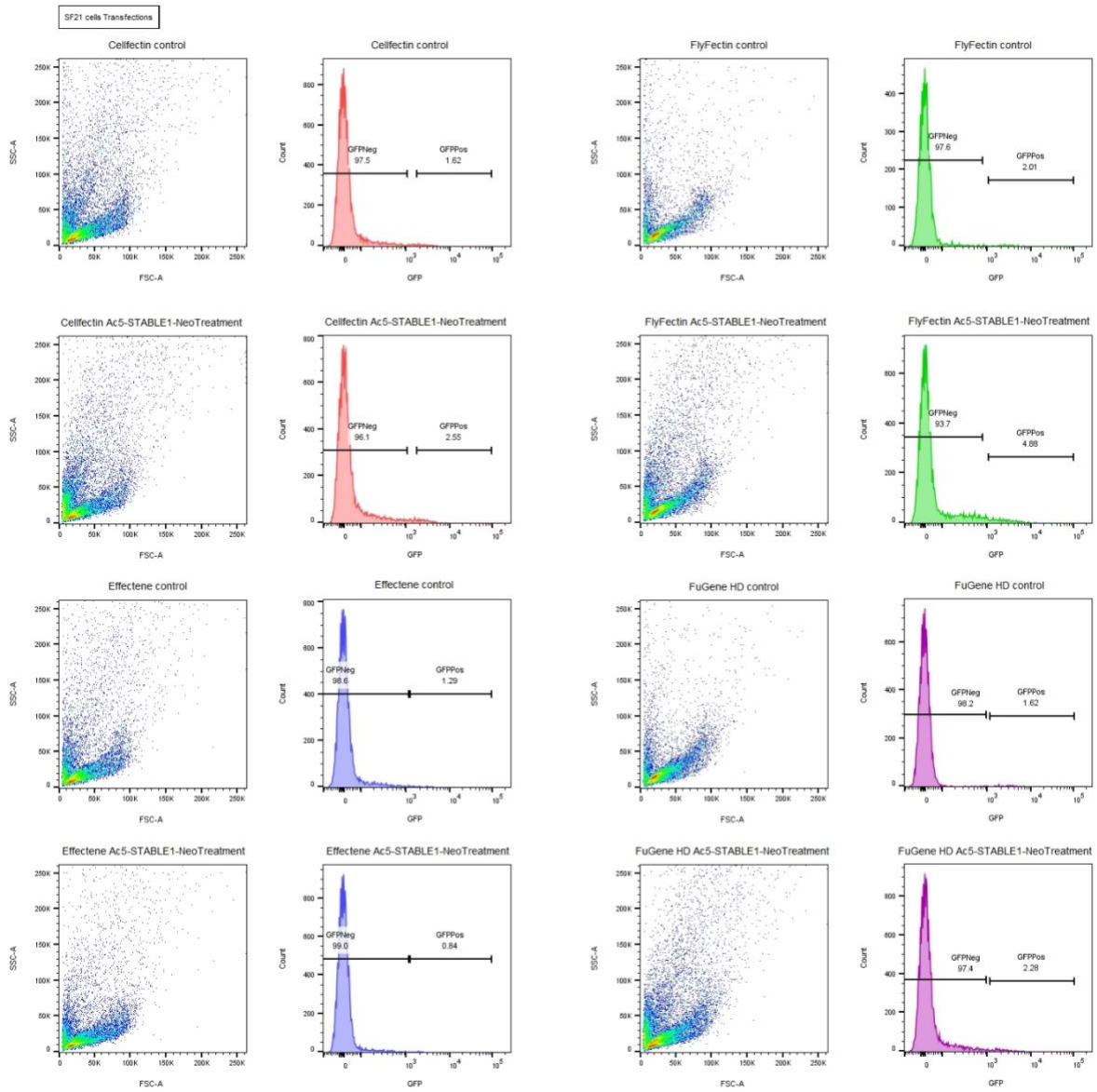
Volume of media with G418 (16mg/ml)(µL)	*Media (µL)*	*Final Concentration (µg/ml)*
250	0	4000
125	125	2000
62.5	187.5	1000
37.5	212.5	600
31.25	218.75	500
15.625	234.375	250
7.8125	242.1875	125
3.90625	246.09375	62.5
0	250	0



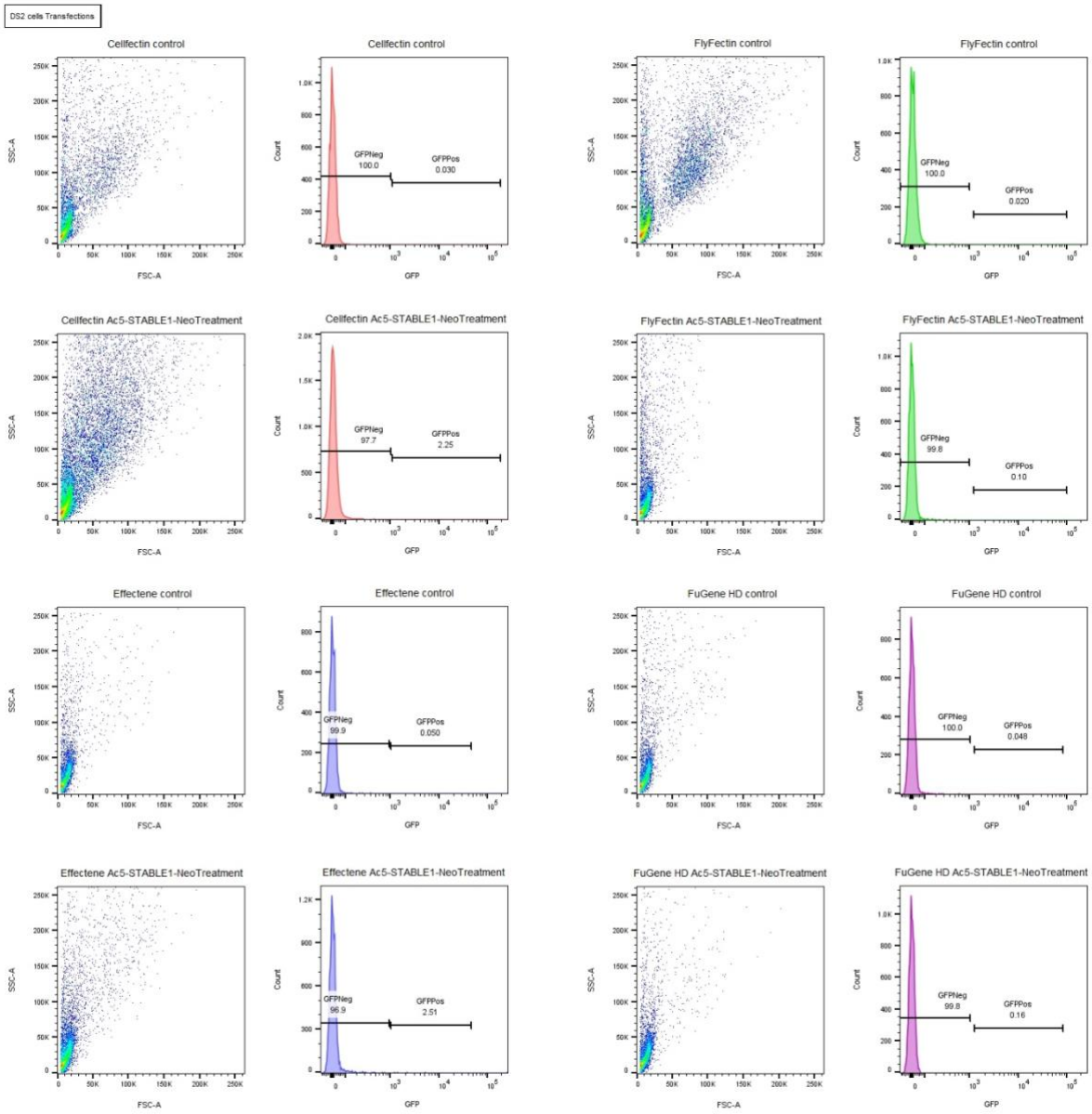
Appendix 8. Illustration of the plasmids used in cell transfections showing key features including fluorescence markers (EGFP, GFP, DsRed1, RFP), promoters (Ac5, CMV, Opie2, Ubiquitin 63e, dU6-2), and Cas9 and gRNA scaffold regions. UbiqCas9.874w and pHome-T are PiggyBac plasmids requiring helper transposase (no shown). pDCC6 is a CRISPR backbone plasmid, used to insert gRNA sequences at the gRNA scaffold region.



Appendix 9. Flow cytometry of **An4a3B cells** transfected with Ac5-STABLE1-Neo. FSC-A/SSC-A plots are shown next to GFP/Count plots with GFP+ and GFP- gates applied. Controls plots are stacked above treatment plots, where the plasmid was used in the transfection. Each histogram colour represents a different transfection reagent used. Cellfectin (red), Flyfectin (green), Effectene (blue), and FuGene HD (purple).



Appendix 10. Flow cytometry of **Sf21 cells** transfected with Ac5-STABLE1-Neo. FSC-A/SSC-A plots are shown next to GFP/Count plots with GFP+ and GFP- gates applied. Controls plots are stacked above treatment plots, where the plasmid was used in the transfection. Each histogram colour represents a different transfection reagent used. Cellfectin (red), Flyfectin (green), Effectene (blue), and FuGene HD (purple).



Appendix 11. Flow cytometry of **DS2 cells** transfected with Ac5-STABLE1-Neo. FSC-A/SSC-A plots are shown next to GFP/Count plots with GFP+ and GFP- gates applied. Controls plots are stacked above treatment plots, where the plasmid was used in the transfection. Each histogram colour represents a different transfection reagent used. Cellfectin (red), Flyfectin (green), Effectene (blue), and FuGene HD (purple).

Chapter 5

In vivo mutagenesis of *Lutzomyia longipalpis* and *Phlebotomus papatasi* using CRISPR and PiggyBac approaches

Introduction

Different genetic modification techniques have been developed to induce targeted genome modifications in cells and whole organisms for a wide range of application including the development of vector control strategies. PiggyBac transposable elements, transcription activator-like effector nucleases, zinc finger nucleases, and clustered regularly interspaced short palindromic repeats (CRISPR), are four major techniques that have been used in modification of insects, however CRISPR has recently superseded other genetic techniques due to its highly targeted nature and relative ease of use, becoming the dominant method for genetic modification in insect vectors. Each of the aforementioned techniques is described in more detail below.

Although genetic modification tools have been applied to major insect vectors of disease including *Aedes*, *Anopheles*, and *Culex* there have been limited attempts to develop the tools within other insect vectors of neglected tropical diseases, including sand fly vectors of Leishmania. Technologies to improve the control of *Lutzomyia longipalpis* and *Phlebotomus papatasi* are lacking, yet many of the gene editing methodologies can be applied to sand flies in principle, with the enormous potential to develop novel vector control approaches. For example, population suppression strategies or replacement gene drives hold great promise (see Chapter 2, *CRISPR-based Gene Drives and Control strategies* for more detail)

Background to gene editing in insects

Zinc Finger Nucleases (ZFNs)

One method for targeted modification of DNA by the generation of double stranded breaks (DSBs) follow by DNA repair is Zinc finger nucleases. ZFNs are synthetically generated proteins, fusing zinc finger binding domains with a *FokI* cleavage domain (Figure 36). Each zinc finger is highly specific, recognising a 3-4 nucleotide sequence. Zinc fingers can be assembled in a modular nature to target a specific sequence, delivering the *FokI* to the desired location. They often contain 3-6 binding domains which can specify a region of 18bp within the target gene. For a DSB to occur in the DNA, ZFN dimers must be formed. Cleavage takes place in the spacer region between the dimerised *FokI* domains (Carroll, 2011).

Theoretically the system is simple, however in application ZFNs approaches are difficult, chromatin structure of the DNA frequently prevents endonuclease cleavage, and non-specific binding if too many zinc fingers are assembled. ZFNs require significant effort to design and construct, however they have been applied successfully in *D. melanogaster* (Beumer, Bhattacharyya, Bibikova, Trautman, & Carroll, 2006; Bibikova, Beumer, Trautman, & Carroll, 2003), and in *Ae. aegypti* to disrupt to *Orco* and *Gr3* genes, altering behavioural responses to odour cues (DeGennaro et al., 2013), and responses to CO_2 respectively (McMeniman, Corfas, Matthews, Ritchie, & Vosshall, 2014).

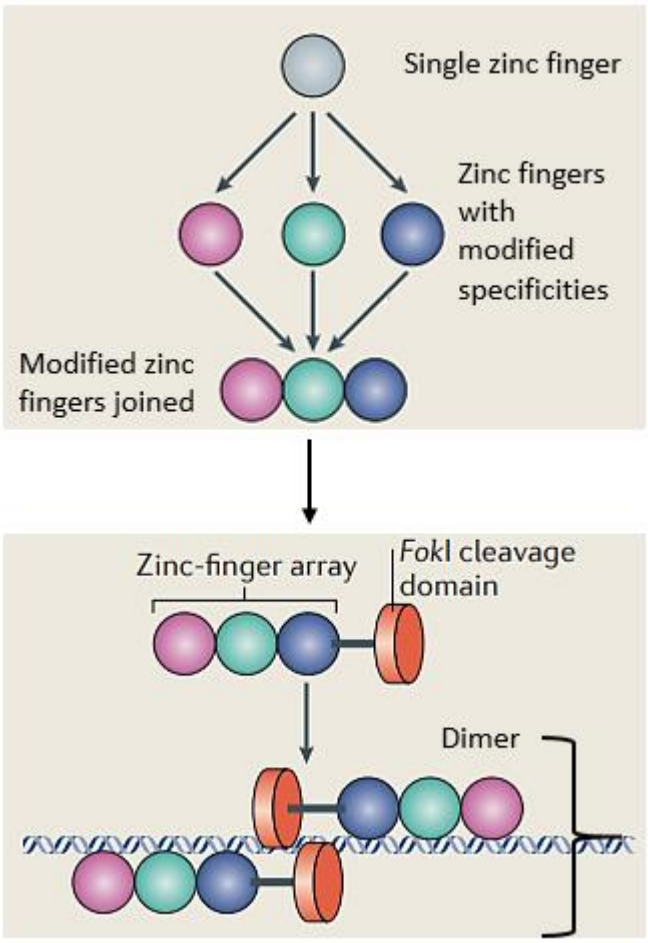


Figure 1. **Zinc Finger Nuclease (ZFN) modification of DNA.** Zinc Fingers are modified to recognises specific nucleotide sequences. These are joined together and fused to a FokI domain. Zinc fingers are represented by different coloured circles. The FokI nuclease is represented by the red disk. Two ZFNs form a dimer and bind to the DNA target. Between the two FokI domains is the spacer region where cleavage occurs. Adapted from Joung *et al.,* (2013).

TALE and TALENs

An alternative approach for DNA modification is the TALE/TALENs system. Transcription activator-like effectors (TALE) are transcriptional activators of genes, found in naturally occurring *Xanthomonas* bacteria. They contain a DNA-binding domain that binds at a tandem repeat sequence (Boch et al., 2009). The DNA-binding domain consists of repeat sequences (33-35 amino acids), containing two hypervariable amino acids (at position 12 and 13 of the repeat) (Figure 2). Each repeat binds to a single specific nucleotide base in sequence (Moscou & Bogdanove, 2009). The TALE binding domains have been fused to the non-specific *FokI* endonuclease. This makes a customisable endonuclease, a TALEN (Cermak et al., 2011). Dimers of TALENs form, bind to DNA, where cleavage occurs in a spacer region between the two TALE binding domains causing a double stranded break (DSB).

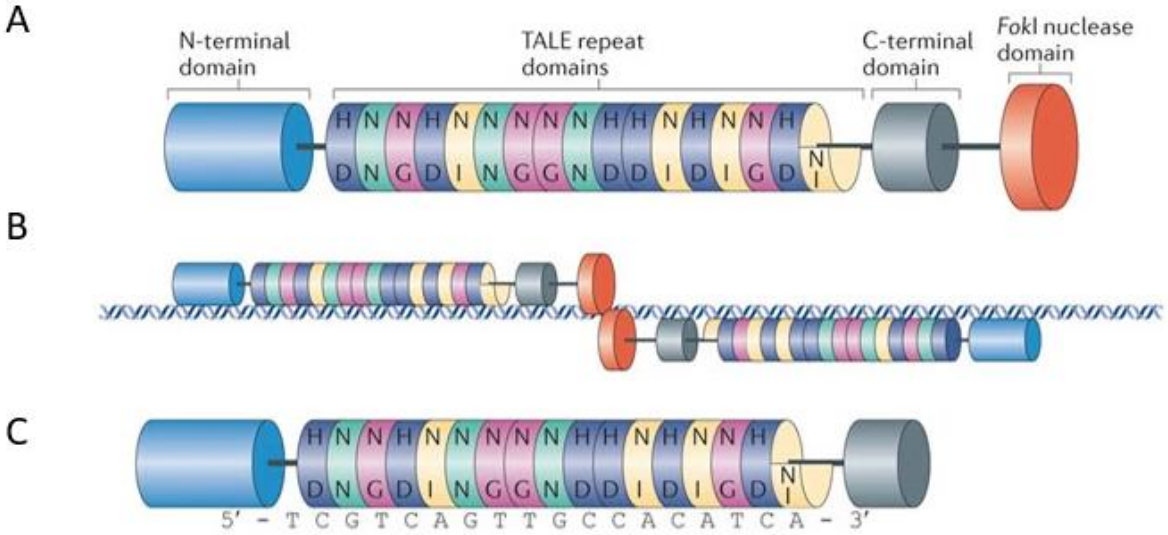


Figure 2. TALEN modification of DNA. A) TALE repeat domains with letters representing the hypervariable amino acid residues. The FokI nuclease is represented by the red disk. B) Two TALENs form a dimer and bind to the DNA target. Between the two FokI domains is the spacer region where cleavage occurs. C) The TALEN binding domain aligned with its specific target DNA sequence. Each repeat domain binds to a single base. Adapted from Joung *et al* (2013).

TALENs have been used successfully in *Ae. aegypti* to knockout the *kmo* eye-pigmentation gene, mutations which were inherited (Aryan, Anderson, Myles, & Adelman, 2013). They have also been demonstrated in *An. gambiae*, disrupting the *TEP1* immunity gene resulting in increased susceptibility to *Plasmodium* parasites and increased parasite infection levels within the gut, properties which were also inherited (Smidler, Terenzi, Soichot, Levashina, & Marois, 2013). While successful in editing vectors, TALENs may be less high-throughput with respect to generating targeted mutations (M. A. E. Anderson et al., 2015), and is time-consuming to construct (Gaj, Gersbach, & Barbas, 2013), resulting in this technique being superseded by other methods.

PiggyBac

One widely used method to integrate exogenous DNA into the genome of insect vectors is the use of transposable elements. A transposable element system was originally developed for *Drosophila melanogaster* (*P* element system). PiggyBac elements, the most commonly used the transposable elements, comprise a 2,472bp genetic element that has the ability to function like a ‘cut-and-paste’; whereby the transposon is excised (cut) from the PiggyBac vector, and integrated by ligation (paste) into genomic DNA.

PiggyBac is a class II (transposase encoding) transposable element isolated from *Trichoplusia ni* (Cabbage looper) cell lines (Fraser, Ciszczon, Elick, & Bauser, 1996). It has single stranded nucleotide sequences (terminal repeat) at each end of the transposable element that are asymmetric; the 5' repeat is 13bp, and the 3' repeat is 19 bp in length. The majority of the PiggyBac element is an open reading frame that encodes functional transposase. The advantage of PiggyBac over other transposons is the ability to integrate elements without changing the nucleotide sequence when removed. Transposons can be integrated into a PiggyBac vector/construct for *in vivo* delivery into organisms genome. PiggyBac constructs integrate DNA cargo sequences into genomic DNA at TTAA sites in the presence of transposase, which is often expressed in the form of a secondary (helper) construct.

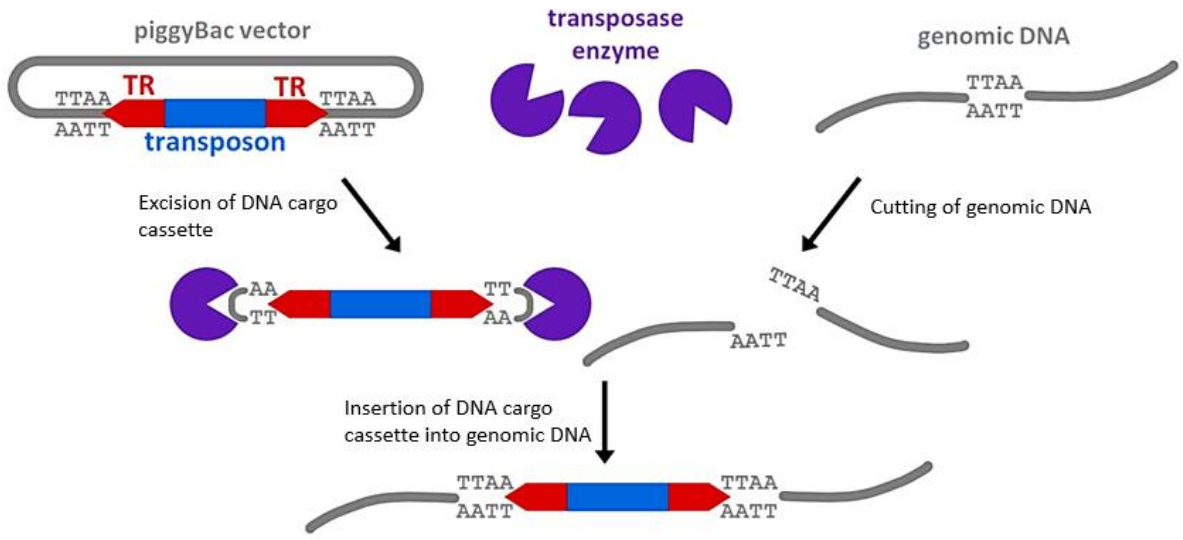


Figure 3. Integration of PiggyBac transposon. Transposase binds to the terminal repeats (TR) and excises the transposon from the PiggyBac vector, and cleaves genomic DNA at the TTAA recognition site. The transposon is inserted into the genomic DNA. Adapted from Vierl *et al* (2021).

The PiggyBac transposon facilitated by helper transposase, binds to the 5' and 3' terminal repeats, excising a DNA cassette (transposon) previously in that position. The transposon recognises a TTAA target site in genomic DNA, a recurring sequence found approximately every 256bp in eukaryotic DNA sequences. Asymmetric cuts occur creating short stretches of unpaired nucleotides at DNA (overhangs), which are complementary to the transposon. The transposon is integrated via ligation into the genomic DNA without the retention or addition of nucleotides (scarring) (Figure 38).

PiggyBac has been successfully applied to mosquito vectors including to express fluorescent markers and ant-parasitic genes: *Anopheles gambiae* (Grossman et al., 2001), *An. stephensi* (Ito, Ghosh, Moreira, & Wimmer, 2002; Nolan, Bower, Brown, Crisanti, & Catteruccia, 2002), *An. albimanus* (Perera, Ii, & Handler, 2002), *Aedes aegypti* (Kokoza, Ahmed, Wimmer, & Raikhel, 2001), *Ae.*

albopictus (Labbé, Nimmo, & Alphey, 2010), and *Ae. fluviatilis* (Rodrigues, Oliveira, Rocha, & Moreira, 2006). Importantly the technique has not successfully demonstrated in sand flies which is a significant omission for such a potentially useful tool.

CRISPR-Cas9

A highly precise gene editing approach is via CRISPR-Cas9. Successful transformations and use applied to vector species is described in detail in Chapter 2 (see *CRISPR – Prospects and opportunities for disease control*). Briefly, a guide RNA (gRNA) and a CRISPR-associated endonuclease 9 (Cas9) form a complex. The gRNA is homologous to the target region of DNA which is to be affected, directing the Cas9 endonuclease to cleave the target DNA sequence, inducing a double stranded break (DSB). One of two cellular DNA repair mechanisms Non-Homologous End Joining (NHEJ) or Homology Directed Repair (HDR). NHEJ is error prone, and results in insertions or deletions (indels), and HDR leads to the insertion of exogenous DNA in the presence of a suitable homology template (Figure 4).

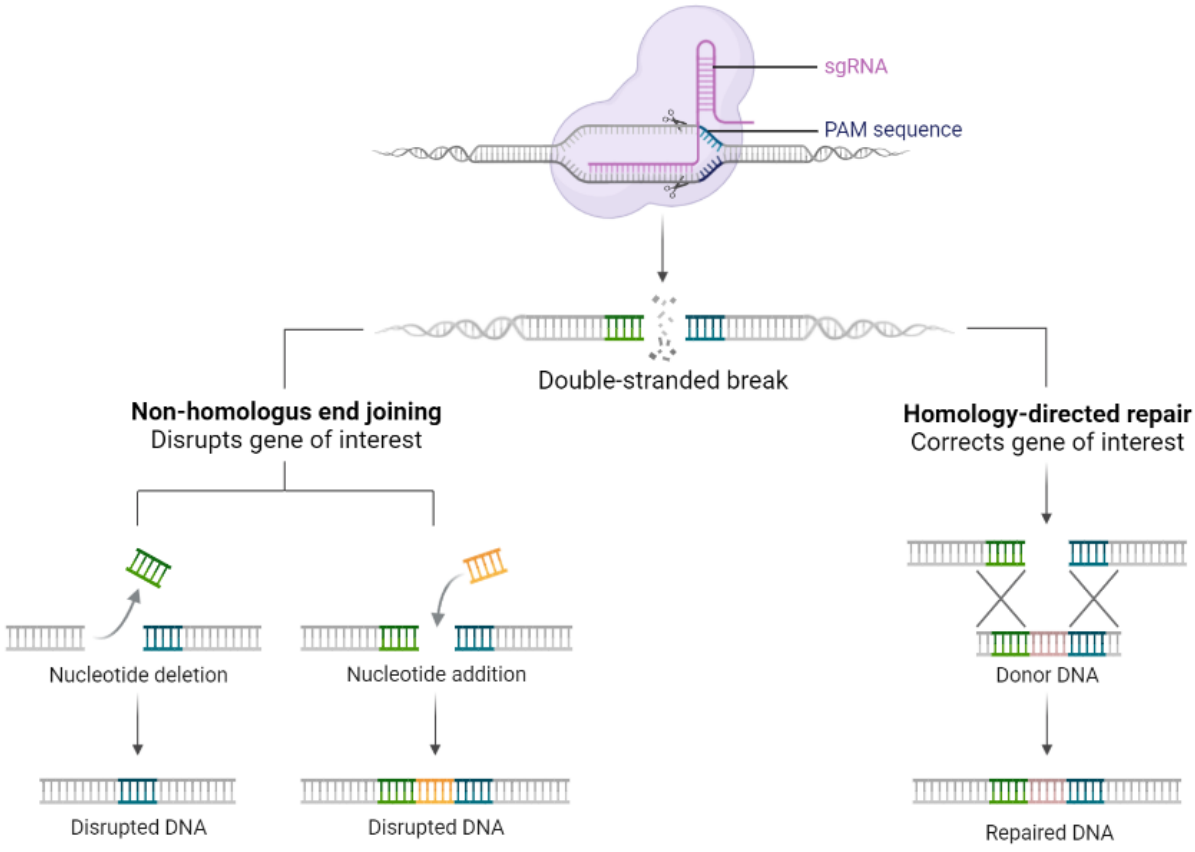


Figure 4. CRISPR-Cas9 modification of DNA and repair mechanisms. The formation of the CRISPR-Cas9 complex consisting of the Cas9 endonuclease and the sgRNA. The complex binds to complementary DNA before a DSB is made The DSB is repaired by NHEJ or by HDR if a homology repair template is provided. Figure from www.biorender.com.

One revolutionary approach is the development of CRISPR-Cas9 mediated gene drives for insect vectors using two different approaches, (1) To reduce reproductive capacity resulting in population elimination, or (2) to affect the ability of mosquitoes to transmit disease by blocking parasite development, for example by expression of anti-parasite peptides (detailed in Chapter 2). Briefly, a reduction in reproductive capacity was first demonstrated in *An. gambiae* by disrupting genes resulting in recessive female sterility (Hammond et al., 2016), and though the development of intersex *An. gambiae* with complete sterility (Kyrou et al., 2018). Disease transmission blocking modifications have been demonstrated by the development of a transgenic strain of *An. stephensi* with anti-*Plasmodium falciparum* effector genes which were induced upon blood feeding (Gantz et al., 2015), and through the expression of exogenous anti-microbial peptides in *An. gambiae* which reduce *Plasmodium* development within the vector (Hoermann et al., 2022). These studies successfully added gene drive elements to spread the CRISPR-Cas9 gene modifications through caged populations.

CRISPR component delivery methods for genome modification

Mutagenesis using CRISPR-Cas9 can be achieved using a number of strategies involving the delivery of Cas9 and gRNAs in various forms. Direct injection of Cas9 mRNA, or recombinant Cas9 protein alongside *in vitro* transcribed gRNAs have been successful in *Anopheles* (Gantz et al., 2015), *Aedes* (Basu et al., 2015; S. Dong et al., 2015; Hall et al., 2015; Kistler, Vosshall, & Matthews, 2015), and *Culex* (M. E. Anderson et al., 2019; M. Li et al., 2020; Purusothaman, Shackleford, Anderson, Harvey-Samuel, & Alphey, 2021), as have injections of plasmids expressing Cas9 alongside plasmids expressing gRNAs (Feng, López Del Amo, et al., 2021; Galizi et al., 2016; Hammond et al., 2016; Kistler et al., 2015; Kyrou et al., 2018; Simoni et al., 2020).

One approach, in terms of insect vectors, is to generate a line of transgenic flies that express the Cas9 protein. Subsequently, vectors expressing Cas9 can be injected (embryo microinjection) with *in vitro* transcribed gRNA or gRNA constructs, or crossed with individuals transgenically expressing gRNAs. This method has shown a high degree of success, resulting in efficient genome engineering. In *Cx. quinquefasciatus*, a homozygous line expressing Cas9 within the germline was derived by injection of CRISPR homology directed repair constructs and subsequent mating crosses. Expression of Cas9 protein was confirmed by injecting gRNAs targeting a specific phenotypic gene (*kh*, white eye) and identifying knockouts (Feng, López Del Amo, et al., 2021). In *Ae. aegypti*, a line transgenically expressing Cas9 was also achieved, providing a useful tool to conduct phenotypic knockouts and to optimise HDR integration studies (Ming Li et al., 2017), and in *An. gambiae* a line was generated and used to knockout the *FREP1*gene, supressing the development of *Plasmodium*, by crossing the Cas9 expressing line with gRNA-expressing transgenic lines (Y. Dong, Simões, Marois, & Dimopoulos, 2018).

Injecting plasmid constructs is less efficient at generating transgenics compared to the direct injection of CRISPR RNPs, however it is by far the simplest method with respect to delivery of required components. Verification of germline expression and inheritance of Cas9 is required, and several crosses may be needed to generate a stable line for future use. However, if successful, this provides a powerful tool for functional genetic studies via directed mutagenesis, and the development of gene drive technologies for controlling vectors of disease.

Genetic Modification in sand flies

Attempts at microinjection of sand fly embryos are few and far between, with only a limited number of examples in the published literature (Jeffries, Rogers, & Walker, 2018; Louradour, Ghosh, Inbar, & Sacks, 2019; Martin-Martin, Aryan, Meneses, Adelman, & Calvo, 2018).

In more detail, Jeffries *et al.* (2018) developed a microinjection methodology to inject the wMel strain of *Wolbachia* purified from *D. melanogaster*. They injected 1,815 *L. longipalpis* eggs, resulting in six (6/1815 = 0.33% survival to G0 female, post-injection) fertile female survivors. Fifty percent of the surviving females had *Wolbachia* infections. G1 progeny from the *Wolbachia* infected females resulted in two lines where maternal transmission had occurred. This transmission was maintained to the fourth generation, after which no *Wolbachia* was detected.

Martin-Martin *et al.* (2018) developed a protocol for microinjection, with limited results, demonstrating a hatching rate of 11.90-14.22% when injecting Cas9 mRNA and sgRNA into 775 (84, 269, and 422 injections per round) *L. longipalpis* eggs. Ten, 38, and 60 larvae hatched from three rounds of injections respectively, (11.90%, 14.13%, and 14.22% respectively), compared to non-injected wild type embryos which had a hatch rate of 64.7%. From 60 hatchlings, 42 survived (20 males, 22 females). Transgenic modification in emergent hatchlings were not detected.

In a key publication Louradour *et al.* (2019) successfully demonstrated CRISPR-Cas9 knockouts in *P. papatasi* via injection of Cas9 recombinant protein and *in vitro* transcribed gRNAs. Loss of function alleles were generated for the *relish* gene, a transcription factor in the immune deficiency pathway, with individuals becoming more permissive to *L. major* development in the gut. The number of adult survivors from injections was low (11/540 = 2.04%), however the rate of mutagenesis was high (8/11 = 72.72%). Mutations were inherited, however knockout lines unstable and lost.

Currently there is no published data on genetic modification of sand fly species (*Lutzomyia longipalpis)* through CRISPR mediated transgenesis or PiggyBac transformation to demonstrate knockin or knockout potential. In this chapter, we demonstrate microinjection of CRISPR constructs and assessment of mutagenesis in the two major sand fly vectors are described, providing the first

demonstration of PiggyBac insertion and inheritance in both species, towards development of transgenic Cas9-expressing lines as a tool for future studies. The use of CRISPR mutagenesis in *L. longipalpis* is demonstrated by microinjection of plasmid constructs containing Cas9 and gRNAs, with inheritance of mutagenesis in phenotypic marker genes, and olfactory genes involved in host seeking verified.

Aims

Functional genetics in sand fly species has only been attempted by a small number of research groups and has been limited by the lack of easy-to-use tools for targeted mutagenesis until recently. The research described in this chapter attempts to modify *L. longipalpis* and also *P. papatasi* sand flies using an *in vivo* high throughput pipeline to inject large number of sand fly eggs. Here we apply PiggyBac and CRISPR approaches (from Chapter 4) to demonstrate targeted mutagenesis. In addition, we sought to demonstrate *in vivo* knockouts of phenotypic marker genes as proof of concept, and also olfactory genes towards potential host seeking behavioural modifications, for incorporation into future gene drives to affect transmission. We validate different methods to detect mutagenesis in emergent transfected sand flies.

Methods

Sand fly colony rearing

The maintenance of successful sand fly colonies followed methods described by Volf and Volfova (2011). *Lutzomyia longipalpis* and *P. papatasi* adults were kept in large (50 x 50 cm), mixed cages, provided with access to 30% sugar solution (soaked cotton wool) placed on top of the cage. Cages were maintained at 25 - 28°C, 80% relative humidity, wrapped in plastic bags to maintain the humidity, and a 14hr:10hr light-dark cycle (Volf & Volfova, 2011).

For larval rearing, gravid females were aspirated into plastic pots (10cm diameter) with a plaster-of-Paris base. The pots were sealed with a mesh cover to prevent escape of adult females. Females lay eggs 6-10 days post blood-feeding (*L. longipalpis* 6-8 days post blood-feed), and were removed once oviposition has occurred. Larvae hatch from eggs 6-10 days later. The larval period lasts ~3 weeks (17-21 days for *L. longipalpis*). A small amount of food (Rabbit faeces and rabbit pellets, air-dried and ground to a powder (Volf & Volfova, 2011)) was applied to the larval pot surface. Larval pots were placed within plastic boxes containing a base of moist sand until they become pupae (Lawyer et al., 2017). The pupal stage lasts for ~7-10 days, before adult eclosion.

Females were offered a bloodmeal 1-2 times a week. They were then left for 24 hours before being transferred to small 20 x 20cm cages for 5-6 days, before transfer to the larval pots for oviposition (**Figure 5**).

L. longipalpis and *P. papatasi* sand fly colonies were reared and maintained at scale in the laboratory of Professor Petr Volf (Charles University, Prague, Czech Republic), and all microinjection studies were conducted within the laboratory.

