Expanding the toolbox for genetic modification in medfly

Thesis submitted for PhD by Alexandra Siddall

University of East Anglia

School of Biological Sciences

December 2023

University of East Anglia

Abstract

Ceratitis capitata, commonly known as the Mediterranean Fruit Fly or medfly, is an invasive Tephritid pest with a wide geographic spread and potential to expand further due to climate change. With the ability to damage over 300 species of fruit and vegetable crops, this agricultural pest is of major economic importance. Control of medfly has previously relied upon pesticides and Sterile Insect Technique (SIT). With the advent of CRIPSR/Cas9 new and improved methods of genetic pest control are possible. To develop new technologies such as gene drives (drivers of a transgene through inheritance bias) and improve existing ones, the toolbox for genetic modification in medfly must be expanded. Here I develop and validate several endogenous, germline specific Cas9 promotors, one of which was subsequently used for the development of the first homing drive (drives which utilise homology-directed repair) in medfly. The sex determination pathway in medfly provides several potential targets for sex conversion, including the conserved genes *transformer* (*tra*) and *doublesex* (*dsx*) and the masculinising signal of *Maleness-on-the-Y* (*MoY*). Guide cassettes were designed to target both *tra* and *dsx,* with two *tra* guide constructs being synthesised and used to generate transgenic lines. Crossing these with Cas9 lines gave no evidence of cutting, and subsequent investigations revealed an error in the guide designs, prompting a redesign of the guides for future use. The use of *MoY* to induce masculinisation is another promising method of sex conversion in medfly, though the lethal effects of overexpression of *MoY* need to be overcome. A construct was designed to package *MoY* into *gfp*-tagged sperm cells by placing it under the control of the spermatogenesis specific promotor of *ß2-tubulin*. This aimed to create a transgenic line that could induce masculinisation through *MoY*-carrying sperm and reduce the temporal expression of *MoY* to a smaller window to avoid lethality. The results revealed post-microinjection embryonic death, indicating overexpression of *MoY* was causing lethality. A further *Cre-Lox* construct was designed for future usage which should allow for the transgenic line to be established without lethality during microinjection. The final part of the genetic toolbox focused on essential genes. Potential essential genes were identified and sequenced across two wildtype, laboratory-maintained populations to check for sequence conservation. Based on these results, guides targeting these genes were designed, alongside rescue versions of these genes the guides should be unable to target. In this research, several important toolbox parts were designed, synthesised and tested in medfly which can contribute to the future development of genetic pest control strategies.

Acknowledgements

Prior to beginning this PhD the research experiences I had which defined my field of interest allowed me to find my interest within this field. I would therefore like to thank all of those who supervised and mentored me on these placements. My supervisory team has been invaluable during my PhD and without whom I doubt I would have completed this work. Dr Philip Leftwich and Prof Tracey Chapman have provided exceptional support and guidance throughout this work for which I have been extremely grateful. I would also like to thank my external supervisor Dr Tim Harvey-Samuel for his contributions to this project. I could not have asked for a more supportive supervisory team and would like to thank them for all they have done. An additional thankyou to Dr Angela Meccariello for her expertise in microinjection and all things medfly who provided invaluable help to this work.

I would also like to thank the other members of the Chapman Lab, past and present, who have always been available for help and have made my PhD enjoyable. To this end I also thank the rest of floor 01 who have not only been helpful colleagues but supportive friends. I am grateful for the members of my cohort who have become close friends over the course of our PhDs. The COVID pandemic made this PhD more arduous and without the support of these people I am unsure if I would have finished this work. The BBSRC NRP DTP provided the funding for this work and I would like to thank them for this contribution.

Finally (but in no way least), I would like to thank my family for their support during this undertaking and their acceptance in me moving to the other side of the country and never making me feel guilty for limiting the time we could spend together. This includes my granny and grandad who I have lost during this time. My partner Peter has encouraged me, come along on countless weekend and holiday lab visits and provided much needed emotional support.

Contents

<u>Attribution statements</u>

The work presented in this thesis is my own with the following acknowledgements:

<u>Chapter 1</u>

This chapter contains work completed as the first draft of the literature review published in (Siddall et al. 2022) which was conceptualised by myself under the supervision of Tracey Chapman, Philip Leftwich and Tim Harvey-Samuel. All additional content was authored by myself with feedback provided by Tracey Chapman and Philip Leftwich.

<u>Chapter 2</u>

I conceived and designed these experiments under the supervision of Philip Leftwich, Tracey Chapman and Tim Harvey-Samuel. Microinjection was completed by Angela Meccariello (Imperial College) who cultured the survivors until pupation, at which point they were sent to UEA. Philip Leftwich assisted with the screening of potential transgenics to establish transgenic lines. The fitness study in Chapter 2 was designed and completed by Joerel Gestopo, an MSci student, under my supervision. The fitness data processing written by Joerel Gestopo was used by myself for analysis in R statistical software. Philip Leftwich provided supervision of the statistical analysis of the fitness data. Emily Fowler assisted with experimental design in the qPCR experiment. The qPCR experiment was designed and worked on concurrently by both myself and Joerel Gestopo until the primer efficiency results were obtained, at which point I finished this experiment independently. All other experimental work and writing of this chapter was completed by me with feedback from Philip Leftwich and Tracey Chapman.

<u>Chapter 3</u>

Myself and Philip Leftwich conceptualised and contributed to the design of the Cas9 promotors. From the point of plasmid synthesis, integration and subsequent testing was completed by Angela Meccariello who kindly provided the qPCR data and wrote the manuscript. The qPCR analysis script was written by Philip Leftwich and modified by myself. The writing of this chapter was completed by me with feedback from Philip Leftwich and Tracey Chapman.

<u>Chapter 4</u>

I conceived and designed these experiments with Philip Leftwich and with support from Tracey Chapman, Tim Harvey-Samuel and Angela Meccariello. Philip Leftwich assisted in the screening of potential transgenics. Microinjection was completed by Angela Meccariello who

cultured the survivors until pupation at which point they were sent to UEA. All other experimental work and the writing of this chapter was completed by me with feedback from Philip Leftwich and Tracey Chapman.

Chapter 5

I conceived and designed these experiments with Philip Leftwich and with support from Tracey Chapman and Tim Harvey-Samuel. Microinjection was completed by Angela Meccariello who cultured the survivors until pupation at which point they were sent to UEA. Philip Leftwich assisted in the screening of potential transgenics to establish a transgenic line. A *tra* guide sequence was kindly provided by Angela Meccariello. All other experimental work and writing of this chapter was completed by me with feedback from Philip Leftwich and Tracey Chapman.

Chapter 6

I conceived and designed these experiments with Philip Leftwich and with support from Tracey Chapman and Tim Harvey-Samuel. All other experimental work and writing of this chapter was completed by me with feedback from Philip Leftwich and Tracey Chapman.

List of Abbreviations

bp – Base Pairs

ClvR – Cleave and Rescue

CRISPR – Clustered Regularly Interspaced Palindromic Repeats

dsx – doublesex

fsRIDL – Female-Specific Release of Insects with Dominant Lethals

gfp – Green Fluorescent Protein

GSS – Genetic Sexing Strain

HDR – Homology Directed Repair

MoY – Maleness-on-the-Y

NHEJ – Non-homologous End Joining

PBS – Phosphate Buffered Saline

PCR – Polymerase Chain Reaction

pgSIT – Precision Guided Sterile Insect Technique

qPCR Quantitative Polymerase Chain Reaction

RACE – Rapid Amplification of cDNA Ends

rfp – Red Fluorescent Protein

RIDL – Release of Insects with Dominant Lethals

RNAi – RNA interference

RO – Reverse Osmosis

RPM – Rotations Per Minute

RT-PCR – Reverse Transcription Polymerase Chain Reaction

RT-qPCR – Real Time Quantitative Polymerase Chain Reaction

SIT – Sterile Insect Technique

TADE – Toxin Antidote Dominant Embryo

TARE – Toxin Antidote Recessive Embryo

tra – transformer

UTR – Untranslated Region

wt - Wildtype

1 Chapter One - Introduction

1.1 *Ceratitis capitata* – a global agricultural pest

Ceratitis capitata (common name Mediterranean Fruit Fly or medfly) is a pest species of global agricultural importance. With a warming climate, the invasive potential of medfly is predicted to increase (Gutierrez et al. 2021; Sultana et al. 2020; De Meyer et al. 2007). The adaptability of medfly to new food sources may arise due to the plasticity of its oviposition behaviour (Malacrida et al. 2007). This makes it a particularly potent threat for invasion into new regions and crops, and thus further expanding its already wide geographical spread (Weldon 2022). Currently, over 350 plant species have been identified as medfly hosts (Liquido et al. 2017). Female flies oviposit eggs directly into ripening or ripe fruits, the crop is then damaged through both the feeding of the larvae and the introduction of plant pathogens (Ordax et al. 2015). Medfly infestation in an area can also make crops unsuitable for export, with trade regulations in place to prevent the export of infested crops to medfly free areas (Lance et al. 2014). Economic loss can be the result of direct damage and the controls on export from regions in which outbreaks have been detected with medfly costing the US alone $2 billion annually (Thomas et al. 2022).