Sand fly embryo microinjection methods

In advance of experiments, *L. longipalpis* and *P. papatasi* sand flies were prepared using the protocol developed by Jeffries *et al.,* (2018). In summary, 3-day-old post-blood-fed females were transferred to laying pots in groups of 15-20, and kept in the dark for one hour. Laying pots (modified 50ml Falcon tubes) contained an agarose surface (2%) for oviposition within the lid (Figure 5). Sand flies were aspirated from the laying pots, after one hour, and placed into a fresh laying pot to allow for high turnover of egg laying. Eggs were collected using a fine paintbrush made damp with sterile water, and transferred to a glass slide. Eggs were aligned against a nitrocellulose membrane kept moist using Whatman filter paper (Figure 5).

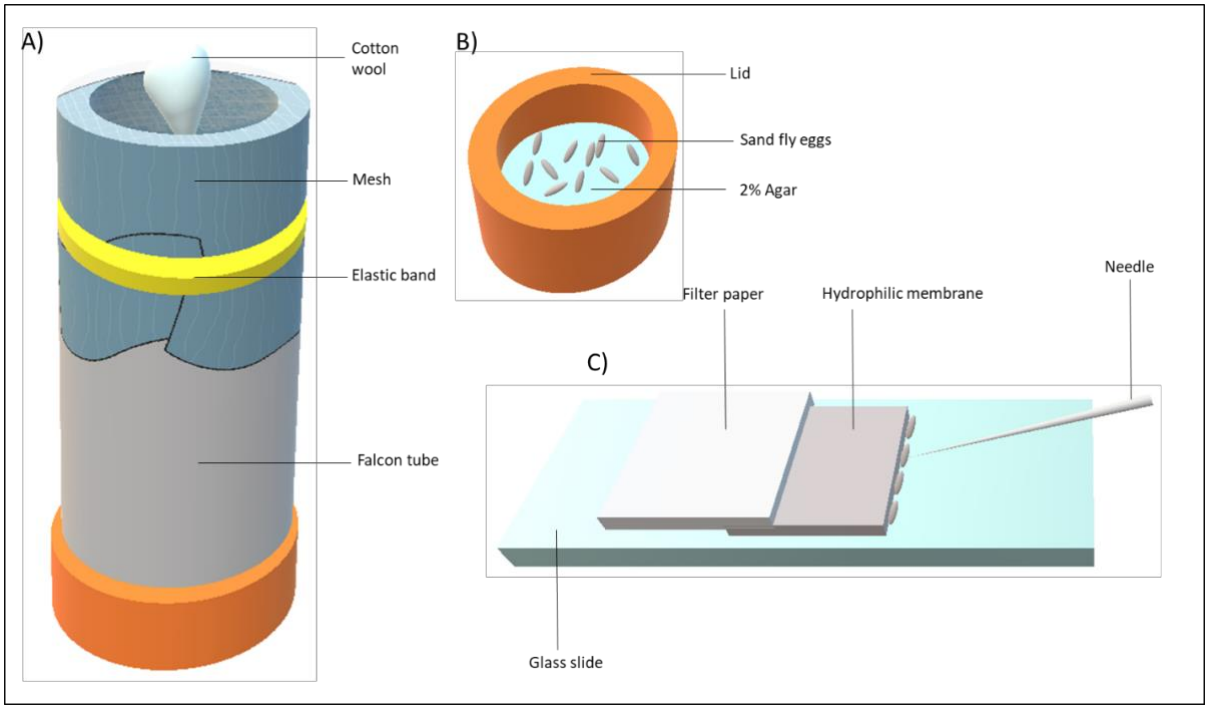


Figure 5. **Diagram of microinjection preparation protocol.** A) Gravid female sand flies are removed from rearing cages using an aspirator and placed into the oviposition chamber via a slit in the mesh, plugged with cotton wool. B) The lid of the oviposition chamber contains 2% agar to provide a substrate for oviposition. C) Eggs are aligned on a glass slide against a hydrophilic membrane, which is kept

moist by filter paper to prevent egg desiccation. Eggs are injected using a fine borosilicate needle containing the injection mixture at ~25% of the length from the posterior pole.

Microinjection Needle (pulling)

Needles were pulled with a Narishige PC-10 needle puller using Narishige G-1(Narishige, Japan) borosilicate glass capillaries (outer diameter 1mm; inner diameter 0.6mm). A two-step pulling method was used with the following settings: Heater No.1 = 65.1; Heater No.2 = 75; upper shutter = 2.5; lower shutter = 5; All weights provided were used. Using these settings the needles have a medium-length sharp point. An aperture at the tip of the needle was opened by bumping the tip against a glass cover slip or by using a beveller set at 20-30°.

Microinjection mixture

A 10µl injection mixture was prepared prior to each injection session and incubated at 37°C for 10-20 minutes, before being placed on ice for 20 minutes prior to injections. 2µl of the mixture was backloaded into the needle, ensuring no air bubbles were present at the tip of the needle.

For generation of PiggyBac transformation lines the mixtures consisted of a PiggyBac plasmid (UbiqCas9.874w or pHome-T, (**Error! Reference source not found.**)) (~200ng/µl) and the IhyPBase transposase helper plasmid (~400ng/µl), giving a 1:2 ratio. The injection mixtures for UbiqCas9.874W (Addgene, USA) plasmid containing GFP and RFP markers, and Cas9 coding region, and pHome-T plasmid containing GFP and RFP markers, were prepared in the same manner (see Appendix 3).

For CRISPR knockout lines, all plasmids targeting the same gene were prepared for simultaneous injection. Three *Rudimentary*-targeting, and three *vestigial*-targeting pDCC6 constructs were combined in an even ratio to derive a final concentration of 317.58ng/µL. Six olfactory-targeting plasmids were combined the injection mixture consisted of an even ratio of all constructs for a final concentration of 316.68ng/µL. Four ebony pDCC6 constructs were combined for a concentration of 278.48ng/µL, and four Caspar pDCC6 constructs were combined for a concentration of 214.48gn/µL (see Appendices 3-5 for protocols, and Appendices 6-7 for injection mixtures).

For CRISPR homology directed repair knockin lines, the injection mixture contained two pDCC6 knockout constructs with gRNAs targeting the *Scarlet* gene (5`LLst pDCC6 and 3`LLst pDCC6) (constructed via restriction cloning), and a donor construct containing regions of homology to the *Scarlet* gene, flanking a DNA insert for the expression of a DsRed-Express fluorescence protein (LLst Ubiq DsRed) (constructed via Gibson Assembly using a backbone pDsRed-attp plasmid (Addgene, Plasmid #51019). The mixture contained 100ng/µl of each construct, giving a final mixture ratio of 1:1:1 (see Appendix 5)

Microinjections

Injections were performed using a manual Narishige IM-9B microinjector and Narishige MMO-4 (Narishige, Japan) micromanipulator mounted on a Leica DMi8 inverted microscope (Leica, Germany). A steady flow of injection mixture was generated from the needle tip by increasing the pressure in the microinjector and each egg was injected between 1/4 and 1/3 of the length of the egg from the posterior pole (Jeffries et al., 2018). Once injected the glass slides containing the eggs were transferred to a petri dish containing moist filter paper, and deposited at 25-28°C, 80% relative humidity. After ~1-3 days egg were brushed from the glass slides into standard oviposition pots.

Identification of transgenic insects

Fluorescent microscopy

Initially emergent transgenic insects were imaged by fluorescence stereo microscope (Leica M205 FA, Germany). Here, L1-L4 hatching larvae were imaged to identify fluorescence. The GFP and EGFP marker are excited by 485nm and 488nm respectively, the DsRed1 and RFP markers viewed at 558nm and 587nm respectively (Table 1).

Potential outcomes for G0 injection survivors of PiggyBac UbiqCas9.874W transfections are red fluorescent mosaicism demonstrating transient plasmid expression (*Opie2* promoter of DsRed), and green fluorescent mosaicism indicating expression of Cas9 due to Cas9 tagged with GFP, separated by T2A ribosome skipping element. G0 survivors from pHome injections will express eye specific expression of GFP driven by *3xP3*, and red fluorescent mosaicism if the plasmid is integrated.

G1 survivors successfully expressing fluorescent proteins will demonstrate integration of plasmid cargo into the sand fly genome, and demonstrate inheritance of inserted sequences This indicates plasmids were delivered early in the development of G0 embryos to affect insertion in germline cells.

Table 1. Fluorescence markers in PiggyBac and non-integrative constructs used for microinjection.

Plasmid	*Fluorescence*	*Promoter*	*Excitation (nm)*	*Emission (nm)*	*Outcome*
UbiqCas9.874W	EGFP	Ubiquitin 63e	488	507	Cas9 expression
	DsRed1	Opie2	558	583	Integration marker
pHOME	GFP	3xP3	485	510	Eyes of adults
	RFP	Actin5c	587	637	Integration marker
Ac5-STABLE-Neo	EGFP	Actin5c	488	507	
	Excitation and emission wavelengths from: www.fpbase.org				

Genomic DNA extraction

50% of larvae were collected to extract DNA and to conduct sequencing. The remaining 50% were allowed to pupate and emerge as adults for evaluation of genetic modification, and subsequent crossing (see *Outcrossing/Backcrossing* below).

Genomic DNA was extracted using a DNeasy Blood and Tissue Kit (Qiagen, Germany) or Monarch Genomic DNA Purification Kit (NEB, USA) following the manufacturer's standard protocol. Briefly, for DNeasy, sand fly larvae or adults were homogenized in a 1.5ml tube using a pestle with the addition of kit-specific buffer and Proteinase K. The homogenate was incubated for 3 hours at 56°C, followed by the addition of further buffers and several spin column centrifugation steps. The spin column was then placed in a 1.5ml tube and 30µl of Elution Buffer was added. Extracted DNA was quantified using a Nanodrop spectrophotometer to determine the concentration.

For the NEB Genomic DNA Purification kit, the manufacturer's Animal Tissue protocol was followed. Sand fly larvae or adults were homogenised with Tissue lysis buffer, followed by Proteinase K. The mixture was incubated at 56°C. Binding buffer was added to the sample and transferred to a spin column for centrifugation and wash steps. 30µl of Elution Buffer was added to the spin column to elute into 1.5ml tubes. A Nanodrop spectrophotometer was used to determine the concentration.

PCR Amplification

PCR amplification of genomic sand fly DNA was conducted in 20µl PCR reaction using 2x Phusion HF Master mix (NEB, USA), or Phusion High Fidelity PCR kit (NEB, USA) with 10 µM forward and reverse primers (see Chapter 4, *In vitro transcription: Confirmation of gRNA sequences*), 1µl DNA, DMSO and up to a total volume with nuclease free water. The samples were run in a thermocycler as follows: 5 minutes 98°C for 30 seconds, followed by 30 cycles of 30 seconds at 98°C for 10 seconds, 60°C for 30 seconds, 72°C for 45 seconds at, followed by 72°C for 5 minutes.

Gel extraction and PCR Cleanup

PCR products were gel extracted or purified using Monarch DNA Gel Extraction Kit (NEB, USA) or Monarch PCR and DNA Cleanup Kit (NEB, USA) following the manufacturer's standard protocol.

For gel extraction, 3-4µl of gel loading dye was added to each PCR product, and 20µl was run on a 1.5-2% agarose gel with relevant controls and ladders (Hyperladder 1kb and Hyperladder 100bp, Bioline). Gels were viewed on a UV trans-illuminator and desired fragments excised. Fragments were then purified using the Gel Extraction Kit, and eluted into 30µl of pre-warmed Elution Buffer.

For PCR cleanup, 1µl of gel loading dye was added to each PCR product, and 5µl was run on 1.5-2% agarose gel by electrophoresis. The gel was then viewed on a gel dock to identify the PCR products for

clean-up. The products were purified using the PCR and DNA Cleanup kit following the Cleanup protocol, and eluted into 10-30µl of pre-warmed Elution Buffer.

Sanger sequencing

Sequencing was conducted by Genewiz Inc (Leipzig, Germany) using 5µl of Purified PCR product (10-50ng/µl) and 5µl of 5mM primer.

Primers were designed to identify the presence of integrated PiggyBac constructs by targeting elements of the plasmid that do not exist within the sand flies without integration (Table *2*).

For the PiggyBac Ubiq-Cas9.874W plasmid primer pairs were designed targeting the 5' and 3' regions of the Cas9 sequence in addition to primers confirming integration of GFP and RFP. For the second PiggyBac pHome-T plasmid primers were designed flanking the GFP and RFP.

Primers were also designed to flank genes of interest that were targeted using pDCC6 CRISPR knockout constructs (Table *2*).

Table 2. **Primers to identify transgenesis in G0 injection survivors and G1 survivors of mating crosses.** DNA was extracted from L4 larvae and pooled in batches of 10, or extracted from adults. Primers amplified the presence or lack of fluorescence insertion, and Cas9 plasmid cargo. Primers flanking Olfactory and non-lethal phenotypic wing gRNA targeting sites were designed.

Plasmid	*Target gene*	*Primer 1*	*Primer 2*	*Fragment size (bp)*
pHome-T	GFP	GFP_001: ATGGTGAGCAAGGGCGAGGA	GFP_002: TGTACAGCTCGTCCATGCCG	715
	RFP	RFP_001: ATGGTGCGCTCCTCCAAGAA	RFP_002: CTACAGGAACAGGTGGTGGC	681
Ac5-STABLE1-Neo	EGFP	EGFP_001: ATGGTGAGCAAGGGCGAGGA	EGFP_002: TTGTACAGCTCGTCCATGCC	716
Ubiq-Cas98.74W	EGFP	EGFP_001: ATGGTGAGCAAGGGCGAGGA	EGFP_002: GTACAGCTCGTCCATGCCGA	714
	DsRed1	DsRed_001: ATGGTGCGCTCCTCCAAGAA	DsRed_002: CTACAGGAACAGGTGGTGGC	681
	Cas9	Cas9_001: CAAGAAGTACAGCATCGGCC	Cas9_002: GTAGCCGTTCTTGCTCTGGT	1072
	Cas9	Cas9_009: ACTGCAGAAGGGAAACGAAC	Cas9_010: TGAACTTGTGGCCGTTTACG	635
Gr2-1-pDCC6 Gr2-2-pDCC6	Gr2	Prim#1: GAGATTTCGTGCAGGTGACA	Prim#2: GGCAAAGATAAAGAGCAGCG	1158
Ir8a-1-pDCC6	Orco	Prim#3: TCACACACAGCATCACGAAA	Prim#4: CGCATCCATACCCAAACTTTA	810

Ir8a-2-pDCC6				
Orco-1 - pDCC6 Orco-2-pDCC6	Ir8a	Prim#1: ACCCGTTGCAAATTCTGAAG	Prim#2: GGGAGACTTGTAGTCGCCAA	634
R1-pDCC6	Rudimentary	r_1L: AATCAAATTCCCATCGAACG	r_1R: GAAACGGCAACAAAGGTGAT	530
R2-pDCC6	Rudimentary	r_2L: AACGCGGAAAAGGGAGTATT	r_2R: GTTGCATGCTTTGGGGATAA	586
R3-pDCC6	Rudimentary	r_3L: GCACAATACGCCCTGAATTT	r_3R: AAGCTAAAGAAATGCCCGTG	534
V1-pDCC6	Vestigial	v_1L: GTTGGTCGTGATGCAATCTG	v_1R: AAATTTTAGCCCGGGAAATG	577
V2-pDCC6	Vestigial	v_2L: TTACCACGCGAGATGAAAGG	v_RL: TATCAAAAGGGCCGAACAAG	567
V3-pDCC6	Vestigial	v_3L: CAAGGAGGCACACAATTGAA	v_RL: TCGCCAAACCTTAGATAACAA	562

Outcrossing/Backcrossing

Multiple mating strategies were used for survivors of microinjections. Following the first round of UbiqCas9.874W injections for *L. longipalpis* and *P. papatasi*, 50% of larvae were allowed to pupate and eclose (50% of larvae had DNA extracted). G0 adults were separated by sex, observed for transient expression of fluorescence, and outcrossed with wildtype adults of the opposite sex. G1 larvae were observed for fluorescent phenotypes by microscopy, and prepared for sequencing if they did not survive to G1 adults.

In the second round of UbiqCas9.874W microinjections, all *L. longipalpis* larvae were allowed to pupate and eclose, with adult survivors outcrossed with wildtype. In the second round of *P. papatasi* microinjections with UbiqCas9.874W, G0 were sibling crossed. Separately, G0 male and females *L. longipalpis* injected with wing-targeting constructs (V1-pDCC6, V2-pDCC6, V3-pDCC6, R1-pDCC6, R2-pDCC6, R3-pDCC6) were sibling crossed (see Figure 8), as were *L. longipalpis* injected with olfactory constructs (Gr2-1-pDCC6, Gr2-2-pDCC6, Ir8a-1-pDCC6, Ir8a-2-pDCC6, Orco-1-pDCC6, and Orco-2-pDCC6).

Full diagrams of these mating schemes are located in the appendices (Appendix 1 and 2).

Detecting mutagenesis by T7 Endonuclease I Heteroduplex assays

To detect the presence of the CRISPR modifications of in transfected sand flies the Engen Mutation Detection Kit (NEB, USA) was used following the manufacturer's protocol. Briefly, a PCR reaction is setup using genomic DNA, primers, DMSO, nuclease free water and 2x Phusion HF master mix. The

reaction conditions were as follows: 98°C for 30 seconds, followed by 35 cycles of 98°C for 5 seconds, 60°C for 10 seconds, 72°C for 20 seconds, followed by an extension step at 72°C for 2 minutes, before the reaction is held a 4°C.

Control PCR reaction consists of Q5 Hot Start HF 2x master mix, control template and primers, and nuclease free water. The reaction is thermocycled as before, with the annealing temperature increased to 65°C.

Next, an annealing reaction is set up using PCR product, 10x NEBbuffer 2 and nuclease-free water. The reaction is thermocycled as follows: 95°C for 5 minutes, 95°C reducing to 85°C at -2°Cper second, 85 reducing to 25°C at -0.1°C per second, hold at 4°C. The heteroduplex digestion reaction is set up with the annealed PCR product and T7 Endonuclease I, incubated at 37°C for 15 minutes, followed by the addition of Proteinase K, and 5 minutes incubation at 37°C to inactivate the T7 Endonuclease I (T7EI). The fragments were separated by gel electrophoresis.

A target region is amplified via PCR. If modification such as insertions or deletions (indels) have occurred, the amplicon pool will contain homoduplexes of unmodified DNA and homoduplexes of mutated DNA. When the PCR amplicons go through rounds of denaturation and annealing, heteroduplexes can form between the amplicons with a modified locus and un-modified locus, in addition to re-formation of homoduplexes. The T7EI recognises mismatches (heteroduplexes) and cleaves both strands (DSB) resulting in two fragments (see results section, *T7 Endonuclease I (T7EI) Heteroduplex Assay and Densitometric Analysis*). Gel electrophoresis is used to separate products of the heteroduplex assay, and the fragments generated can be used to determine the efficiency of genome editing at the region of interest.

Densitometric Analysis

Subsequent to T7EI heteroduplex assays, mutation efficiency was determined using Densitometric analysis of gel bands to quantify DNA mass of the T7EI cleavage products. Images taken on a gel doc were analysed in Image Lab software (Bio-Rad, USA). Bands of interest, including the un-cleaved DNA, and fragments generated by T7EI cleavage, were defined using the Lane Profile tool with the background excluded. The Analysis Table function was used to generate the Adjusted volume (background excluded volume) and the Band Percent (percent of the adjusted volume of the selected band, relative to the other selected bands in the lane (e.g. % density of the sum of the Adjusted volume)). The following formula (NEB, USA) was used to calculate estimated percentage modification of DNA:

Estimated % modification: $100 \times \left[1 - (1 - X)^{1/2}\right]$

$$X = Conc.digestedproduct \div (Conc.digestedproducts + Conc.undigestedband)$$

Where Adjusted volume values generated by Image Lab software represent concentration of digested and undigested products.

Algorithmic deconvolution analysis of Sanger sequence data - ICE Analysis

CRISPR-Cas9 validation strategies to determine the success of gene editing included the use of next generation sequencing (NGS), mismatch detection assays (e.g. T7EI), or Sanger sequencing followed by algorithmic deconvolution analysis. The gold standard is amplicon sequencing/next generation sequencing (NGS) (followed by algorithmic analysis), which provides the most resolution in the context of characterising gene edits, however this is a costly and time consuming process that is not accessible in all research settings. Sanger sequencing is a simpler alternative, however the sequence are a mixture of the possible genotypes present within the DNA that has been extracted, the dominant variant is likely the one observed in the Sanger sequence data. The sequence metadata can be analysed and resolved into its constituent elements (deconvoluted) to identify the potential edits that may have occurred by use of different algorithms. These include Sanger sequence data (TIDE (Brinkman, Chen, Amendola, & Van Steensel, 2014), DECODR (Bloh et al., 2021), CRISP-ID (Dehairs, Talebi, Cherifi, & Swinnen, 2016), ICE (Conant et al., 2022)) or to analyse Amplicon sequence data (TIDER (Brinkman et al., 2018) and CRISPResso (Pinello et al., 2016))

ICE analysis (Synthego Performance Analysis, ICE Analysis. 2019. V3.0. Synthego; Oct. 2022) was selected after identification of its use in multiple invertebrate species to identify CRISPR edits ((M. E. Anderson et al., 2019; Chen et al., 2021; Sharma et al., 2022). Sanger sequence data in ab1 format are uploaded to the webtool (https://ice.synthego.com/#/) comprising a control file for the sequence of interest, and the experimental file to be analysed alongside a guide sequence (17-23 nucleotide RNA sequence without the PAM).

Graphical outputs from ICE analysis include a Trace file, a Discordance graph, and an Indel size-frequency histogram (Figure *6*). The Trace file is a comparison of the edited sample and the control Sanger sequence. The gRNA sequence is underlined in the control sample, the predicted Cas9 cut site is shown with a vertical dotted line, and the PAM sequence is underlined with a red dotted line.

The Discordance graph shows the difference between the control Sanger sequence trace file (orange) and the putative edited Sanger sequence trace file (green). Within the Discordance graph, a dotted black line represents the cut site. A pink line on the graph shows the alignment window, where the control and putative edited sequences is used to align the two sequences. An orange line on the graph (Interference window) is a region of the Sanger sequence traces around the cut site where the ICE

algorithm infers the difference between the control and edited Sanger sequence traces. The Indel size-frequency histogram shows the size of indels determined by the ICE algorithm and their respective frequencies.

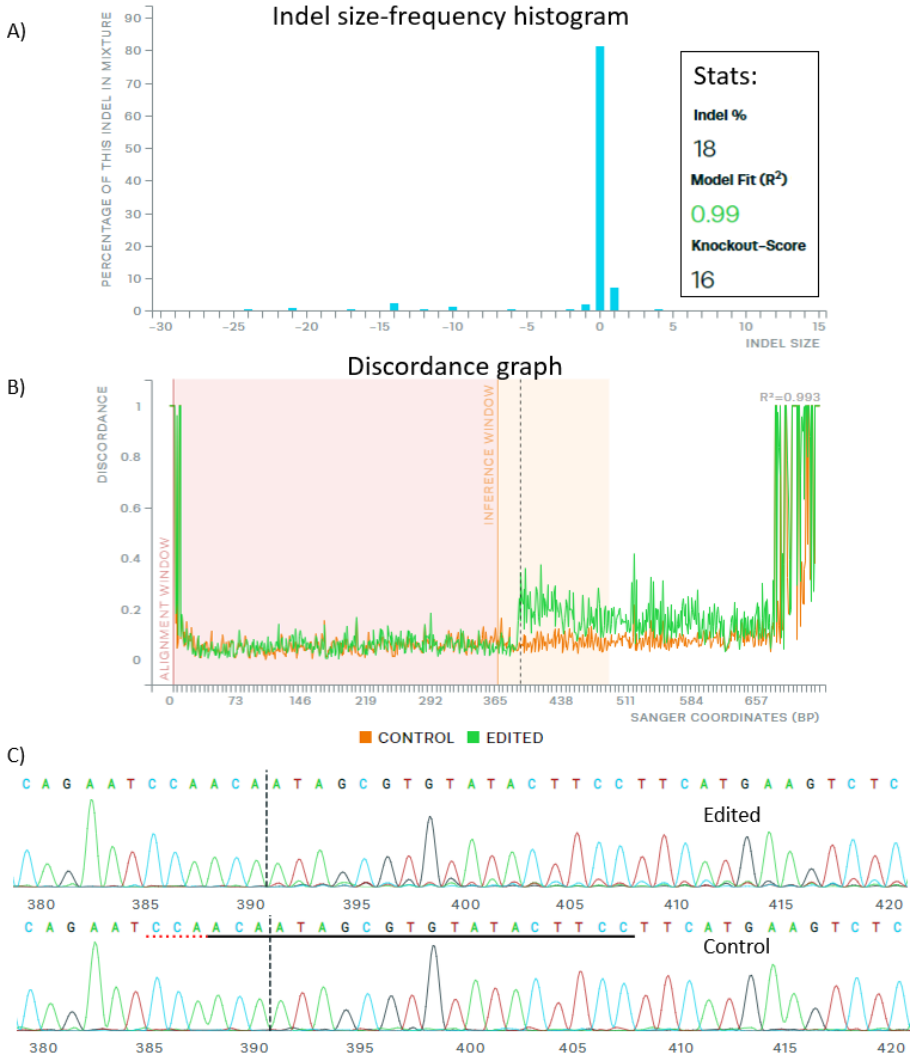


Figure 6. **ICE analysis outputs from example data**. A) Indel size-frequency histogram, showing predicted indel sizes, and statistical outputs. B) Discordance graph, the difference between the control and edited sequence traces C) Trace files comparing the control sequence to the edited sequence, showing the location of the PAM sequence (red dotted line), and the Cas9 cut site (black dotted line).

The Statistical Outputs from ICE Analysis (of relevance to the knockout experiments) include ICE, KO-score, and R^2 values. ICE (indel %) is the proportion of genotypes that contain an indel. The KO-score is a measure of the genotype variants that are likely to result in functional protein knockout via a frame shift. The R^2 value is the model fit. This value is the Pearson correlation coefficient. It is a measure of how well the ICE analysis algorithm and indels predicted by the algorithm fit the edited sequence data that is observed. If the $R^2 = 1$, the model perfectly fits the indels in the edited sequence data. An R^2 value <1, for example $R^2 = 0.7$ shows 70% of the indels in the sequence data were predicted

by the model, and 30% of indels are unexplained by the model. The higher the R^2 value the more confidence there is that the ICE values and KO-score are robust. An R^2 value ≥ 0.8 demonstrates robust analysis.

Results

The results in this chapter are separated into sections based on the analysis method used (Sanger sequence alignment, heteroduplex assays and densitometric analysis, and ICE analysis), and are described in detail. A subsequent summary of the results section (see *Summary Result*) considers the implications of the combined outputs..

Microinjections

Constructs were designed to target phenotypic marker genes and genotypic targets in both *L. longipalpis* and *P. papatasi* (see Chapter 3). For *L. longipalpis,* constructs were designed to target wing development genes, *Rudimentary* and *Vestigial*, and olfactory genes *Gr2*, *IR8a*, and *Orco*. Three constructs were built for each of the wing development genes, and two constructs were built to target each olfactory gene. For *P. papatasi* constructs were designed to target the *Ebony* gene involved in cuticle pigmentation, and the *Caspar* gene involved in sand fly gut immunity (Table *3*).

Separately PiggyBac constructs UbiqCas9.874W and pHOME were also microinjectioned as was the Ac5-STABLE1-Neo non-integrative plasmid. Numbers of microinjections for each experiment follows.

Table 3. **pDCC6 CRISPR knockout constructs targeting olfactory, wing development, pigmentation and immunity genes in *L. longipalpis* and *P. papatasi*.** gRNA (green) with overhangs (black) for Gibson assembly into the pDCC6 backbone. AAAC and AATT are overhangs for *L. longipalpis*, ands AATT and AAAC are overhangs for *P. papatasi*.

Species	*Gene*	*gRNA rank*	*gRNA sequence*	*Construct name*
L. longipalpis	Gr2	2	**AAAC** ATAAAGAGGCAATACGTGG **C**	Gr2-1-pDCC6
		4	**AATT G** GCTGTGTACAAGACAATGTGGGG	Gr2-2-pDCC6
	Orco	1	**AATT G** TGTGAGATACATGACCAACAAGG	Orco-1-pDCC6
		3	**AATT G** TACAGCAATCAAGTATTGGGTGG	Orco-2-pDCC6
	Ir8a	2	**AAAC** AGCAATAGCGAGACCTGCGAGGG **C**	Ir8a-1-pDCC6
		3	**AATT G** CCCTGTCGGGATTTTACACTCAA	Ir8a-2-pDCC6
	Rudimentary	27	**AATT** AGCATTGGAAGACACAGGGT	R1-pDCC6
		4	**AATT** TTGTAGCCCATCTTCACGA	R2-pDCC6
		1	**AATT** GGAACTATGGCATTCCGTG	R3-pDCC6
	Vestigial	2	**AATT** TCATCATTACGGTTCCTACG	V1-pDCC6
		16	**AATT** GGAAAATTTCTCGCCGACAT	V2-pDCC6
		38	**AATT** TCGCGGACACGTATTGTGCT	V3-pDCC6
P. papatasi	Ebony	2	**ATTT** TCGCATTCAGCACATCCTTG	Ebony1-pDCC6

		3	ATTT AAAGTGCATGGTAATCAGGA	Ebony2-pDCC6
		4	ATTT CCTGGCCATATGGAAATGTG	Ebony3-pDCC6
		5	ATTT TCGGACAACTTGATAGCCAC	Ebony4-pDCC6
	Caspar	1	ATTT GGATTCTGAGAGTTCCATGG	Caspar1-pDCC6
		2	ATTT TATCCTCAAGAATCTCAATG	Caspar2-pDCC6
		5	ATTT CCCTACAGCGTTGCTTCATC	Caspar3-pDCC6
		6	ATTT TGAGGCTATCTATTTGCTAG	Caspar4-pDCC6

Three rounds of embryo microinjections took place rotating between the *L. longipalpis* and *P. papatasi* colonies (Table *4*). In total 10,749 eggs were injected (5,375 and 5,374 for *L. longipalpis* and *P. papatasi*, respectively), with the overall adult survival for both species as a percentage of larvae of ~ 66.30%, and adult survival as a percentage of eggs injected of 1.70% (see Appendix 7).

For PiggyBac injections with UbiqCas9.874W in *L. longipalpis*, low mixture concentration (61.41 ng/µl) resulted in the highest larval survival (6.6%). Approximately tripling the concentration (186.29 ng/µl) reduced larval survival to 1.13%. In *P. papatasi*, low UbiqCas9.874W mixture concentration (59.86 ng/µl) resulted in no larval survivors. Doubling the concentration of the injection mixture (120.26 ng/µl) resulted in 3.47% larval survival (Table 4). For PiggyBac injections with the pHome plasmid at a low mixture concentration (61.66 ng/µl), a higher proportion of larvae survived in *L. longipalpis* (7.61%), compared to *P. papatasi* (5.25%).

Injections using the non-integrative Ac5-STABLE-Neo construct at 98.35 ng/µl resulted in a higher survival (8.67%) in *L. longipalpis* (Table 4), however there were no survivors in *P. papatasi*. None of the survivors expressed GFP when viewed under a fluorescence stereo microscope.

Injections of *L. longipalpis* using a mixture of CRISPR plasmids targeting *Rudimentary* and *Vestigial* wing genes (317.58 ng/µl) resulted in larval survival of 1.04%. Larval survival rate with the plasmids targeting Olfactory genes (*Gr2*, *IR8a*, and *Orco*) at 316.68 ng/µl was 1.74% (Table 4).

Injections of *P. papatasi* using *Ebony* pigmentation targeting CRISPR plasmids resulted in 0.77% larval survival. When targeting the immunity gene *Caspar*, larval survival from injections was 0.33%. Injections of *L. longipalpis* with HDR plasmids (201.69 ng/µl) resulted in a low larval survival rate of 0.43% (Table 4).

Table 4. Injection survival results for injections of *L. longipalpis* and *P. papatasi* eggs injected with multiple plasmids. Counting larvae is complicated with sand flies resulting in undercounting. This can lead to adult survival as % of larvae being greater than 100% (indicated by *†).*

Injection round	***Species***	***Plasmid(s)***	***Injected eggs***	***Larval survival (%)***	***Adult survival Male***	***Adult survival Female***	***Adult survival as% of larvae***	***Adult survival as% of eggs injected***
1	*L. longipalpis*	UbiqCas9.874W + IhyPBase	1031	68 (**6.6**)	24	12	52.94	3.49
	P. papatasi	UbiqCas9.874W + IhyPBase	1267	44 (**3.47**)	14	9	52.27	1.82
	L. longipalpis	pHome + IhyPBase	749	57 (**7.61**)	15	21	63.16	4.81
	P. papatasi	pHome + IhyPBase	800	42 (**5.25**)	12	9	50.0	2.63
	L. longipalpis	Ac5-STABLE-Neo	173	15 (**8.67**)	0	0	0	0
	P. papatasi	Ac5-STABLE-Neo	197	0 (**0**)	0	0	0	0
2	*L. longipalpis*	pDCC6: Gr2-1,Gr2-2,Ir8a-1,Ir8a-2,Orco 12, Orco 34	576	10 (**1.74**)	9	2	110†	1.91
	P. papatasi	pDCC6: Ebony1, Ebony2, Ebony3, Ebony4	1034	8 (**0.77**)	6	6	150†	1.16
	L. longipalpis	pDCC6: R1, R2, R3, V1, V2, V3	579	6 (**1.04**)	3	3	100	1.04
	P. papatasi	pDCC6:* Caspar1, Caspar2, Caspar3, Caspar4	922	3 (**0.33**)	0	2	66.67	0.21
	L. longipalpis	HDR pDsRED, 3' St pDCC6, 5' St pDCC6	701	3 (**0.43**)	1	2	100	0.43
	L. longipalpis	PiggyBac: UbiqCas9.874w + IhyPBase	533	6 (**1.13**)	3	2	83.33	0.94
3	*P. papatasi*	pDCC6: Caspar1, Caspar2, Caspar3, Caspar4	594	9 (1.52)	2	2	44.44	0.67
	L. longipalpis	HDR pDsRED only	539	4 (0.74)	2	2	100	0.74
	P. papatasi	PiggyBac: UbiqCas9.874w + IhyPBase	560	n/a	15	4	n/a	3.39

	L. longipalpis	PiggyBac: UbiqCas9.874w + IhyPBase	494	1 (0.20)	0	1	100	0.20

Phenotypic analysis of immature and mature G0 and G1 sand flies

Injection survivors were screened for fluorescence and for phenotypic markers in L4 larval stage. Screening G0 and G1 survivors from UbiqCas9.874W and pHome-T PiggyBac plasmids, no EGFP/GFP, DsRed or RFP (under Ubiquitin63E, 3xP3, Opie2 and Actin5c promoters respectively) fluorescence was observed.

Images of wildtype *P. papatasi* pupae at different time points post pupation were compared to G0 pupal survivors of Ebony pDCC6 construct injections. Wildtype pupae have pale yellow pupal casing one day post pupation and become a darker yellow by 4 days post pupation (Figure *7*). Two *Ebony* G0 pupae were identified with dark banding patterns on the pupal case, not previously observed (personal communication with Professor Petr Volf, Charles University, Czech Republic).

On pupal eclosion, two G0 *P. papatasi* adults were much darker when compared to the wildtype colony *P. papatasi*, which are normally pale yellow/brown in colour. No images were taken to prevent injury or death of these individuals prior to mating crosses.

G0 survivors from injections with *Rudimentary* and *Vestigial* targeting pDCC6 constructs had no difference in wing morphology when viewed under a microscope. However, three male and three female G0 were sibling crossed, and in the resulting G1 progeny two males and one female had modified wing morphology (Figure *8*). One of these males and the female (G1-M6-D and G1-F3-C, respectively) had wings with pointed tips, and one individual (G1-M2-Z) had only one wing, which might be expected for a knockout phenotype targeting these wing development genes.

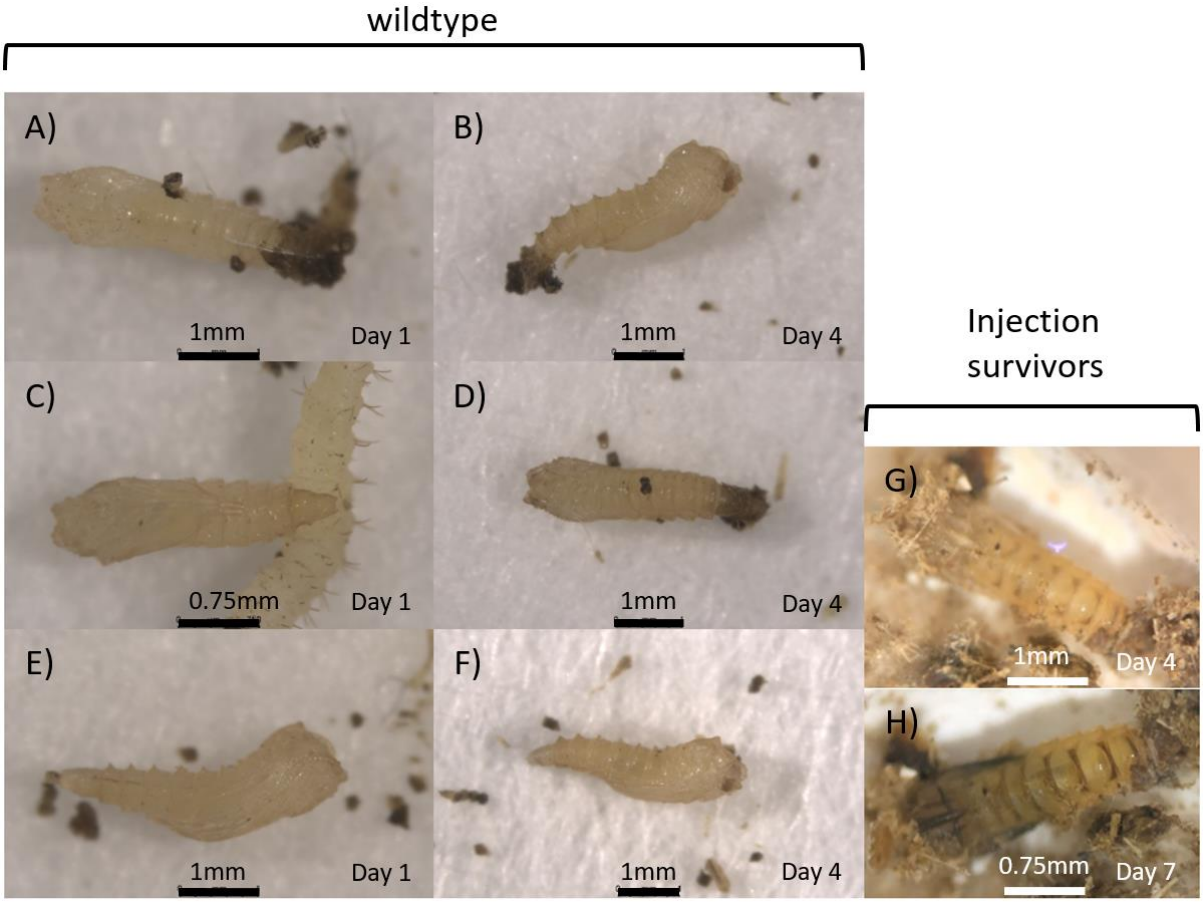


Figure 7. Comparison of wildtype *P. papatasi* pupae and pupal survivors after injection with *Ebony* knockout plasmid constructs at different time points. Wildtype images (A-F) were taken on approximately day 1, and day 4 of pupation using brightfield, and injection survivor images were taken at approximately day 4 (G) and 1 week (H).

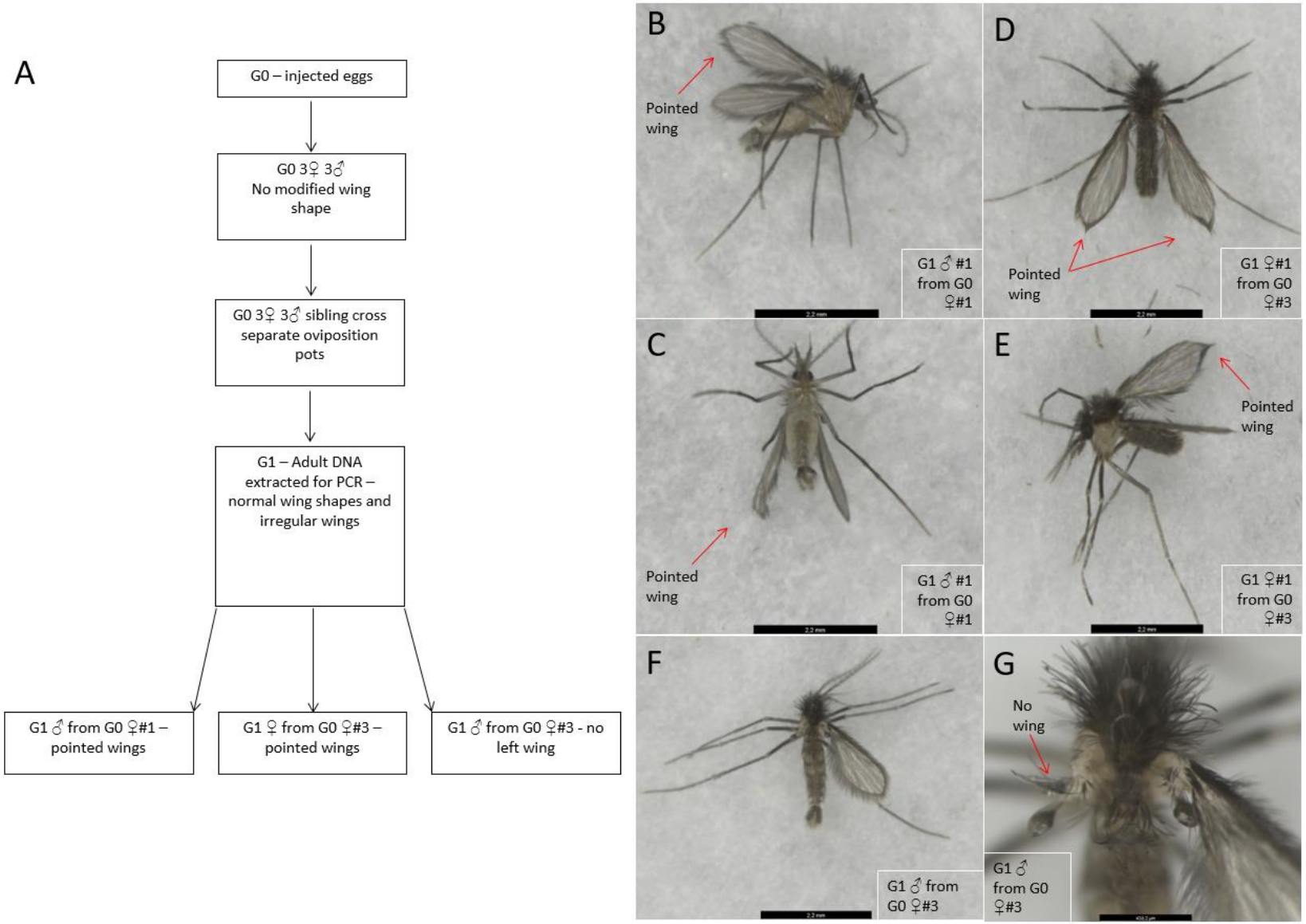


Figure 8. Wing phenotype targets mating strategy and putative positive G1 survivors. A) Mating strategy used for survivors of wing targeting construct microinjections, including phenotypic characteristics if observed. B) and C) show G1 male individual #1 (sample G1-m6-Z), highlighting putative modified wing phenotype. D) and E) show G1 female individual #1 (sample G1-f3-C), highlighting putative modified wing phenotype. F) and G) show G1 male missing left wing (samples G1-m2-Z). Scale bars 2.2mm.

Genotypic analysis of knockout sand flies via Sanger sequencing

Ac5-STABLE1-Neo

L. longipalpis and *P. papatasi* injected with Ac5-STABLE1-Neo (173 and 197 injected, respectively) resulted in 15 larval survivors in *L. longipalpis* and 0 survivors in *P. papatasi*. No adults survived and none of the pooled samples were positive for GFP expression.

UbiqCas9.874W PiggyBac

G0 *L. longipalpis* injected with UbiqCas9.874W and IhyPBase (1,031 injected), three pooled samples of 10 larvae, and one pooled sample of adults (3 females) were positive for Cas9, confirmed by sequencing (Figure *9*).

G1 larval survivors from crossing G0 with WT individuals were also analysed in pools of 10 for presence of Cas9. Fifty-two of 62 samples were tested, 33 were positive via PCR and gel electrophoresis, with 15 of these were confirmed by Sanger sequencing (Figure *9*).

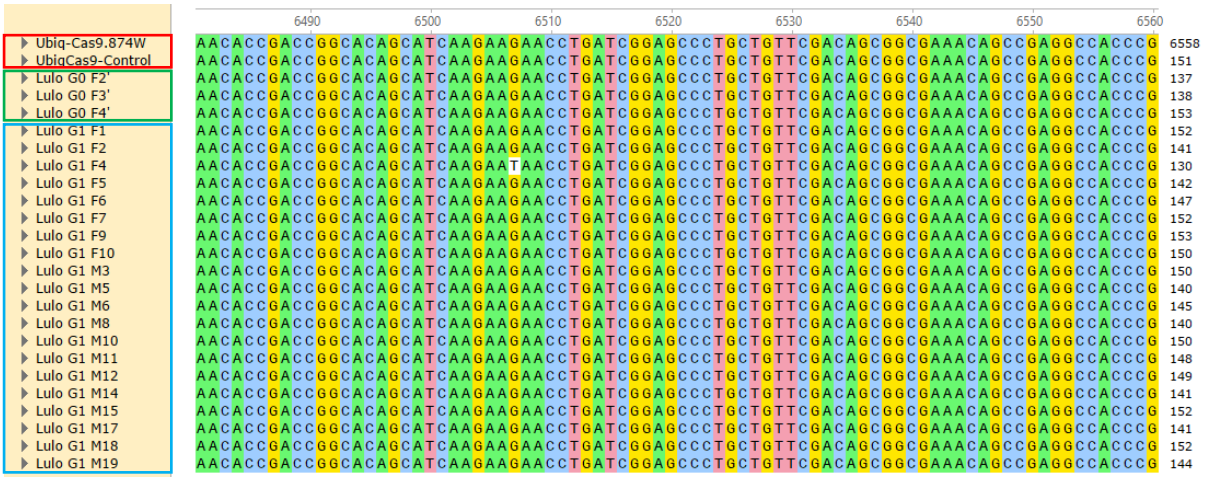


Figure 9. **Cas9 sequence alignment of G0 and G1 *L. longipalpis* samples injected with PiggyBac plasmid UbiqCas9.874W**. G0 samples (green box) and G1 samples (blue box) from G0 and wildtype crosses. Sequences aligned on SnapGene and aligned with Cas9 sequence from the UbiqCas9.874W plasmid map, and the sequence amplified using the primers used with the samples (red box).

For G0 *P. papatasi* injected with UbiqCas9.874W and IhyPBase (1,267 injected), two pooled samples of 10 larvae and two adult samples were PCR positive for Cas9, and two single adult samples were also positive for Cas9 via PCR and sequencing, out of a total of 21 samples.

pHome PiggyBac

For *L. longipalpis* G1 survivors from crossing G0 injected with pHome and IhyPBase (749 injected) and wildtype, 9/60 samples (10 pooled larvae) were GFP positive via PCR, and 7 of these 9 samples

were confirmed via sequencing. *P. papatasi* G1 samples (pooled larvae) demonstrated 7/14 positive for GFP via PCR and sequencing.

CRISPR Cas9 mediated constructs targeting non-lethal phenotypic wing targets (Rudimentary and Vestigial).

Six G0 *L. longipalpis* survivors (6/579) of injection with wing-targeting CRISPR plasmids had amplicons at the expected sizes for the regions of interest around the gRNAs (Figure *10*). All amplicons were gel extracted and sequenced. Two samples (G0-M3, and G0-F1) at the *R2* cut site had mismatches in the sequence alignment compared to control sequences (G0-M3 also positive for indels via ICE analysis, see below). Two samples (G0-M1, and G0-F1) had mismatches around the *R3* cut site (both had indels via ICE analysis, see below). Quality of *R1*, *V1*, *V2* and *V3* amplicons were too poor to sequence.

Twenty-six G1 *L. longipalpis* samples were analysed via PCR and sequence aligned at the expected wing gene cut sites. For the *R1* cut site 12/26 aligned with wildtype, 16/26 aligned at the *R2* cut site (Figure *11*B), and 7/26 aligned at the *R3* cut site. None of these samples demonstrated indels via this analysis method within the *Rudimentary* gene. All samples aligned at the *V1* cut site, with 16 having full alignment with the control sequence. Eight samples (G1-D, G, I, N, P, Q, R and U) had a SNP within the *V1* gRNA sequence (17bp upstream of the PAM site, *G→A*), and for the *V2* cut site, all samples aligned, except six (G1-D, I, J, P, Q and U), with no modifications around the cut site (Figure *12*C and D).

CRISPR Cas9 mediated constructs targeting olfactory Gr2, Ir8a, and Orco genes

Eleven G0 *L. longipalpis* survivors (11/576) of injection with olfactory gene targeting CRISPR plasmids had amplicons of the expected sizes for the regions of interest flanking the gRNAs for *Gr2* (Figure *13*), *Ir8a*, and *Orco* (Figure *14*).

At the *Gr2-1* gRNA region, sample sequences aligned with the VectorBase database sequence, however the wildtype had two single nucleotide insertions within the gRNA sequence, upstream from the PAM site (TGG/ reverse complement ACC) (Figure *13*B). All samples had multiple SNPs upstream of the gRNA, and 1-3 bp deletions upstream of the gRNA compared to the wildtype sequence. Sample G0-M1 could not be aligned, and G0-M5 had no SNPs around the gRNA region. Around the second gRNA region within the *GR2* gene (*Gr2-2*), all samples (except G0-M1) aligned with both the VectorBase sequence and the wildtype sequence with, no SNPs (Figure *13*C).

At the *Ir8a-1* gRNA, no indels were observed in any of the 11 G0 samples, and one sample (G0 M8) could not be aligned. Alignment was not possible in the region of the second gRNA (*Ir8a-2*). At the *Orco-1* gRNA all sequences aligned with no indels (except G0-M3, F1, and F2), and G0-M7 had a large indel starting 42bp upstream from the start of the gRNA. All male samples M1-M9 were aligned without

indels at the *Orco-2* gRNA, and three samples had three SNPs of 38-72bp downstream from the PAM site (TGG).

Thirty-six G1 *L. longipalpis* had DNA extracted (21 males and 15 females). At *Gr2-1* gRNA, thirteen samples (G1-M4, 5, 6, 8, 12, and MA-MI (except MB)) aligned with the control sequence with no indels or SNPs within the gRNA sequence. Two samples (G1-MA and MC) had SNPs upstream of the gRNA, four samples (G1-M1, 3, 9, and 10) had insertions and SNPs within the gRNA sequence, and one (G1-M7) could not be aligned. G1-M1, 3, 9, and 10 had small insertions ~30bp upstream of the gRNA. Nine female samples also aligned (FA-FK, except FE and FI) with the control sequence. Four female samples had SNPs and small indels upstream of the gRNA (FA, FF, FI, and FK). At *Gr2-1* gRNA, sequences aligned well, except for six (G1-M7, M11, FE, FH, MB, and 9F), which could not be aligned, and M5 which had two SNPs within the gRNA sequence.

At *Ir8a-1*, samples M1-12, MA-MI, and FA-FL align with no indels or SNPs around the gRNA. At *Orco-1*, nine samples (MC, MD, ME, MH, FD, FF, FH, FI, and FJ) did not align at the gRNA sequence region. No indels or SNPs were observed in the samples that aligned with the wildtype sequence.

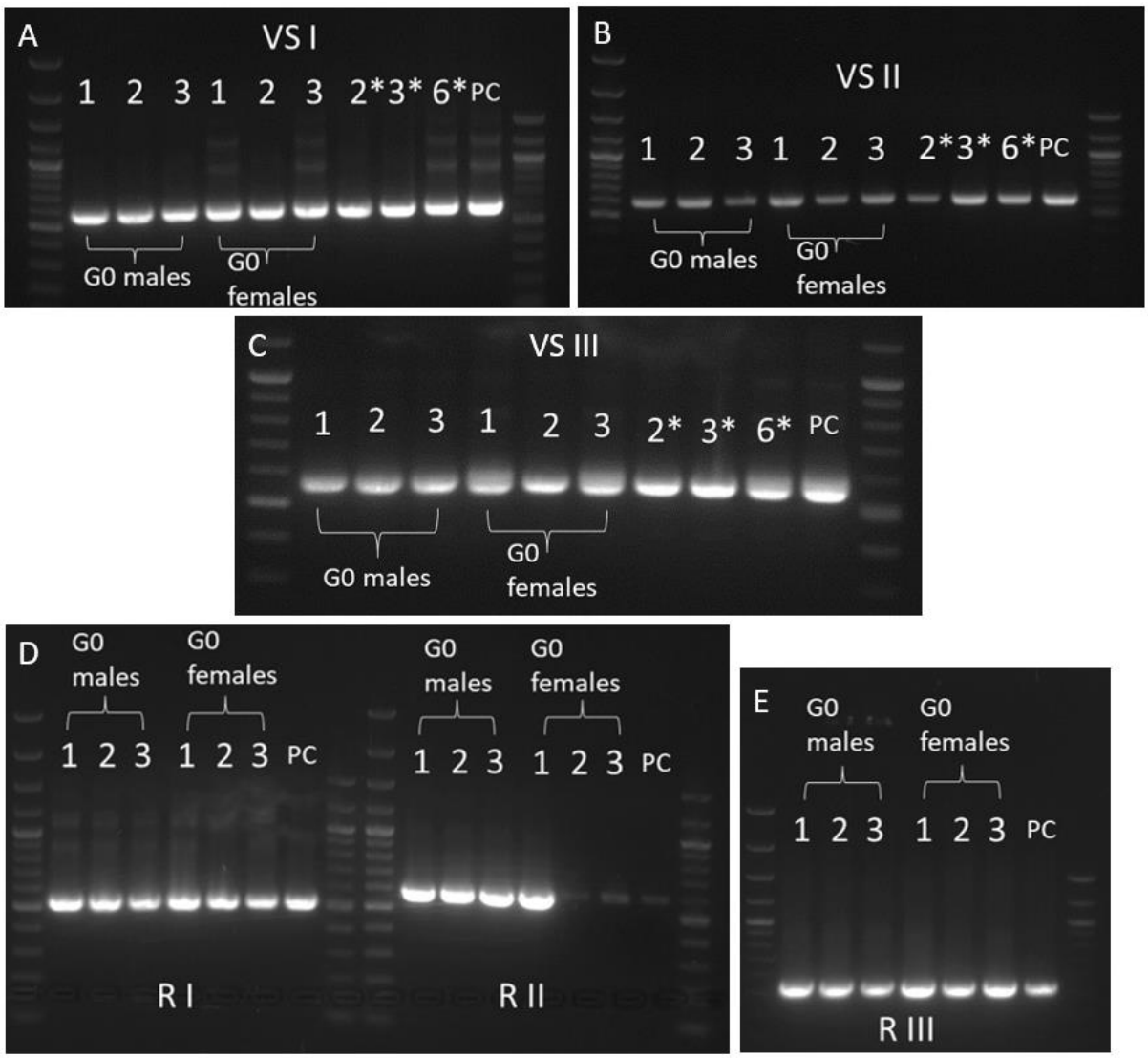


Figure 10. ***L. longipalpis*** **G0 injection survivors of wing targeting genes.** *Vestigial* (VS) and *Rudimentary* (R). Three pairs of primers were used for each gene to amplify around the gRNA sequences. A) Vestigial 1 amplicon, expected size 577bp. B) Vestigial 2 amplicon expected size 567bp. C) Vestigial 3 amplicon, expected size 562bp. D) Rudimentary1 and 2 amplicons, expected sizes 530bp and 586bp respectively. E) Rudimentary 3 amplicon, expected size 534bp. PC = positive control. G1 samples are also included with an asterisk.

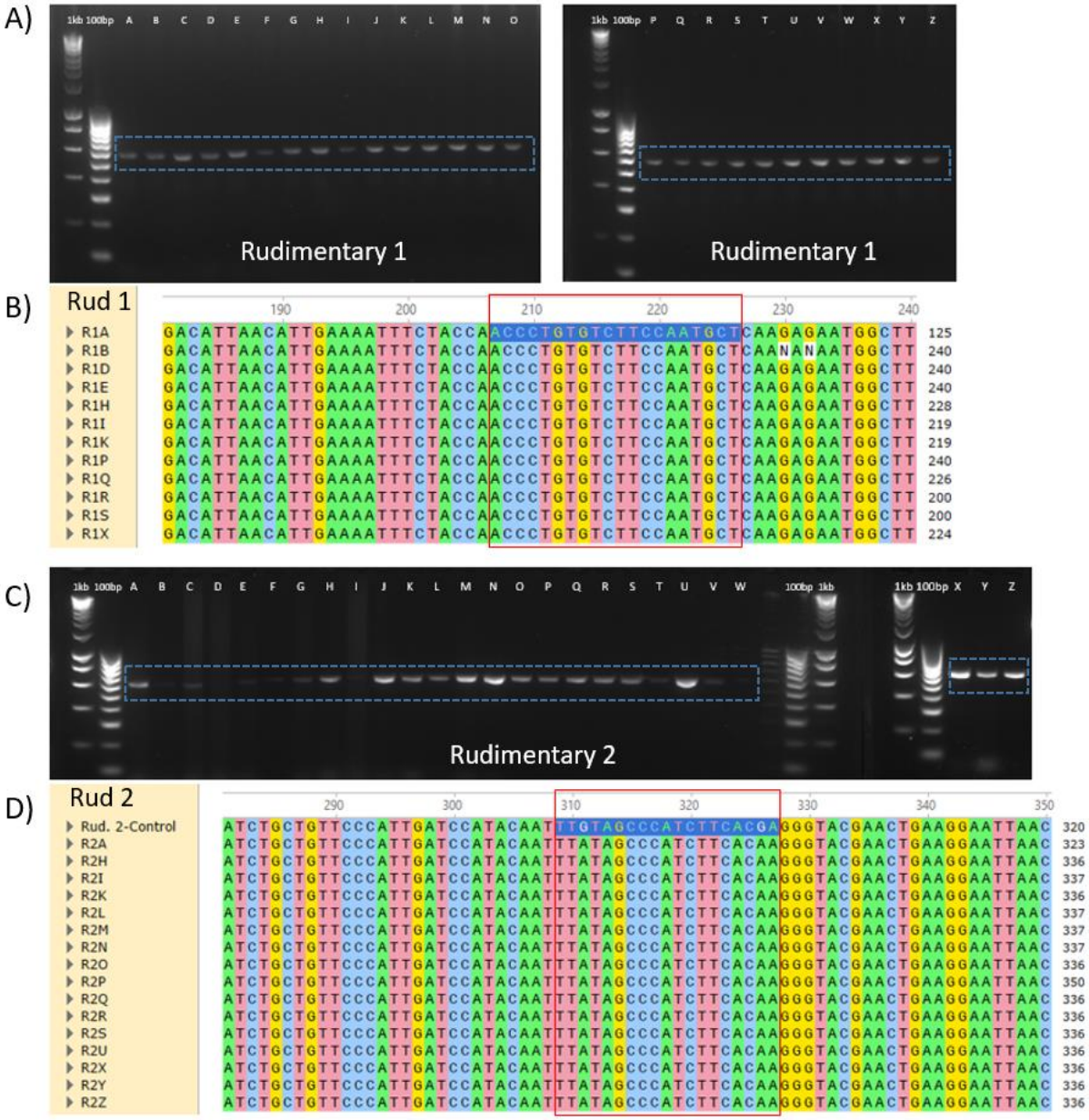


Figure 11. **PCR and Sanger sequence alignment for *L. longipalpis* G1 adults.** G1 adults (A-Z) aligned at the R1 and R2 cut sites. A) All samples have R1 amplicons at the expected size (530bp, blue box) demonstrating no large indels at the cut site. B) R1 amplicon sequences were aligned showing no indels (red box). Not all sequences were high enough quality for alignment. C) R2 amplicons were at the expected size (586bp, blue box) demonstrating no large indels at the cut site. D) R2 sequences were aligned, showing no indels (red box). Not all sequences were high enough quality for alignment. Sequence alignments conducted using SnapGene.

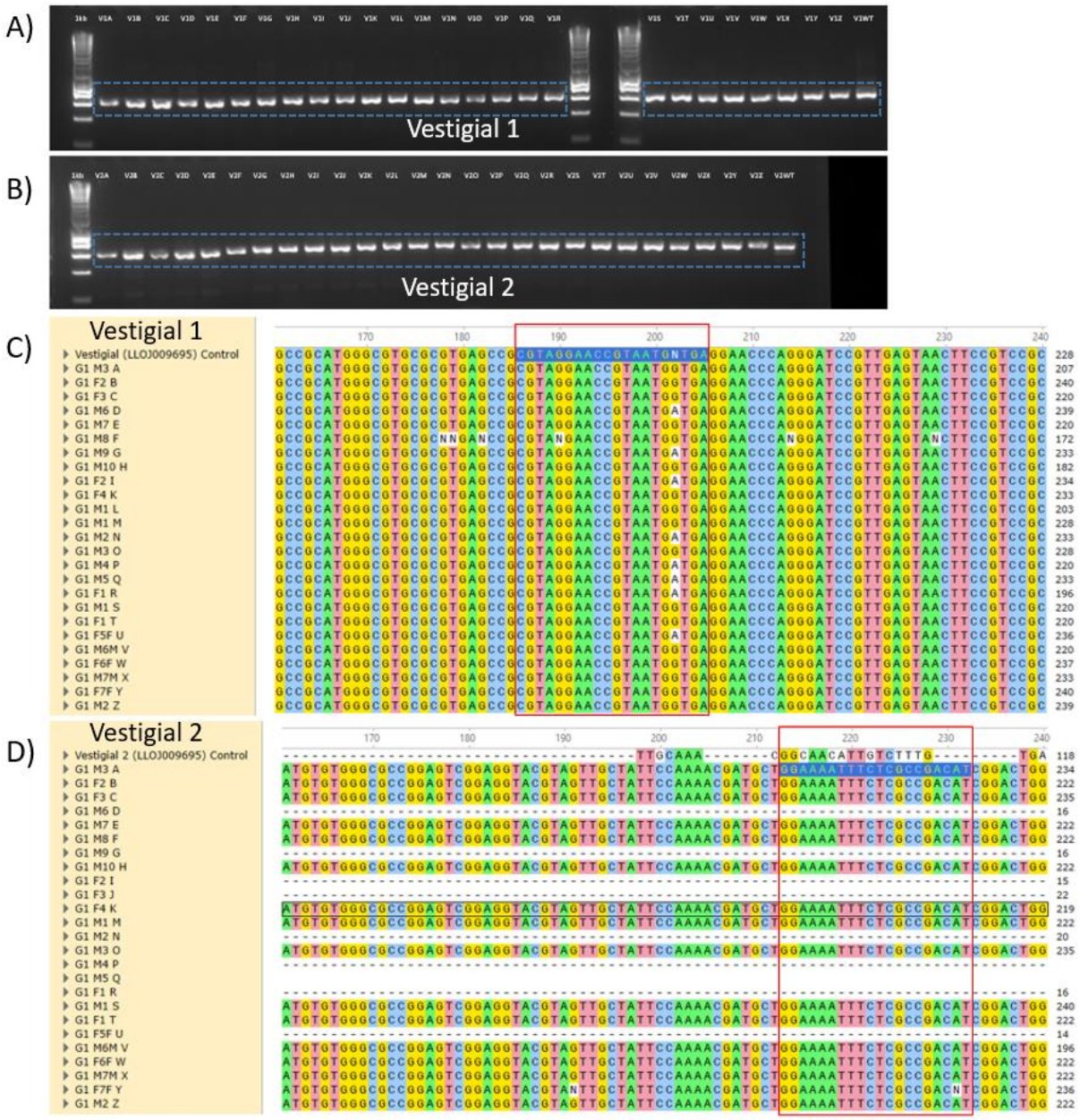


Figure 12. **PCR and Sanger sequence alignment for *L. longipalpis* G1 adults at the V1 and V2 cut sites.** G1 adults (A-Z) aligned A) All samples have V1 amplicons at the expected sizes (577bp, blue box) demonstrating no large indels at the cut sites. B) All samples have V2 amplicons at the expected sizes (567bp, blue box) demonstrating no large indels at the cut sites. C) V1 sequence alignment, and (D) V2 sequence alignment conducted using SnapGene, with gRNA sequence highlighted (red boxes).

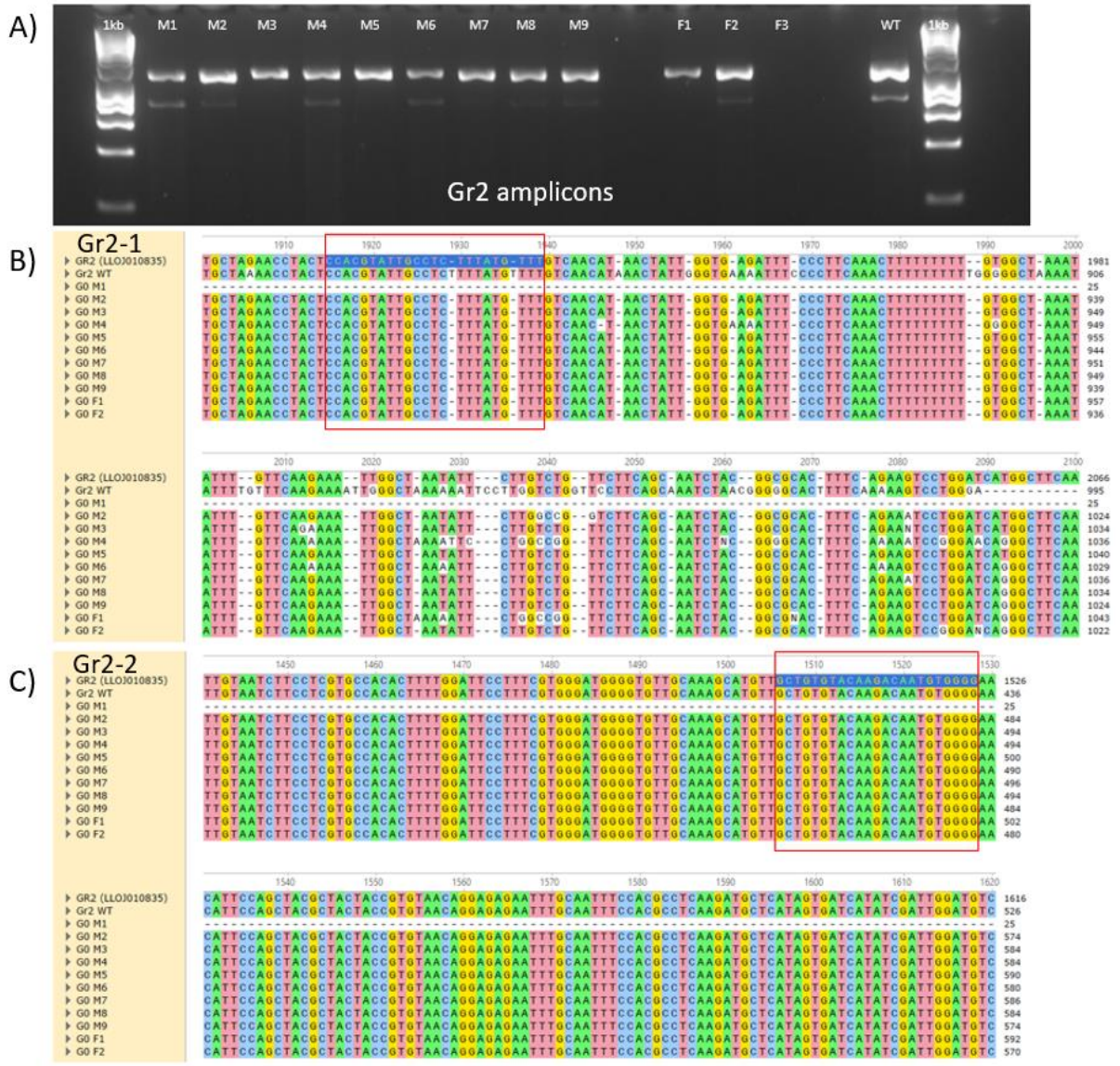


Figure 13. PCR and Sanger sequence alignment for *L. longipalpis* G0 adults (M1-9 and F1-3) Gr2 amplicon at the Gr2-1 and Gr2-2 cut sites. A) All samples have amplicons at the expected size (1158bp, except F3) demonstrating no large indels at the cut sites. B) Sequence alignment at the Gr2-1 cut site and Gr2-2 cut site (B) were conducted using SnapGene, with gRNA sequence highlighted (red boxes).

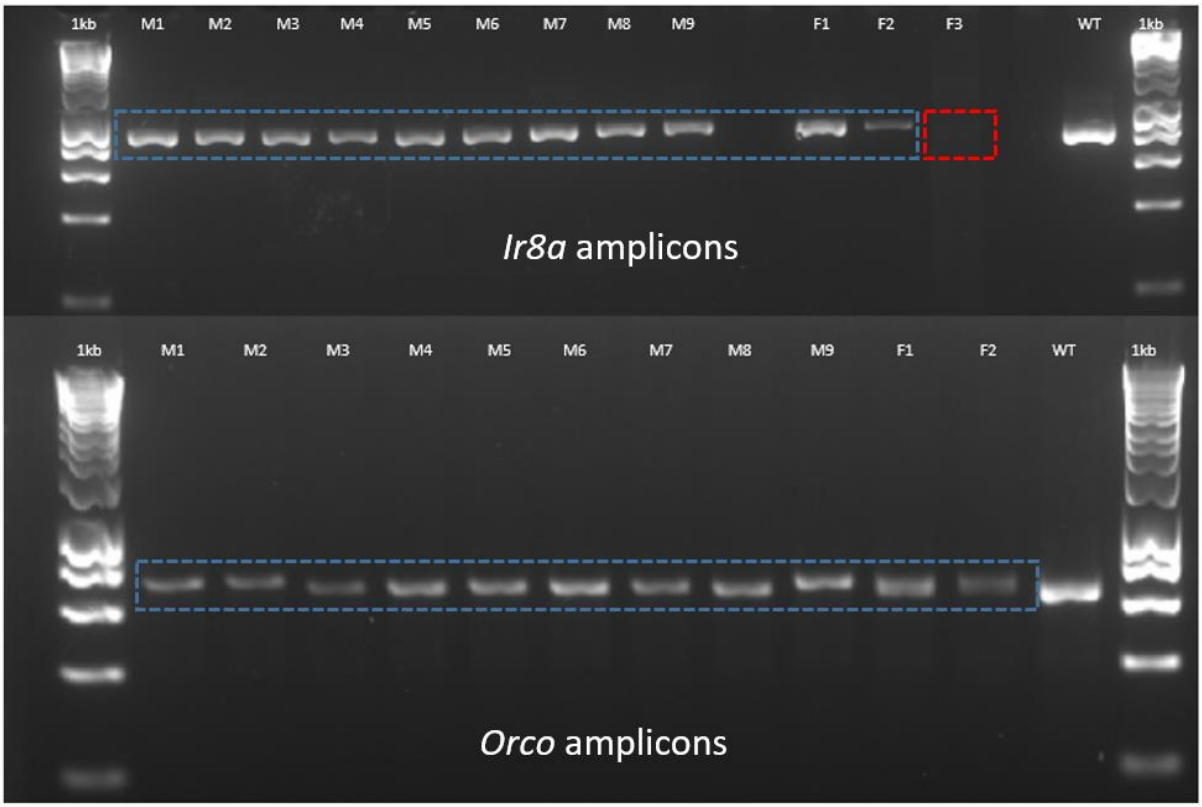


Figure 14. PCR for *L. longipalpis* G0 adults (M1-9 and F1-3) *Ir8a* and *Orco.* All samples have amplicons at the expected sizes, with 810bp for Ir8a, and 634bp for Orco (blue boxes), except sample F3 (red box), demonstrating no large indels at the cut sites.