Due to its prevalence as a pest, medfly has been well studied in several different contexts. For example, it has been used as a model to investigate global population spread, due to its documented invasion history and wealth of genetic information available (Diamantidis et al. 2008). Medfly display complex but extensively studied courtship behaviour and mating preferences, and its sex determination pathway is also fairly well characterised (Salvemini et al. 2009; Saccone et al. 2008). They have been much studied in the context of sterile insect technique (SIT) implementation, and genetic sexing strains (GSS) (Robinson 2002) have also been used successfully for population suppression (Hendrichs et al. 2002). The recent identification of the male determining element *Maleness-on-the-Y* (*MoY*) uncovered the upstream elements of the sex determination pathway, which opens up the opportunity for additional studies of Y chromosome evolution. The important status of medfly as a pest coupled with the wealth of life history knowledge and genomics tools available make this species an ideal candidate for the development of new genetic control methods. Existing and newer control strategies are described below.

1.2   Methods of genetic control for insect pests

1.2.1   Self-limiting genetic pest control strategies

1.2.1.1   Sterile Insect Technique

Exploiting our knowledge of sex determination in insects became a significantly more promising field for developing novel methods of pest control with the advent of CRISPR/Cas9. Prior to this, the most effective and commonly used method of sex-based pest control was the Sterile Insect Technique (Hendrichs et al. 1995; Lees et al. 2015; Hendrichs and Robinson 2009). The mass release of irradiated, sterilised males in SIT programmes results in a lower proportion of fertilised females due to mating with the sterilised males as opposed to the fertile, wild-type males. Males are mass-reared and sterilised through a small dose of ionising radiation at late-stage pupal development. The dose must be precise and well timed to ensure male sterility without effecting mating competitiveness (Chen et al. 2023). The number of insects that must be released for the technique to be effective may be ten or many tens of times larger than the wild population size itself (Burt and Deredec 2018). Parameters such as population density, mating frequency and male dispersal all effect neccersary release number and because of this modelling is often relied upon to predict release numbers (White et al. 2010; Oléron Evans and Bishop. 2014). Due to this, maintaining a constant and sufficient supply of sterile male insects is costly. In addition to the large numbers needed, control is more effective if only males are released as it drives males to mate with wildtype females as oppose to the sterile, released females. This prompts the development of sex sorting methods. Though successful in some notable cases, techniques such as SIT are not effective in all pest species, for example those that have longer lifespans, occur at low density and engage in multiple matings (Gordillo 2015). This is because females can continue mating until they produce viable offspring and are therefore more likely to encounter fertile wildtype males.

1.2.1.2   Genetic Sexing Strains

To address the issues with rearing large numbers of males for SIT, genetic sexing strains (GSS) were developed, which differentiated between males and females by using selective markers such as pupal colour, or conditional temperature-sensitive and female-specific lethality (Robinson 2002). Kandul et al. (2020) recently described the development of a transgenic *D. melanogaster* line generated with CRISPR/Cas9 which can produce entirely male or female

offspring depending upon the diet on which flies developed. The strategy utilised the alternative splicing mechanisms of the conserved sex determination genes *doublesex* (*dsx*) and *transformer* (*tra*) and inserted sex-specific introns into antibiotic resistance genes which then generated sex-specific antibiotic resistance as the functional antibiotic resistance gene would only be formed when the male specific splicing pattern was followed. This interesting strategy has yet to be tested in any pest species, but could be useful in the future in addition to the other sex sorting mechanisms. It also demonstrates the potential of Cas9 to expand the armoury of control methods available. GSS techniques allow sex sorting for sterile insect release to be done at large scale in a short time frame. However, many GSS strains carry fitness costs and can show evidence of instability due to chromosome translocations (Hendrichs and Robinson 2009; Robinson 2002).

1.2.1.3   RIDL

The release of insects carrying a dominant lethal (RIDL) is a transgenic-based alternative to SIT and was first developed in *Drosophila melanogaster* (Thomas et al. 2000). RIDL utilises a dominant gene lethal in larval development which is repressed in the lab population usually through a dietary supplement. Once the lab reared carriers of this are released, their offspring inherit this dominant lethal gene which causes fatality without the dietary supplement. A female-specific version of RIDL was developed in which a female lethal gene is controlled by conditional repression, usually in the form of tetracycline or doxycycline provided in the food during lab rearing, which is not available for consumption in the wild. To use this for population control, males carrying a female-specific lethal gene would be released to mate. The resulting offspring should be entirely male, each of which carries a copy of the female lethal gene. Upon maturation and mating, these male offspring then have 50% fewer female offspring and 50% of their male offspring would also carry this gene. The female killing gene would decrease with each subsequent generation allowing for a self-limiting method of control. Modelling showed the use of female-specific RIDL (fsRIDL) releases to be significantly more effective at population reduction than the release of sterilised males (Alphey et al. 2009) especially when optimised effectively. The fsRIDL method also removes the need for both the sex sorting and irradiation steps that are standard in SIT, thus reducing costs and complexity. RIDL has been developed in several pest species (Valdez et al. 2011; Labbé et al. 2012; Facchinelli et al. 2013), caged field trials have been successful at demonstrating population control in medfly (Leftwich et al. 2014) and commercial usage

has been developed in several species including *Aedes* and medfly with Oxitec Fiendly™ products (Oxitec 2024).

1.2.2   Threshold-dependent self-sustaining pest control strategies

1.2.2.1   Homing Drives

The idea to use homing endonuclease genes for genetically engineering populations was first proposed by Burt in 2003 (Burt 2003). Homing endonuclease genes encode an endonuclease which forms double-stranded breaks to copy itself to specific genomic locations, ensuring its propagation. The advent of CRISPR/Cas9 made this possible. Esvelt et al. (2014) and Gantz and Bier (2015) both proposed the development of endonuclease gene drives using CRISPR to trigger homology-directed repair (HDR). The advantages of this type of strategy are the ability to spread throughout a population even despite fitness costs to individuals and the requirement for only low introduction frequencies. Since then, several different drive types utilising CRISPR to provide drive have been proposed. By modifying individuals to contain both the modification and the means to express the CRISPR/Cas9 complex in early development, 100% inheritance of the gene can be ensured as the second unmodified

chromosome inherited from a non-transgenic parent will be modified to contain the transgene (*Figure 1.1*).
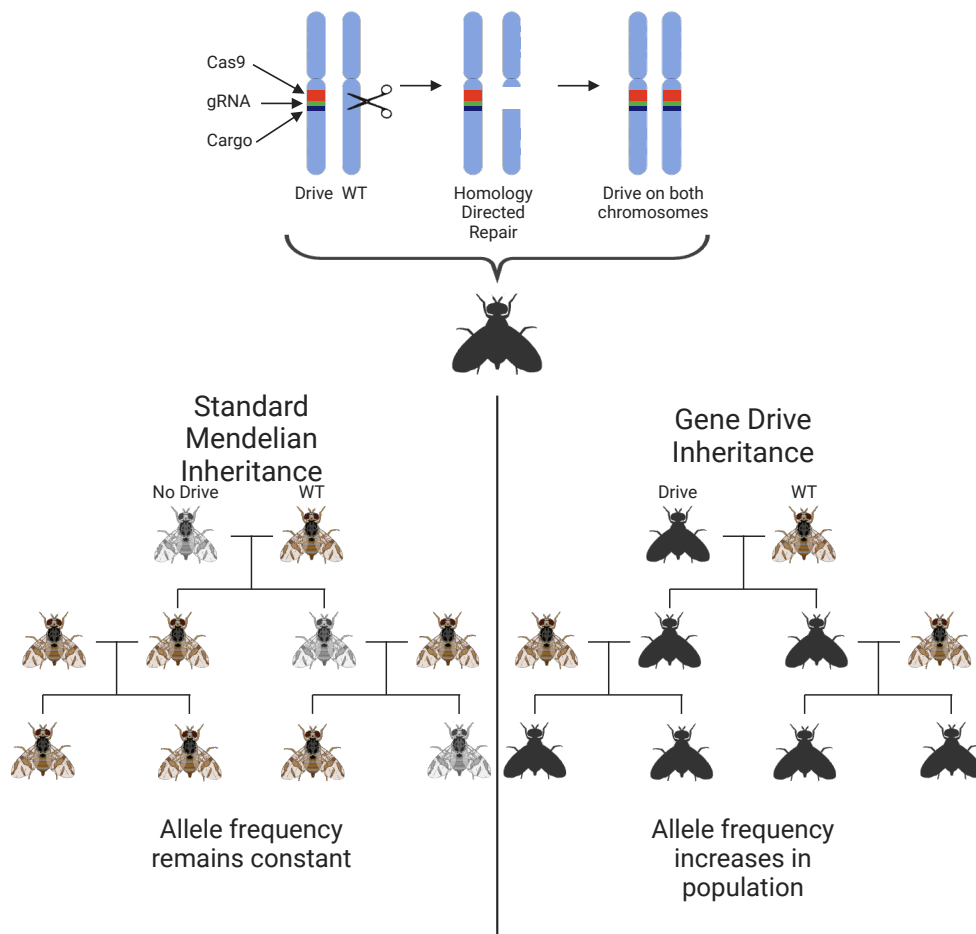
Figure 1.1 The genetic principle and biased inheritance of a homing gene drive. Homing drives ensure their own transmission to the next generation through homology-directed repair. The Cas9 element of the construct, coupled with specific guides cut the wild type chromosome at a precise genomic location triggering a double-strand break. This break is then repaired using homology-directed repair with the remaining chromosome being used as a template. The drive construct is therefore present on both chromosome copies ensuring 100% offspring inheritance. A cargo element is also coupled with the drive elements which can spread a desired trait through a population. This biases the usual 50% inheritance of standard Mendelian genetics in which allele frequency remains relatively constant and results in the 100% inheritance of the drive in offspring. The drive frequency in the population will therefore increase with subsequent populations. Generated with Biorender and published in Siddall et al (2022).