Ebony and Caspar constructs in P. papatasi

Ten G0 *Ebony* survivors had DNA sequenced at the four gRNA regions (six males, and six females). All male samples demonstrated PCR amplicons at the expected size, and four of five female samples analysed via PCR demonstrated amplicons at the expected size (Figure *15*C). Sequence alignments showed no indels or SNPs at any of the gRNAs. Two female samples (samples 7 and 10) could not be aligned. Analysis via the ICE algorithm showed no indels.

Six G0 *Caspar* survivors had DNA extracted, and four had amplicons at the expected size (Figure *15*A and B). Sequence alignments showed no SNPs or Indels related to the gRNA regions.

HDR constructs

For the analysis of survivors of HDR construct injections (see Methods, *Microinjection mixture*) a reverse primer was designed downstream of the right homology arm (RHA) and the forward primer was designed within the HDR insert. Amplification of a 1,234 bp band would occur if insertion had taken place; however, amplification cannot occur if the HDR cassette has not been inserted into the genome.

Six *L. longipalpis* G0 and four G1 adults had DNA extracted and PCR amplification attempted. All six samples showed no amplification, therefore no insertion of the HDR cassette. The negative control was wildtype DNA amplified using the same primers. No amplification occurred, as expected (Figure 51).

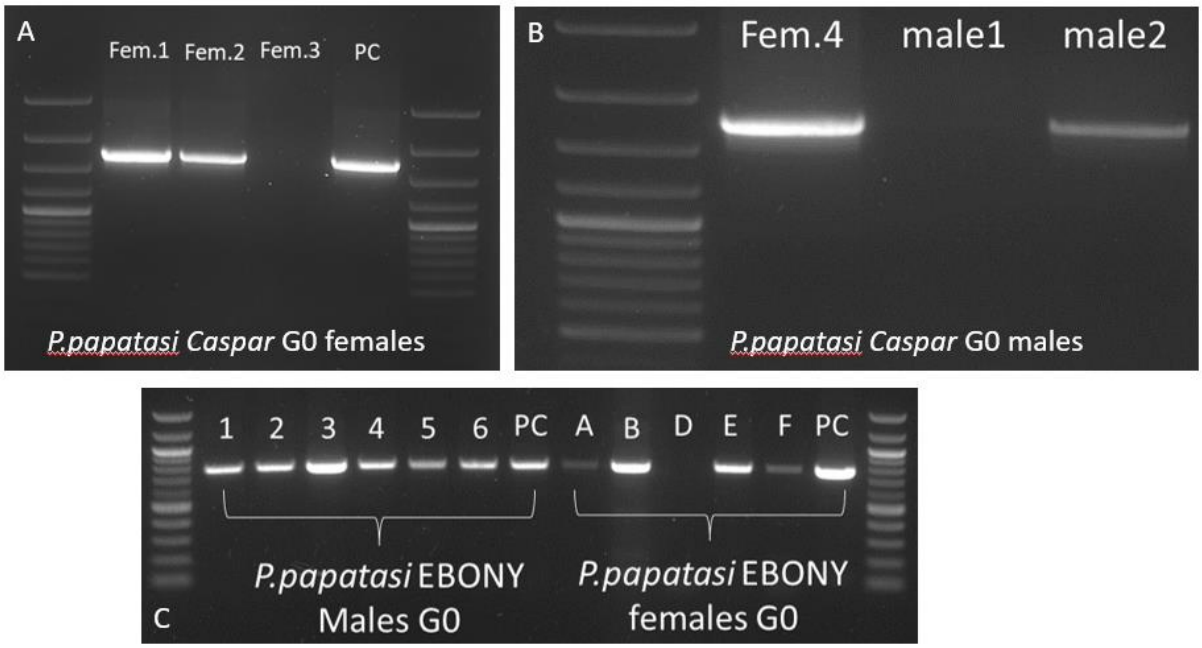


Figure 15. PCR amplification of regions of interest in *P. papatasi* G0 survivors of *Caspar* and *Ebony* CRISPR plasmid injections.

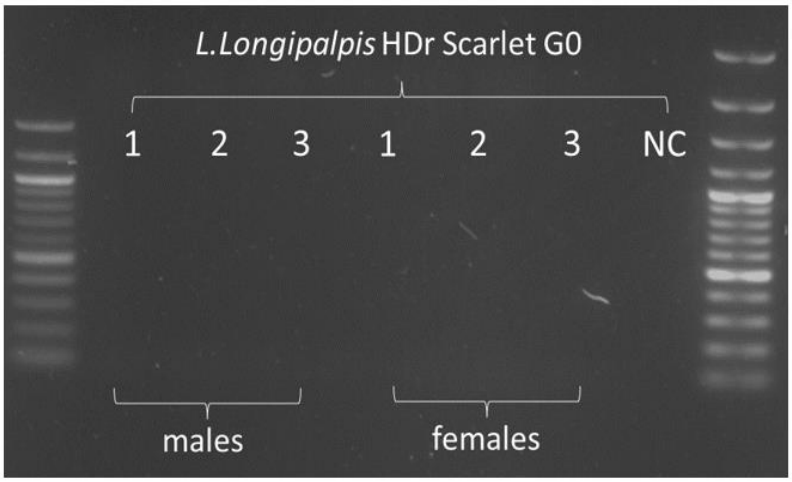


Figure 16. PCR for *L. longipalpis* G0 adult survivors of HDR construct microinjections. NC = negative control.

T7 Endonuclease I (T7EI) Heteroduplex Assay and Densitometric Analysis

Wing Phenotype G1

In parallel with sequencing approaches heteroduplex assays to detect mutagenesis was conducted with 26 samples (A-Z) of G1 *L. longipalpis* analysed the *R1* cut site. Faint amplicons were observed above the expected size for the *R1* PCR amplicon (530bp) for samples A-N (excluding A, I, K and N, where no amplicon was visible). Samples O-P had amplicon bands of the expected size (excluding sample T, which had no amplicon visible). None of the samples A-Z had T7EI activity, therefore no densitometric analysis was conducted.

At the *R2* cut site, amplicons of the expected size (586bp) were observed for samples A-R (excluding B, C, D, E, F, G, and I, where no amplicons were visible). None of the samples A-R, had to T7EI activity, therefore no densitometric analysis was conducted.

At the *R3* cut site, amplicons of the expected size (534bp) were observed for all samples. Successful digestion of the amplicon with T7EI would result in products at 244bp and 290bp. Six samples (R3C*, R3M*, R3N*, R3U*, R3V*, and R3Z*) had T7EI activity, however only a single band was observed at ~400bp. Densitometric analysis (see Methods, *Densitometric Analysis*, and Appendix 8) estimated modification rates between 0.36% and 6.13%. Additionally, controls for 14 samples (C, D, E, H, M, P, Q, R, S, U, V, X, Y, and Z) had endonuclease activity with faint bands also observed at ~400bp. Densitometric analysis estimated modification rates between 0.15% and 4.43%. The heteroduplex positive controls had an estimated modification rate of 2.4%.

Samples analysed at the *V1* cut site had amplicons at the expected size (577bp) for all samples. Successful digestion of the amplicon with T7EI would result in products at 222bp and 355bp. Six samples (V1M*, V1P*, V1R*, V1T*, V1U* and V1Y*) had T7EI activity with bands observed between 200-400bp. Densitometric analysis (see appendix 9) estimated modification rates between 2.98% – 18.51%.

At the *V2* cut site, amplicons were observed at the expected size (567bp) for all samples, excluding samples J and L (Figure 17, see appendix 10 for data tables). Successful digestion of the amplicon with T7EI would result in products at 261bp and 306bp. Nine samples (V2D*, V2G*, V2I*, V2K*, V2O*, V2Q*, V2R*, V2U*, and V2Y*) had T7EI activity with bands observed between 200-400bp. Densitometric analysis estimated modification rates between 0.64% – 15.06%. Controls for samples D, E, M, and N had endonuclease activity with bands also observed at ~400bp. Densitometric analysis estimated modification rates between 0.05% and 1.44%.

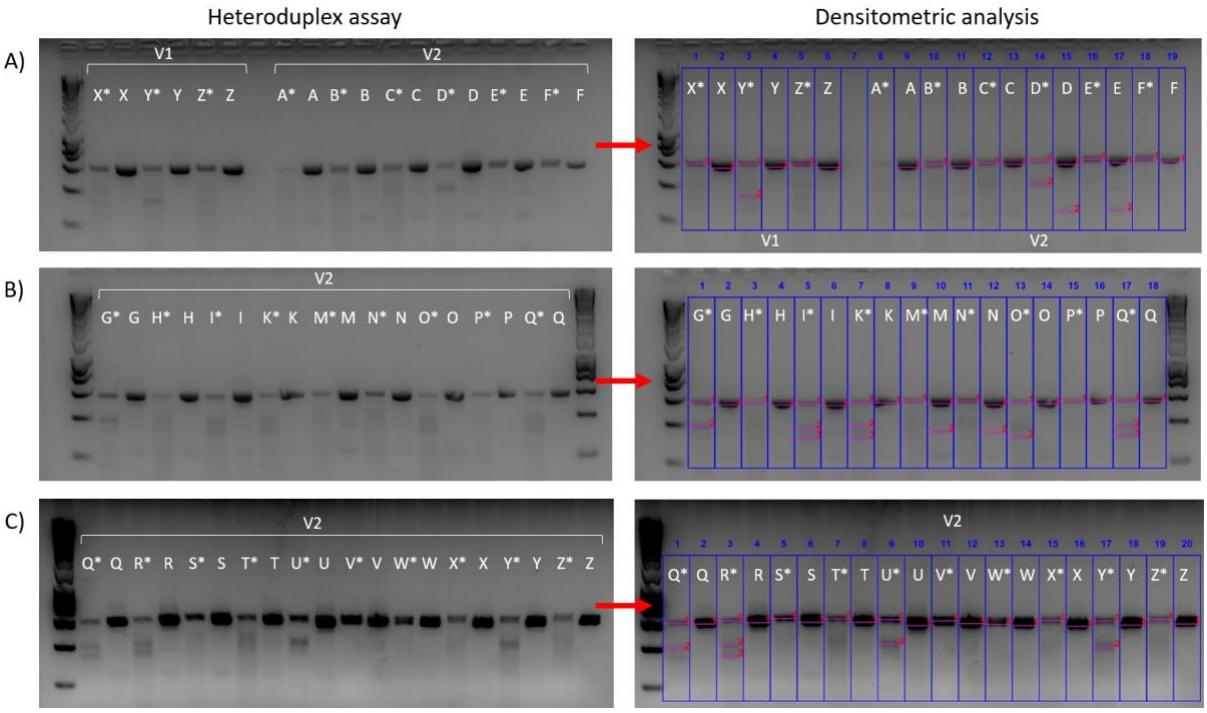


Figure 17. T7 Endonuclease I Heteroduplex assay and densitometric analysis on *L. longipalpis* G1 samples at the V1 and V2 gRNA regions. A) Samples V1X-V1Z followed by V2A-V2F, B) samples V2G-V2Q, and C) samples V2Q-V2Z. Asterisks (*) indicates PCR product treated with Endonuclease. No asterisk indicates controls, untreated with Endonuclease. In Densitometric analysis the blue box indicates the area of each lane used for density analysis, and pink lines identify the bands analysed within each lane.

Olfactory Genotype G0

Nine male and two female G0 *L. longipalpis* samples were analysed for modification of the *Gr2* gene. Faint amplicons were observed of the expected size for the *Gr2* amplicon (1158bp) for all samples. Six of the samples (G1*, G2*, G4*, G6*, G9*, and G12*) had T7EI activity. Only a single band was observed between 600-800bp. Densitometric analysis (see appendix 11) estimated modification rates between 0.68% and 7.52%. Additionally, six controls (G1, G2, G4, G6, G9, and G12) had endonuclease activity with faint bands also observed between 600-800bp. Densitometric analysis estimated modification rates between 1.12 and 10.11%.

For the *Ir8a* gene amplicons were observed at the expected size (810bp). Four samples had T7EI activity. Sample I2* had three bands between 200-800bp, I7* had two bands between 200-600bp, I8* had one band between 400-600bp, and I9* had two bands around 400bp. Densitometric analysis (see appendix 12) estimated modification rates of 0.48% – 3.91%.

Olfactory Genotype G1

Twelve G1 *L. longipalpis* male samples (A-L) were analysed for modification of the *Gr2* gene. Faint bands were observed the expected size for (1158bp) for all samples. Three samples (GE*, GI*, and

GJ*) had T7EI activity. Only a single band was observed between 600-800bp. Densitometric analysis (see appendix 13) estimated modification rates between 0.19% and 2.36%. Controls for eight samples (GD, GE, GF, GG, GI, GJ, GK, and GL) had endonuclease activity with faint bands also observed between 600-800bp. Densitometric analysis estimated modification rates between 0.92% and 5.51%.

On further G1 male and female samples (MA-MI and FA-FL, respectively), all male controls had bands at the expected size (1158bp) for the *Gr2* amplicon, and all samples had five bands between 400-1200bp, suggesting T7EI activity. Densitometric analysis (see Appendix 14) estimated modification rates between 1.42% and 57.64%. Female controls also had bands at the expected size for the *Gr2* gene, except FH and FL, which did not show PCR amplification. For FA* modification rate was estimated to be 2.17% and for FC*, modification rate was 2.13%.

For the *Ir8a* gene, amplicons were observed at the expected size (810bp) for all samples (IA-IL). No samples had T7EI activity, therefore densitometric analysis was not conducted. For all male samples (MA-MI), endonuclease activity did occur, with estimated modification rates between 0.22% and 2.05%. All female samples (FA-FL) had endonuclease activity, except FH and FL. The estimated modification rates were between 0.15% and 3.32% (see Appendix 15).

Analysing modification of the *Orco* gene, the G1 *L. longipalpis* samples (MA-ML, and FA-FL) had amplicons at the expected size (634bp) for all samples, except FG and FH. In the heteroduplex assay, multiple samples had T7EI activity. Nine male and nine female samples had bands between 200-600bp. Densitometric analysis (see Appendix 16) estimated modification rates of between 0.44% and 9.00% in male samples, and 1.09 and 9.95% in female samples. Additionally, the control for sample G1-FL had endonuclease activity, and an estimated mutation rate of 10.31%.

In summary, the T7E1 assay and densitometric analysis indicated gene editing within wing development genes (R3, V1, V2 cut sites), with estimated modification rates between 0.15% and 18.51%. The analysis also indicated gene editing in G0 samples within the olfactory genes (*Gr2* and *Ir8a*, 0.48% - 10.11%), and G1 samples (*Gr2*, *Ir8a*, and *Orco*, 0.15%-57.64%) (see Chapter5 Appendices 9-16).

In silico Sanger sequence deconvolution: ICE Analysis

As previously described, Sanger sequence data can be analysed using *in silico* methods (see Methods, *Algorithmic deconvolution analysis of Sanger sequence data*) to detect the presence of gene edits by conducting statistical analysis comparing un-edited sequence traces (controls) to putatively edited sequence traces. Statistical outputs include indel % (ICE score), likely knockout of protein function via frame shift (KO-score), and an indication of the model fit to the edited sequence data (R^2) – with an R^2 value >0.8 indicating strong fit.

The results described (below) are separated by genes targeted (wing development genes and olfactory genes), and by generation (G0, survivors of microinjection, and G1, survivors of mating crosses); overall results are outlined in the *Summary of results* section.

Wing phenotype G0

Six *L. longipalpis* survived microinjection with an injection mixture containing three plasmids targeting the *Rudimentary* gene and three plasmids targeting the *Vestigial* gene (*R1-pDCC6, R2-pDCC6, R3-pDCC6, V1-pDCC6, V2-pDCC6, and V3-pDCC6*). DNA was extracted from these individuals and Sanger sequence data generated for each of the six potential editing sites. These were analysed via the ICE algorithm (see Appendix 17).

Four of the six samples generated ICE data, with having indels around the *R3* cut site and two samples having no indels detected. One sample identified indels around the *R2* cut site (Table *5* & Figure 53B). Poor Sanger sequence data precluded analysis at the other cut sites.

Table 5. ICE analysis summary statistics for G0 and G1 survivors of *Rudimentary* gene knockouts at cut sites R2 and R3, positive for indels. ICE = the proportion of genotypes that contain an indel (indel %), KO-Score = measure of the variants that are likely result in functional protein knockout via a frame shift, R2 = model fit. Expanded data tables can be found in the Chapter5 Appendices.

Samples	*G0 Rudimentary 2 (R2) and Rudimentary 3 (R3)*			
	Algorithm Success	*ICE*	*KO-score*	R^2
G0-M3_F (R2)	Indels	23	19	0.54
G0-F1_R (R3)	Indels	55	55	0.55
G0-F3_R (R3)	Indels	57	0	0.57
	G1 Rudimentary 2 (R2) and Rudimentary 3 (R3)			
G1 M3_A (R2)	Indels	22	17	0.53
G1 M10_H (R2)	Indels	22	18	0.53
G1 F2_I (R2)	Indels	22	18	0.53
G1 F4_K (R2)	Indels	22	18	0.53
G1 M1_L (R2)	Indels	22	18	0.53
G1 M1_M (R2)	Indels	22	18	0.53
G1 M2_N (R2)	Indels	22	18	0.53
G1 M3_O (R2)	Indels	22	18	0.53
G1 M4_P (R2)	Indels	21	17	0.52
G1 M5_Q (R2)	Indels	21	17	0.52
G1 F1_R (R2)	Indels	22	18	0.53
G1 M1_S (R2)	Indels	22	18	0.53
G1 F5F_U (R2)	Indels	22	18	0.53
G1 M7M_X (R2)	Indels	22	18	0.53
G1 F7F_Y (R2)	Indels	22	18	0.53
G1 M2_Z (R2)	Indels	22	17	0.53
G1 M8_F (R3)	Indels	54	54	0.54
G1 M9_G (R3)	Indels	55	55	0.55

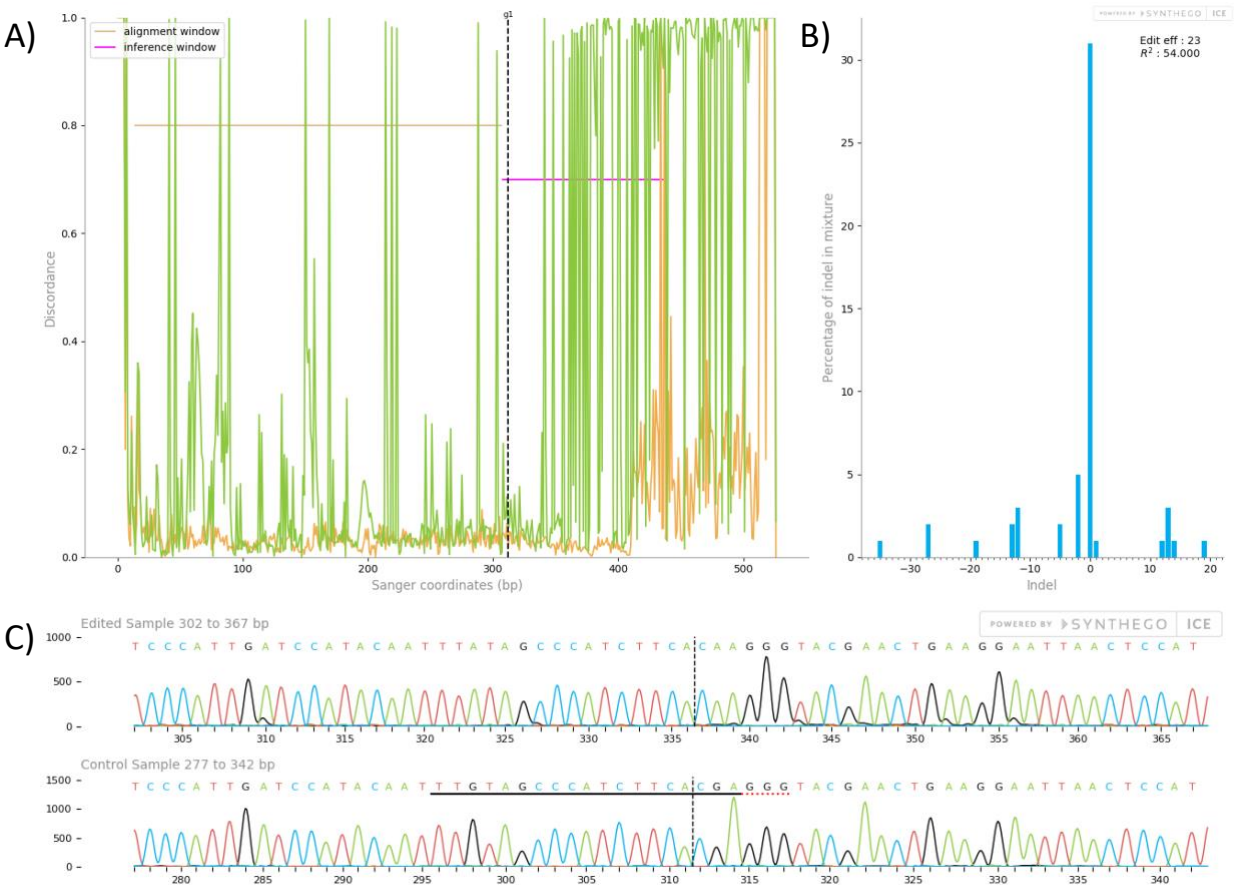


Figure 18. ICE analysis output from software for a guide sequence targeting the *Rudimentary* gene (using R2-pDCC6 construct) for sample G0-M3_f. (A) Discordance graph showing the difference between control Sanger sequence trace file (orange) and the putative edited Sanger sequence trace file. Within the Discordance graph the cut site is represented by a dotted black line. A pink line (horizontal) shows the alignment window, where the high quality of the control and putative edited sequences is used to align the two. The Interference window (orange horizontal line) is a region of the Sanger sequence traces around the cut site where the algorithm infers the difference between the control and edited Sanger sequence traces. (B) Indel size-frequency histogram. Shows the size of indels determined by the ICE algorithm and their respective frequencies. (C) Trace file generated by the ICE algorithm spanning the cut site from the control and the edited samples. The guide sequence is underlined in black, and the protospacer adjacent motif (PAM) sequence is underlined by a dotted red line in the control sample. Vertical dotted lines denote the expected cut site.

Wing phenotype G1

For G1 survivors from mating between G0 injection survivors and wildtype *L .longipalpis*, 16/26 samples had indels at the *R2* cut site. All samples had similar ICE statistics with ICE scores of ~ 22 and KO-scores of ~ 18, however these indel calls were not a good fit for the ICE algorithm (R^2 = ~0.53). Two of the 26 samples demonstrated indels at the *R3* cut site (R^2= ~0.54) (Table *5*).

The ICE algorithm was unable to produce statistics for these samples around the *R1*, *V1*, *V2* and *V3* cut sites (see appendix 18 and 19).

Olfactory genotypes G0

Eleven *L. longipalpis* (nine males, and two females) survived microinjection with an injection mixture containing two plasmids targeting the each of the *Gr2, Ir8a*, and *Orco* genes (Gr2-1-pDCC6, Gr2-2-pDCC6, Ir8a-1-pDCC6, Ir8a-2-pDCC6, Orco-1-pDCC6, and Orco-2-pDCC6). DNA was extracted from these individuals and Sanger sequence data was generated for each of the six potential editing sites.

Of the 11 samples, four had indels detected for the *Gr2-1* cut site (*AAACATAAAGAGGCAATACG*). Seven of the 11 samples indels detected around the *Gr2-2* cut site (*GCTGTGTACAAGACAATGTG*) (Table *6*).

In ten of the 11 injection survivors, no indels were detected around the *Ir8a-1* cut site (*AGCAATAGCGAGACCTGCGA*), with similar degrees of robustness (ICE = 0, KO = 0, and R^2 = 0.98 for six samples; ICE = 0, KO = 0, and R^2 = 1 for four samples), and 1/11 had indels detected (ICE = 33, KO = 33, and R^2 = 0.76) at this cut site (Figure *19*B).

One of the 11 G0 injection survivors had indels at the *Orco-1* cut site in both the forward and reverse configurations of the ICE algorithm. This sample, G0-m7, had ICE = 39 KO = 39 and R^2 = 0.84 when the forward guide sequence was used (forward guide sequence: *TGTGAGATACATGACCAACA*), and had ICE = 14 KO = 14 and R^2 = 0.87 when the reverse guide sequence was used in the ICE analysis (reverse guide: *TACAGCAATCAAGTATTGGG*) (Figure *19*A). The remaining 10 samples had no indels detected at the *Orco-1* cut site, with similar degree of robustness (R^2 = 0.94 - 0.99). Nine of the 11 samples had no indels detected around the *Orco-2* cut site (*TACAGCAATCAAGTATTGGG*) (R^2 = 0.94 - 1).

Table 6. ICE analysis summary statistics for G0 and G1 survivors of Olfactory gene knockouts at cut sites Gr2-1, Gr2-2, Ir8a-1, and Orco-1, positive for indels. ICE = the proportion of genotypes that contain an indel (indel %), KO-Score = measure of the variants that are likely result in functional protein knockout via a frame shift, R2 = model fit. Asterisk represents R^2 values with a robust fit to the model.

Samples	*Gr2-1 (1) and Gr2-2 (2)*			
	Algorithm Success	*ICE*	*KO-score*	R^2
G0 m2_f (1)	Indels	60	60	0.76
G0 m5_f (1)	Indels	26	22	0.63
G0 m7_f (1)	Indels	36	34	0.36
G0 m9_f (1)	Indels	27	26	0.52
G0 m1_f (2)	Indels	16	16	0.77
G0 m2_r (2)	Indels	57	0	0.57
G0 m3_r (2)	Indels	1	1	0.98*
G0 m6_r (2)	Indels	2	2	0.98*
G0 m7_r (2)	Indels	2	2	0.98*
G0 m8_r (2)	Indels	1	0	0.97*
G0 m9_r (2)	Indels	3	3	0.98*
	Ir8a-1			
G0 M8	Indels	33	33	0.76*
	Orco-1			
G1 m7_f	Indels	1	1	0.98*
G1 m7_r	Indels	59	59	0.59

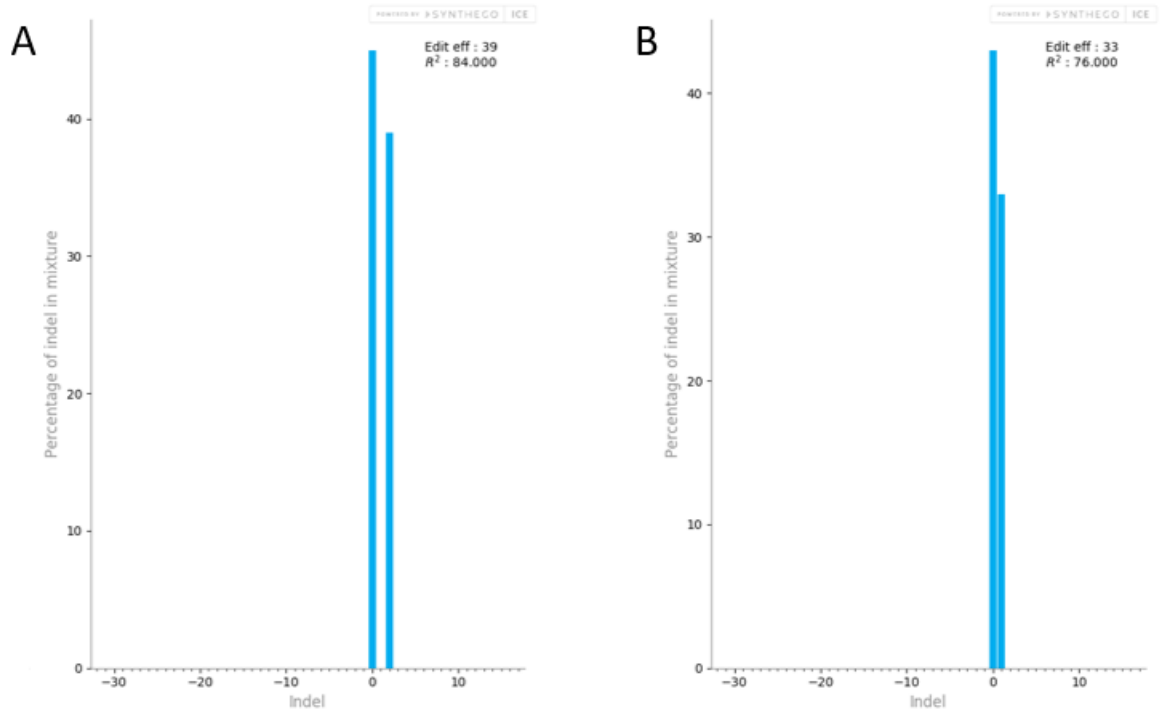


Figure 19. Indel size-frequency histogram for (A) *L. longipalpis* G0 survivors of Orco-1-pDCC6 plasmid injection (sample G0 M7), and (B) *L. longipalpis* G0 survivor (sample G0 M8) of Ir8a-1-pDCC6 plasmid injection. Shows the size of indels determined by the ICE algorithm and their respective frequencies around the Orco-1 (TGTGAGATACATGACCAACA) and Ir8a-1 (AGCAATAGCGAGACCTGCGA) cut sites.

Olfactory genotypes G1

For G1 survivors from mating between G0 injection survivors and wildtype *L .longipalpis* 36 individuals survived. Of the DNA extracted from these samples, 35/36 samples were used in ICE analysis (see appendix 21).

Sixteen of 35 samples were successfully analysed by ICE. Fifteen of these 16 samples had indels detected around the *Gr2-1* cut site (AAACATAAAGAGGCAATACG). All showed a good fit to the model (R^2 = 0.85 – 0.89) (Table *7*).

Thirty of 35 samples were successfully analysed by ICE for the *Ir8a-1* cut site (AGCAATAGCGAGACCTGCGA). Three samples had indels, with a good fit to the model (R^2 = 0.91 – 0.96) (Table *7*). Twenty-seven of the 30 samples analysed had no indels detected.

Twenty-four of 35 samples were successfully analysed by ICE for the *Orco-1* cut site (TGTGAGATACATGACCAACA). Six samples had indels, with a good fit to the model (R^2 = 0.98 – 0.99), and three samples (G1 m2_r, G1 m9_r, G1 m10_r) had indels detected with a less robust fit to the ICE model (R^2 = 0.59, 0.62, and 0.62 respectively) (Table *7*).

Table 7. ICE analysis summary statistics for G1 survivors of *Gr2*, *Ir8a*, and *Orco* gene knockouts. ICE value is the proportion of genotypes that contain an indel (indel %), KO-Score is measure of the variants that are likely result in functional protein knockout via a frame shift, R^2 is the model fit. Asterisk (*) represents R^2 values with a robust fit to the model.

Samples	*Gr2*			
	Algorithm Success	*ICE*	*KO-score*	R^2
G1 FA	Indels	18	18	0.86*
G1 FC	Indels	4	2	0.87*
G1 FD	Indels	6	3	0.85*
G1 FG	Indels	4	3	0.88*
G1 FI	Indels	7	6	0.85*
G1 FJ	Indels	8	5	0.85*
G1 FK	Indels	2	1	0.89*
G1 MA	Indels	4	2	0.88*
G1 MC	Indels	3	1	0.88*
G1 MD	Indels	6	4	0.85*
G1 ME	Indels	6	2	0.86*
G1 MF	Indels	6	2	0.86*
G1 MG	Indels	7	4	0.85*
G1 MH	Indels	7	4	0.85*
G1 MI	Indels	5	2	0.85*
	Ir8a			
G1 FC	Indels	2	2	0.96*
G1 MB	Indels	5	5	0.95*
G1 ME	Indels	16	16	0.91*
	Orco			
G1 m2_f	Indels	1	1	0.98*
G1 m2_r	Indels	59	59	0.59
G1 m3_f	Indels	1	1	0.98*
G1 m4_f	Indels	1	1	0.98*

G1 m5_f	Indels	1	1	0.98*
G1 m9_r	Indels	62	62	0.62
G1 m10_r	Indels	62	62	0.62
G1 m11_f	Indels	1	1	0.98*
G1 MA	Indels	1	1	0.99*

Summary of Results

Five methods were used to assess modifications in microinjection G0 survivors, and G1 offspring of G0 crosses with wildtype: phenotypic analysis, Sanger sequencing, T7 Endonuclease I heteroduplex analysis, Densitometric analysis to attempt quantification of modifications, and Algorithmic deconvolution of Sanger sequence data to predict indels. These methods indicated mutagenesis in isolation, and in combination for a number of samples (described below).

Overall, there were six G0 *L. longipalpis* survivors (3 males and 3 females) of microinjection with wing phenotype targeting constructs. After crossing these with wildtype individuals to produce G1 sand flies, DNA was extracted for analysis of modification. Four of the six G0 survivors had predicted Indels via ICE analysis. These indels only occurred around the *R2* and *R3* cut sites. Alternative methods were not able to determine if modifications had occurred.

For G1 survivors, 26 were tested for modifications (16 males and 10 females). Three individuals were identified with modified wing morphology compared to wildtype (Figure 8). One male and one female had pointed wings and one male had one wing only, clearly indicative of phenotypic mutagenesis. Sixteen samples had indels predicted by ICE analysis at the *R2* cut site. Importantly, this includes the adult male with a single wing. Six samples showed heteroduplex cleavage at the *R3* cut site including the male with one wing and the female with pointed wing phenotype, however neither of these had indels predicted via ICE analysis. Overall, only two of 26 samples showed indels via ICE at the *R3* cut site. Six of 26 samples had heteroduplex cleavage at the *V1* cut site and nine of 26 samples had heteroduplex cleavage at the *V2* cut site.

For G0 *L. longipalpis* survivors of microinjections with olfactory gene targeting constructs, there were 11 survivors (ten males and two females). After crossing these with wildtype individuals to produce G1 sand flies, DNA was extracted for analysis of modification. In total, eight of 11 samples had predicted indels via ICE analysis. These indels occurred in the *Gr2*, *Ir8a*, and *Orco* genes, with the majority occurring in the *Gr2* gene. Four of the eight samples with predicted indels also showed heteroduplex cleavage.

For G1 survivors, 34 were analysed for modifications (21 males and 13 females). Seventeen of the 34 samples demonstrated modification via heteroduplex or ICE analysis in the *Gr2* gene. Of these, 10/17 had modification via both heteroduplex and ICE analysis.

Seventeen G1 samples had indels via heteroduplex or ICE analysis at the *IR8a* gene, and three of these 17 were positive for modifications via heteroduplex and ICE analysis.

Twenty-five G1 samples were positive for modifications at the *Orco* gene via heteroduplex or ICE analysis. Importantly, one of the 25 samples (G1-MA) was positive via both heteroduplex and ICE (0.44% estimated modification rate and R2= 0.99, KO=1, respectively).

Overall, the combined methods (T7EI, densitometric analysis, and ICE sequence deconvolution) for evaluation of gene editing provide evidence of mutagenesis in both G0 and G1 *L. longipalpis* after microinjection with CRISPR constructs targeting wing phenotype and olfactory genes. Taken individually, all methods were successful in identify a degree of mutagenesis in a number of samples, and mutagenesis predicted by both T7EI and ICE in the same sample, was demonstrated on 18 occasions - giving robust support for the occurrence of gene editing in these sand flies.

Discussion

The first evidence of *in vivo* PiggyBac supporting evidence for CRISPR-based targeted mutagenesis in *L. longipalpis* and *P. papatasi* vectors of *Leishmania* parasites through microinjection of plasmid constructs, and subsequent inheritance, is presented. We show evidence supporting CRISPR-Cas mediated mutagenesis under the control of the Ubiquitin-63E promoter, and specific gRNA expression under the control of U6 promoters detected by heteroduplex assays and ICE analysis. applied to phenotypic marker genes, and olfactory genes involved in host detection is shown.

We also present evidence of phenotypic mutation resulting in morphological change in a small number of individual sand flies (see Results, *Phenotypic analysis of immature and mature G0 and G1 sand flies*) namely alterations in wing morphology and pigmentation in pupae.

Microinjection

Successful embryo microinjection relies on several factors that vary between different insect species. In sand flies, the melanisation state of the egg can both prevent microinjection needles from penetrating the chorion, or conversely render them prone to rupture when embryos are un-melanised. The melanisation state of *L. longipalpis* embryos is unpredictable, with some being laid fully melanised alongside un-melanised embryos (Jeffries et al., 2018).

An important aspect for microinjection into insect eggs is injection into the right location, and at the right time, with respect to embryogenesis. For modification using PiggyBac and CRISPR it is vital to introduce constructs during early embryogenesis. This increases the chance of modifications occurring in germline cells before extensive cell division and membrane formation has taken place (M. Li et al., 2020). The posterior pole cells have been identified as precursors to germline cells, and therefore the optimal region for microinjection in mosquitoes eggs to maximise likelihood of inheritance (Biedler,

Hu, Tae, & Tu, 2012; M. Li et al., 2020). In *P. papatasi* germline cells have been identified at the posterior pole in the pre-blastoderm stage, and are present up to 36 hours post oviposition (Abbassy, Helmy, Osman, Cope, & Presley, 1995). This provides a longer window for microinjections to take place compared to the faster embryo development seen in mosquitoes (2-3 hours) (Biedler et al., 2012). In this study, sand fly embryos were collected and processed for injection under 3 hours from oviposition, ensuring that injections took place prior to blastoderm formation. The toughness of the cuticle was a key factor in successful injections. With experience, identification of embryos that were melanised to a degree that allowed smooth entry and exit of the needle without clogging or becoming stuck, was possible. Unfortunately, identification of the posterior pole in *L. longipalpis* and *P. papatasi* was difficult, and ultimately not viable as part of the injection protocol if embryos were to be injected before they became un-injectable.

Another aspect for successful microinjection is the concentration of the plasmids making up the injection mixture. We aimed to identify concentrations of injection mixture that could be delivered without causing significant mortality. Standard PiggyBac mixture concentrations used in mosquito vectors consist of 200ng/µL PiggyBac plasmid and 400ng/µL of helper transposase (Galizi et al., 2016; Simoni et al., 2020). For CRISPR approaches, mixtures consist of up to 600ng/µL. Using these as a starting point, lower concentrations were initially utilised due to the size disparity between sand fly and mosquito eggs, and the previous lack of success with microinjections within sand flies. A ten times lower concentration (~ 60ng/µL) was used in PiggyBac injections, which was increased to a maximum of 3.15 times less (~190ng/µL) than previously used. The results for *L. longipalpis* survival decreased as mixture concentration increased, however the reverse was observed with *P. papatasi*. For CRISPR construct injections approximately 2.75- 1.89 times less (~ 214 to 318ng/µL) was delivered. Survival rate was low in *L. longipalpis*, comparable to previous microinjection survival in sand flies (Louradour et al., 2019), although in *P. papatasi* survival was significantly lower.

These observations are unlikely to be a result of the increased concentration of the mixtures as previous *P. papatasi* injections have used mixture concentrations up to 380ng/µL (with Cas9 recombinant protein and sgRNA), showing survival to adults of 2.03% (Louradour et al., 2019), and 430ng/µL in *L. longipalpis* with 9.95% surviving to adult (Martin-Martin et al., 2018). It is more likely a result of the volume of mixture injected, which was not controlled, damage caused by microinjection needles, and the timing of the injections. Low survival rates in general are not surprising as these insects are non-model organisms, and methods, concentrations, and volumes for delivering constructs had to be adapted from more established insect models. In the future, injection mixtures could be further optimised.

Phenotypic analysis

Phenotypic analysis involved visually screening emergent survivors for the presence of fluorescence (after injection with UbiqCas9.874W and pHome-T PiggyBac plasmids) that should indicate construct

integration. Initial microinjection attempts were not successful, and no G0 PiggyBac transient individuals as determined by transient GFP and RFP expression were recovered.

In UbiqCas9.874W putative transgenics it was expected that DsRed1 fluorescence would indicate integration of the PiggyBac cargo, and EGFP fluorescence would indicate Cas9 expression. However, DsRed1 and EGFP expression was not observed in any samples, which may be explained by the construct design. The Cas9 in this construct is tagged by EGFP and linked via a T2A ribosome-skipping element that results in two separate proteins, rather than a hybrid protein. The use of this element has been shown to result in a 70% reduction in the expression of the protein that follows the T2A (Liu et al., 2017). Additionally, the absence of EGFP expression does not necessarily indicate the absence of Cas9 expression. This was borne out in subsequent results demonstrating the presence of the Cas9 (see below, *Molecular and in silico analysis*).

In pHome transgenics, RFP expression was expected as an indication of cargo integration (promoted by *D. melanogaster* Actin 5C), alongside GFP expression promoted by the eye-specific *3xP3* in adults or immature life cycle stages, as seen in *Drosophila* (Horn, Jaunich, & Wimmer, 2000). No RFP fluorescence was observed, potentially due to the Actin 5C promoter being non-specific to sand fly species, nor was GFP observed within larval or adult stages of injection survivors. However, positive integration of PiggyBac pHome-T constructs was again identified on a molecular level (see below, *Molecular and in silico analysis*).

Phenotypic wing modification was identified in *L. longipalpis* post microinjection with CRISPR knockin constructs (pDCC6: V1, V2, V3, R1, R2, R3) and mating crosses, with three G1 flies having mutations indicated by the presence of malformed wing development and shape (no wing, or pointed wings) (see Figure 8). This was a very low number of individuals (3), and potentially low survival could be due to modification of wing development having a knock on effect on pupal development, or eclosion from pupal casing, as seen in G0 *Vestigial* knockouts in *Ae. aegypti* (Ming Li et al., 2017). The low numbers in our study overall are not inconsistent with microinjections previously attempted in sand flies, where 2.04% survived to adulthood (Louradour et al., 2019). Additionally, G0 flies with undeveloped wings may be unable to fly properly, or to mate, thus reducing the chances of observing G1 flies with wing deformation.

The putative dark pupae phenotypes observed in two G0 survivors of CRISPR constructs targeting *Ebony P. papatasi* (Figure 7) is very similar to that seen in *Ae. aegypti* (Ming Li et al., 2017) and *Cx. quinquefasciatus* (Feng, Kambic, et al., 2021) when orthologues of this gene were interrupted. This phenotype occurs because loss of function of the *Ebony* gene increases black cuticle pigmentation (Ming Li et al., 2017). The low number of immature stage insects with an *Ebony* phenotype may be in part due to previous findings suggesting *Ebony* G0s are difficult to identify from wild type (Feng, Kambic, et al., 2021). Identification of alternative non-lethal phenotypic marker genes to target may

increase the chances of successfully identifying mutant G0 or G1 sand flies. Potential markers include the *kh* gene which if knocked out results in a white-eyed mutant (Aryan et al., 2013), or the *yellow* gene which if knocked out alters cuticular pigmentation. Both have orthologous genes in sand flies.

Molecular and *in silico* analysis

Phenotypic analysis was suggestive of induced phenotypic change in a small number of emergent individuals (described above). Parallel molecular analysis to detect possible mutagenesis at the level of genes was particularly important as the fluorescence markers were not observed in the case of PiggyBac and HDR inserts. Molecular analysis is also important in knockout targets that do not result in obvious physical phenotypes.

The gold standard to identify CRISPR mutations is amplicon sequencing (NGS), which allows identification of all alleles that have occurred within an individual or pooled DNA sample, and provides the frequency data for affected alleles. This method has been used consistently to identify gene disruption or identify the development of resistance alleles in insect vectors (Feng, Kambic, et al., 2021; Hammond et al., 2021; Ming Li et al., 2021). However, this method requires complex preparation steps and is expensive, therefore it was not possible to include in this research. Five alternative molecular and *in silico* methods for analysis of gene editing were identified, and successfully demonstrated in this thesis. These included Sanger sequencing, alongside heteroduplex assays and densitometric analysis, and algorithmic sequence deconvolution (ICE), which to our knowledge have not previously been attempted in *L. longipalpis* and *P. papatasi.*

Sanger sequencing identified Cas9 positive G0 in *L. longipalpis* and *P. papatasi* in pooled larval DNA samples, and individual adults DNA samples post UbiqCas9.874W injections (1 pool and 3 adults, 2 pools and 2 adults, respectively). In addition, pooled larval samples from *L. longipalpis* and *P. papatasi* confirmed the presence of GFP after pHome-T injections (7/60 and 7/14 pooled samples were positive, respectively. These results clearly demonstrate the first successful use of PiggyBac mediated gene editing in these sand fly species, providing a foundation for future editing attempts. In contrast, gene editing mediated by CRISPR constructs was more difficult to ascertain using Sanger sequencing, as a result of the PCR and sequencing method amplifying the dominant alleles within the DNA samples. In our microinjections of sand fly eggs using CRISPR, it is likely that only a small number of cells within the sand fly embryo were modified, resulting in mosaicism. Therefore, the majority of cells were unmodified, leading to the dominant sequencing calls masking the minority of modified alleles. Sanger sequencing alone was not sufficient to identify whether CRISPR-Cas9 modification had occurred with the phenotypic and olfactory genes selected.

We utilised heteroduplex assays (T7E1) as an alternative to assess CRISPR modifications within sand fly DNA samples. T7E1 enabled detection of modifications, with more sensitivity than the phenotypic

visualisation as evidenced by cleavage of DNA samples extracted from G1 survivors targeted by the wing development CRISPR constructs (6 samples with R3-pDCC6, 6 with V1-pDCC6, and 9 with V2-pDCC6). In addition, the assay enabled detection of modifications in the olfactory gene targets in G0 and G1 sand flies (6 G0 samples with Gr2-pDCC6, 4 with Ir8a-1-pDCC6, three G1 with Gr2-pDCC6, and 18 G1 with Orco-pDCC6) which were not possible to identify by Sanger sequencing. Densitometric analysis was able to quantify the modifications detected by the T7E1 assays, providing estimated percentages of modification from 0.15% up to 57.64%. The combination of T7E1 and Densitometric analysis require further optimisation, yet provide a simple and inexpensive method to identify potential samples of interest to be analysed further.

A range of bioinformatics tools are now available to allow elucidation of modification using sequence data. Advanced tools can analyse amplicon sequences, however these cannot be used to analyse Sanger sequence data; alternative algorithms have been designed for use with Sanger sequence data. Algorithms have been used to identify modifications in disease vectors including mosquitos and ticks (Sharma et al., 2022). Anderson *et al* (2019) used the ICE algorithm (Synthego, USA) alongside an alternative tool, InDelphi, which were able to predict the most abundant indels generated by gRNAs, although not for every gRNA used in the study with *Cx. quinquefasciatus*. We successfully utilised the ICE algorithm to analyse our Sanger sequence data, identifying indels in G0 and G1 for wing target genes and olfactory target genes. In particular, we demonstrated indels in G0 Orco genes and G1 Gr2, Ir8a and Orco genes ($R^2 = 0.87$, $R^2 = 0.85 – 0.89$, $R^2 = 0.91 – 0.96$, and $R^2 = 0.98 – 0.99$, respectively), with a strong fit to the ICE model. However, there were multiple samples (see Appendix 17-21) where the quality of the Sanger data was insufficient for the ICE algorithm to function, despite electropherogram traces being clear within other software (SnapGene, Genious, BioEdit).

The use of these *in silico* tools is imperfect, however with greater use in the context of insect vector modification the algorithms are likely to improve with increased datasets, making them an invaluable tool for validating CRISPR gene editing, especially when amplicon sequencing access is limited.

Overall, multiple samples (see Results, *In silico Sanger sequence deconvolution* and *Summary of Results*) analysed via Sanger sequencing, heteroduplex assays, and the ICE algorithm demonstrated consensus positivity indicative of mutagenesis across all of these analysis methods.

The results presented in this chapter demonstrate robust evidence for the first demonstration of *in vivo* gene editing in *Lutzomyia longipalpis* through the application of the PiggyBac based transposon system. Furthermore, evidence is provided for CRISPR based knockouts of phenotypic and olfaction targets evidenced by sequencing, associated ICE analysis, and further supported by heteroduplex detection assays. CRISPR induced phenotypic mutagenesis is also evidenced in a small number of affected

individual sand fly larvae and adults. Together the results demonstrate the ability perform *in vivo* microinjections at scale (10,749) and perform downstream processing. This in spite of COVID induced interruptions that severely impacted planned experiments and access to insectaries at collaborator facilities and also wet lab facilities. Together the bioinformatics pathway (see Chapter 3) and description of an *in vitro* platform (see Chapter 4) in tandem with the *in vitro* pipeline provide a platform to investigate further gene editing approaches in the context of interrupting transmission of leishmaniasis.

References

Abbassy, M. M., Helmy, N., Osman, M., Cope, S. E., & Presley, S. M. (1995). Embryogenesis of the Sand Fly Phlebotomus papatasi (Diptera: Psychodidae): Cell Cleavage, Blastoderm Formation, and Gastrulation. *Annals of the Entomological Society of America*, *88*(6), 815–820. https://doi.org/10.1093/aesa/88.6.815

Anderson, M. A. E., Samuel, G. H., Adelman, Z. N., Basu, S., Myles, K. M., Aryan, A., … Dahlem, T. J. (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. https://doi.org/10.1073/pnas.1502370112

Anderson, M. E., Mavica, J., Shackleford, L., Flis, I., Fochler, S., Basu, S., & Alphey, L. (2019). CRISPR/Cas9 gene editing in the West Nile Virus vector, Culex quinquefasciatus Say. *PLoS ONE*, *14*(11), 1–10. https://doi.org/10.1371/journal.pone.0224857

Aryan, A., Anderson, M. A. E., Myles, K. M., & Adelman, Z. N. (2013). TALEN-Based Gene Disruption in the Dengue Vector Aedes aegypti. *PLoS ONE*, *8*(3). https://doi.org/10.1371/journal.pone.0060082

Basu, S., Aryan, A., Overcash, J. M., Samuel, G. H., Anderson, M. A. E., Dahlem, T. J., … 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 of the United States of America*, *112*(13), 4038–4043. https://doi.org/10.1073/pnas.1502370112

Beumer, K., Bhattacharyya, G., Bibikova, M., Trautman, J. K., & Carroll, D. (2006). Efficient gene targeting in Drosophila with zinc-finger nucleases. *Genetics*, *172*(4), 2391–2403. https://doi.org/10.1534/genetics.105.052829

Bibikova, M., Beumer, K., Trautman, J. K., & Carroll, D. (2003). Enhancing gene targeting with designed zinc finger nucleases. *Science*, *300*(5620), 764. https://doi.org/10.1126/science.1079512

Biedler, J. K., Hu, W., Tae, H., & Tu, Z. (2012). Identification of early zygotic genes in the yellow fever mosquito Aedes aegypti and discovery of a motif involved in early zygotic genome activation. *PLoS ONE*, *7*(3), 6–14. https://doi.org/10.1371/journal.pone.0033933

Bloh, K., Kanchana, R., Bialk, P., Banas, K., Zhang, Z., Yoo, B. C., & Kmiec, E. B. (2021). Deconvolution of Complex DNA Repair (DECODR): Establishing a Novel Deconvolution Algorithm for Comprehensive Analysis of CRISPR-Edited Sanger Sequencing Data. *CRISPR Journal*, *4*(1), 120–131. https://doi.org/10.1089/crispr.2020.0022

Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., … Bonas, U. (2009). Breaking the code of DNA binding specificity of TAL-type III effectors. *Science*, *326*(5959), 1509–1512. https://doi.org/10.1126/science.1178811

Brinkman, E. K., Chen, T., Amendola, M., & Van Steensel, B. (2014). Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Research*, *42*(22), 1–8. https://doi.org/10.1093/nar/gku936

Brinkman, E. K., Kousholt, A. N., Harmsen, T., Leemans, C., Chen, T., Jonkers, J., & Van Steensel, B. (2018). Easy quantification of template-directed CRISPR/Cas9 editing. *Nucleic Acids Research*, *46*(10), E58. https://doi.org/10.1093/nar/gky164

Carroll, D. (2011). Genome engineering with zinc-finger nucleases. *Genetics*, *188*(4), 773–782.

https://doi.org/10.1534/genetics.111.131433

Cermak, T., Doyle, E. L., Christian, M., Wang, L., Zhang, Y., Schmidt, C., … Voytas, D. F. (2011). Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Research*, *39*(12), 1–11. https://doi.org/10.1093/nar/gkr218

Chen, Z., Traniello, I. M., Rana, S., Cash-Ahmed, A. C., Sankey, A. L., Yang, C., & Robinson, G. E. (2021). Neurodevelopmental and transcriptomic effects of CRISPR/Cas9-induced somatic orco mutation in honey bees. *Journal of Neurogenetics*, *35*(3), 320–332. https://doi.org/10.1080/01677063.2021.1887173

Conant, D., Hsiau, T., Rossi, N., Oki, J., Maures, T., Waite, K., … Stoner, R. (2022). Inference of CRISPR Edits from Sanger Trace Data. *CRISPR Journal*, *5*(1), 123–130. https://doi.org/10.1089/crispr.2021.0113

DeGennaro, M., McBride, C. S., Seeholzer, L., Vosshall, L. B., Jasinskiene, N., James, A. A., … Dennis, E. J. (2013). orco mutant mosquitoes lose strong preference for humans and are not repelled by volatile DEET. *Nature*, *498*(7455), 487–491. https://doi.org/10.1038/nature12206

Dehairs, J., Talebi, A., Cherifi, Y., & Swinnen, J. V. (2016). CRISP-ID: Decoding CRISPR mediated indels by Sanger sequencing. *Scientific Reports*, *6*(August 2015), 1–5. https://doi.org/10.1038/srep28973

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), 1–13. https://doi.org/10.1371/journal.pone.0122353

Dong, Y., Simões, M. L., Marois, E., & Dimopoulos, G. (2018). CRISPR/Cas9 -mediated gene knockout of Anopheles gambiae FREP1 suppresses malaria parasite infection. *PLoS Pathogens*, *14*(3), 1–16. https://doi.org/10.1371/journal.ppat.1006898

Feng, X., Kambic, L., Nishimoto, J. H. K., Reed, F. A., Denton, J. A., Sutton, J. T., & Gantz, V. M. (2021). Evaluation of Gene Knockouts by CRISPR as Potential Targets for the Genetic Engineering of the Mosquito Culex quinquefasciatus. *CRISPR Journal*, *4*(4), 595–608. https://doi.org/10.1089/crispr.2021.0028

Feng, X., López Del Amo, V., Mameli, E., Lee, M., Bishop, A. L., Perrimon, N., & Gantz, V. M. (2021). Optimized CRISPR tools and site-directed transgenesis towards gene drive development in Culex quinquefasciatus mosquitoes. *Nature Communications*, *12*(1). https://doi.org/10.1038/s41467-021-23239-0

Fraser, M. J., Ciszczon, T., Elick, T., & Bauser, C. (1996). Precise excision of TTAA-specific lepidopteran transposons piggyBac (IFP2) and tagalong (TFP3) from the baculovirus genome in cell lines from two species of Lepidoptera. *Insect Molecular Biology*, *5*(2), 141–151. https://doi.org/10.1111/j.1365-2583.1996.tb00048.x

Gaj, T., Gersbach, C. A., & Barbas, C. F. (2013). ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends in Biotechnology*, *31*(7), 397–405. https://doi.org/10.1016/j.tibtech.2013.04.004

Galizi, R., Hammond, A., Kyrou, K., Taxiarchi, C., Bernardini, F., O'Loughlin, S. M., … Crisanti, A. (2016). A CRISPR-Cas9 sex-ratio distortion system for genetic control. *Scientific Reports*, *6*(April), 2–6. https://doi.org/10.1038/srep31139

Gantz, V. M., Tatarenkova, O., Macias, V. M., James, A. A., Fazekas, A., Bier, E., & Jasinskiene, N. (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. https://doi.org/10.1073/pnas.1521077112

Grossman, G. L., Rafferty, C. S., Clayton, J. R., Stevens, T. K., Mukabayire, O., & Benedict, M. Q. (2001). *Germline transformation of the malaria vector , Anopheles gambiae , with the piggyBac transposable element*. *10*(May), 597–604.

Hall, A. B., Basu, S., Jiang, X., Qi, Y., Timoshevskiy, V. A., Biedler, J. K., … Tu, Z. (2015). A male-determining factor in the mosquito Aedes aegypti. *Science*, *348*(6240), 1268–1270. https://doi.org/10.1126/science.aaa2850

Hammond, A., Galizi, R., Kyrou, K., Simoni, A., Siniscalchi, C., Katsanos, D., … Nolan, T. (2016). A CRISPR-Cas9 gene drive system targeting female reproduction in the malaria mosquito vector Anopheles gambiae. *Nature Biotechnology*, *34*(1), 78–83. https://doi.org/10.1038/nbt.3439

Hammond, A., Pollegioni, P., Persampieri, T., North, A., Minuz, R., Trusso, A., … Crisanti, A. (2021). Gene-drive suppression of mosquito populations in large cages as a bridge between lab and field. *Nature Communications*, *12*(1), 1–9. https://doi.org/10.1038/s41467-021-24790-6

Hoermann, A., Habtewold, T., Selvaraj, P., Del Corsano, G., Capriotti, P., Inghilterra, M. G., … Windbichler, N. (2022). Gene drive mosquitoes can aid malaria elimination by retarding Plasmodium sporogonic development. *Science Advances*, *8*(38), 1–9. https://doi.org/10.1126/sciadv.abo1733

Horn, C., Jaunich, B., & Wimmer, E. A. (2000). Highly sensitive, fluorescent transformation marker for Drosophila transgenesis. *Development Genes and Evolution*, *210*(12), 623–629. https://doi.org/10.1007/s004270000111

Ito, J., Ghosh, A., Moreira, L. A., & Wimmer, E. A. (2002). *Transgenic anopheline mosquitoes impaired in transmission of a malaria parasite*. *417*(May).

Jeffries, C. L., Rogers, M. E., & Walker, T. (2018). Establishment of a method for Lutzomyia longipalpis sand fly egg microinjection: The first step towards potential novel control strategies for leishmaniasis. *Wellcome Open Research*, *3*(0), 55. https://doi.org/10.12688/wellcomeopenres.14555.2

Kistler, K. E., Vosshall, L. B., & Matthews, B. J. (2015). Genome engineering with CRISPR-Cas9 in the mosquito aedes aegypti. *Cell Reports*, *11*(1), 51–60. https://doi.org/10.1016/j.celrep.2015.03.009

Kokoza, V., Ahmed, A., Wimmer, E. A., & Raikhel, A. S. (2001). Efficient transformation of the yellow fever mosquito Aedes aegypti using the piggyBac transposable element vector pBac[3xP3-EGFP afm]. *Insect Biochemistry and Molecular Biology*, *31*(12), 1137–1143. https://doi.org/10.1016/S0965-1748(01)00120-5

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

Labbé, G. M. C., Nimmo, D. D., & Alphey, L. (2010). Piggybac- and PhiC31-mediated genetic transformation of the Asian tiger mosquito, Aedes albopictus (Skuse). *PLoS Neglected Tropical Diseases*, *4*(8). https://doi.org/10.1371/journal.pntd.0000788

Lawyer, P., Volf, P., Rowton, E., Rowland, T., Killick-Kendrick, M., Rowland, T., … Volf, P. (2017). Laboratory colonization and mass rearing of phlebotomine sand flies (Diptera, Psychodidae). *Parasite*, *24*, 42. https://doi.org/10.1051/parasite/2017041

Li, M., Li, T., Liu, N., Raban, R. R., Wang, X., & Akbari, O. S. (2020). Methods for the generation of

heritable germline mutations in the disease vector Culex quinquefasciatus using clustered regularly interspaced short palindrome repeats-associated protein 9. *Insect Molecular Biology*, *29*(2), 214–220. https://doi.org/10.1111/imb.12626

Li, Ming, Bui, M., Yang, T., Bowman, C. S., White, B. J., & Akbari, O. S. (2017). Germline Cas9 expression yields highly efficient genome engineering in a major worldwide disease vector, Aedes aegypti. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(49), E10540–E10549. https://doi.org/10.1073/pnas.1711538114

Li, Ming, Yang, T., Bui, M., Gamez, S., Wise, T., Kandul, N. P., … Akbari, O. S. (2021). Suppressing mosquito populations with precision guided sterile males. *Nature Communications*, *12*(1). https://doi.org/10.1038/s41467-021-25421-w

Liu, Z., Chen, O., Wall, J. B. J., Zheng, M., Zhou, Y., Wang, L., … Liu, J. (2017). Systematic comparison of 2A peptides for cloning multi-genes in a polycistronic vector. *Scientific Reports*, *7*(1), 1–9. https://doi.org/10.1038/s41598-017-02460-2

Louradour, I., Ghosh, K., Inbar, E., & Sacks, D. L. (2019). CRISPR/Cas9 Mutagenesis in Phlebotomus papatasi: the Immune Deficiency Pathway Impacts Vector Competence for Leishmania major. *MBio*, *10*(4), 1–14.

Martin-Martin, I., Aryan, A., Meneses, C., Adelman, Z. N., & Calvo, E. (2018). Optimization of sand fly embryo microinjection for gene editing by CRISPR/Cas9. *PLoS Neglected Tropical Diseases*, *12*(9), 1–18. https://doi.org/10.1371/journal.pntd.0006769

McMeniman, C. J., Corfas, R. A., Matthews, B. J., Ritchie, S. A., & Vosshall, L. B. (2014). Multimodal integration of carbon dioxide and other sensory cues drives mosquito attraction to humans. *Cell*, *156*(5), 1060–1071. https://doi.org/10.1016/j.cell.2013.12.044

Moscou, M. J., & Bogdanove, A. J. (2009). A simple cipher governs DNA recognition by TAL effectors. *Science*, *326*(5959), 1501. https://doi.org/10.1126/science.1178817

Nolan, T., Bower, T. M., Brown, A. E., Crisanti, A., & Catteruccia, F. (2002). piggyBac-mediated germline transformation of the malaria mosquito Anopheles stephensi using the red fluorescent protein dsRED as a selectable marker. *Journal of Biological Chemistry*, *277*(11), 8759–8762. https://doi.org/10.1074/jbc.C100766200

Perera, O. P., Ii, R. A. H., & Handler, A. M. (2002). *Germ-line transformation of the South American malaria vector , Anopheles albimanus , with a piggyBac / EGFP transposon vector is routine and highly efficient*. *11*, 291–297.

Pinello, L., Canver, M. C., Hoban, M. D., Orkin, S. H., Kohn, D. B., Bauer, D. E., & Yuan, G. C. (2016). Analyzing CRISPR genome-editing experiments with CRISPResso. *Nature Biotechnology*, *34*(7), 695–697. https://doi.org/10.1038/nbt.3583

Purusothaman, D. K., Shackleford, L., Anderson, M. A. E., Harvey-Samuel, T., & Alphey, L. (2021). CRISPR/Cas-9 mediated knock-in by homology dependent repair in the West Nile Virus vector Culex quinquefasciatus Say. *Scientific Reports*, *11*(1), 1–8. https://doi.org/10.1038/s41598-021-94065-z

Rodrigues, F. G., Oliveira, S. B., Rocha, B. C., & Moreira, L. A. (2006). Germline transformation of Aedes fluviatilis (Diptera:Culicidae) with the piggyBac transposable element. *Memorias Do Instituto Oswaldo Cruz*, *101*(7), 755–757. https://doi.org/10.1590/S0074-02762006000700008

Sharma, A., Pham, M. N., Reyes, J. B., Chana, R., Yim, W. C., Heu, C. C., … Gulia-Nuss, M. (2022). Cas9-mediated gene editing in the black-legged tick, Ixodes scapularis, by embryo injection and ReMOT Control. *IScience*, *25*(3), 103781. https://doi.org/10.1016/j.isci.2022.103781

Simoni, A., Hammond, A. M., Beaghton, A. K., Galizi, R., Taxiarchi, C., Kyrou, K., … Crisanti, A. (2020). A male-biased sex-distorter gene drive for the human malaria vector Anopheles gambiae. *Nature Biotechnology*, *38*(9), 1054–1060. https://doi.org/10.1038/s41587-020-0508-1

Smidler, A. L., Terenzi, O., Soichot, J., Levashina, E. A., & Marois, E. (2013). Targeted Mutagenesis in the Malaria Mosquito Using TALE Nucleases. *PLoS ONE*, *8*(8), 1–9. https://doi.org/10.1371/journal.pone.0074511

Volf, P., & Volfova, V. (2011). Establishment and maintenance of sand fly olonies. *Journal of Vector Ecology*, *36*(SUPPL.1), 1–9. https://doi.org/10.1111/j.1948-7134.2011.00106.x

Anderson, M. A. E., Samuel, G. H., Adelman, Z. N., Basu, S., Myles, K. M., Aryan, A., … Dahlem, T. J. (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. https://doi.org/10.1073/pnas.1502370112

Anderson, M. E., Mavica, J., Shackleford, L., Flis, I., Fochler, S., Basu, S., & Alphey, L. (2019). CRISPR/Cas9 gene editing in the West Nile Virus vector, Culex quinquefasciatus Say. *PLoS ONE*, *14*(11), 1–10. https://doi.org/10.1371/journal.pone.0224857

Aryan, A., Anderson, M. A. E., Myles, K. M., & Adelman, Z. N. (2013). TALEN-Based Gene Disruption in the Dengue Vector Aedes aegypti. *PLoS ONE*, *8*(3). https://doi.org/10.1371/journal.pone.0060082

Basu, S., Aryan, A., Overcash, J. M., Samuel, G. H., Anderson, M. A. E., Dahlem, T. J., … 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 of the United States of America*, *112*(13), 4038–4043. https://doi.org/10.1073/pnas.1502370112

Beumer, K., Bhattacharyya, G., Bibikova, M., Trautman, J. K., & Carroll, D. (2006). Efficient gene targeting in Drosophila with zinc-finger nucleases. *Genetics*, *172*(4), 2391–2403. https://doi.org/10.1534/genetics.105.052829

Bibikova, M., Beumer, K., Trautman, J. K., & Carroll, D. (2003). Enhancing gene targeting with designed zinc finger nucleases. *Science*, *300*(5620), 764. https://doi.org/10.1126/science.1079512

Biedler, J. K., Hu, W., Tae, H., & Tu, Z. (2012). Identification of early zygotic genes in the yellow fever mosquito Aedes aegypti and discovery of a motif involved in early zygotic genome activation. *PLoS ONE*, *7*(3), 6–14. https://doi.org/10.1371/journal.pone.0033933

Bloh, K., Kanchana, R., Bialk, P., Banas, K., Zhang, Z., Yoo, B. C., & Kmiec, E. B. (2021). Deconvolution of Complex DNA Repair (DECODR): Establishing a Novel Deconvolution Algorithm for Comprehensive Analysis of CRISPR-Edited Sanger Sequencing Data. *CRISPR Journal*, *4*(1), 120–131. https://doi.org/10.1089/crispr.2020.0022

Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., … Bonas, U. (2009). Breaking the code of DNA binding specificity of TAL-type III effectors. *Science*, *326*(5959), 1509–1512. https://doi.org/10.1126/science.1178811

Brinkman, E. K., Chen, T., Amendola, M., & Van Steensel, B. (2014). Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Research*, *42*(22), 1–8. https://doi.org/10.1093/nar/gku936

Brinkman, E. K., Kousholt, A. N., Harmsen, T., Leemans, C., Chen, T., Jonkers, J., & Van Steensel, B. (2018). Easy quantification of template-directed CRISPR/Cas9 editing. *Nucleic Acids Research*,

46(10), E58. https://doi.org/10.1093/nar/gky164

Carroll, D. (2011). Genome engineering with zinc-finger nucleases. *Genetics*, *188*(4), 773–782. https://doi.org/10.1534/genetics.111.131433

Cermak, T., Doyle, E. L., Christian, M., Wang, L., Zhang, Y., Schmidt, C., … Voytas, D. F. (2011). Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Research*, *39*(12), 1–11. https://doi.org/10.1093/nar/gkr218

Clement K, Rees H, Canver MC, Gehrke JM, Farouni R, Hsu JY, … Pinello L. (2019). CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nature Biotechnology*, *37*(3), 224–226.

Conant, D., Hsiau, T., Rossi, N., Oki, J., Maures, T., Waite, K., … Stoner, R. (2022). Inference of CRISPR Edits from Sanger Trace Data. *CRISPR Journal*, *5*(1), 123–130. https://doi.org/10.1089/crispr.2021.0113

DeGennaro, M., McBride, C. S., Seeholzer, L., Vosshall, L. B., Jasinskiene, N., James, A. A., … Dennis, E. J. (2013). orco mutant mosquitoes lose strong preference for humans and are not repelled by volatile DEET. *Nature*, *498*(7455), 487–491. https://doi.org/10.1038/nature12206

Dehairs, J., Talebi, A., Cherifi, Y., & Swinnen, J. V. (2016). CRISP-ID: Decoding CRISPR mediated indels by Sanger sequencing. *Scientific Reports*, *6*(August 2015), 1–5. https://doi.org/10.1038/srep28973

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), 1–13. https://doi.org/10.1371/journal.pone.0122353

Dong, Y., Simões, M. L., Marois, E., & Dimopoulos, G. (2018). CRISPR/Cas9 -mediated gene knockout of Anopheles gambiae FREP1 suppresses malaria parasite infection. *PLoS Pathogens*, *14*(3), 1–16. https://doi.org/10.1371/journal.ppat.1006898

Feng, X., Kambic, L., Nishimoto, J. H. K., Reed, F. A., Denton, J. A., Sutton, J. T., & Gantz, V. M. (2021). Evaluation of Gene Knockouts by CRISPR as Potential Targets for the Genetic Engineering of the Mosquito Culex quinquefasciatus. *CRISPR Journal*, *4*(4), 595–608. https://doi.org/10.1089/crispr.2021.0028

Feng, X., López Del Amo, V., Mameli, E., Lee, M., Bishop, A. L., Perrimon, N., & Gantz, V. M. (2021). Optimized CRISPR tools and site-directed transgenesis towards gene drive development in Culex quinquefasciatus mosquitoes. *Nature Communications*, *12*(1). https://doi.org/10.1038/s41467-021-23239-0

Fraser, M. J., Ciszczon, T., Elick, T., & Bauser, C. (1996). Precise excision of TTAA-specific lepidopteran transposons piggyBac (IFP2) and tagalong (TFP3) from the baculovirus genome in cell lines from two species of Lepidoptera. *Insect Molecular Biology*, *5*(2), 141–151. https://doi.org/10.1111/j.1365-2583.1996.tb00048.x

Galizi, R., Hammond, A., Kyrou, K., Taxiarchi, C., Bernardini, F., O'Loughlin, S. M., … Crisanti, A. (2016). A CRISPR-Cas9 sex-ratio distortion system for genetic control. *Scientific Reports*, *6*(April), 2–6. https://doi.org/10.1038/srep31139

Gantz, V. M., Tatarenkova, O., Macias, V. M., James, A. A., Fazekas, A., Bier, E., & Jasinskiene, N. (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. https://doi.org/10.1073/pnas.1521077112

Grossman, G. L., Rafferty, C. S., Clayton, J. R., Stevens, T. K., Mukabayire, O., & Benedict, M. Q.

(2001). *Germline transformation of the malaria vector , Anopheles gambiae , with the piggyBac transposable element*. *10*(May), 597–604.

Hall, A. B., Basu, S., Jiang, X., Qi, Y., Timoshevskiy, V. A., Biedler, J. K., … Tu, Z. (2015). A male-determining factor in the mosquito Aedes aegypti. *Science*, *348*(6240), 1268–1270. https://doi.org/10.1126/science.aaa2850

Hammond, A., Galizi, R., Kyrou, K., Simoni, A., Siniscalchi, C., Katsanos, D., … Nolan, T. (2016). A CRISPR-Cas9 gene drive system targeting female reproduction in the malaria mosquito vector Anopheles gambiae. *Nature Biotechnology*, *34*(1), 78–83. https://doi.org/10.1038/nbt.3439

Hammond, A., Pollegioni, P., Persampieri, T., North, A., Minuz, R., Trusso, A., … Crisanti, A. (2021). Gene-drive suppression of mosquito populations in large cages as a bridge between lab and field. *Nature Communications*, *12*(1), 1–9. https://doi.org/10.1038/s41467-021-24790-6

Horn, C., Jaunich, B., & Wimmer, E. A. (2000). Highly sensitive, fluorescent transformation marker for Drosophila transgenesis. *Development Genes and Evolution*, *210*(12), 623–629. https://doi.org/10.1007/s004270000111

Ito, J., Ghosh, A., Moreira, L. A., & Wimmer, E. A. (2002). *Transgenic anopheline mosquitoes impaired in transmission of a malaria parasite*. *417*(May).