Homing drives (Esvelt et al. 2014) are those which utilise HDR to drive the construct throughout a population. By including additional guides linked to the driving machinery, genes responsible for female development can be targeted to reduce the number of viable female offspring (AGAP007280 targeted in Hammond et al. 2016; McFarlane et al. 2018). As the percentage of individuals in the population carrying the drive increases, the number of female offspring will decrease, skewing the population sex ratio. Given that females are responsible for the reproductive potential of a population, this gradual decline in females will eventually lead to a population crash – unless resistance alleles develop as females carrying a pre-existing resistance allele would have a fitness advantage over those without. Resistance allele formation is one of the major drawbacks of homing drives (Hammond et al. 2017) as a pre-existing or *de-novo* mutation at the target site can prevent guide recognition if the sequence is no longer complementary to the guide. Multiplexing (the use of multiple guides as oppose to a single guide) of the guide RNAs could help to overcome resistance allele formation (Oberhofer et al. 2018; Champer et al. 2018), though there are limitations to this as multiplexing cannot protect against NHEJ deletions. The optimal number of guides varies depending on drive type and the drive performance parameters such as homing efficiency (Champer et al. 2020b). This adds a further layer of complexity as additional guides can often decrease homing efficiency (Oberhofer et al. 2018). Controlling the expression timing of Cas9 by limiting it to early oogenesis improves the function of the drive and reduces resistance allele formation by minimising the levels of Cas9 deposition in the embryo (Hammond et al. 2018). The first CRISPR homing drive was developed as a proof of principle in *Drosophila melanogaster* (Gantz and Bier 2015) and has since been developed in several pests, including several species of mosquito (Hammond et al. 2016; Gantz et al. 2015; Carballar-Lejarazú et al. 2020; Kyrou et al. 2018; Verkuijl et al. 2022) and in *Drosophila suzukii* (Yadav et al. 2023).

1.2.2.2   Sex chromosome shredders

A strategy to using additional guides to disrupt a female fertility gene, is to target the X chromosome as a whole. Such 'X-shredder' strategies (Galizi et al. 2014; 2016) use an endonuclease to target the X chromosome at several points close to the centromere resulting in its complete degradation. Expression of this endonuclease during spermatogenesis ensures that fathers cannot produce female offspring, as they are only capable of providing a Y chromosome. Given that the Y chromosome will be inherited in all offspring, if the construct is located on the Y chromosome, it will also be present in all offspring resulting in a potentially global drive with minimal need for continuous releases. If the construct is placed

within an autosome, it would result in a much less invasive method of control as the shredder will not drive. This could be used as a test of X-shredder effectiveness prior to the release of a global drive, or as a localised, self-limiting control. Although this is a potentially highly effective sex conversion drive, it is only effective in XY heterogametic species. X shredders developed in *A. gambiae* (Galizi et al. 2016) were shown to have a strong sex bias towards male offspring (>83% male offspring) and are now being considered for release (Alcalay et al. 2019). X shredders located on an autosome are expected to need a continuous release to maintain high levels of suppression, these drives would be significantly more invasive if attached to the Y chromosome. However, the probability of Y localisation occurring randomly in the wild is minimal (Alcalay et al. 2019). An X shredder attached to the Y chromosome would create a highly invasive, self-perpetuating drive as recently discussed by Simoni (Simoni et al. 2020). Although this seems a perfect design, expression from the Y chromosome is usually very limited and often unpredictable due to meiotic sex chromosome inactivation, expression variation, unpredictability and the rapid evolution of Y chromosomes (Hall et al. 2016; Gamez et al. 2021). This could create a feasibility challenge when it comes to constructing a functional drive, relying on expression from the Y chromosome.

Despite the number of reproducing females being directly indicative of the reproductive potential and spread of a population, Y shredders have also been developed with the idea to skew the population towards females in order to cause population suppression due to the unavailability of males (Prowse et al. 2019). For use in pest insect control, this would most likely be significantly less effective than a female suppressing drive as males are able to have multiple matings and are less directly related to the reproductive potential of a population.

When trying to introduce this principle to a ZW heterogametic species, the idea of shredding the chromosome resulting in female offspring remains the same although it is slightly simpler as females are the heterogametic sex. Given that females carry the sex specific chromosome (W), a W shredder active during egg development would result in a decrease in female offspring. Depending on the timing of expression of Cas9 the number of viable offspring will be variable. Early expression during oogenesis will allow the non-viable female offspring to be replaced by viable male offspring, keeping the total number of offspring to its usual level. Later expression, although successful in preventing female offspring from developing, does not allow male offspring to develop in their place, resulting in 50% fewer offspring. A further option when employing W shredders is to have them express in somatic tissues causing female death, though this may not be as successful at population control (Holman 2019).

Depending upon the method of sex determination involved in insects, some drives will be more effective in certain species. A sex conversion principle that would be highly successful in an XX/XY system might be non-functional in a Z/W or X0 system as the sex-specific upstream regulation is not conserved (Siddall et al. 2022). The level of control needed over the drive is also an important factor to consider. Local drives (limited to single populations) are generally preferable, especially in the case of drives designed to decrease the population. This is because the release of a global drive would not only be difficult to justify ethically but could also be a legislative minefield as country borders will not prevent the spread of transgenic insects under such a system. The degree to which these concerns remain is also affected by the severity of the (usually human) disease that is being targeted (Committee on Gene Drive Research in Non-Human Organisms: Recommendations for Responsible Conduct et al. 2016.).

Homing drives, although promising do have several challenges to overcome. Homing rates can be both difficult to predict and are often much lower than needed for successful drive propagation. As the Cas9 is driving homing, its expression must be very tightly controlled and restricted to the germline to prevent mosaicism and repair through non-homologous end joining.

1.2.2.3   Toxin/Antidote Drives

To avoid issues with resistance allele formation and the specificity of timing needed with homing, CRISPR/Cas9 drives that do not utilise homing have been proposed. Toxin/antidote systems are well documented throughout biology and ensure their propagation by causing toxicity if they are not inherited. A well utilised and documented example of this is the *Medea* gene (Beeman and Friesen 1999). Toxin/antidote refers to the same construct causing both the toxicity through targeting an essential gene and the antidote by reversing this effect. Toxin/antidote drives are those which target a haplosufficient gene (a single copy of the gene is sufficient to restore function) for their toxin element, in addition to a recoded version of this gene, which is not effected by the toxin, to act as the rescue element. Using CRISPR to target the essential genes during early development will result in individuals which have no functional copies of the gene other than that provided as a rescue. The result is that offspring will only be viable if they inherit the drive construct therefore propagating both itself and the linked cargo throughout the population. TARE (Toxin-antidote recessive embryo) (Champer et al. 2020a) and ClvR (Cleave and Rescue) (Oberhofer et al. 2019) are two such drive systems. ClvR drives outlined such a system with the guides and Cas9 being located on a separate

section of the genome to the rescue construct (Oberhofer et al. 2019). The design of TARE drives places the entire construct in the same location on the genome however the two drives function in much the same way.