Jeffries, C. L., Rogers, M. E., & Walker, T. (2018). Establishment of a method for Lutzomyia longipalpis sand fly egg microinjection: The first step towards potential novel control strategies for leishmaniasis. *Wellcome Open Research*, *3*(0), 55. https://doi.org/10.12688/wellcomeopenres.14555.2

Kistler, K. E., Vosshall, L. B., & Matthews, B. J. (2015). Genome engineering with CRISPR-Cas9 in the mosquito aedes aegypti. *Cell Reports*, *11*(1), 51–60. https://doi.org/10.1016/j.celrep.2015.03.009

Kokoza, V., Ahmed, A., Wimmer, E. A., & Raikhel, A. S. (2001). Efficient transformation of the yellow fever mosquito Aedes aegypti using the piggyBac transposable element vector pBac[3xP3-EGFP afm]. *Insect Biochemistry and Molecular Biology*, *31*(12), 1137–1143. https://doi.org/10.1016/S0965-1748(01)00120-5

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

Labbé, G. M. C., Nimmo, D. D., & Alphey, L. (2010). Piggybac- and PhiC31-mediated genetic transformation of the Asian tiger mosquito, Aedes albopictus (Skuse). *PLoS Neglected Tropical Diseases*, *4*(8). https://doi.org/10.1371/journal.pntd.0000788

Lawyer, P., Volf, P., Rowton, E., Rowland, T., Killick-Kendrick, M., Rowland, T., … Volf, P. (2017). Laboratory colonization and mass rearing of phlebotomine sand flies (Diptera, Psychodidae). *Parasite*, *24*, 42. https://doi.org/10.1051/parasite/2017041

Li, M., Li, T., Liu, N., Raban, R. R., Wang, X., & Akbari, O. S. (2020). Methods for the generation of heritable germline mutations in the disease vector Culex quinquefasciatus using clustered regularly interspaced short palindrome repeats-associated protein 9. *Insect Molecular Biology*, *29*(2), 214–220. https://doi.org/10.1111/imb.12626

Li, Ming, Bui, M., Yang, T., Bowman, C. S., White, B. J., & Akbari, O. S. (2017). Germline Cas9 expression yields highly efficient genome engineering in a major worldwide disease vector, Aedes aegypti. *Proceedings of the National Academy of Sciences of the United States of America*,

114(49), E10540–E10549. https://doi.org/10.1073/pnas.1711538114

Li, Ming, Yang, T., Bui, M., Gamez, S., Wise, T., Kandul, N. P., … Akbari, O. S. (2021). Suppressing mosquito populations with precision guided sterile males. *Nature Communications*, *12*(1). https://doi.org/10.1038/s41467-021-25421-w

Liu, Z., Chen, O., Wall, J. B. J., Zheng, M., Zhou, Y., Wang, L., … Liu, J. (2017). Systematic comparison of 2A peptides for cloning multi-genes in a polycistronic vector. *Scientific Reports*, *7*(1), 1–9. https://doi.org/10.1038/s41598-017-02460-2

Louradour, I., Ghosh, K., Inbar, E., & Sacks, D. L. (2019). *CRISPR/Cas9 Mutagenesis in Phlebotomus papatasi: the Immune Deficiency Pathway Impacts Vector Competence for Leishmania major*. *10*(4), 1–14.

Martin-Martin, I., Aryan, A., Meneses, C., Adelman, Z. N., & Calvo, E. (2018). Optimization of sand fly embryo microinjection for gene editing by CRISPR/Cas9. *PLoS Neglected Tropical Diseases*, *12*(9), 1–18. https://doi.org/10.1371/journal.pntd.0006769

McMeniman, C. J., Corfas, R. A., Matthews, B. J., Ritchie, S. A., & Vosshall, L. B. (2014). Multimodal integration of carbon dioxide and other sensory cues drives mosquito attraction to humans. *Cell*, *156*(5), 1060–1071. https://doi.org/10.1016/j.cell.2013.12.044

Moscou, M. J., & Bogdanove, A. J. (2009). A simple cipher governs DNA recognition by TAL effectors. *Science*, *326*(5959), 1501. https://doi.org/10.1126/science.1178817

Nolan, T., Bower, T. M., Brown, A. E., Crisanti, A., & Catteruccia, F. (2002). piggyBac-mediated germline transformation of the malaria mosquito Anopheles stephensi using the red fluorescent protein dsRED as a selectable marker. *Journal of Biological Chemistry*, *277*(11), 8759–8762. https://doi.org/10.1074/jbc.C100766200

Perera, O. P., Ii, R. A. H., & Handler, A. M. (2002). *Germ-line transformation of the South American malaria vector , Anopheles albimanus , with a piggyBac / EGFP transposon vector is routine and highly efficient*. *11*, 291–297.

Pinello, L., Canver, M. C., Hoban, M. D., Orkin, S. H., Kohn, D. B., Bauer, D. E., & Yuan, G. C. (2016). Analyzing CRISPR genome-editing experiments with CRISPResso. *Nature Biotechnology*, *34*(7), 695–697. https://doi.org/10.1038/nbt.3583

Purusothaman, D. K., Shackleford, L., Anderson, M. A. E., Harvey-Samuel, T., & Alphey, L. (2021). CRISPR/Cas-9 mediated knock-in by homology dependent repair in the West Nile Virus vector Culex quinquefasciatus Say. *Scientific Reports*, *11*(1), 1–8. https://doi.org/10.1038/s41598-021-94065-z

Rodrigues, F. G., Oliveira, S. B., Rocha, B. C., & Moreira, L. A. (2006). Germline transformation of Aedes fluviatilis (Diptera:Culicidae) with the piggyBac transposable element. *Memorias Do Instituto Oswaldo Cruz*, *101*(7), 755–757. https://doi.org/10.1590/S0074-02762006000700008

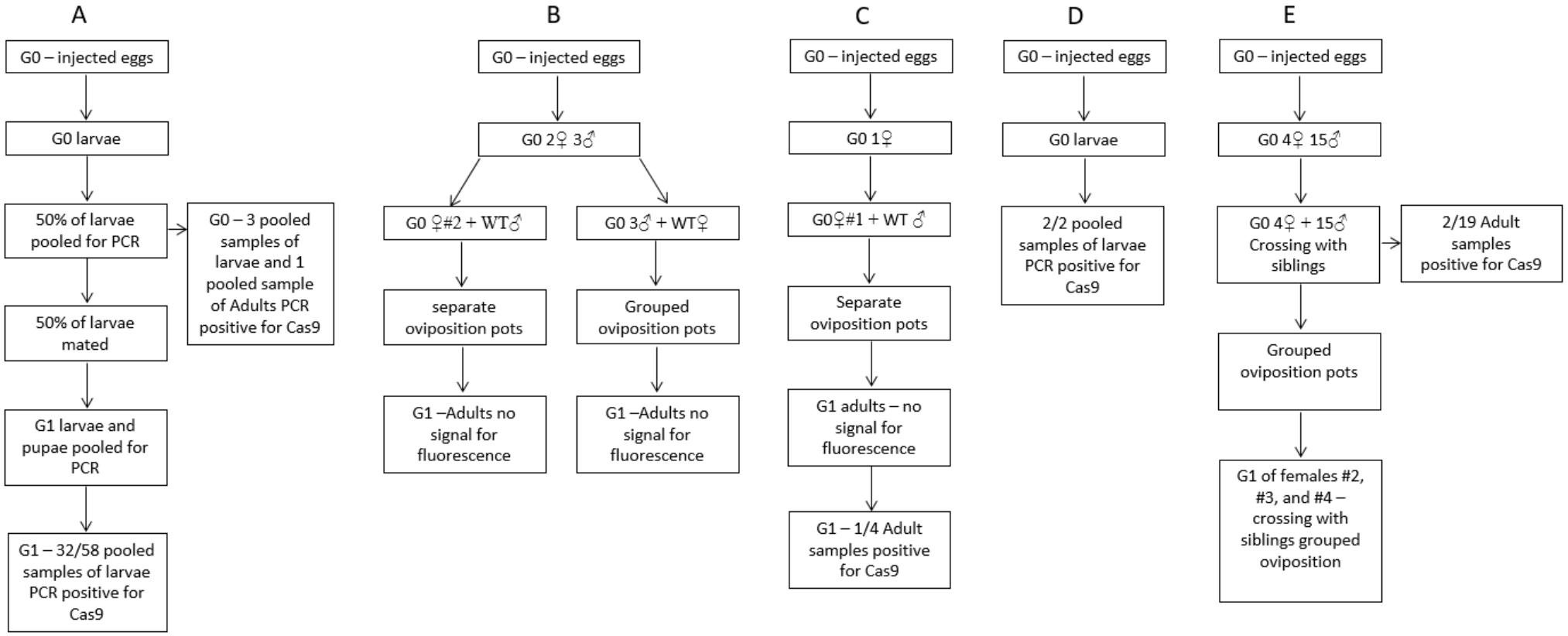
Simoni, A., Hammond, A. M., Beaghton, A. K., Galizi, R., Taxiarchi, C., Kyrou, K., … Crisanti, A. (2020). A male-biased sex-distorter gene drive for the human malaria vector Anopheles gambiae. *Nature Biotechnology*, *38*(9), 1054–1060. https://doi.org/10.1038/s41587-020-0508-1

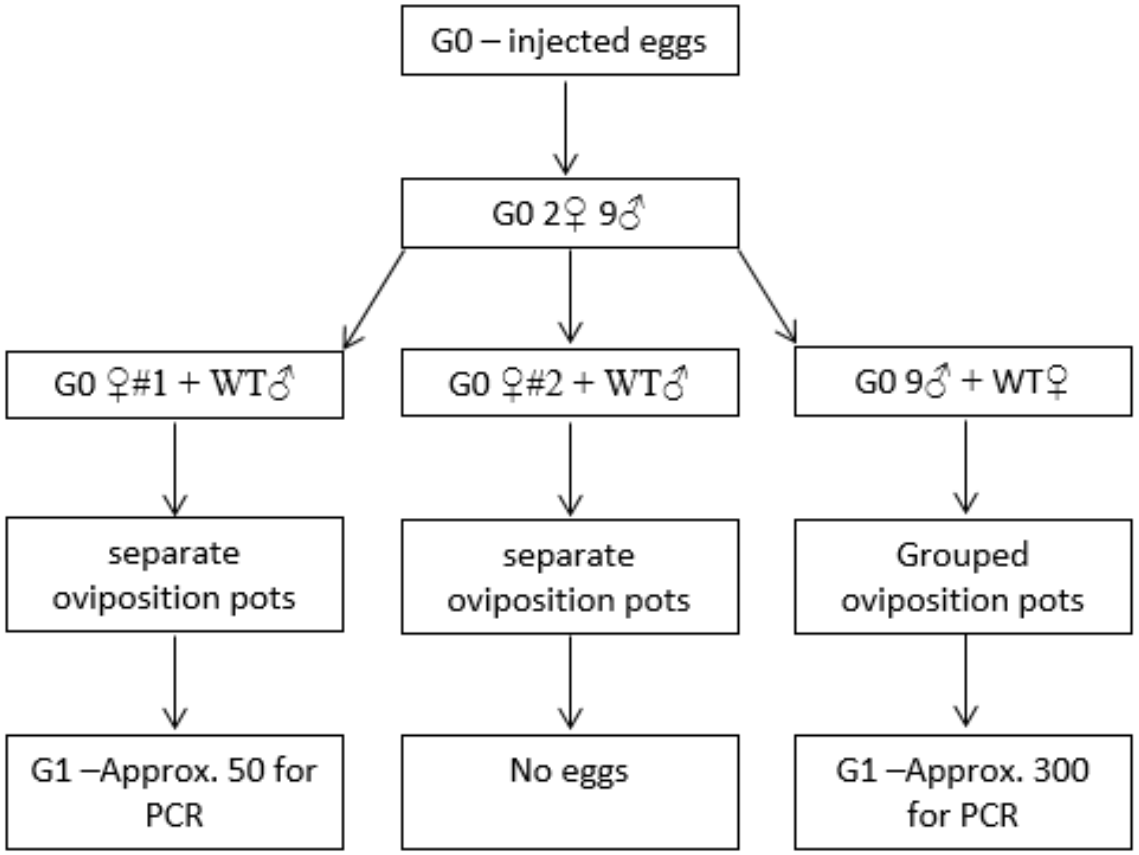
Smidler, A. L., Terenzi, O., Soichot, J., Levashina, E. A., & Marois, E. (2013). Targeted Mutagenesis in the Malaria Mosquito Using TALE Nucleases. *PLoS ONE*, *8*(8), 1–9. https://doi.org/10.1371/journal.pone.0074511

Volf, P., & Volfova, V. (2011). Establishment and maintenance of sand fly olonies. *Journal of Vector Ecology*, *36*(SUPPL.1), 1–9. https://doi.org/10.1111/j.1948-7134.2011.00106.x

Chapter 5 appendices

Appendix 1. *L. longipalpis* and *P. papatasi* UbiqCas9.874W injection survivors mating strategy and results. A) *L. longipalpis* round 1 injection survivors, with G0 and G1 PCR results indicated. B) *L. longipalpis* round 2 injection survivors. C) *L. longipalpis* round 3 injection survivors, with G1 PCR results indicated. D) *P. papatasi* round 1 injection survivors, with G0 PCR results indicated. E) *P. papatasi* round 2 injection survivors, with G0 PCR results indicated.





Appendix 2. ***L. longipalpis*** **Olfactory pDCC6 (Gr2-1, Gr2-2, Ir8a-1, Ir8a-2, Orco-1, and Orco-2) injection survivors mating strategy.**

Appendix 3. PiggyBac and Helper Transposase plasmids injection mixture protocol.

PiggyBac and Helper Transposase plasmids injection mixture protocol

1. **Calculate the volume of the PiggyBac plasmid needed to give the desired concentration.**

   = Desired PiggyBac conc. (ng/µl) / Actual PiggyBac conc. (ng/µl)
   = Volume of PiggyBac (µl)

   e.g. = 200ng/µl / 449ng/µl = 0.45µl

2. **Calculate the volume of the Helper plasmid needed to give the desired concentration.**
   = Desired Helper conc. (ng/µl) / Actual Helper conc. (ng/µl)
   = Volume of Helper (µl)

   e.g. = 400ng/µl / 538.7ng/µl = 0.74µl

3. **Determine the desired final volume of your injection mixture**

   e.g. 5µl

4. **Calculate the multiplication factor for the piggyBac and helper plasmids**
   = Final mixture volume / (Volume of PiggyBac + Volume of Helper)

   e.g. = 5µl / (0.45µl +0.74µl) = 4.21

5. **Calculate the Final volume of the PiggyBac and Helper plasmids needed to make up an injection mixture of your desired final volume.**
   = Final volume of PiggyBac x multiplication factor AND Final volume of Helper x multiplication factor

   e.g. = 0.45µl x 4.21 = 1.87µl AND = 0.74µl x 4.21 = 3.13µl

6. **Determine the final concentration of the plasmids within the final mixture**
   = (Actual PiggyBac conc. x Final volume of PiggyBac)/ (Final mixture volume)

   e.g. = (449ng/µl x 1.87µl)/ 5µl = 168.36ng/µl
   e.g. = (538.7ng/µl x 3.13µl)/ 5µl = 336.71ng/µl

7. **Determine the Ratio of the plasmids**

   Example using pHome (449ng/µl) and IhypBase (538.7ng/µl).

Appendix 4. CRISPR knockout plasmids injection mixture protocol

CRISPR knockout plasmids injection mixture protocol

1. **Calculate the volume of the pDCC6 plasmid needed to give the desired concentration.**

   = Desired pDCC6 conc. (ng/µl) / Actual pDCC6 conc. (ng/µl)
   = Volume of pDCC6 (µl)

   e.g. = 100ng/µl / 314ng/µl = 0.32µl
   e.g. = 100ng/µl / 614ng/µl = 0.61µl
   e.g. = 100ng/µl / 473ng/µl = 0.21µl
   e.g. = 100ng/µl / 337ng/µl = 0.30µl

2. **Determine the desired final volume of your injection mixture**

   e.g. 5µl

3. **Calculate the multiplication factor for the pDCC6 plasmids**

   = Final mixture volume / (Sum of Volume of pDCC6(s))

   e.g. = 5µl / (0.32µl + 0.61µl + 0.21µl + 0.30µl) = 3.48

4. **Calculate the Final volume of the pDCC6 plasmids needed to make up an injection mixture of your desired final volume.**
   = Final volume of pDCC6(s) x multiplication factor

   e.g. = 0.32µl x 3.48 = 1.11µl
   e.g. = 0.61µl x 3.48 = 2.12µl
   e.g. = 0.21µl x 3.48 = 0.74µl
   e.g. = 0.30µl x 3.48 = 1.03µl

5. **Determine the final concentration of the plasmids within the final mixture**
   = (Actual pDCC6 conc. x Final volume of pDCC6)/ (Final mixture volume)

   e.g. = (314ng/µl x 1.11µl)/ 5µl = 69.62ng/µl
   e.g. = (614ng/µl x 2.12µl)/ 5µl = 69.62ng/µl
   e.g. = (473ng/µl x 0.74µl)/ 5µl = 69.62ng/µl
   e.g. = (337ng/µl x 1.03µl)/ 5µl = 69.62ng/µl

6. **Determine the Ratio of the plasmids**

   Example using pDCC6 plasmids: Ebony gRNA1 (314ng/µl), Ebony gRNA2 (164ng/µl), Ebony gRNA3 (473ng/µl), Ebony gRNA4 (337ng/µl)

Appendix 5. HDR CRISPR knockin plasmids injection mixture protocol

HDR CRISPR knockin plasmids injection mixture protocol

1. **Calculate the volume of the HDR donor plasmid needed to give the desired concentration.**

   = Desired HDR conc. (ng/µl) / Actual HDR conc. (ng/µl)
   = Volume of HDR (µl)

   e.g. = 100ng/µl / 176.7ng/µl = 0.57µl

2. **Calculate the volume of the pDCC6 knockout plasmid(s) needed to give the desired concentration.**

   = Desired pDCC6 conc. (ng/µl) / Actual pDCC6 conc. (ng/µl)
   = Volume of pDCC6 (µl)

   e.g. = 100ng/µl / 243.3ng/µl = 0.41µl
   e.g. = 100ng/µl / 195.93ng/µl = 0.51µl

3. **Determine the desired final volume of your injection mixture**

   e.g. 5µl

4. **Calculate the multiplication factor for the pDCC6 plasmids**

   = Final mixture volume / (Sum of Volume of HDR and pDCC6(s))

   e.g. = 5µl / (0.57µl + 0.41µl + 0.51µl) = 3.36

5. **Calculate the Final volume of the pDCC6 plasmids needed to make up an injection mixture of your desired final volume.**

   = Final volume of HDR and pDCC6(s) x multiplication factor

   e.g. = 0.57µl x 3.36= 1.90µl
   e.g. = 0.41µl x 3.36= 1.38µl
   e.g. = 0.51µl x 3.36= 1.72µl

6. **Determine the final concentration of the plasmids within the final mixture**

   = (Actual pDCC6 conc. x Final volume of pDCC6)/ (Final mixture volume)

   e.g. = (176.7ng/µl x 1.90µl)/ 5µl = 67.23ng/µl
   e.g. = (243.3ng/µl x 1.38µl)/ 5µl = 67.23ng/µl
   e.g. = (195.93ng/µl x 1.72µl)/ 5µl = 67.23ng/µl

7. **Determine the Ratio of the plasmids**

Appendix 6. Second round Injection mixtures used for microinjection of *L. longipalpis* and *P. papatasi* eggs.

Species injected	*Plasmid(s)*	*Plasmid conc. (ng/µl)*	*plasmid vol. (µl)*	*dH2O vol. (µl)*	*Final plasmid conc. in mixture (ng/µl)*	*Plasmid Ratio*
L. longipalpis	pDCC6: Gr2-1 Gr2-2 Ir8a-1 Ir8a-2 Orco12 Orco 34	 335.5 342.3 278.0 345.6 265.2 359.5	 0.30 0.29 0.36 0.29 0.38 0.28	0	 52.78 52.78 52.78 52.78 52.78 52.78	1:1:1:1:1:1
P. papatasi	pDCC6: Ebony gRNA1 Ebony gRNA2 Ebony gRNA3 Ebony gRNA4	 314.0 164.0 473.0 337.0	 0.32 0.61 0.21 0.30	0	 69.62 69.62 69.62 69.62	1:1:1:1
L. longipalpis	pDCC6: R1 R2 R3 V1 V2 V3	 299.1 335.6 370.6 324.7 405.3 231.2	 0.33 0.30 0.27 0.31 0.25 0.43	0	 52.93 52.93 52.93 52.93 52.93 52.93	1:1:1:1:1:1
P. papatasi	pDCC6: Caspar1 Caspar2* Caspar3 Caspar4	 242 87 200 262			 68.14 10.24 68.14 68.14	1:0.15:1:1
L. longipalpis	HDR: pDSRED 3' St pDCC6 5' St pDCC6	 176.7 243.3 195.9			 67.23 67.23 67.23	1:1:1
L. longipalpis	PiggyBac: UbiqCas9.874w IHypBase	 80.7 538.7			 62.10 124.19	1:2

**CASPAR mixture: concentration for Caspar gRNA2 was too low and would lower down the overall concentrations of the three alternative Caspar-pDCC6 plasmids. Volumes for Caspar1,Caspar3, and Caspar4 were calculated, and 2ul of Caspar2 was added, making a final volume of 17µl. Final concentrations were 68.14ng/µl for each of Caspar1,Caspar3, and Caspar4 and 10.24ng/µl for Caspar2.*

Appendix 7. Injection mixtures used for microinjection of *L. longipalpis* (LL) *and P. papatasi* (PP) eggs, total number injected and survival rate from injection.

Plasmid	*Actual conc. (ng/µl)*	*Final Conc. (ng/µl)*	*Mixture conc.(ng/µl)*	*Ratio of plasmids*	*Injected*	*Larvae number*	*Survival (%)*
UbiqCas9.874W	82.9	19.98					
IhyPBase	414.3	41.43	61.41	1;2	1031(LL)	68	6.6
UbiqCas9.874W	98.5	20.00					
IhyPBase	538.7	39.86	59.86	1;2	609(PP)	0	0
UbiqCas9.874W	98.5	39.99					
IhyPBase	538.7	80.27	120.26	1;2	658(PP)	13	2.13
pHOME	449.5	20.23			749(LL)	46	6.14
IhyPBase	414.3	41.43	61.66	1;2	664(PP)	18	2.71
pHOME	449.5	40.46					
IhyPBase	414.3	82.86	123.32	1;2	136(PP)	0	0
Ac5-STABLE-Neo	196.7	98.35	98.35	n/a	173(LL) 197(PP)	15 0	8.67 0
R1	299.1	52.93					
R2	335.6	52.93					
R3	370.6	52.93	317.58		579(LL)	6	1.04
V1	324.7	52.93		1;1;1;1;1;1			
V2	405.3	52.93					
V3	231.2	52.93					
Gr2-1	335.5	52.78					
Gr2-2	342.3	52.78					
Ir8a-1	278	52.78	316.68		576(LL)	10	1.74
Ir8a-2	345.6	52.78		1;1;1;1;1;1			
Orco 12	265.2	52.78					
Orco 34	359.5	52.78					
Ebony gRNA 1	314	69.62					
Ebony gRNA 2	164	69.62	278.48		1034(PP)	8	0.77
Ebony gRNA 3	473	69.62		1;1;1;1			
Ebony gRNA 4	337	69.62					
Caspar gRNA 1	242	68.14					
Caspar gRNA2	87	10.24	214.66		922(PP)	3	0.33
Caspar gRNA 3	200	68.14		1;0.15;1;1			
Caspar gRNA 4	262	68.14					
UbiqCas9.874w	80.7	62.10	186.29	1;2	533(LL)	6	1.13
IhypBase (2)	538.7	124.19					
LLst Ubiq DsRed	176.7	67.23					
5`LLst pDCC6	243.3	67.23	201.69	1;1;1	701(LL)	3	0.43
3`LLst pDCC6	195.9	67.23					

Appendix 8. Densitometric Analysis of G1 *L. longipalpis* samples A-Z to estimate modification rate of the Rudimentary gene at the R3-pDCC6 cut site. Asterisk represents sample amplicons digested with T7 Endonuclease I in a heteroduplex assay. Undigested amplicon controls have no asterisk.

Band No.	Band Label	Adj. Volume (Int)	Band %	Lane %	X	Est% mod
1	R3C*	10822955	94.82	86.84	0.05	2.62
2	R3C*	590720	5.18	4.74		
1	R3C	26890110	98.30	93.36	0.02	0.85
2	R3C	465530	1.70	1.62		
1	R3D	22751690	99.70	95.16	0.00	0.15
2	R3D	69225	0.30	0.29		
1	R3E	16810560	95.82	91.60	0.04	2.11
2	R3E	732810	4.18	3.99		
1	R3H	18339100	97.20	92.59	0.03	1.41
2	R3H	527995	2.80	2.67		
1	R3M*	4572282	88.20	68.43	0.12	6.08
2	R3M*	611490	11.80	9.15		
1	R3M	21069510	94.66	69.96	0.05	2.71
2	R3M	1188462	5.34	3.95		
1	R3N*	7259538	94.59	75.95	0.05	2.74
2	R3N*	414876	5.41	4.34		
1	R3P	26813160	95.19	85.95	0.05	2.44
2	R3P	1355706	4.81	4.35		
1	R3Q	22427658	98.17	80.85	0.02	0.92
2	R3Q	419034	1.83	1.51		
1	R3R	26262786	98.17	86.12	0.02	0.92
2	R3R	489456	1.83	1.60		
1	R3S	27139926	97.20	76.92	0.03	1.41
2	R3S	781242	2.80	2.21		
1	R3U*	10377576	90.67	79.51	0.09	4.78
2	R3U*	1068210	9.33	8.18		
1	R3U	33294690	92.67	85.99	0.07	3.73
2	R3U	2632344	7.33	6.80		
1	R3V*	7020750	88.11	79.49	0.12	6.13
2	R3V*	947430	11.89	10.73		
1	R3V	15980712	91.34	80.69	0.09	4.43
2	R3V	1515690	8.66	7.65		
1	R3X	18002688	95.16	78.67	0.05	2.45
2	R3X	915156	4.84	4.00		

1	R3Y	19045422	97.41	81.49	0.03	1.31
2	R3Y	507012	2.59	2.17		
1	R3Z*	17486574	99.27	69.89	0.01	0.36
2	R3Z*	127806	0.73	0.51		
1	R3Z	56987706	96.33	89.87	0.04	1.85
2	R3Z	2171172	3.67	3.42		
1	PJET	43389066	98.68	79.79	0.01	0.66
2	PJET	579666	1.32	1.07		
1	PJET delta	59016588	99.83	84.13	0.00	0.09
2	PJET delta	101490	0.17	0.14		
1	PJET/delta	47051682	89.47	82.73	0.05	2.40
2	PJET/delta	2498490	4.75	4.39		
3	PJET/delta	3039090	5.78	5.34		

Appendix 9. Densitometric Analysis of G1 *L. longipalpis* samples A-Z to estimate modification rate of the Vestigial gene at the V1-pDCC6 cut site. Asterisk represents sample amplicons digested with T7 Endonuclease I in a heteroduplex assay. Undigested amplicon controls have no asterisk.

Lane	Band No.	Band Label	Adj. Volume (Int)	Band %	Lane %	X	Est% mod
1	1	V1M*	2562522	94.13	64.83	0.06	2.98
1	2	V1M*	159804	5.87	4.04		
7	1	V1P*	2239671	68.17	51.33	0.32	17.44
7	2	V1P*	1045764	31.83	23.97		
11	1	V1R*	2694933	66.41	55.25	0.34	18.51
11	2	V1R*	1362888	33.59	27.94		
15	1	V1T*	3684600	92.19	78.72	0.07	3.67
15	2	V1T*	287799	7.20	6.15		
15	3	V1T*	24495	0.613	0.52		
17	1	V1U*	4962687	78.16	66.49	0.22	11.59
17	2	V1U*	1386555	21.84	18.58		
3	1	V1Y*	6637593	81.10	63.89	0.19	9.94
3	2	V1Y*	1546635	18.90	14.89		

Appendix 10. Densitometric Analysis of G1 *L. longipalpis* samples A-Z to estimate modification rate of the Vestigial gene at the V2-pDCC6 cut site. Asterisk represents sample amplicons digested with T7 Endonuclease I in a heteroduplex assay. Undigested amplicon controls have no asterisk.

Lane	Band No.	Band Label	Adj. Volume (Int)	Band %	Lane %	X	Est% mod
14	1	V2D*	4338720	94.57	69.73	0.05	2.75
14	2	V2D*	249159	5.43	4.00		
15	1	V2D	50443692	99.91	88.32	0.00	0.05
15	2	V2D	47196	0.09	0.08		
17	1	V2E	28223001	97.14	85.16	0.03	1.44
17	2	V2E	829725	2.86	2.50		
1	1	V2G*	7745640	73.13	62.65	0.27	14.48
1	2	V2G*	2846130	26.87	23.02		
5	1	V2I*	5382440	78.14	60.30	0.13	6.91
5	2	V2I*	918540	13.33	10.29		
5	3	V2I*	587440	8.53	6.58		
7	1	V2K*	5104960	97.53	62.09	0.01	0.66
7	2	V2K*	69370	1.33	0.84		
7	3	V2K*	59990	1.15	0.73		
10	1	V2M	41572930	98.74	90.49	0.01	0.63
10	2	V2M	531650	1.26	1.16		
12	1	V2N	37341080	99.48	88.97	0.01	0.26
12	2	V2N	193550	0.52	0.46		
13	1	V2O*	2540510	92.54	74.03	0.07	3.80
13	2	V2O*	204680	7.46	5.96		
17	1	V2Q*	3224550	66.94	57.72	0.16	8.16
17	2	V2Q*	753690	15.65	13.49		
17	3	V2Q*	838530	17.41	15.01		
1	1	V2Q*	3823844	83.70	62.79	0.16	8.51
1	2	V2Q*	744736	16.30	12.23		

3	1	V2R*	4124880	92.33	77.23	0.01	0.64
3	2	V2R*	57188	1.28	1.07		
3	3	V2R*	285600	6.39	5.35		
9	1	V2U*	9419768	79.37	68.31	0.21	10.91
9	2	V2U*	2447932	20.63	17.75		
17	1	V2Y*	6227644	72.15	61.23	0.28	15.06
17	2	V2Y*	2403324	27.85	23.63		

Appendix 11. Densitometric Analysis to estimate modification rate of the Gr2 gene of G0 *L. longipalpis* male samples 1-9 and female samples 11 and 12. Asterisk (*) represents sample amplicons digested with T7 Endonuclease I in a heteroduplex assay. Undigested amplicon controls have no asterisk.

Lane	Band No.	Band Label	Adj. Volume (Int)	Band %	Lane %	X	Est% mod
1	1	G1	5032764	80.81	66.79	0.19	10.11
1	2	G1	1195392	19.19	15.86		
2	1	G1*	10721370	85.53	70.96	0.14	7.52
2	2	G1*	1813218	14.47	12.00		
3	1	G2	7746354	97.10	82.58	0.03	1.46
3	2	G2	231198	2.90	2.46		
4	1	G2*	12938904	97.62	88.12	0.02	1.20
4	2	G2*	314952	2.38	2.14		
7	1	G4	5115330	91.60	73.26	0.08	4.29
7	2	G4	468930	8.40	6.72		
8	1	G4*	10327680	93.67	85.75	0.06	3.22
8	2	G4*	697884	6.33	5.79		
11	1	G6	4199778	91.53	71.75	0.08	4.33
11	2	G6	388740	8.47	6.64		
12	1	G6*	5786682	98.65	87.23	0.01	0.68
12	2	G6*	79464	1.35	1.20		
17	1	G9	10195416	97.78	81.22	0.02	1.12
17	2	G9	231990	2.22	1.85		
18	1	G9*	7972074	97.33	87.33	0.03	1.35
18	2	G9*	218922	2.67	2.40		
21	1	G12	7543206	96.68	76.66	0.03	1.67

21	2	G12	258786	3.32	2.63		
22	1	G12*	14905308	96.88	90.69	0.03	1.57
22	2	G12*	479358	3.12	2.92		

Appendix 12. Densitometric Analysis to estimate modification rate of the Ir8a gene of G0 *L. longipalpis* male samples 1-9 and female samples 11 and 12. Asterisk (*) represents sample amplicons digested with T7 Endonuclease I in a heteroduplex assay. Undigested amplicon controls have no asterisk.

Lane	Band No.	Band Label	Adj. Volume (Int)	Band %	Lane %	X	Est% mod
4	1	I2*	7475322	83.12	74.12	0.08	3.91
4	2	I2*	690345	7.68	6.85		
4	3	I2*	644460	7.17	6.39		
4	4	I2*	183402	2.04	1.82		
14	1	I7*	4543098	95.54	84.87	0.03	1.66
14	2	I7*	156492	3.29	2.92		
14	3	I7*	55545	1.17	1.04		
16	1	I8*	4586223	99.04	75.72	0.01	0.48
16	2	I8*	44574	0.96	0.74		
18	1	I9*	2054613	89.75	67.03	0.05	2.33
18	2	I9*	105639	4.61	3.45		
18	3	I9*	128961	5.63	4.21		

Appendix 13. Densitometric Analysis to estimate modification rate of the Gr2 gene of G1 *L. longipalpis* male samples A-L. Asterisk (*) represents sample amplicons digested with T7 Endonuclease I in a heteroduplex assay. Undigested amplicon controls have no asterisk.

Lane	Band No.	Band Label	Adj. Volume (Int)	Band %	Lane %	X	Est% mod
7	1	GD	2446572	97.18	75.02	0.03	1.42
7	2	GD	70992	2.82	2.18		
9	1	GE	1921476	97.92	72.87	0.02	1.05
9	2	GE	40800	2.08	1.55		
10	1	GE*	2880208	95.34	78.80	0.05	2.36
10	2	GE*	140692	4.66	3.85		
11	1	GF	2867424	98.17	82.67	0.02	0.92
11	2	GF	53584	1.83	1.55		
13	1	GG	1070184	95.45	53.35	0.05	2.30
13	2	GG	51000	4.55	2.54		
17	1	GI	1869932	94.39	68.94	0.06	2.85
17	2	GI	111180	5.61	4.10		
18	1	GI*	4247960	97.48	80.88	0.03	1.27
18	2	GI*	109684	2.52	2.09		
19	1	GJ	2783444	97.30	75.65	0.03	1.36
19	2	GJ	77316	2.70	2.10		
20	1	GJ*	3378376	99.62	82.96	0.00	0.19
20	2	GJ*	12988	0.38	0.32		
21	1	GK	1701836	89.29	67.58	0.11	5.51
21	2	GK	204136	10.71	8.11		
23	1	GL	1773236	97.23	63.44	0.03	1.39
23	2	GL	50456	2.77	1.81		

Appendix 14. Densitometric Analysis to estimate modification rate of the Gr2 gene of G1 *L. longipalpis* male samples MA-MI and female samples FA-FL. Asterisk (*) represents sample amplicons digested with T7 Endonuclease I in a heteroduplex assay. Undigested amplicon controls have no asterisk.

Lane	Band No.	Band Label	Adj. Volume (Int)	Band %	Lane %	X	Est% mod
2	1	MA*	12759890	66.76	64.07	0.10	5.13
2	2	MA*	1911390	10.00	9.60		
2	3	MA*	1005680	5.26	5.05		
2	4	MA*	2004925	10.49	10.07		
2	5	MA*	1431560	7.49	7.19		
4	1	MB*	8887125	76.56	70.22	0.08	3.92
4	2	MB*	892320	7.69	7.05		
4	3	MB*	295555	2.55	2.34		
4	4	MB*	703755	6.06	5.56		
4	5	MB*	828555	7.14	6.55		
5	1	MC	8886930	94.58	82.37	0.05	2.75
5	2	MC	508820	5.42	4.72		
6	1	MC*	11316955	84.31	79.59	0.14	7.34
6	2	MC*	1898520	14.14	13.35		
6	3	MC*	84695	0.63	0.60		
6	4	MC*	63700	0.47	0.45		
6	5	MC*	58565	0.44	0.41		
8	1	MD*	17339010	92.99	89.16	0.05	2.74
8	2	MD*	1007630	5.40	5.18		
8	3	MD*	82875	0.44	0.43		
8	4	MD*	104910	0.56	0.54		

8	5	MD*	110760	0.59	0.57		
10	1	ME*	15209025	82.06	75.11	0.82	57.64
10	2	ME*	1081470	5.83	5.34		
10	3	ME*	425100	2.29	2.10		
10	4	ME*	1120275	6.04	5.53		
10	5	ME*	698685	3.77	3.45		
12	1	MF*	16514940	85.73	80.09	0.07	3.61
12	2	MF*	1365000	7.09	6.62		
12	3	MF*	412100	2.14	2.00		
12	4	MF*	886795	4.60	4.30		
12	5	MF*	84500	0.44	0.41		
14	1	MG*	15331550	95.58	89.30	0.03	1.42
14	2	MG*	453115	2.82	2.64		
14	3	MG*	224120	1.40	1.31		
14	4	MG*	18200	0.11	0.11		
14	5	MG*	13000	0.08	0.08		
16	1	MH*	18871515	92.96	88.51	0.05	2.65
16	2	MH*	1060800	5.23	4.98		
16	3	MH*	270400	1.33	1.27		
16	4	MH*	35165	0.17	0.16		
16	5	MH*	63375	0.31	0.30		
18	1	MI*	14922050	79.62	74.77	0.08	4.32
18	2	MI*	1584635	8.45	7.94		
18	3	MI*	482430	2.57	2.42		
18	4	MI*	927485	4.95	4.65		

18	5	MI*	825695	4.41	4.14		
2	1	FA*	5410116	91.94298	75.44991	0.05	2.71
2	2	FA*	315168	5.356167	4.395358		
2	3	FA*	158924	2.700856	2.216367		
4	1	FC*	5402880	90.31145	69.32361	0.04	2.13
4	2	FC*	579617	9.688546	7.436986		

Appendix 15. Densitometric Analysis to estimate modification rate of the Ir8a gene of G1 *L. longipalpis* male samples MA-MI and female samples FA-FL. Asterisk (*) represents sample amplicons digested with T7 Endonuclease I in a heteroduplex assay. Undigested amplicon controls have no asterisk.

Lane	Band No.	Band Label	Adj. Volume (Int)	Band %	Lane %	X	Est% mod
1	1	MA*	14951586	93.69	85.23	0.03	1.31
1	2	MA*	416003	2.61	2.37		
1	3	MA*	524744	3.29	2.99		
1	4	MA*	66263	0.42	0.38		
3	1	MB*	24293530	99.56	85.74	0.00	0.22
3	2	MB*	106597	0.44	0.38		
5	1	MC*	10013351	91.12	84.33	0.04	2.05
5	2	MC*	445282	4.05	3.75		
5	3	MC*	320729	2.92	2.70		
5	4	MC*	209911	1.91	1.77		
7	1	MD*	15803223	95.90	88.68	0.01	0.59
7	2	MD*	194434	1.18	1.09		

7	3	MD*	366758	2.23	2.06		
7	4	MD*	113833	0.69	0.64		
9	1	ME*	24430277	99.28	85.39	0.01	0.36
9	2	ME*	176612	0.72	0.62		
11	1	MF*	16994818	97.86	90.09	0.02	1.08
11	2	MF*	371783	2.14	1.97		
13	1	MG*	18108291	97.44	92.04	0.02	0.93
13	2	MG*	343643	1.85	1.75		
13	3	MG*	22311	0.12	0.11		
13	4	MG*	73432	0.40	0.37		
13	5	MG*	36783	0.20	0.19		
15	1	MH*	16294467	97.84	90.36	0.01	0.48
15	2	MH*	160599	0.96	0.89		
15	3	MH*	199861	1.20	1.11		
17	1	MI*	17828231	98.67	85.28	0.01	0.59
17	2	MI*	212792	1.18	1.02		
17	3	MI*	27738	0.15	0.13		
1	1	FA*	10276296	86.40	73.23	0.06	2.84
1	2	FA*	665788	5.60	4.74		
1	3	FA*	951524	8.00	6.78		
3	1	FC*	8681356	96.12	84.11	0.04	1.96
3	2	FC*	350880	3.88	3.40		
5	1	FD*	8325648	99.70	82.08	0.00	0.15
5	2	FD*	25364	0.30	0.25		
7	1	FE*	2361708	96.40	67.86	0.04	1.82

7	2	FE*	88128	3.60	2.53		
9	1	FF*	16967020	99.65	86.00	0.00	0.18
9	2	FF*	60316	0.35	0.31		
15	1	FI*	3349884	98.30	67.90	0.01	0.52
15	2	FI*	35088	1.03	0.71		
15	3	FI*	22848	0.67	0.46		
17	1	FJ*	10756308	82.49	76.70	0.07	3.32
17	2	FJ*	852312	6.54	6.08		
17	3	FJ*	1396992	10.71	9.96		
17	4	FJ*	33388	0.26	0.24		
19	1	FK*	6138224	99.53	81.38	0.00	0.23
19	2	FK*	28696	0.47	0.38		

Appendix 16. Densitometric Analysis to estimate modification rate of the Orco gene of G1 *L. longipalpis* male and female samples MA-ML and FA-FL. Asterisk (*) represents sample amplicons digested with T7 Endonuclease I in a heteroduplex assay. Undigested amplicon controls have no asterisk.

Lane	Band No.	Band Label	Adj. Volume (Int)	Band %	Lane %	X	Est% mod
2	1	MA*	24269075	94.44	81.45	0.01	0.44
2	2	MA*	223914	0.87	0.75		
2	3	MA*	990997	3.86	3.33		
2	4	MA*	212926	0.83	0.71		
4	1	MB*	23504002	95.10	83.22	0.03	1.74
4	2	MB*	853245	3.45	3.02		
4	3	MB*	358852	1.45	1.27		
6	1	MC*	31286119	97.19	88.02	0.03	1.39
6	2	MC*	890430	2.77	2.51		
6	3	MC*	15477	0.05	0.04		
8	1	MD*	19132989	94.79	87.05	0.03	1.75
8	2	MD*	700485	3.47	3.19		
8	3	MD*	351884	1.74	1.60		
10	1	ME*	25236019	95.02	84.95	0.02	0.81
10	2	ME*	429269	1.62	1.45		
10	3	ME*	572314	2.15	1.93		
10	4	ME*	321131	1.21	1.08		
11	1	MF	18217300	98.83	88.96		
11	2	MF	215003	1.17	1.05		
12	1	MF*	19745637	94.54	83.42	0.05	2.63
12	2	MF*	1083524	5.19	4.58		
12	3	MF*	56012	0.27	0.24		
14	1	MG*	15044247	86.78	79.48	0.10	5.21
14	2	MG*	1758549	10.14	9.29		
14	3	MG*	533052	3.07	2.82		

16	1	MH*	17676677	88.63	81.98	0.07	3.78
16	2	MH*	1480834	7.43	6.87		
16	3	MH*	786245	3.94	3.65		
18	1	MI*	15332950	81.22	73.77	0.17	9.00
18	2	MI*	3247021	17.20	15.62		
18	3	MI*	299356	1.59	1.44		
4	1	FC*	3039124	81.09	57.22	0.19	9.95
4	2	FC*	708764	18.91	13.34		
6	1	FD*	14132848	97.83	85.55	0.02	1.09
6	2	FD*	313140	2.17	1.90		
8	1	FE*	6410632	92.62	69.06	0.07	3.48
8	2	FE*	473892	6.85	5.10		
8	3	FE*	37264	0.54	0.40		
10	1	FF*	17542164	97.24	83.60	0.02	1.26
10	2	FF*	450704	2.50	2.15		
10	3	FF*	46580	0.26	0.22		
12	1	FG*	6675968	92.21	76.79	0.07	3.79
12	2	FG*	538220	7.43	6.19		
12	3	FG*	25432	0.35	0.29		
16	1	FI*	2507908	82.88	59.19	0.17	8.96
16	2	FI*	518092	17.12	12.23		
18	1	FJ*	10297988	85.77	65.96	0.14	7.02
18	2	FJ*	1626356	13.55	10.42		
18	3	FJ*	82620	0.69	0.53		
20	1	FK*	27843824	93.81	79.21	0.05	2.50
20	2	FK*	1467236	4.94	4.17		
20	3	FK*	370532	1.25	1.05		
21	1	FL	3809360	80.44	61.51	0.20	10.31

21	2	FL	926228	19.56	14.95		
22	1	FL*	5445440	75.33	58.57	0.25	13.21
22	2	FL*	1783368	24.67	19.18		

Appendix 17. ICE analysis of G0 survivors of Rudimentary and Vestigial wing phenotype knockouts via algorithmic deconvolution of Sanger sequence data to estimate editing outcomes. Gen = Generation, Alg.Succ =Algorithm Success. Dashes represent where sequence data was not present (gRNA column), or no data was generated by the ICE algorithm. ICE = the proportion of genotypes that contain an indel (indel %), KO-Score = measure of the variants that are likely result in functional protein knockout via a frame shift, R^2 = model fit.

Gen	Sample	Rudimentary 1					Rudimentary 2					Rudimentary 3				
		Guide sequence	Alg.Succ	ICE	KO-Score	R^2	Guide sequence	Alg.Succ.	ICE	KO-Score	R^2	Guide sequence	Alg.Succ.	ICE	KO-Score	R^2
G0	G0-M1_F	-	-	-	-	-	-	-	-	-	-	-	-			
	G0-M1_R	-	-	-	-	-	-	-	-	-	-	GGAACTATGGCATTCCGTG	No Indels	0	0	0.94
	G0-M2_F	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	G0-M2_R	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	G0-M3_F	-	-	-	-	-	TTGTAGCCCATCTTCACGA	Indels	23	19	0.54	-	-	-	-	-
	G0-M3_R	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	G0-F1_F	-	-	-	-	-	TTGTAGCCCATCTTCACGA	Failed	-	-	-	GGAACTATGGCATTCCGTG	No Indels	0	0	0.91
	G0-F1_R	-	-	-	-	-	-	-	-	-	-	GGAACTATGGCATTCCGTG	Indels	55	55	0.55
	G0-F2_F	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	G0-F2_R	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	G0-F3_F	-	-	-	-	-	-	-	-	-	-	GGAACTATGGCATTCCGTG	Failed	-	-	-
	G0-F3_R	-	-	-	-	-	-	-	-	-	-	GGAACTATGGCATTCCGTG	Indels	57	0	0.57

Gen	Sample	Vestigial 1					Vestigial 2					Vestigial 3				
		Guide sequence	Alg.Succ.	ICE	KO-Score	R^2	Guide sequence	Alg.Succ.	ICE	KO-Score	R^2	Guide sequence	Alg.Succ.	ICE	KO-Score	R^2
G0	G0-M1_F	-	-	-	-	-	-	-	-	-	-	TCGCGGACACGTATTGTGCT	Failed	-	-	-
	G0-M1_R	-	-	-	-	-	-	-	-	-	-	TCGCGGACACGTATTGTGCT	Failed	-	-	-
	G0-M2_F	-	-	-	-	-	-	-	-	-	-	TCGCGGACACGTATTGTGCT	Failed	-	-	-
	G0-M2_R	-	-	-	-	-	-	-	-	-	-	TCGCGGACACGTATTGTGCT	Failed	-	-	-
	G0-M3_F	-	-	-	-	-	-	-	-	-	-	TCGCGGACACGTATTGTGCT	Failed	-	-	-
	G0-M3_R	-	-	-	-	-	-	-	-	-	-	TCGCGGACACGTATTGTGCT	Failed	-	-	-
	G0-F1_F	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	G0-F1_R	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	G0-F2_F	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	G0-F2_R	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	G0-F3_F	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	G0-F3_R	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Appendix 18. ICE analysis of G1 survivors of Rudimentary phenotype knockouts via algorithmic deconvolution of sanger sequence data to estimate editing outcomes. Gen = Generation, Alg.Succ =Algorithm Success. Dashes represent where sequence data was not present (gRNA column), or no data was generated by the ICE algorithm. ICE = the proportion of genotypes that contain an indel (indel %), KO-Score = measure of the variants that are likely result in functional protein knockout via a frame shift, R2 = model fit.

Gen	Sample	Rudimentary 1					Rudimentary 2					Rudimentary 3				
		Guide sequence	Alg.Succ	ICE	KO-Score	R^2	Guide sequence	Alg.Succ	ICE	KO-Score	R^2	Guide sequence	Alg.Succ	ICE	KO-Score	R^2
G1	G1 M3_A	AGCATTGGAAGACACAGGGT	Failed	-	-	-	TTGTAGCCCATCTTCACGA	Indels	22	17	0.53	GGAACTATGGCATTCCGTG	Failed	-	-	-
	G1 F2_B	AGCATTGGAAGACACAGGGT	Failed	-	-	-	-	-	-	-	-	GGAACTATGGCATTCCGTG	No indels	0	0	0.87
	G1 F3_C	AGCATTGGAAGACACAGGGT	Failed	-	-	-	-	-	-	-	-	GGAACTATGGCATTCCGTG	No indels	0	0	0.85
	G1 M6_D	AGCATTGGAAGACACAGGGT	Failed	-	-	-	-	-	-	-	-	GGAACTATGGCATTCCGTG	No indels	0	0	0.85
	G1 M7_E	AGCATTGGAAGACACAGGGT	Failed	-	-	-	TTGTAGCCCATCTTCACGA	Failed	-	-	-	GGAACTATGGCATTCCGTG	No indels	0	0	0.85
	G1 M8_F	AGCATTGGAAGACACAGGGT	Failed	-	-	-	-	-	-	-	-	GGAACTATGGCATTCCGTG	Indels	54	54	0.54
	G1 M9_G	AGCATTGGAAGACACAGGGT	Failed	-	-	-	TTGTAGCCCATCTTCACGA	Failed	-	-	-	GGAACTATGGCATTCCGTG	Indels	55	55	0.55
	G1 M10_H	AGCATTGGAAGACACAGGGT	Failed	-	-	-	TTGTAGCCCATCTTCACGA	Indels	22	18	0.53	GGAACTATGGCATTCCGTG	Failed	-	-	-
	G1 F2_I	AGCATTGGAAGACACAGGGT	Failed	-	-	-	TTGTAGCCCATCTTCACGA	Indels	22	18	0.53	GGAACTATGGCATTCCGTG	Failed	-	-	-
	G1 F3_J	AGCATTGGAAGACACAGGGT	Failed	-	-	-	TTGTAGCCCATCTTCACGA	Failed	-	-	-	GGAACTATGGCATTCCGTG	Failed	-	-	-
	G1 F4_K	AGCATTGGAAGACACAGGGT	Failed	-	-	-	TTGTAGCCCATCTTCACGA	Indels	22	18	0.53	GGAACTATGGCATTCCGTG	Failed	-	-	-
	G1 M1_L	AGCATTGGAAGACACAGGGT	Failed	-	-	-	TTGTAGCCCATCTTCACGA	Indels	22	18	0.53	GGAACTATGGCATTCCGTG	Failed	-	-	-
	G1 M1_M	AGCATTGGAAGACACAGGGT	Failed	-	-	-	TTGTAGCCCATCTTCACGA	Indels	22	18	0.53	GGAACTATGGCATTCCGTG	Failed	-	-	-
	G1 M2_N	AGCATTGGAAGACACAGGGT	Failed	-	-	-	TTGTAGCCCATCTTCACGA	Indels	22	18	0.53	GGAACTATGGCATTCCGTG	Failed	-	-	-
	G1 M3_O	AGCATTGGAAGACACAGGGT	Failed	-	-	-	TTGTAGCCCATCTTCACGA	Indels	22	18	0.53	GGAACTATGGCATTCCGTG	Failed	-	-	-
	G1 M4_P	CATTGGAAGACACAGGGT	Failed	-	-	-	TTGTAGCCCATCTTCACGA	Indels	21	17	0.52	GGAACTATGGCATTCCGTG	Failed	No indels	0	0
	G1 M5_Q	CATTGGAAGACACAGGGT	Failed	-	-	-	TTGTAGCCCATCTTCACGA	Indels	21	17	0.52	GGAACTATGGCATTCCGTG	Failed	-	-	-
	G1 F1_R	AGCATTGGAAGACACAGGGT	Failed	-	-	-	TTGTAGCCCATCTTCACGA	Indels	22	18	0.53	GGAACTATGGCATTCCGTG	Failed	-	-	-
	G1 M1_S	CATTGGAAGACACAGGGT	Failed	-	-	-	TTGTAGCCCATCTTCACGA	Indels	22	18	0.53	GGAACTATGGCATTCCGTG	Failed	-	-	-
	G1 F1_T	AGCATTGGAAGACACAGGGT	Failed	-	-	-	-	-	-	-	-	GGAACTATGGCATTCCGTG	Failed	No indels	0	0
	G1 F5F_U	AGCATTGGAAGACACAGGGT	Failed	-	-	-	TTGTAGCCCATCTTCACGA	Indels	22	18	0.53	GGAACTATGGCATTCCGTG	Failed	-	-	-
	G1 M6M_V	AGCATTGGAAGACACAGGGT	Failed	-	-	-	TTGTAGCCCATCTTCACGA	Failed	-	-	-	GGAACTATGGCATTCCGTG	Failed	-	-	-
	G1 F6F_W	AGCATTGGAAGACACAGGGT	Failed	-	-	-	-	-	-	-	-	GGAACTATGGCATTCCGTG	Failed	-	-	-
	G1 M7M_X	AGCATTGGAAGACACAGGGT	Failed	-	-	-	TTGTAGCCCATCTTCACGA	Indels	22	18	0.53	GGAACTATGGCATTCCGTG	Failed	-	-	-
	G1 F7F_Y	AGCATTGGAAGACACAGGGT	Failed	-	-	-	TTGTAGCCCATCTTCACGA	Indels	22	18	0.53	GGAACTATGGCATTCCGTG	Failed	-	-	-
	G1 M2_Z	AGCATTGGAAGACACAGGGT	Failed	-	-	-	TTGTAGCCCATCTTCACGA	Indels	22	17	0.53	GGAACTATGGCATTCCGTG	Failed	-	-	-

Appendix 19. ICE analysis of G1 survivors of Vestigial phenotype knockouts via algorithmic deconvolution of sanger sequence data to estimate editing outcomes. Gen = Generation, Alg.Succ =Algorithm Success. Dashes represent where sequence data was not present (gRNA column), or no data was generated by the ICE algorithm. ICE = the proportion of genotypes that contain an indel (indel %), KO-Score = measure of the variants that are likely result in functional protein knockout via a frame shift, R2 = model fit.

Gen	Sample	Vestigial 1					Vestigial 2					Vestigial 3				
		Guide sequence	Alg.Succ	ICE	KO-Score	R^2	Guide sequence	Alg.Succ	ICE	KO-Score	R^2	Guide sequence	Alg.Succ	ICE	KO-Score	R^2
G1	G1 M3_A	AGCATTGGAAGACACAGGGT	Failed	-	-	-	GGAAAATTTCTCGCCGACAT	Failed	-	-	-	-	-	-	-	-
	G1 F2_B	AGCATTGGAAGACACAGGGT	Failed	-	-	-	GGAAAATTTCTCGCCGACAT	Failed	-	-	-	-	-	-	-	-
	G1 F3_C	AGCATTGGAAGACACAGGGT	Failed	-	-	-	GGAAAATTTCTCGCCGACAT	Failed	-	-	-	-	-	-	-	-
	G1 M6_D	AGCATTGGAAGACACAGGGT	Failed	-	-	-	GGAAAATTTCTCGCCGACAT	Failed	-	-	-	-	-	-	-	-
	G1 M7_E	AGCATTGGAAGACACAGGGT	Failed	-	-	-	GGAAAATTTCTCGCCGACAT	Failed	-	-	-	-	-	-	-	-
	G1 M8_F	AGCATTGGAAGACACAGGGT	Failed	-	-	-			-	-	-	-	-	-	-	-
	G1 M9_G	AGCATTGGAAGACACAGGGT	Failed	-	-	-	GGAAAATTTCTCGCCGACAT	Failed	-	-	-	-	-	-	-	-
	G1 M10_H	AGCATTGGAAGACACAGGGT	Failed	-	-	-	GGAAAATTTCTCGCCGACAT	Failed	-	-	-	-	-	-	-	-
	G1 F2_I	AGCATTGGAAGACACAGGGT	Failed	-	-	-	GGAAAATTTCTCGCCGACAT	Failed	-	-	-	-	-	-	-	-
	G1 F3_J	AGCATTGGAAGACACAGGGT	Failed	-	-	-	GGAAAATTTCTCGCCGACAT	Failed	-	-	-	-	-	-	-	-
	G1 F4_K	AGCATTGGAAGACACAGGGT	Failed	-	-	-	GGAAAATTTCTCGCCGACAT	Failed	-	-	-	-	-	-	-	-
	G1 M1_L	AGCATTGGAAGACACAGGGT	Failed	-	-	-	GGAAAATTTCTCGCCGACAT	Failed	-	-	-	-	-	-	-	-
	G1 M1_M	AGCATTGGAAGACACAGGGT	Failed	-	-	-	GGAAAATTTCTCGCCGACAT	Failed	-	-	-	-	-	-	-	-
	G1 M2_N	AGCATTGGAAGACACAGGGT	Failed	-	-	-	GGAAAATTTCTCGCCGACAT	Failed	-	-	-	-	-	-	-	-
	G1 M3_O	AGCATTGGAAGACACAGGGT	Failed	-	-	-	GGAAAATTTCTCGCCGACAT	Failed	-	-	-	-	-	-	-	-
	G1 M4_P	AGCATTGGAAGACACAGGGT	Failed	-	-	-	GGAAAATTTCTCGCCGACAT	Failed	-	-	-	-	-	-	-	-
	G1 M5_Q	AGCATTGGAAGACACAGGGT	Failed	-	-	-	GGAAAATTTCTCGCCGACAT	Failed	-	-	-	-	-	-	-	-
	G1 F1_R	AGCATTGGAAGACACAGGGT	Failed	-	-	-	GGAAAATTTCTCGCCGACAT	Failed	-	-	-	-	-	-	-	-
	G1 M1_S	AGCATTGGAAGACACAGGGT	Failed	-	-	-	GGAAAATTTCTCGCCGACAT	Failed	-	-	-	-	-	-	-	-
	G1 F1_T	AGCATTGGAAGACACAGGGT	Failed	-	-	-	GGAAAATTTCTCGCCGACAT	Failed	-	-	-	-	-	-	-	-
	G1 F5F_U	AGCATTGGAAGACACAGGGT	Failed	-	-	-	GGAAAATTTCTCGCCGACAT	Failed	-	-	-	-	-	-	-	-
	G1 M6M_V	AGCATTGGAAGACACAGGGT	Failed	-	-	-	GGAAAATTTCTCGCCGACAT	Failed	-	-	-	-	-	-	-	-
	G1 F6F_W	AGCATTGGAAGACACAGGGT	Failed	-	-	-	GGAAAATTTCTCGCCGACAT	Failed	-	-	-	-	-	-	-	-
	G1 M7M_X	AGCATTGGAAGACACAGGGT	Failed	-	-	-	GGAAAATTTCTCGCCGACAT	Failed	-	-	-	-	-	-	-	-
	G1 F7F_Y	-	-	-	-	-	GGAAAATTTCTCGCCGACAT	Failed	-	-	-	-	-	-	-	-
	G1 M2_Z	-	-	-	-	-	GGAAAATTTCTCGCCGACAT	Failed	-	-	-	TCGCGGACACGTATTGTGCT	Failed	-	-	-

Appendix 20. ICE analysis of G0 survivors of Olfactory knockouts via algorithmic deconvolution of sanger sequence data to estimate editing outcomes. Gen = Generation, Alg.Succ =Algorithm Success. Dashes represent where sequence data was not present (gRNA column), or no data was generated by the ICE algorithm. ICE = the proportion of genotypes that contain an indel (indel %), KO-Score = measure of the variants that are likely result in functional protein knockout via a frame shift, R^2 = model fit.

Gen	Sample	GR2-1 / GR2-2					IR8A					ORCO-1 / ORCO-2				
		Guide Sequence	Alg. Succ	ICE	KO-Score	R^2	Guide Sequence	Alg. Succ	ICE	KO-Score	R^2	Guide Sequence	Alg. Succ	ICE	KO-Score	R^2
G0	G0m1	AAACATAAAGAGGCAATACGGCTGTGTACAAGACAATGTG	No Indels Indels	0 16	0 16	0.98 0.77	TCGCAGGTCTCGCTATTGC	Failed	-	-	-	TGTGAGATACATGACCAACATACAGCAATCAAGTATTGGG	No Indels No Indels	0 0	0 0	0.97 0.97
	G0m1_r	-	-	-	-	-	AGCAATAGCGAGACCTGCGA	No Indels	0	0	0.98	TGTGAGATACATGACCAACATACAGCAATCAAGTATTGGG	No Indels No Indels	0 0	0 0	0.94 0.94
	G0m2_f	AAACATAAAGAGGCAATACGGCTGTGTACAAGACAATGTG	Indels No Indels	60 0	60 0	0.76 1	TCGCAGGTCTCGCTATTGC	Failed	-	-	-	TGTGAGATACATGACCAACATACAGCAATCAAGTATTGGG	No Indels No Indels	0 0	0 0	0.97 0.97
	G0m2_r	AAACATAAAGAGGCAATACGGCTGTGTACAAGACAATGTG	Indels Indels	59 57	0 0	0.59 0.57	AGCAATAGCGAGACCTGCGA	No Indels	0	0	1	TGTGAGATACATGACCAACATACAGCAATCAAGTATTGGG	No Indels No Indels	0 0	0 0	0.94 0.94
	G0m3_f	-	-	-	-	-	TCGCAGGTCTCGCTATTGC	Failed	-	-	-	TGTGAGATACATGACCAACATACAGCAATCAAGTATTGGG	No Indels No Indels	0 0	0 0	0.99 1
	G0m3_r	AAACATAAAGAGGCAATACGGCTGTGTACAAGACAATGTG	No Indels Indels	0 1	0 1	0.98 0.98	AGCAATAGCGAGACCTGCGA	No Indels	0	0	0.98	TGTGAGATACATGACCAACATACAGCAATCAAGTATTGGG	No Indels No Indels	0 0	0 0	0.98 0.94
	G0m4_f	-	-	-	-	-	TCGCAGGTCTCGCTATTGC	Failed	-	-	-	TGTGAGATACATGACCAACATACAGCAATCAAGTATTGGG	No Indels No Indels	0 0	0 0	0.99 1
	G0m4_r	AAACATAAAGAGGCAATACGGCTGTGTACAAGACAATGTG	No Indels No Indels	0 0	0 0	0.98 0.98	AGCAATAGCGAGACCTGCGA	No Indels	0	0	0.98	TGTGAGATACATGACCAACATACAGCAATCAAGTATTGGG	No Indels No Indels	0 0	0 0	0.98 0.94
	G0m5_f	AAACATAAAGAGGCAATACGGCTGTGTACAAGACAATGTG	Indels No Indels	26 0	22 0	0.63 1	TCGCAGGTCTCGCTATTGC	Failed	-	-	-	TGTGAGATACATGACCAACATACAGCAATCAAGTATTGGG	No Indels No Indels	0 0	0 0	0.99 1
	G0m5_r	AAACATAAAGAGGCAATACGGCTGTGTACAAGACAATGTG	No Indels No Indels	0 0	0 0	0.98 0.98	AGCAATAGCGAGACCTGCGA	No Indels	0	0	0.98	TGTGAGATACATGACCAACATACAGCAATCAAGTATTGGG	No Indels No Indels	0 0	0 0	0.98 0.94
	G0m6_f	-	-	-	-	-	TCGCAGGTCTCGCTATTGC	Failed	-	-	-	TGTGAGATACATGACCAACATACAGCAATCAAGTATTGGG	No Indels No	0 0	0 0	0.99 1

													Indels			
	G0m6_r	AAACATAAAGAGGCAATACGGCTGTGTACAAGACAATGTG	No Indels Indels	02	02	0.980.98	AGCAATAGCGAGACCTGCGA	No Indels	0	0	0.98	TGTGAGATACATGACCAACATACAGCAATCAAGTATTGGG	No Indels No Indels	00	00	0.980.94
	G0m7_f	AAACATAAAGAGGCAATACGGCTGTGTACAAGACAATGTG	Indels No Indels	360	340	0.361	TCGCAGGTCTCGCTATTGC	Failed	-	-	-	TGTGAGATACATGACCAACATACAGCAATCAAGTATTGGG	Indels No Indels	390	390	0.841
	G0m7_r	AAACATAAAGAGGCAATACGGCTGTGTACAAGACAATGTG	No Indels Indels	02	02	0.980.98	AGCAATAGCGAGACCTGCGA	No Indels	0	0	1	TGTGAGATACATGACCAACATACAGCAATCAAGTATTGGG	Indels No Indels	140	140	0.870.94
	G0m8_f	-	-	-	-	-	TCGCAGGTCTCGCTATTGC	Failed	-	-	-	TGTGAGATACATGACCAACATACAGCAATCAAGTATTGGG	No Indels No Indels	00	00	0.991
	G0m8_r	AAACATAAAGAGGCAATACGGCTGTGTACAAGACAATGTG	No Indels Indels	01	00	0.980.97	AGCAATAGCGAGACCTGCGA	Indels	33	33	0.76	TGTGAGATACATGACCAACATACAGCAATCAAGTATTGGG	No Indels No Indels	00	00	0.970.94
	G0m9_f	AAACATAAAGAGGCAATACGGCTGTGTACAAGACAATGTG	Indels No Indels	270	260	0.521	TCGCAGGTCTCGCTATTGC	Failed	-	-	-	TGTGAGATACATGACCAACATACAGCAATCAAGTATTGGG	No Indels No Indels	00	00	0.970.97
	G0m9_r	AAACATAAAGAGGCAATACGGCTGTGTACAAGACAATGTG	No Indels Indels	03	03	0.980.98	AGCAATAGCGAGACCTGCGA	No Indels	0	0	1	TGTGAGATACATGACCAACATACAGCAATCAAGTATTGGG	No Indels No Indels	00	00	0.940.94
	G0F1	GCTGTGTACAAGACAATGTG	Failed	-	-	-	AGCAATAGCGAGACCTGCGA	No Indels	0	0	0.98	TGTGAGATACATGACCAACA	Failed	-	-	-
	G0F2	GCTGTGTACAAGACAATGTG	Failed	-	-	-	AGCAATAGCGAGACCTGCGA	No Indels	0	0	1	TGTGAGATACATGACCAACA	Failed	-	-	-

Appendix 21. CE analysis of G1 survivors of Olfactory knockouts via algorithmic deconvolution of sanger sequence data to estimate editing outcomes. Gen = Generation, Alg.Succ =Algorithm Success. Dashes represent where sequence data was not present (gRNA column), or no data was generated by the ICE algorithm. ICE = the proportion of genotypes that contain an indel (indel %), KO-Score = measure of the variants that are likely result in functional protein knockout via a frame shift, R^2 = model fit.