1.3 Developing CRISPR/Cas9 in *Ceratitis* for genetic control

The new strategies proposed for genetic methods of pest control principally rely upon modification using the CRISPR/Cas9 system both to improve current genetic sexing strains (Nguyen et al. 2021) and in the development of gene drives (McFarlane et al. 2018). The expression of Cas9 is an important part of these designs, and transgenes are often placed under the control of viral promotors to ensure high levels of expression. However, this is not suitable for Cas9 expression as high levels of ubiquitous expression can cause issues with mosaicism and resistance allele formation as the necessary homology directed repair is only favoured during certain cell cycle stages (Raban et al. 2020). Cas9 expression outside of the germline will induce mutations that will be repaired through non-homologous end joining, potentially forming resistance alleles. To avoid this, Cas9 expression must be tightly controlled.

There are several requirements for a Cas9 promotor which vary upon the intended usage. The most common requirement is for a germline specific promotor. Expression within the germline is important as it allows for mutations to be induced early in development. This reduces the chances of mosaicism as cell division and differentiation will occur after the sequence has been modified, resulting in all cells having the modified sequence (Kondo and Ueda 2013). The use of endogenous promotors is the most effective way of restricting expression. This has been identified as a necessity for gene drive designs in medfly (Carrami et al. 2018) to reduce chimerism and increase editing efficiency (Wang et al. 2015; Mao et al. 2016; Lei et al. 2021; Feng et al. 2021). Several germline-specific genes have been used as promotors (the most commonly used including *nanos*, *vasa* and *zpg*), though it is not yet known whether their activity will be consistent across species.

One of the most common problems with germline promotors is leaky expression. Transegnic constructs under the expression of germline specific promotors often also cause a small amount of somatic expression even in instances in which the original gene has no known expression outside the germline. Somatic expression at low levels is often at too low a level to cause high levels of mosiacism. However, it does drive the formation of resistance alleles in constructs which rely on homing. Homing is only favoured over non-homologous end

joining during certain stages of cell development such as in the germline (Raban et al. 2020). Somatic expression of Cas9, even at low levels, can induce mutations which are repaired through non-homologous end joining, which can then result in the formation of resistance alleles (Champer et al. 2018). In a similar vein, Cas9 deposition into embryos can also result in the formation of resistance alleles in homing drives. This however can be beneficial in designs which do not rely on homing as it can allow a dominant knockout effect beneficial for precision guided SIT (pgSIT) (Li et al. 2021) or the accumulation of broken essential gene alleles in toxin/antidote designs (Champer et al. 2020a; Oberhofer et al. 2019).

At the start of this project there were no Cas9 constructs driven by endogenous regulatory elements in medfly. The use of a native 5' UTR from a germline specific gene (active in both male germline and female germline is desirable with female germline expression being utilised in more drive designs) is a necessity for the development of gene drives in medfly (Carrami et al. 2018). However, whether a native 3' UTR is also required is less clear. Transgenes often deploy viral 3'UTRs as they allow for the upregulation of expression. However, this can lead to higher levels of Cas9 deposition as more protein is present, resulting in persistence. The 3' UTR has an essential role in localisation (Gavis et al. 1996) and its use in transgene expression has been shown to more closely mimic native gene expression (Ebron and Shukla 2016). This means its use could lead to a tighter level of restriction to the germline making it important in homing drive designs but less so in designs which do not rely on the homing pathway.

These differences in requirements across the genetic pest control methods outlined above highlights the need for the development and testing of multiple Cas9 promotors. *vasa* and *nanos* are two of the most commonly used germline-specific Cas9 promotors in the early developments in other species (Champer et al. 2017; 2018; Kyrou et al. 2018; Adolfi et al. 2020; Feng et al. 2021). However, a promotor which behaves perfectly has yet to be found despite extensive testing in *Drosophila* (Du et al. 2023) and promoters that avoid somatic expression often result in problems with the deposition of Cas9 into embryoes (Champer et al. 2018). Replication of promotor results across species is often low or highly variable (Du et al. 2023). Nevertheless, previous studies in other species can provide a starting point for identifying candidate genes to drive Cas9 expression in medfly. However, without testing these in medfly it is impossible to know which is most effective in this species.

 This informed the starting point of this research by identifying the most pressing need for the development of genetic pest control in medfly as being a functional Cas9 promotor. In

Chapter 2 and Chapter 3 I worked towards the development of several Cas9 promotors for strategies which both do and do not utilise homing.

1.4 Manipulating the sex determination pathway in *Ceratitis* for control strategies

The sex determination pathway provides some of the most promising targets for genetic pest control especially in medfly. At the start of my PhD, I completed a literature review discussing the depth of sex determination systems in pest species and how these could be utilised in pest control strategies (Siddall et al. 2022). Here I focus on the sex determination system in medfly as components of this pathway are focused on extensively in this thesis.

The sex determination system in medfly can be briefly summarised (*Figure 1.2*).

Figure 1.2 A summary of the sex determination pathway in medfly. The presence or absence of *transformer* (*tra*) results in the alternative splicing of *doublesex (dsx)* into the male or female specific isoforms. This subsequently drives male or female development. Generated with Biorender and published in Siddall et al. (2022).

The primary male determining element which triggers this pathway in medfly was recently identified as a Y carried gene *Maleness-on-the-Y (MoY)* (Meccariello et al. 2019). This acts by repressing the production of *transformer (tra)*, the absence of which results in male development. The action of *tra* in sex determination is variable across species. However, in

medfly it has been shown to upregulate its own expression (and the expression of *tra-2*) which subsequently act as splice enhancers for the female-specific isoform of *dsx* (Salvemini et al. 2009). The alternative splicing of *doublesex* (Kyriacou 1992) is a well-conserved mechanism of sex determination across the insect taxa (Geuverink and Beukeboom 2014; Price et al. 2015). In medfly the retention of exon 4 is specific to the female isoform and this is retained when the appropriate splicing enhancers and machinery is recruited to the weak splice site (Saccone et al. 2008). These three key elements of the sex determination pathway in medfly have all been identified as promising choices for targeting to induce sex conversion.

The injection of *MoY* and the repression through RNAi of *tra* has the ability to trigger male development in XX individuals to generate XX pseuodmales (Meccariello et al. 2019; Primo et al. 2020). An unexpected feature of these XX pseudomales was the discovery that they are fertile despite the lack of a Y chromosome. This indicates the Y chromosome does not carry genes necessary for male fertility. This makes the manipulation of the sex determination pathway in medfly promising for control, as the ability to generate fertile pseudomales should reduce fitness costs associated with carrying a drive. It also raises interesting questions about the role of the sex determination pathway in dosage compensation - as the induction of dosage compensation has often been associated with the sex determination pathway (Hopkins and Kopp 2021). This makes both *MoY* and *tra* usable targets for sex conversion in medfly for use in suppression strategies and genetic sexing strains. They do, however, have different mechanisms of action: *MoY* would have to be provided to induce male development whereas *tra* would have to be repressed. In Chapter 4, I investigate how *MoY* could be integrated as a stable insertion to allow for sex conversion. Chapter 5 focuses on the targeted knockout of *tra* to induce male development.

The targeting of *dsx* is more of an unknown. Masculinisation has been achieved by providing the male isoform of *dsx* (Saccone et al. 2008). However, full sex conversion did not occur. The knocking out of *dsx* would not result in conversion as there would be no male-specific *dsx* produced, resulting in death or sterility (Kyrou et al. 2018). Targeting the downstream elements of the sex determination pathway is less likely to induce full conversion as the elements upstream of this often a have role outside of the cascade. *Doublesex* does still have other potential uses besides full conversion, especially in the exploitation of sex-specific splice sites (Labbé et al. 2012). *Dsx* targeting in medfly is also explored in Chapter 5.

1.5   Thesis summary

In this thesis I investigated the tools needed in medfly for developing methods of genetic pest control. Several strategies of genetic control were considered, with the primary focus being on the creation of the tools necessary for the development of gene drive. However, several of the components investigated have applicability to other fields of research. At the start of this work there were no tested endogenous Cas9 promotors in medfly, which are necessary for the development of CRISPR/Cas9-utilising methods of genetic control. Several Cas9 promotors were constructed in my thesis work and all were shown to drive expression of Cas9. To use these promotors for population suppression, the most promising targets in medfly were found to be those in the sex determination pathway. *MoY*, *tra* and *dsx* were all explored as potential targets for sex conversion, as examples of genes at varying stages of the sex determination pathway. Further important components of genetic control such as the use of essential genes was also examined, to identify suitability for use as rescue components.

In Chapter 2, I identified several germline specific genes present in medfly which could be used as Cas9 promotors. Through RT-PCR, the expression pattern of these genes across multiple tissues (ovaries, testes, somatic and eggs) was obtained to determine how well expression was restricted to the germline. From these results, suitable candidates were identified and through a combination of RACE sequencing and genome annotations, the 5' UTR was identified and amplified from *Ceratitis* genomic DNA. Two genes, *bgcn* (*benign gonial cell neoplasm)* and *mei-W68* were successfully integrated into Cas9 promotor constructs and established as transgenic lines through the microinjection of *piggybac* flanked constructs. Real-time quantitative PCR was used to determine the expression levels of Cas9 in these transgenic lines across multiple tissue types (ovaries, testes and somatic). A small-scale fitness study was completed to determine if these transgenic lines displayed any fitness cost.

Chapter 3 encompasses the work I completed to contribute towards the development of the first homing drive in medfly. Two Cas9 constructs were designed which utilised the regulatory elements (including both the 5'UTR and 3' UTR) of *vasa* and *bgcn.* This required the modification of the *vasa* 5'UTR to remove intronic sequence, reducing the size to allow its inclusion in a transgenic construct. After attempts to assemble these through Gibson assembly were unsuccessful these parts were synthesised externally. The *vasa* construct was

subsequently used by A. Meccariello (Imperial College) in a homing drive targeting *white eye* to test homing rates of several Cas9 promotors. The *vasa* promotor was identified as the most promising and was subsequently used for a homing drive targeting the sex determination gene *tra*.

In Chapter 4, I examined the suitability of expressing a *MoY* carrying transgene to induce sex conversion of XX females to pseudomales. The integration of MoY under a ubiquitous promotor was shown to be lethal (A. Meccariello, personal communication). Therefore, to restrict the expression of *MoY*, I attempted to package *MoY* into sperm cells. *MoY* was placed under the promotor of the male germline specific gene *ß2-tubulin* (the expression pattern of which was explored in Chapter 2) with a *gfp* tag to allow for the visualisation of packaging into sperm cells. After microinjection with integration through randomised *piggybac* insertion, high levels of embryonic death were observed. Following this, no transgenics were recovered, indicating construct toxicity. Further designs utilising the *Cre-Lox* system are discussed which would further restrict the expression of MoY to establish a stable transgenic line.

Chapter 5 introduces the targeting of sex determination genes to generate XX pseudomales. Guides were designed to target both *dsx* and *tra,* to force male development. Guides targeting *dsx,* were designed to remove the splicing site as well as enhancers which allow for the splicing of the female isoform of *dsx*. *Tra* guides were designed to induce a knockout of *tra,* the absence of which should result in male development. Multiplexed guide cassettes containing 3 guides were designed and the *tra* cassette was synthesised externally. After integration into the *piggybac* backbone for microinjection, the cassette recombined resulting in only a single guide being retained. This construct was successfully integrated through microinjection and subsequently crossed to the Cas9 lines previously developed. The sex ratio of the offspring was measured for two generations, though no sex bias was observed. To test whether this was an issue with the guide, a further guide cassette was synthesised with a single *tra* guide that had been previously tested and shown to induce conversion. The issue with these guides was identified and all constructs were redesigned to correct this.

Chapter 6 identifies essential genes in *Ceratitis* which have potential use as toxin/antidote targets. Of the genes identified, all were sequenced in the two wildtype populations to assess how well conserved these sequences were and identify any population specific SNPs. Guides which would induce a knockout of these genes were designed and recoded rescue elements were designed to which these guides could not bind.

1.6   References

Adolfi, Adriana, Valentino M. Gantz, Nijole Jasinskiene, Hsu-Feng Lee, Kristy Hwang, Gerard Terradas, Emily A. Bulger, et al. 2020. 'Efficient Population Modification Gene-Drive Rescue System in the Malaria Mosquito *Anopheles stephensi*'. *Nature Communications* 11 (1): 5553. https://doi.org/10.1038/s41467-020-19426-0.

Alcalay, Yehonatan, Silke Fuchs, Roberto Galizi, Federica Bernardini, Roya Elaine Haghighat-Khah, Douglas B Rusch, Jeffrey R Adrion, Matthew W Hahn, Pablo Tortosa, and Philippos Aris Papathanos. 2019. 'The Potential for a Released Autosomal X-Shredder Becoming a Driving-Y Chromosome and Invasively Suppressing Wild Populations of Malaria Mosquitoes'. *bioRxiv*. https://doi.org/10.1101/860551.

Alphey, Nina, Michael B. Bonsall, and Luke Alphey. 2009. 'Combining Pest Control and Resistance Management: Synergy of Engineered Insects With Bt Crops'. *Journal of Economic Entomology* 102 (2): 717–32. https://doi.org/10.1603/029.102.0233.

Beeman, Richard W. and Friesen, Kenlee S. 1999. 'Properties and Natural Occurrence of Maternal-effect Selfish Genes ('Medea'Factors) in the Red Flour Beetle, Tribolium Castaneum.' *Heredity* 82 (5): 529–34. https://doi.org/10.1046/j.1365-2540.1999.00515.x.

Burt, Austin. 2003. 'Site-Specific Selfish Genes as Tools for the Control and Genetic Engineering of Natural Populations'. *Proc. Biol. Sci.* 270 (1518): 921–28. https://doi.org/10.1098/rspb.2002.2319.

Burt, Austin, and Anne Deredec. 2018. 'Self-Limiting Population Genetic Control with Sex-Linked Genome Editors'. *Proc. Biol. Sci.* 285 (1883). https://doi.org/10.1098/rspb.2018.0776.

Carballar-Lejarazú, Rebeca, Christian Ogaugwu, Taylor Tushar, Adam Kelsey, Thai Binh Pham, Jazmin Murphy, Hanno Schmidt, Yoosook Lee, Gregory C. Lanzaro, and Anthony A. James. 2020. 'Next-Generation Gene Drive for Population Modification of the Malaria Vector Mosquito, *Anopheles gambiae*'. *Proceedings of the National Academy of Sciences* 117 (37): 22805–14. https://doi.org/10.1073/pnas.2010214117.

Carrami, Eli M., Kolja N. Eckermann, Hassan M. M. Ahmed, Héctor M. Sánchez C., Stefan Dippel, John M. Marshall, and Ernst A. Wimmer. 2018. 'Consequences of Resistance Evolution in a Cas9-Based Sex Conversion-Suppression Gene Drive for Insect Pest Management'. *Proceedings of the National Academy of Sciences* 115 (24): 6189–94. https://doi.org/10.1073/pnas.1713825115.

Champer, Jackson, Esther Lee, Emily Yang, Chen Liu, Andrew G Clark, and Philipp W Messer. 2020a. 'A Toxin-Antidote CRISPR Gene Drive System for Regional Population Modification'. *Nat. Commun.* 11 (1): 1082. https://doi.org/10.1038/s41467-020-14960-3.

Champer, Jackson, Jingxian Liu, Suh Yeon Oh, Riona Reeves, Anisha Luthra, Nathan Oakes, Andrew G Clark, and Philipp W Messer. 2018. 'Reducing Resistance Allele Formation in CRISPR Gene Drive'. *Proc. Natl. Acad. Sci. U. S. A.* 115 (21): 5522–27. https://doi.org/10.1073/pnas.1720354115.

Champer, Jackson, Riona Reeves, Suh Yeon Oh, Chen Liu, Jingxian Liu, Andrew G. Clark, and Philipp W. Messer. 2017. 'Novel CRISPR/Cas9 Gene Drive Constructs Reveal Insights into Mechanisms of Resistance Allele Formation and Drive Efficiency in Genetically Diverse Populations'. Edited by Harmit S. Malik. *PLOS Genetics* 13 (7): e1006796. https://doi.org/10.1371/journal.pgen.1006796.

Champer, Samuel E, Suh Yeon Oh, Chen Liu, Zhaoxin Wen, Andrew G Clark, Philipp W Messer, and Jackson Champer. 2020b. 'Computational and Experimental Performance of CRISPR Homing Gene Drive Strategies with Multiplexed gRNAs'. *Sci Adv* 6 (10): eaaz0525. https://doi.org/10.1126/sciadv.aaz0525.

Chen, Chao, Robert L. Aldridge, Seth Gibson, Jedidiah Kline, Vindhya Aryaprema, Whitney Qualls, Rui-de Xue, Leigh Boardman, Kenneth J. Linthicum, and Daniel A. Hahn. 2023. 'Developing the Radiation-based Sterile Insect Technique (SIT) for Controlling *Aedes Aegypti* : Identification of a Sterilizing Dose'. *Pest Management Science* 79 (3): 1175–83. https://doi.org/10.1002/ps.7303.

Committee on Gene Drive Research in Non-Human Organisms: Recommendations for Responsible Conduct, Board on Life Sciences, Division on Earth and Life Studies, and Engineering National Academies of Sciences and Medicine. 2016. *Gene Drives on the Horizon: Advancing Science, Navigating Uncertainty, and Aligning Research with Public Values*. Washington (DC): National Academies Press (US). https://doi.org/10.17226/23405.