Gen	Sample	GR2-1 / GR2-2					IR8A					ORCO-1 / ORCO-2				
		Guide Sequence	Algorithm success	ICE	KO-Score	R^2	Guide Sequence	Algorithm success	ICE	KO-Score	R^2	Guide Sequence	Algorithm success	ICE	KO-Score	R^2
G1	G1 m1_f	CGTATTGCCTCTTTATGTTT	Failed	-	-	-	AGCAATAGCGAGACCTGCGA	No Indels	0	0	0.98	TGTGAGATACATGACCAACA	Failed	-	-	-
	G1 m1_r	-	-	-	-	-	-	-	-	-	-	TGTGAGATACATGACCAACA	Failed	-	-	-
	G1 m2_f	CGTATTGCCTCTTTATGTTT	Failed				AGCAATAGCGAGACCTGCGA	No Indels	0	0	0.98	TGTGAGATACATGACCAACA	Indels	1	1	0.98
	G1 m2_r	-	-	-	-	-	-	-	-	-	-	TGTGAGATACATGACCAACA	Indels	59	59	0.59
	G1 m3_f	CGTATTGCCTCTTTATGTTT	Failed				AGCAATAGCGAGACCTGCGA	No Indels	0	0	0.98	TGTGAGATACATGACCAACA	Indels	1	1	0.98
	G1 m3_r	-	-	-	-	-	-	-	-	-	-	TGTGAGATACATGACCAACA	No Indels	0	0	0.98
	G1 m4_f	CGTATTGCCTCTTTATGTTT	Failed				AGCAATAGCGAGACCTGCGA	No Indels	0	0	0.98	TGTGAGATACATGACCAACA	Yes	1	1	0.98
	G1 m4_r	-	-	-	-	-	-	-	-	-	-	TGTGAGATACATGACCAACA	No Indels	0	0	0.97
	G1 m5_f	CGTATTGCCTCTTTATGTTT	Failed				AGCAATAGCGAGACCTGCGA	No Indels	0	0	0.98	TGTGAGATACATGACCAACA	Indels	1	1	0.98
	G1 m5_r	-	-	-	-	-	-	-	-	-	-	TGTGAGATACATGACCAACA	No Indels	0	0	0.98
	G1 m6_f	CGTATTGCCTCTTTATGTTT	Failed				AGCAATAGCGAGACCTGCGA	No Indels	0	0	0.98	TGTGAGATACATGACCAACA	Failed	-	-	-
	G1 m6_r	-	-	-	-	-	-	-	-	-	-	TGTGAGATACATGACCAACA	Failed	-	-	-
	G1 m7_f	CGTATTGCCTCTTTATGTTT	Failed				AGCAATAGCGAGACCTGCGA	No Indels	0	0	0.98	TGTGAGATACATGACCAACA	Failed	-	-	-
	G1 m7_r	-	-	-	-	-	-	-	-	-	-	TGTGAGATACATGACCAACA	Failed	-	-	-
	G1 m8_f	CGTATTGCCTCTTTATGTTT	Failed				AGCAATAGCGAGACCTGCGA	No Indels	0	0	0.98	TGTGAGATACATGACCAACA	Failed	-	-	-
	G1 m8_r	-	-	-	-	-	-	-	-	-	-	TGTGAGATACATGACCAACA	No Indels	0	0	0.97
	G1 m9_f	CGTATTGCCTCTTTATGTTT	Failed				AGCAATAGCGAGACCTGCGA	No Indels	0	0	0.98	TGTGAGATACATGACCAACA	No Indels	0	0	0.98
	G1 m9_r	-	-	-	-	-	-	-	-	-	-	TGTGAGATACATGACCAACA	Indels	62	62	0.62
	G1 m10_f	CGTATTGCCTCTTTATGTTT	Failed				AGCAATAGCGAGACCTGCGA	No Indels	0	0	0.98	TGTGAGATACATGACCAACA	No Indels	0	0	0.98
	G1 m10_r	-	-	-	-	-	-	-	-	-	-	TGTGAGATACATGACCAACA	Indels	62	62	0.62
	G1 m11_f	CGTATTGCCTCTTTATGTTT	Failed				AGCAATAGCGAGACCTGCGA	No Indels	0	0	0.98	TGTGAGATACATGACCAACA	Indels	1	1	0.98
	G1 m11_r	-	-	-	-	-	-	-	-	-	-	TGTGAGATACATGACCAACA	No Indels	0	0	0.97
	G1 m12_f	CGTATTGCCTCTTTATGTTT	Failed				AGCAATAGCGAGACCTGCGA	No Indels	0	0	0.98	TGTGAGATACATGACCAACA	Failed	-	-	-
	G1 m12_r	-	-	-	-	-	-	-	-	-	-	TGTGAGATACATGACCAACA	Failed	-	-	-
	G1 FA	AAACATAAAGAGGCAATACG	Indels	18	18	0.86	AGCAATAGCGAGACCTGCGA	No Indels	0	0	1	TGTGAGATACATGACCAACA	No Indels	0	0	0.99
	G1 FC	AAACATAAAGAGGCAATACG	Indels	4	2	0.87	AGCAATAGCGAGACCTGCGA	Indels	2	2	0.96	TGTGAGATACATGACCAACA	No Indels	0	0	0.99
	G1 FD	AAACATAAAGAGGCAATACG	Indels	6	3	0.85	AGCAATAGCGAGACCTGCGA	Failed				TGTGAGATACATGACCAACA	Failed			
	G1 FE	AAACATAAAGAGGCAATACG	Failed				AGCAATAGCGAGACCTGCGA	No Indels	0	0	1	TGTGAGATACATGACCAACA	No Indels	0	0	1
	G1 FF	AAACATAAAGAGGCAATACG	No Indels	0	0	0.97	AGCAATAGCGAGACCTGCGA	No Indels	0	0	0.98	TGTGAGATACATGACCAACA	Failed			
	G1 FG	AAACATAAAGAGGCAATACG	Indels	4	3	0.88	AGCAATAGCGAGACCTGCGA	No Indels	0	0	0.97	TGTGAGATACATGACCAACA	No Indels	0	0	1
	G1 FH	AAACATAAAGAGGCAATACG	Failed				AGCAATAGCGAGACCTGCGA	No Indels	0	0	1	TGTGAGATACATGACCAACA	Failed	-	-	-
	G1 FI	AAACATAAAGAGGCAATACG	Indels	7	6	0.85	AGCAATAGCGAGACCTGCGA	No Indels	0	0	1	TGTGAGATACATGACCAACA	Failed	-	-	-
	G1 FJ	AAACATAAAGAGGCAATACG	Indels	8	5	0.85	AGCAATAGCGAGACCTGCGA	No Indels	0	0	1	TGTGAGATACATGACCAACA	Failed	-	-	-
	G1 FK	AAACATAAAGAGGCAATACG	Indels	2	1	0.89	AGCAATAGCGAGACCTGCGA	No Indels	0	0	1	TGTGAGATACATGACCAACA	No Indels	0	0	1
	G1 FL	-	-	-	-	-	-	-	-	-	-	TGTGAGATACATGACCAACA	No Indels	0	0	0.99
	G1 MA	AAACATAAAGAGGCAATACG	Indels	4	2	0.88	AGCAATAGCGAGACCTGCGA	No Indels	0	0	1	TGTGAGATACATGACCAACA	Indels	1	1	0.99
	G1 MB	AAACATAAAGAGGCAATACG	Failed				AGCAATAGCGAGACCTGCGA	Indels	5	5	0.95	TGTGAGATACATGACCAACA	Failed	-	-	-
	G1 MC	AAACATAAAGAGGCAATACG	Indels	3	1	0.88	AGCAATAGCGAGACCTGCGA	No Indels	0	0	0.98	TGTGAGATACATGACCAACA	Failed	-	-	-
	G1 MD	AAACATAAAGAGGCAATACG	Indels	6	4	0.85	AGCAATAGCGAGACCTGCGA	No Indels	0	0	1	TGTGAGATACATGACCAACA	Failed	-	-	-
	G1 ME	AAACATAAAGAGGCAATACG	Indels	6	2	0.86	AGCAATAGCGAGACCTGCGA	Indels	16	16	0.91	TGTGAGATACATGACCAACA	Failed	-	-	-
	G1 MF	AAACATAAAGAGGCAATACG	Indels	6	2	0.86	AGCAATAGCGAGACCTGCGA	No Indels	0	0	0.99	TGTGAGATACATGACCAACA	No Indels	0	0	0.99
	G1 MG	AAACATAAAGAGGCAATACG	Indels	7	4	0.85	AGCAATAGCGAGACCTGCGA	No Indels	0	0	1	TGTGAGATACATGACCAACA	No Indels	0	0	1
	G1 MH	AAACATAAAGAGGCAATACG	Indels	7	4	0.85	AGCAATAGCGAGACCTGCGA	No Indels	0	0	0.99	TGTGAGATACATGACCAACA	Failed	-	-	-
	G1 MI	AAACATAAAGAGGCAATACG	Indels	5	2	0.85	AGCAATAGCGAGACCTGCGA	No Indels	0	0	1	TGTGAGATACATGACCAACA	Failed	-	-	-
	G1 9F	AAACATAAAGAGGCAATACG	Failed	-	-	-	AGCAATAGCGAGACCTGCGA	Failed	-	-	-	TGTGAGATACATGACCAACA	Failed	-	-	-

Chapter 6

Discussion and Future directions

Discussion

Leishmaniasis is a neglected tropical disease that is globally widespread, causing 1-2 million cases, and 70,000 deaths each year (Burza, Croft, & Boelaert, 2018). Treatments for the disease exist, however these often have poor efficacy and severe side effects, and vaccines have yet to be generated as an effective method to prevent disease (Alvar et al., 2013). Vector control plays an important role in the reduction of disease transmission, mostly relying on insecticides in the form of long lasting insecticide nets (LLINs) and indoor residual spraying (IRS). These have demonstrated a large effect on reducing the disease burden (Bhatt et al., 2015; Gunay et al., 2014), however vector resistance to insecticides is increasingly an issue of concern (Alexander et al., 2009; Dinesh et al., 2010). Recently genetic modification approaches have gained traction in the vector control sphere, with a range of population suppression and population replacement strategies having been designed (discussed at length in Chapter 2). However, research in the context of sand flies the vectors of leishmaniasis, are severely lacking.

This thesis describes the development of sand fly specific *in vitro* and *in vivo* research platform. We outline in more detail the bioinformatics approach to identification and rationalisation of genes that elicit a non-lethal phenotypic response, along with genes implicated in olfaction. Subsequently we apply a suite of molecular tools and analytics to perform, asses, and characterise CRISPR and PiggyBac mediated mutagenesis. We then pursue constructs developed for *in vivo* assessment at scale. We suggest that the approach described forms a robust scaffold for further CRISPR mediated study of gene editing in sand flies. Through application of the platform we also demonstrate robust evidence indicating gene editing via PiggyBac based approaches (*in vitro* and *in vivo*) and also by the use of CRISPR based mutagenesis *in vivo*. These observations are novel for the *Leishmania* vectors *Lutzomyia longipalpis* and *Phlebotomus papatasi*.

Currently, few proof-of-concept studies have attempted to modify the genome of sand flies, or to develop genetic control within these neglected vectors (Louradour, Ghosh, Inbar, & Sacks, 2019; Martin-Martin, Aryan, Meneses, Adelman, & Calvo, 2018). The research described in this thesis seeks to build on these pioneering studies to develop a foundation for novel sand fly control through gene drives.

Initially we have reviewed olfactory host detection in mosquito vectors of disease, and compared the findings to olfaction in sand fly vectors (see Chapter 2). Key odours produced by humans that elicit attractive responses in *Anopheles*, *Aedes* and *Culex*, were identified, with overlaps in *L. longipalpis* and *P. papatasi* being identified (CO_2 , L-(+) lactic acid, hexanoic acid, nonanal, 1-octen-3-ol, Ammonia).

Recent genome annotation in *L. longipalpis* and *P. papatasi* identified the presence of receptors and associated genes that generate the ligand-gated ion channels that bind odour molecules, leading to

responses to specific human odours. We demonstrate that genes previously identified for their involvement in mosquito host detection (Chapter 3, *Phylogenetic analysis of candidate Olfactory genes*) are analogous to those in sand flies. In particular, orthologues of the *Orco* gene involved in nonanal and 1-octen-3-ol detection, *Ir8a* involved in L-(+) lactic acid detection, and two of the three GRs (*Gr1* and *Gr2*) linked to the detection of CO_2 were identified (Hickner et al., 2020). Our phylogenetic analysis of peptide sequences for these genes revealed a high degree of conservation between sand flies and mosquitoes, in addition to *D. melanogaster* and the significantly more distantly related silk moth, *B. mori*. Encouragingly, gene knockout studies in a variety of insects had resulted in significantly modified olfactory behaviour (Koutroumpa et al., 2016; Y. Li et al., 2016; Q. Liu et al., 2017), with host-seeking affected in mosquito vectors species (DeGennaro et al., 2013; H. Liu et al., 2016; Sun, Liu, Ye, Baker, & Zwiebel, 2020; Wang et al., 2022). These phylogenetic similarities and responses post gene-knockout, suggested that targeting these genes in sand flies can elicit similarly altered host-seeking behaviour. However, the *Orco*, *Ir8a* and *Gr2* have not been explored in sand flies to verify if they have the same function to those in other vectors. This provides an opportunity to explore these key genes and determine their impact on behaviour. As suggested by olfactory knockout studies in vectors, there is a huge potential to manipulate vector-host contact, and ultimately disease transmission, by modifying olfactory behaviour.

A bioinformatics pipeline was outlined in this thesis to go from gene target identification to the design and generation of CRISPR-based modification tools for sand flies. This pipeline was initially ustilised to identify genes that could be used as non-lethal markers of transgenesis, to aid the development of tools for modification within sand flies. Cuticular pigmentation (*Yellow* and *Ebony*,) and eye pigmentation genes (*Scarlet* and *Cinnabar*), along with wing-development genes (*Rudimentary* and *Vestigial*) were identified as potential targets using bioinformatics approaches. Phylogenetic analysis (based on the genomes available) was used to determine conservation/relatedness of protein orthologues (in other species) with known function, to assess the likelihood of conserved function (Chapter 3, *Phylogenetic analysis of Phenotypic marker genes*). This conserved function suggests that a range of putative phenotypic marker genes in sand flies are appropriate targets for proof-of-concept knockouts using CRISPR methods, with similar phenotypic effects anticipated. The pipeline was also used to identify key olfactory gene targets (*Orco*, *Gr2*, and *Ir8a*) in *L. longipalpis* and *P. papatasi* that have been demonstrated to lead to reduced or altered behavioural responses in vector species, towards human hosts. CRISPR tools (including *in vitro* transcribed sgRNAs and CRISPR constructs containing gRNAs and Cas9 sequences) were developed for all of the putative phenotypic and olfactory gene targets using design software (see Chapter 3). These tools were constructed successfully following standard approaches outlined in Chapter 4, and assessed in both *in vitro* and *in vivo* studies within cell lines and sand flies (Chapter 4 and Chapter 5 respectively).

In this research we outline the first use of novel sand fly cell lines (LLE40 and LLE45) using chemo-transfection reagents. A variety of plasmids were delivered to LLE40 and LLE45 cells, to determine whether a range of promoters function effectively, with a view to using these cell lines as an intermediate step for molecular tool validation, prior to *in vivo* research. The importance of developing an *in vitro* platform for validation of gene editing cannot be understated, with respect to sand flies. Few laboratories have access to sand fly colonies (Lawyer et al., 2017) to perform *in vivo* experiments; therefore, cell lines provide a practical and accessible alternative.

We demonstrated that chemo-transfection was possible within these novel cell lines using pMaxGFP, Ac5-STABLE1-Neo, PiggyBac UbiqCas9.874W, and PiggyBac pHome-T plasmids, containing *Actin5C*, *CMV*, *hsp70*, *3x3p*, and *Ubiquitin 63e* promoters, which worked within the sand fly cells. This was evidenced by fluorescence after transfections with all of the plasmids described, and is an essential component for developing CRISPR gene editing tool; identifying promoters that are able to express gene editing components or exogenous genes.

Transfection efficiencies achieved were low in the sand fly cell lines (qualitatively assessed by Microscopy), which was confirmed via flow cytometry (see Chapter 4). Using four different transfection reagents (Cellfectin, Effectene, Flyfectin, and FuGene HD), transfection efficiencies of no more than 1.13% in LLE45 cells were achieved. However, the transfection efficiencies using the same transfection reagents and plasmid (Ac5-STABLE1-Neo) in alternative insect cell lines (DS2, Sf21 and An4a3B) were similar, with the lowest transfection efficiency at 0.083%, and the highest of 2.55% (excluding one outlier (4.88%)). With respect to sand fly cell lines, Cellfectin and Effectene were identified as the most efficient for plasmid transfections via flow cytometric analysis (Odds Ratios of 3.08 and 3.04 for Cellfectin and Effectene, respectively).

The successful use of these transfections reagents in sand fly cell lines is a positive step towards the validation of tools for transgenesis, prior to use in an *in vivo* system, and will be of use to the sand fly research community. Additionally, the promoters confirmed to function within the LLE40 and LLE45 cells are applicable for use in CRISPR constructs for sand flies, as they have been proven in several mosquito modification studies. Identification of alternative sand fly specific promoters for the expression of Cas9, is an appropriate next step, to restrict targeted endonuclease activity to the germline, enhancing the likelihood of germline transmission of CRISPR induced modifications.

The attempt to generate a Cas9-expressing cell line for validating CRISPR components (Chapter 3) was ultimately unsuccessful. However, a modified plasmid (UbiqCas9-NeoR) was constructed to allow for antibiotic selection, and preliminary data was collected to establish the approximate concentrations of the G418 antibiotic for use against the sand fly cells. Antibiotic concentrations identified for the LLE40 and LLE45 cell lines (84.15µg/mL and 259.9µg/mL, respectively) were similar for those used in

selection of other insect cell lines (González et al., 2011). Further optimisation of both the transfection protocols and G418 selection assays would allow the selection of transgenic Cas9 cells. For example, further analysis of G418 concentrations would give a more accurate IC50 for selection. Additionally, selection of transfected cells could be expedited using fluorescence-activated cell sorting (FACS). The generation of Cas9-expressing sand fly cell lines will provide a valuable resource for rapid screening of sgRNAs, without the requirement for recombinant Cas9 protein.

The *in vitro* transcription of gRNAs was optimised, however time constraints prevented the validation of these within LLE40 and LLE45 cell lines, to confirm their ability to cleave DNA. Transfection reagents are available (CRISPRMax, Invitrogen, USA) which would allow for validation, however optimisation steps will be required to ensure these transfections are efficient. CRISPR plasmids were also constructed successfully, however these were not tested in LLE40 and LLE45 cell lines. The knockout plasmids were designed to target both phenotypic marker genes (*Vestigial*, *Rudimentary*, and *Ebony*), and olfactory genes (*Gr2*, *Ir8a*, and *Orco*) in sand flies, which would require transfection followed by DNA extraction and sequencing, to confirm whether cleavage occurs. Assessment of these constructs within the sand fly cell lines would elucidate which plasmids have the most efficient cleavage, allowing rationalised decisions on which plasmids to advance to *in vivo* studies (Chapter 5).

The *in vivo* research presented here provides the first evidence of PiggyBac insertion (to our knowledge), and presents evidence of CRISPR-based targeted mutagenesis in *L. longipalpis* and *P. papatasi* via microinjection of plasmid constructs. In total 10,749 eggs were injected giving an average adult survival (as a percentage of larval survivors) of 66.30%, delivering a total of 24 different plasmids (Chapter 5).

Importantly, Cas9 expression and target-specific gRNAs expression was achieved, and confirmed via heteroduplex cleavage assays and algorithmic analysis of Sanger sequence data (Chapter 5). CRISPR construct delivery targeting *Rudimentary* and *Vestigial* wing development genes resulted in modifications to both G0 and G1, with three flies identified with altered wing shape. This suggested that *Rudimentary* and *Vestigial* genes were successfully cleaved, and that the induced modifications were inherited. Further modifications were observed to the olfactory genes targeted (*Gr2*, *Orco*, *Ir8a*), again demonstrating modifications and inheritance in genes putatively linked to sand fly host seeking. These results demonstrate that key components required for CRISPR-based gene drives function within sand flies successfully targeting phenotypic marker genes, and olfactory genes that may be useful for the development of a population replacement gene drive. In addition, UbiqCas9.874W PiggyBac transfections provides a stepping-stone for the development of Cas9 expressing sand fly lines. This tool provides the foundations for the development of split-drive technologies (M. Li et al., 2020), where mutagenesis can be induced by mating the Cas9 line with individuals expressing gRNAs in the germline.

In addition, these *in vivo* results demonstrated the application and utility of different mutagenesis detection methods within sand flies. T7E1 heteroduplex assays, combined with densitometric analyses were able to rapidly detect the occurrence and estimated modification rate, although these methods unable to identify specific mutations. Further, algorithmic analysis (ICE) of Sanger sequence data was able to detect putative indels, however alternative approaches such as Amplicon sequencing would allow high throughput characterisation of mutations. Without, access to Amplicon sequencing, T7E1 assays, densitometric analysis, and Sanger sequence algorithmic analysis provide complementary and cost effective approaches to detect gene editing.

Successful modification of insects through microinjection of PiggyBac or CRISPR constructs is not universally achievable, regardless of previous microinjection experience. Attempts made at PiggyBac delivery have been unsuccessful in *Cx. quinquefasciatus*, as were attempts using *in vitro* transcribed mRNA, with no evidence of transgenic individuals (M. E. Anderson et al., 2019). Recent attempts at CRISPR construct delivery were also unsuccessful at generating transgenic *L. longipalpis* (Martin-Martin et al., 2018). Therefore, the successful modification of *L. longipalpis* and *P. papatasi* via microinjection achieved in this research is an advancement which creates the foundations for the development of new genetic vector control strategies.

Future Directions

The foundations for CRISPR modification of sand flies targeting key olfactory genes putatively involved in host seeking have been presented in this research. However, there is substantial scope for optimisation of LLE40 and LLE45 cell lines to fully establish an *in vitro* platform for CRISPR component validation. In addition, there is need for optimisation of *in vivo* sand fly modification and assessment, prior to the development of fully functioning gene drive systems. Future avenues of research are set out here.

The sand fly cell lines used in this thesis are currently un-characterised. It would be beneficial to determine more about the cells, such as whether they contain male and female chromosomes, and functional features such as sex determination molecular pathways. Identification of genes such as *doublesex* (*dsx*) involved in sex determination will be useful for the generation of gene drives in sand flies. Therefore, future work could attempt to determine the karyotype characteristics of the cells, by arresting dividing cells in metaphase to count chromosome numbers. Having characterised cell lines provides a resource that allows better genetic studies and CRISPR tool development in cell-based assays.

Ideally, assessment of gRNAs is performed within the cell lines by transfection alongside recombinant Cas9 Protein, however alternative approaches not requiring cells can be utilised. We tested gRNAs via DNA cleavage assay (see Chapter 4) to identify those that were able to cleave extracted DNA as

confirmation of functionality, however it is beneficial to confirm their function in living cells. Protocols and transfections reagents are available for sgRNA transfections such as Lipofectamine CRISPRMax (Invitrogen, USA), which can be used to rapidly assess sgRNAs within the LLE40 and LLE45 cells lines to rationalise those to be used *in vivo*. This would be an elegant use of the cell line platform. In addition, the development of a Cas9-expressing cell line could expedite this gRNA screening process. Preliminary data was collected during the research reported in this thesis, on the generation of plasmids containing Cas9 and a drug resistance marker (NeoR), along with the identification of antibiotic concentrations that could be used to select cells transfected with this new plasmid (UbiqCas9-Neo).

Another direction for future research is the optimisation of gRNA expression in plasmid constructs. To achieve this, a range of Pol III gRNA promoters could be inserted into the pDCC6 plasmid backbone as described in Chapter 4. This should include a comparison of sand fly specific U6 promoters and other Pol III promoters that have been identified (M. A. E. Anderson et al., 2020) to determine whether any of these are more efficient. It has been shown in three insect cell lines and identified a range of Polymerase III promoters which are used to promote gRNAs in CRISPR constructs function when they are not specific to the cell line being used for construct validation. Using U6 promoters, still leads to expression of gRNAs, however levels of transcription varies which may have an impact of the efficiency of the modifications attempted (M. A. E. Anderson et al., 2020). It has also been shown that the ability to drive modifications varies between different U6 promoters from the same species (Port, Chen, Lee, & Bullock, 2014). One version of *L. longipalpis* U6 promoters was used in this thesis, however there are at least two alternative U6 genes have been identified (not described in this work). Alternative Poll III promoters have also been identified that may be of use in taking sand fly research forward. The *7SK* promoter contains similar features to U6, such as a TATA box and a proximal sequence element (PSE) and has greater activity than U6 promoters (M. A. E. Anderson et al., 2020). With respect to the Cas9, the *hsp70* promoter was used in the research reported, however when it comes to generating gene drives, which require inheritance, there is a need to ensure that the Cas9 endonuclease is expressed in the germline. Several germline promoters have been identified as germline-specific, including *nanos*, *vasa* and *zpg*, and have demonstrated expression. *Vasa* functions as a promoter in both males and female germlines, and the *nanos* promoters has been found to be substantially more specific, with less leakage of expression into somatic cells (Hammond et al., 2016). *Zpg* has also been found to promote expression more restrictively to the germline (Simoni et al., 2020).

With respect to further research into sand fly host-seeking behaviours, olfactory knockout lines (sand flies) should be generated, and used to investigate responses to human specific odours. Methods for testing olfactory behaviours in insect vectors have been well-established. Initially single sensillum electrophysiology could be used to identify responses of modified sand flies to key human odours (McMeniman, Corfas, Matthews, Ritchie, & Vosshall, 2014). Following this, uni-port olfactometer studies could be used to test taxis of mutant sand flies in response to constant odour streams (Raji et al.,

2019), and two-port olfactometer studies could be conducted to determine if responses to complete human odour profiles is affected (DeGennaro et al., 2013).

The generation of Cas9-expressing sand fly line would provide a useful tool for improved CRISPR-Cas9 modification. It has been demonstrated that the transgenic provision of Cas9 can result in greater transmission of modifications to subsequent generations (M. Li et al., 2017; Port et al., 2014). To achieve a stably-expressing Cas9 line in sand flies, constructs containing Cas9 driven by a germline specific promoter (*nanos*, *vasa*, *zpg*) could be integrated (Port et al., 2014). The assessment of Cas9 transmission to progeny would be required, and once established, the line could then be used to assess the function of gRNAs targeting phenotypic marker genes, or other genes of interest. In addition, HDR-based constructs could be used to knockin desired markers into the sand fly genome. Constructs could be designed to contain fluorescent marker gene for insertion into non-lethal genes. The generation of fluorescence-expressing sand fly lines could then be used to screen alternative knockins targeted to disrupt this fluorescence. Specifically the research of this thesis could be expanded beyond targeting olfaction, with HDR approaches being used to insert effector molecules that could interrupt Leishmania parasite transmission. Prior to thesis submission, HDR knockin constructs were developed in collaboration with colleagues (at London School of Hygiene and Tropical Medicine). Cell transfections with these plasmids resulted in successful expression of fluorescence markers, opening further research opportunities.

Modification of olfaction in mosquito vectors has elucidated the relationship between human odours, olfactory genes, and host-seeking behavioural responses (DeGennaro et al., 2013; Erdelyan, Mahood, Bader, & Whyard, 2012; F. Liu, Ye, Baker, Sun, & Zwiebel, 2020; H. Liu et al., 2016; McMeniman et al., 2014; Raji et al., 2019; Sun et al., 2020; Wang et al., 2022; Ye et al., 2022). These targeted gene knockouts have demonstrated reduced attraction to human hosts, suggesting that disruption of these genes could be incorporated into control strategies. Now that orthologues of key mosquito olfactory genes (*Orco*, *Gr2*, and *Ir8a*) have been identified in sand flies, the development of gene drives to spread olfactory modifications could form the basis of a population replacement strategy. In theory, the spread of modified olfactory genes would result in reduced interaction with humans, and reduced transmission of *Leishmania*. To achieve this aim and validate the strategy, multiple steps will be required: generation of homozygous knockout lines and behavioural experiments to determine responses to human odours (DeGennaro et al., 2013; McMeniman et al., 2014; Raji et al., 2019), evaluation of survival and mating fitness, and incorporation of a gene drive elements followed by cage studies (Hammond et al., 2016). Potential limitations to this strategy exist which require further consideration, such as the robustness of olfactory host-seeking ablation in the presence of alternative host-seeking cues (host body temperature and visual cues) (McMeniman et al., 2014), and the implications of this potential control strategy for other vertebrate hosts. Overall, this strategy is worth pursuing, initially to characterise the relationship

between sand fly olfactory genes and host-seeking, prior to consideration of gene drive control strategies.

Further annotation of the sand fly genomes alongside phylogenetic analysis are key steps moving forward to identify target genes that are potentially useful for control strategies beyond olfactory behaviour modification. Potential future targets include genes affecting female fertility such as *doublesex*, *fruitless* and *Nix* (Hall et al., 2015; Kyrou et al., 2018; Simoni et al., 2020). These genes are regulators of sexual differentiation in mosquitoes and other Diptera (Burtis & Baker, 1989), and it is yet to be determined if these genes are present in sand flies. Alternative targets to pursue could include anti-parasite effector genes which may be mobilised to block transmission of Leishmania parasite from sand flies to humans. Examples of genes linked to Leishmania parasite transmission could be immunity-related genes such as *Relish* (Louradour et al., 2019) and anti-microbial peptides such as *attacin* and *defensin* (Kykalová, Tichá, Volf, & Telleria, 2021).

In summary, gene editing in sand flies is limited and our approach helps address this omission. Taken together we describe a sand fly specific gene editing platform addressing identification of targets to affect, through both *in vitro* and *in vivo* assessment, at scale, to detect mutagenesis. Key outputs are novel, with evidence provided for *in vitro* and *in vivo* gene editing for the first time in *Lutzomyia longipalpis and Phlebotomus papatasi*. Our findings may be of wider interest to the sand fly research community, and increases the prospects for gene editing components and constructs to be developed and tested in the context of interrupting disease transmission and reducing an enormous burden of human suffering.

References

Alexander, B., Barros, V. C., De Souza, S. F., Barros, S. S., Teodoro, L. P., Soares, Z. R., … Reithinger, R. (2009). Susceptibility to chemical insecticides of two Brazilian populations of the visceral leishmaniasis vector Lutzomyia longipalpis (Diptera: Psychodidae). *Tropical Medicine and International Health*, *14*(10), 1272–1277. https://doi.org/10.1111/j.1365-3156.2009.02371.x

Alvar, J., Croft, S. L., Kaye, P., Khamesipour, A., Sundar, S., & Reed, S. G. (2013). Case study for a vaccine against leishmaniasis. *Vaccine*, *31*(SUPPL2), B244–B249. https://doi.org/10.1016/j.vaccine.2012.11.080

Anderson, M. A. E., Purcell, J., Verkuijl, S. A. N., Norman, V. C., Leftwich, P. T., Harvey-Samuel, T., & Alphey, L. S. (2020). Expanding the CRISPR Toolbox in Culicine Mosquitoes: In Vitro Validation of Pol III Promoters. *ACS Synthetic Biology*, *9*(3), 678–681. https://doi.org/10.1021/acssynbio.9b00436

Anderson, M. E., Mavica, J., Shackleford, L., Flis, I., Fochler, S., Basu, S., & Alphey, L. (2019). CRISPR/Cas9 gene editing in the West Nile Virus vector, Culex quinquefasciatus Say. *PLoS ONE*, *14*(11), 1–10. https://doi.org/10.1371/journal.pone.0224857

Bhatt, S., Weiss, D. J., Cameron, E., Bisanzio, D., Mappin, B., Dalrymple, U., … Gething, P. W. (2015). The effect of malaria control on Plasmodium falciparum in Africa between 2000 and 2015. *Nature*, *526*(7572), 207–211. https://doi.org/10.1038/nature15535

Burtis, K. C., & Baker, B. S. (1989). Drosophila doublesex gene controls somatic sexual differentiation by producing alternatively spliced mRNAs encoding related sex-specific polypeptides. *Cell*, *56*(6), 997–1010. https://doi.org/10.1016/0092-8674(89)90633-8

Burza, S., Croft, S. L., & Boelaert, M. (2018). Leishmaniasis. *The Lancet*, *392*(10151), 951–970. https://doi.org/10.1016/S0140-6736(18)31204-2

DeGennaro, M., McBride, C. S., Seeholzer, L., Vosshall, L. B., Jasinskiene, N., James, A. A., … Dennis, E. J. (2013). orco mutant mosquitoes lose strong preference for humans and are not repelled by volatile DEET. *Nature*, *498*(7455), 487–491. https://doi.org/10.1038/nature12206

Dinesh, D. S., Das, M. L., Picado, A., Roy, L., Rijal, S., Singh, S. P., … Coosemans, M. (2010). Insecticide susceptibility of Phlebotomus argentipes in visceral leishmaniasis endemic districts in India and Nepal. *PLoS Neglected Tropical Diseases*, *4*(10), 1–5. https://doi.org/10.1371/journal.pntd.0000859

Erdelyan, C. N. G., Mahood, T. H., Bader, T. S. Y., & Whyard, S. (2012). Functional validation of the carbon dioxide receptor genes in Aedes aegypti mosquitoes using RNA interference. *Insect Molecular Biology*, *21*(1), 119–127. https://doi.org/10.1111/j.1365-2583.2011.01120.x

González, M., Martín-Ruíz, I., Jiménez, S., Pirone, L., Barrio, R., & Sutherland, J. D. (2011). Generation of stable Drosophila cell lines using multicistronic vectors. *Scientific Reports*, *1*. https://doi.org/10.1038/srep00075

Gunay, F., Karakus, M., Oguz, G., Dogan, M., Karakaya, Y., Ergan, G., … Alten, B. (2014). Evaluation of the efficacy of Olyset® Plus in a village-based cohort study in the Cukurova Plain, Turkey, in an area of hyperendemic cutaneous leishmaniasis. *Journal of Vector Ecology*, *39*(2), 395–405. https://doi.org/10.1111/jvec.12115

Hall, A. B., Basu, S., Jiang, X., Qi, Y., Timoshevskiy, V. A., Biedler, J. K., … Tu, Z. (2015). A male-determining factor in the mosquito Aedes aegypti. *Science*, *348*(6240), 1268–1270.

https://doi.org/10.1126/science.aaa2850

Hammond, A., Galizi, R., Kyrou, K., Simoni, A., Siniscalchi, C., Katsanos, D., … Nolan, T. (2016). A CRISPR-Cas9 gene drive system targeting female reproduction in the malaria mosquito vector Anopheles gambiae. *Nature Biotechnology*, *34*(1), 78–83. https://doi.org/10.1038/nbt.3439

Hickner, P. V., Timoshevskaya, N., Nowling, R. J., Labbé, F., Nguyen, A. D., McDowell, M. A., … Syed, Z. (2020). Molecular signatures of sexual communication in the phlebotomine sand flies. *PLoS Neglected Tropical Diseases*, *14*(12), 1–18. https://doi.org/10.1371/journal.pntd.0008967

Koutroumpa, F. A., Monsempes, C., François, M. C., De Cian, A., Royer, C., Concordet, J. P., & Jacquin-Joly, E. (2016). Heritable genome editing with CRISPR/Cas9 induces anosmia in a crop pest moth. *Scientific Reports*, *6*(February), 1–9. https://doi.org/10.1038/srep29620

Kykalová, B., Tichá, L., Volf, P., & Telleria, E. L. (2021). Phlebotomus papatasi antimicrobial peptides in larvae and females and a gut-specific defensin upregulated by leishmania major infection. *Microorganisms*, *9*(11). https://doi.org/10.3390/microorganisms9112307

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

Lawyer, P., Volf, P., Rowton, E., Rowland, T., Killick-Kendrick, M., Rowland, T., … Volf, P. (2017). Laboratory colonization and mass rearing of phlebotomine sand flies (Diptera, Psychodidae). *Parasite*, *24*, 42. https://doi.org/10.1051/parasite/2017041

Li, M., Bui, M., Yang, T., Bowman, C. S., White, B. J., & Akbari, O. S. (2017). Germline Cas9 expression yields highly efficient genome engineering in a major worldwide disease vector, Aedes aegypti. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(49), E10540–E10549. https://doi.org/10.1073/pnas.1711538114

Li, M., Yang, T., Kandul, N. P., Bui, M., Gamez, S., Raban, R., … Akbari, O. S. (2020). Development of a confinable gene drive system in the human disease vector aedes aegypti. *ELife*, *9*, 1–22. https://doi.org/10.7554/eLife.51701

Li, Y., Zhang, J., Chen, D., Yang, P., Jiang, F., Wang, X., & Kang, L. (2016). CRISPR/Cas9 in locusts: Successful establishment of an olfactory deficiency line by targeting the mutagenesis of an odorant receptor co-receptor (Orco). *Insect Biochemistry and Molecular Biology*, *79*, 27–35. https://doi.org/10.1016/j.ibmb.2016.10.003

Liu, F., Ye, Z., Baker, A., Sun, H., & Zwiebel, L. J. (2020). Gene editing reveals obligate and modulatory components of the CO2 receptor complex in the malaria vector mosquito, Anopheles coluzzii. *Insect Biochemistry and Molecular Biology*, *127*(August), 1–9. https://doi.org/10.1016/j.ibmb.2020.103470

Liu, H., Liu, T., Xie, L., Wang, X., Deng, Y., Chen, C. H., … Chen, X. G. (2016). Functional analysis of Orco and odorant receptors in odor recognition in Aedes albopictus. *Parasites and Vectors*, *9*(1), 1–10. https://doi.org/10.1186/s13071-016-1644-9

Liu, Q., Liu, W., Zeng, B., Wang, G., Hao, D., & Huang, Y. (2017). Deletion of the Bombyx mori odorant receptor co-receptor (BmOrco) impairs olfactory sensitivity in silkworms. *Insect Biochemistry and Molecular Biology*. https://doi.org/10.1016/j.ibmb.2017.05.007

Louradour, I., Ghosh, K., Inbar, E., & Sacks, D. L. (2019). CRISPR/Cas9 Mutagenesis in Phlebotomus papatasi: the Immune Deficiency Pathway Impacts Vector Competence for Leishmania major. *MBio*, *10*(4), 1–14.

Martin-Martin, I., Aryan, A., Meneses, C., Adelman, Z. N., & Calvo, E. (2018). Optimization of sand fly embryo microinjection for gene editing by CRISPR/Cas9. *PLoS Neglected Tropical Diseases*, *12*(9), 1–18. https://doi.org/10.1371/journal.pntd.0006769

McMeniman, C. J., Corfas, R. A., Matthews, B. J., Ritchie, S. A., & Vosshall, L. B. (2014). Multimodal integration of carbon dioxide and other sensory cues drives mosquito attraction to humans. *Cell*, *156*(5), 1060–1071. https://doi.org/10.1016/j.cell.2013.12.044

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

Raji, J. I., Melo, N., Castillo, J. S., Gonzalez, S., Saldana, V., Stensmyr, M. C., & DeGennaro, M. (2019). Aedes aegypti Mosquitoes Detect Acidic Volatiles Found in Human Odor Using the IR8a Pathway. *Current Biology*, *29*(8), 1253-1262.e7. https://doi.org/10.1016/j.cub.2019.02.045

Simoni, A., Hammond, A. M., Beaghton, A. K., Galizi, R., Taxiarchi, C., Kyrou, K., … Crisanti, A. (2020). A male-biased sex-distorter gene drive for the human malaria vector Anopheles gambiae. *Nature Biotechnology*, *38*(9), 1054–1060. https://doi.org/10.1038/s41587-020-0508-1

Sun, H., Liu, F., Ye, Z., Baker, A., & Zwiebel, L. J. (2020). Mutagenesis of the orco odorant receptor co-receptor impairs olfactory function in the malaria vector Anopheles coluzzii. *Insect Biochemistry and Molecular Biology*, *127*(October). https://doi.org/10.1016/j.ibmb.2020.103497

Wang, Y., He, X., Qiao, L., Yu, Z., Chen, B., & He, Z. (2022). CRISPR/Cas9 mediates efficient site-specific mutagenesis of the odorant receptor co-receptor (Orco) in the malaria vector Anopheles sinensis. *Pest Management Science*, (April). https://doi.org/10.1002/ps.6954

Ye, Z., Liu, F., Sun, H., Ferguson, S. T., Baker, A., Ochieng, S. A., & Zwiebel, L. J. (2022). Discrete roles of Ir76b ionotropic coreceptor impact olfaction, blood feeding, and mating in the malaria vector mosquito Anopheles coluzzii. *Proceedings of the National Academy of Sciences of the United States of America*, *119*(23). https://doi.org/10.1073/pnas.2112385119