De Meyer, M., M. P. Robertson, A. T. Peterson, and M. W. Mansell. 2007. 'Ecological Niches and Potential Geographical Distributions of Mediterranean Fruit Fly (*Ceratitis capitata*) and Natal Fruit Fly (*Ceratitis rosa*)'. *Journal of Biogeography* 0 (0): 070821084123001-??? https://doi.org/10.1111/j.1365-2699.2007.01769.x.

Diamantidis, A. D., J. R. Carey, and N. T. Papadopoulos. 2008. 'Life-History Evolution of an Invasive Tephritid'. *Journal of Applied Entomology* 132 (9–10): 695–705. https://doi.org/10.1111/j.1439-0418.2008.01325.x.

Du, Jie, Weizhe Chen, Xihua Jia, Xuejiao Xu, Emily Yang, Ruizhi Zhou, Yuqi Zhang, Matt Metzloff, Philipp W. Messer, and Jackson Champer. 2023. 'New Germline Cas9 Promoters Show Improved Performance for Homing Gene Drive'. Preprint. Genetics. https://doi.org/10.1101/2023.07.16.549205.

Ebron, Jey Sabith, and Girish C. Shukla. 2016. 'Molecular Characterization of a Novel Androgen Receptor Transgene Responsive to MicroRNA Mediated Post-Transcriptional Control Exerted via 3'-Untranslated Region: Characterization of a Novel Androgen Receptor Reporter'. *The Prostate* 76 (9): 834–44. https://doi.org/10.1002/pros.23174.

Esvelt, Kevin M, Andrea L Smidler, Flaminia Catteruccia, and George M Church. 2014. 'Concerning RNA-Guided Gene Drives for the Alteration of Wild Populations'. *Elife* 3 (July). https://doi.org/10.7554/eLife.03401.

Facchinelli, Luca, Laura Valerio, Janine M Ramsey, Fred Gould, Rachael K Walsh, Guillermo Bond, Michael A Robert, et al. 2013. 'Field Cage Studies and Progressive Evaluation of Genetically-Engineered Mosquitoes'. *PLoS Negl. Trop. Dis.* 7 (1): e2001. https://doi.org/10.1371/journal.pntd.0002001.

Feng, Xuechun, Víctor López Del Amo, Enzo Mameli, Megan Lee, Alena L. Bishop, Norbert Perrimon, and Valentino M. Gantz. 2021. 'Optimized CRISPR Tools and Site-Directed Transgenesis towards Gene Drive Development in *Culex quinquefasciatus* Mosquitoes'. *Nature Communications* 12 (1): 2960. https://doi.org/10.1038/s41467-021-23239-0.

Galizi, Roberto, Lindsey A Doyle, Miriam Menichelli, Federica Bernardini, Anne Deredec, Austin Burt, Barry L Stoddard, Nikolai Windbichler, and Andrea Crisanti. 2014. 'A Synthetic Sex Ratio Distortion System for the Control of the Human Malaria Mosquito'. *Nat. Commun.* 5 (June): 3977. https://doi.org/10.1038/ncomms4977.

Galizi, Roberto, Andrew Hammond, Kyros Kyrou, Chrysanthi Taxiarchi, Federica Bernardini, Samantha M O'Loughlin, Philippos-Aris Papathanos, Tony Nolan, Nikolai Windbichler, and Andrea Crisanti. 2016. 'A CRISPR-Cas9 Sex-Ratio Distortion System for Genetic Control'. *Sci. Rep.* 6 (August): 31139. https://doi.org/10.1038/srep31139.

Gamez, Stephanie, Duverney Chaverra-Rodriguez, Anna Buchman, Nikolay P. Kandul, Stelia C. Mendez-Sanchez, Jared B. Bennett, Héctor M. Sánchez C., et al. 2021. 'Exploiting a Y Chromosome-Linked Cas9 for Sex Selection and Gene Drive'. *Nature Communications* 12 (1): 7202. https://doi.org/10.1038/s41467-021-27333-1.

Gantz, Valentino M, and Ethan Bier. 2015. 'Genome Editing. The Mutagenic Chain Reaction: A Method for Converting Heterozygous to Homozygous Mutations'. *Science* 348 (6233): 442–44. https://doi.org/10.1126/science.aaa5945.

Gantz, Valentino M, Nijole Jasinskiene, Olga Tatarenkova, Aniko Fazekas, Vanessa M Macias, Ethan Bier, and Anthony A James. 2015. 'Highly Efficient Cas9-Mediated Gene Drive for Population Modification of the Malaria Vector Mosquito *Anopheles stephensi*'. *Proc. Natl. Acad. Sci. U. S. A.* 112 (49): E6736-43. https://doi.org/10.1073/pnas.1521077112.

Gavis, Elizabeth R., Daniel Curtis, and Ruth Lehmann. 1996. 'Identification Ofcis-Acting Sequences That Control nanos RNA Localization'. *Developmental Biology* 176 (1): 36–50. https://doi.org/10.1006/dbio.1996.9996.

Geuverink, E., and L.W. Beukeboom. 2014. 'Phylogenetic Distribution and Evolutionary Dynamics of the Sex Determination Genes *Doublesex* and *Transformer* in Insects'. *Sexual Development* 8 (1–3): 38–49. https://doi.org/10.1159/000357056.

Gordillo, Luis F. 2015. 'Pest Persistence and Eradication Conditions in a Deterministic Model for Sterile Insect Release'. *Journal of Biological Dynamics* 9 (sup1): 64–78. https://doi.org/10.1080/17513758.2014.942393.

Gutierrez, Andrew Paul, Luigi Ponti, Markus Neteler, David Maxwell Suckling, and José Ricardo Cure. 2021. 'Invasive Potential of Tropical Fruit Flies in Temperate Regions under Climate Change'. *Communications Biology* 4 (1): 1141. https://doi.org/10.1038/s42003-021-02599-9.

Hall, Andrew Brantley, Philippos-Aris Papathanos, Atashi Sharma, Changde Cheng, Omar S. Akbari, Lauren Assour, Nicholas H. Bergman, et al. 2016. 'Radical Remodeling of the Y Chromosome in a Recent Radiation of Malaria Mosquitoes'. *Proceedings of the National Academy of Sciences* 113 (15). https://doi.org/10.1073/pnas.1525164113.

Hammond, Andrew, Roberto Galizi, Kyros Kyrou, Alekos Simoni, Carla Siniscalchi, Dimitris Katsanos, Matthew Gribble, et al. 2016. 'A CRISPR-Cas9 Gene Drive System Targeting Female Reproduction in the Malaria Mosquito Vector *Anopheles gambiae*'. *Nat. Biotechnol.* 34 (1): 78–83. https://doi.org/10.1038/nbt.3439.

Hammond, Andrew, Xenia Karlsson, Ioanna Morianou, Kyros Kyrou, Andrea Beaghton, Matthew Gribble, Nace Kranjc, et al. 2018. 'Regulation of Gene Drive Expression Increases Invasive Potential and Mitigates Resistance'. *bioRxiv*. https://doi.org/10.1101/360339.

Hammond, Andrew M, Kyros Kyrou, Marco Bruttini, Ace North, Roberto Galizi, Xenia Karlsson, Nace Kranjc, et al. 2017. 'The Creation and Selection of Mutations Resistant to a Gene Drive over Multiple Generations in the Malaria Mosquito'. *PLoS Genet.* 13 (10): e1007039. https://doi.org/10.1371/journal.pgen.1007039.

Hendrichs, J, G Franz, and P Rendon. 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): 371–77. https://doi.org/10.1111/j.1439-0418.1995.tb01303.x.

Hendrichs, J, A.S. Robinson, J. P. Cayol, and W. Enkerlin. 2002. 'MEDFLY AREAWIDE STERILE INSECT TECHNIQUE PROGRAMMES FOR PREVENTION, SUPPRESSION OR ERADICATION: THE IMPORTANCE OF MATING BEHAVIOR STUDIES'. *Florida Entomologist* 85 (1): 1–13. https://doi.org/10.1653/0015-4040(2002)085[0001:MASITP]2.0.CO;2.

Hendrichs, Jorge, and Alan Robinson. 2009. 'Sterile Insect Technique'. In *Encyclopedia of Insects*, 953–57. Elsevier. https://doi.org/10.1016/b978-0-12-374144-8.00252-6.

Holman, Luke. 2019. 'Evolutionary Simulations of -Linked Suppression Gene Drives'. *Proc. Biol. Sci.* 286 (1912): 20191070. https://doi.org/10.1098/rspb.2019.1070.

Hopkins, Ben R, and Artyom Kopp. 2021. 'Evolution of Sexual Development and Sexual Dimorphism in Insects'. *Current Opinion in Genetics & Development* 69 (August): 129–39. https://doi.org/10.1016/j.gde.2021.02.011.

Kandul, Nikolay P, Junru Liu, Alexander D Hsu, Bruce A Hay, and Omar S Akbari. 2020. 'A Drug-Inducible Sex-Separation Technique for Insects'. *Nature Communications* 11 (1). https://doi.org/10.1038/s41467-020-16020-2.

Kondo, Shu, and Ryu Ueda. 2013. 'Highly Improved Gene Targeting by Germline-Specific Cas9 Expression in *Drosophila*'. *Genetics* 195 (3): 715–21. https://doi.org/10.1534/genetics.113.156737.

Kyriacou, C. 1992. 'Sex Variations'. *Trends in Genetics* 8 (8): 261–63. https://doi.org/10.1016/0168-9525(92)90241-U.

Kyrou, Kyros, Andrew M Hammond, Roberto Galizi, Nace Kranjc, Austin Burt, Andrea K Beaghton, Tony Nolan, and Andrea Crisanti. 2018. 'A CRISPR–Cas9 Gene Drive Targeting Doublesex Causes Complete Population Suppression in Caged *Anopheles gambiae* Mosquitoes'. *Nature Biotechnology* 36 (11): 1062–66. https://doi.org/10.1038/nbt.4245.

Labbé, Geneviève M C, Sarah Scaife, Siân A Morgan, Zoë H Curtis, and Luke Alphey. 2012. 'Female-Specific Flightless (fsRIDL) Phenotype for Control of *Aedes albopictus*'. *PLoS Negl. Trop. Dis.* 6 (7): e1724. https://doi.org/10.1371/journal.pntd.0001724.

Lance, D. R., W. M. Woods, and M. Stefan. 2014. 'Invasive Insects in Plant Biosecurity: Case Study – Mediterranean Fruit Fly'. In *The Handbook of Plant Biosecurity*, edited by Gordon Gordh and Simon McKirdy, 447–84. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-007-7365-3_15.

Lees, Rosemary Susan, Jeremie R L Gilles, Jorge Hendrichs, Marc J B Vreysen, and Kostas Bourtzis. 2015. 'Back to the Future: The Sterile Insect Technique against Mosquito Disease Vectors'. *Current Opinion in Insect Science* 10: 156–62. https://doi.org/10.1016/j.cois.2015.05.011.

Leftwich, Philip T, Martha Koukidou, Polychronis Rempoulakis, Hong-Fei Gong, Antigoni Zacharopoulou, Guoliang Fu, Tracey Chapman, Aris Economopoulos, John Vontas, and Luke Alphey. 2014. 'Genetic Elimination of Field-Cage Populations of Mediterranean Fruit Flies'. *Proc. Biol. Sci.* 281 (1792). https://doi.org/10.1098/rspb.2014.1372.

Lei, Jianfeng, Peihong Dai, Jiyang Li, Mi Yang, Xiuqin Li, Wanqi Zhang, Guantong Zhou, WangzhenGuo, and Xiaodong Liu. 2021. 'Tissue-Specific CRISPR/Cas9 System of Cotton Pollen with GhPLIMP2b and GhMYB24 Promoters'. *Journal of Plant Biology* 64 (1): 13–21. https://doi.org/10.1007/s12374-020-09272-4.

Li, Ming, Ting Yang, Michelle Bui, Stephanie Gamez, Tyler Wise, Nikolay P. Kandul, Junru Liu, et al. 2021. 'Suppressing Mosquito Populations with Precision Guided Sterile Males'. *Nature Communications* 12 (1): 5374. https://doi.org/10.1038/s41467-021-25421-w.

Liquido, NJ, McQuate, GT, Hanlin, MA, and Suiter, KA. 2017. 'Host Plants of the Mediterranean Fruit Fly, *Ceratitis capitata* (Wiedemann), Version 3.5'. *USDA Compendium of Fruit Fly Host Information*.

Malacrida, A. R., L. M. Gomulski, M. Bonizzoni, S. Bertin, G. Gasperi, and C. R. Guglielmino. 2007. 'Globalization and Fruitfly Invasion and Expansion: The Medfly Paradigm'. *Genetica* 131 (1): 1–9. https://doi.org/10.1007/s10709-006-9117-2.

Mao, Yanfei, Zhengjing Zhang, Zhengyan Feng, Pengliang Wei, Hui Zhang, José Ramón Botella, and Jian-Kang Zhu. 2016. 'Development of Germ-Line-Specific CRISPR-Cas9 Systems to Improve the Production of Heritable Gene Modifications in *Arabidopsis*'. *Plant Biotechnology Journal* 14 (2): 519–32. https://doi.org/10.1111/pbi.12468.

McFarlane, Gus R, C Bruce A Whitelaw, and Simon G Lillico. 2018. 'CRISPR-Based Gene Drives for Pest Control'. *Trends Biotechnol.* 36 (2): 130–33. https://doi.org/10.1016/j.tibtech.2017.10.001.

Meccariello, Angela, Marco Salvemini, Pasquale Primo, Brantley Hall, Panagiota Koskinioti, Martina Dalíková, Andrea Gravina, et al. 2019. '*Maleness-on-the-Y* (*MoY*) Orchestrates Male Sex Determination in Major Agricultural Fruit Fly Pests'. *Science* 365 (6460): 1457–60. https://doi.org/10.1126/science.aax1318.

Nguyen, Thu N. M., Amanda Choo, and Simon W. Baxter. 2021. 'Lessons from *Drosophila*: Engineering Genetic Sexing Strains with Temperature-Sensitive Lethality for Sterile

Insect Technique Applications'. *Insects* 12 (3): 243. https://doi.org/10.3390/insects12030243.

Oberhofer, Georg, Tobin Ivy, and Bruce A Hay. 2018. 'Behavior of Homing Endonuclease Gene Drives Targeting Genes Required for Viability or Female Fertility with Multiplexed Guide RNAs'. *Proc. Natl. Acad. Sci. U. S. A.* 115 (40): E9343–52. https://doi.org/10.1073/pnas.1805278115.

Oberhofer, Georg, Tobin Ivy, Bruce A Hay. 2019. 'Cleave and Rescue, a Novel Selfish Genetic Element and General Strategy for Gene Drive'. *Proc. Natl. Acad. Sci. U. S. A.* 116 (13): 6250–59. https://doi.org/10.1073/pnas.1816928116.

Oléron Evans, Thomas P., and Steven R. Bishop. 2014. 'A Spatial Model with Pulsed Releases to Compare Strategies for the Sterile Insect Technique Applied to the Mosquito Aedes Aegypti'. *Mathematical Biosciences* 254 (August): 6–27. https://doi.org/10.1016/j.mbs.2014.06.001.

Ordax, Mónica, Jaime E. Piquer-Salcedo, Ricardo D. Santander, Beatriz Sabater-Muñoz, Elena G. Biosca, María M. López, and Ester Marco-Noales. 2015. 'Medfly *Ceratitis capitata* as Potential Vector for Fire Blight Pathogen *Erwinia amylovora*: Survival and Transmission'. Edited by Guy Smagghe. *PLOS ONE* 10 (5): e0127560. https://doi.org/10.1371/journal.pone.0127560.

Oxitec. 2024. 'Technology'. Oxitec.com. Accessed May 8, 2024.

https://www.oxitec.com/en/our-technology.

Price, Dana C, Andrea Egizi, and Dina M Fonseca. 2015. 'The Ubiquity and Ancestry of Insect *Doublesex*'. *Sci. Rep.* 5 (August): 13068. https://doi.org/10.1038/srep13068.

Primo, Pasquale, Angela Meccariello, Maria Grazia Inghilterra, Andrea Gravina, Giuseppe Del Corsano, Gennaro Volpe, Germano Sollazzo, et al. 2020. 'Targeting the Autosomal *Ceratitis capitata* Transformer Gene Using Cas9 or dCas9 to Masculinize XX Individuals without Inducing Mutations'. *BMC Genetics* 21 (S2): 150. https://doi.org/10.1186/s12863-020-00941-4.

Prowse, Thomas Aa, Fatwa Adikusuma, Phillip Cassey, Paul Thomas, and Joshua V Ross. 2019. 'A Y-Chromosome Shredding Gene Drive for Controlling Pest Vertebrate Populations'. *Elife* 8 (February). https://doi.org/10.7554/eLife.41873.

Raban, Robyn R., John M. Marshall, and Omar S. Akbari. 2020. 'Progress towards Engineering Gene Drives for Population Control'. Edited by Michael H. Dickinson, Leslie B. Vosshall, and Julian A. T. Dow. *Journal of Experimental Biology* 223 (Suppl_1): jeb208181. https://doi.org/10.1242/jeb.208181.

Robinson, A S. 2002. 'Genetic Sexing Strains in Medfly, *Ceratitis capitata*, Sterile Insect Technique Programmes'. *Genetica* 116 (1): 5–13. https://doi.org/10.1023/a:1020951407069.

Saccone, Giuseppe, Marco Salvemini, Attilio Pane, and Lino C. Polito. 2008. 'Masculinization of XX *Drosophila* Transgenic Flies Expressing the *Ceratitis capitata DoublesexM* Isoform'. *The International Journal of Developmental Biology* 52 (8): 1051–57. https://doi.org/10.1387/ijdb.082657gs.

Salvemini, Marco, Mark Robertson, Benjamin Aronson, Peter Atkinson, Lino C Polito, and Giuseppe Saccone. 2009. '*Ceratitis capitata Transformer-2* Gene Is Required to Establish and Maintain the Autoregulation of *Cctra*, the Master Gene for Female Sex Determination'. *Int. J. Dev. Biol.* 53 (1): 109–20. https://doi.org/10.1387/ijdb.082681ms.

Schliekelman, P, and F Gould. 2000. 'Pest Control by the Release of Insects Carrying a Female-Killing Allele on Multiple Loci'. *J. Econ. Entomol.* 93 (6): 1566–79. https://doi.org/10.1603/0022-0493-93.6.1566.

Siddall, Alex, Tim Harvey-Samuel, Tracey Chapman, and Philip T. Leftwich. 2022. 'Manipulating Insect Sex Determination Pathways for Genetic Pest Management: Opportunities and Challenges'. *Frontiers in Bioengineering and Biotechnology* 10 (June): 867851. https://doi.org/10.3389/fbioe.2022.867851.

Simoni, Alekos, Andrew M Hammond, Andrea K Beaghton, Roberto Galizi, Chrysanthi Taxiarchi, Kyros Kyrou, Dario Meacci, et al. 2020. 'A Male-Biased Sex-Distorter Gene Drive for the Human Malaria Vector *Anopheles Gambiae*'. *Nat. Biotechnol.* 38 (9): 1054–60. https://doi.org/10.1038/s41587-020-0508-1.

Sultana, Sabira, John B. Baumgartner, Bernard C. Dominiak, Jane E. Royer, and Linda J. Beaumont. 2020. 'Impacts of Climate Change on High Priority Fruit Fly Species in Australia'. Edited by Nikos T. Papadopoulos. *PLOS ONE* 15 (2): e0213820. https://doi.org/10.1371/journal.pone.0213820.

Thomas, D D, C A Donnelly, R J Wood, and L S Alphey. 2000. 'Insect Population Control Using a Dominant, Repressible, Lethal Genetic System'. *Science* 287 (5462): 2474–76. https://doi.org/10.1126/science.287.5462.2474.

Thomas, Michael C., John B. Heppner, Robert E. Woodruff, Howard V. Weems, Jr., Gary J. Steck, and Thomas R. Fasulo. 2022. 'Mediterranean Fruit Fly, *Ceratitis capitata* (Wiedemann) (Insecta: Diptera: Tephritidae)'. *EDIS* 2004 (8). https://doi.org/10.32473/edis-in371-2001.

Valdez, Megan R Wise de, Derric Nimmo, John Betz, Hong-Fei Gong, Anthony A James, Luke Alphey, and William C Black. 2011. 'Genetic Elimination of Dengue Vector Mosquitoes'. *Proceedings of the National Academy of Sciences* 108 (12): 4772–75. https://doi.org/10.1073/pnas.1019295108.

Verkuijl, Sebald A. N., Estela Gonzalez, Ming Li, Joshua X. D. Ang, Nikolay P. Kandul, Michelle A. E. Anderson, Omar S. Akbari, Michael B. Bonsall, and Luke Alphey. 2022. 'A CRISPR Endonuclease Gene Drive Reveals Distinct Mechanisms of Inheritance Bias'. *Nature Communications* 13 (1): 7145. https://doi.org/10.1038/s41467-022-34739-y.

Wang, Zhi-Ping, Hui-Li Xing, Li Dong, Hai-Yan Zhang, Chun-Yan Han, Xue-Chen Wang, and Qi-Jun Chen. 2015. 'Egg Cell-Specific Promoter-Controlled CRISPR/Cas9 Efficiently Generates Homozygous Mutants for Multiple Target Genes in *Arabidopsis* in a Single Generation'. *Genome Biology* 16 (1): 144. https://doi.org/10.1186/s13059-015-0715-0.

Weldon, C. 2022. '*Ceratitis capitata* (Mediterranean Fruit Fly)'. https://doi.org/10.1079/cabicompendium.12367.

White, Steven M., Pejman Rohani, and Steven M. Sait. 2010. 'Modelling Pulsed Releases for Sterile Insect Techniques: Fitness Costs of Sterile and Transgenic Males and the Effects on Mosquito Dynamics'. *Journal of Applied Ecology* 47 (6): 1329–39. https://doi.org/10.1111/j.1365-2664.2010.01880.x.

Yadav, Amarish K., Cole Butler, Akihiko Yamamoto, Anandrao A. Patil, Alun L. Lloyd, and Maxwell J. Scott. 2023. 'CRISPR/Cas9-Based Split Homing Gene Drive Targeting *Doublesex* for Population Suppression of the Global Fruit Pest *Drosophila suzukii*'. *Proceedings of the National Academy of Sciences* 120 (25): e2301525120. https://doi.org/10.1073/pnas.2301525120.

1.7 Appendices

1.7.1 Published literature review

Manipulating Insect Sex Determination Pathways for Genetic Pest Management: Opportunities and Challenges

Alex Siddall[1†], Tim Harvey-Samuel[2†], Tracey Chapman[1] and Philip T. Leftwich[1]*

[1]School of Biological Sciences, University of East Anglia, Norwich, United Kingdom, [2]Arthropod Genetics, The Pirbright Institute, Pirbright, United Kingdom

Sex determination pathways in insects are generally characterised by an upstream primary signal, which is highly variable across species, and that regulates the splicing of a suite of downstream but highly-conserved genes (*transformer*, *doublesex* and *fruitless*). In turn, these downstream genes then regulate the expression of sex-specific characteristics in males and females. Identification of sex determination pathways has and continues to be, a critical component of insect population suppression technologies. For example, "first-generation" transgenic technologies such as fsRIDL (Female-Specific Release of Insects carrying Dominant Lethals) enabled efficient selective removal of females from a target population as a significant improvement on the sterile insect technique (SIT). Second-generation technologies such as CRISPR/Cas9 homing gene drives and precision-guided SIT (pgSIT) have used gene editing technologies to manipulate sex determination genes *in vivo*. The development of future, third-generation control technologies, such as Y-linked drives, (female to male) sex-reversal, or X-shredding, will require additional knowledge of aspects of sexual development, including a deeper understanding of the nature of primary signals and dosage compensation. This review shows how knowledge of sex determination in target pest species is fundamental to all phases of the development of control technologies.

Keywords: gene drive, sex conversion, release of insects carrying a dominant lethal, sterile insect technique (SIT), doublesex (dsx), *tra*, dosage compensation, sex determination'

INTRODUCTION

Insect pests cause enormous damage to human health (through the transmission of diseases such as dengue fever and malaria) and agriculture (through damage to crops or livestock). Existing control methods include pesticides, biological control, and integrated pest and habitat management. However, while many of these approaches have been highly successful, they also have limitations. For example, the use of pesticides can select strongly for resistance, damage non-pest populations (Hawkins et al., 2018) and the environment. The success of biological control and integrated management programmes may also depend upon whether efficient natural enemies are available and the specific ecological setting. As a result, existing control strategies, particularly chemical control, are likely to become increasingly restricted while simultaneously becoming less effective. Global climate change is also predicted to increase the

range and the number of insect pests (Deutsch et al., 2018; Gomez-Zavaglia et al., 2020; Sultana et al., 2020). Therefore, it is clear that there are significant challenges for the future in controlling insect pests, to safeguard against disease and maintain global food security.

In light of these concerns, there has been considerable investment in new and alternative technologies, such as genetics-based approaches to pest control, to protect health and food security (Alphey, 2014; Burt and Crisanti, 2018; Raban et al., 2020). Genetic Pest Management (GPM) aims to harness the natural mating system of the pest species to introduce into the target population traits that will reduce fitness and ultimately lead to a reduction of numbers or elimination. The most widely used GPM systems for suppression to date have been variants of the sterile insect technique (SIT) (Klassen, 2005), including the Wolbachia incompatible insect technique (IIT) (Atyame et al., 2011; De Barro et al., 2011; Zheng et al., 2019) and genetic engineering (Phuc et al., 2007). GPM systems that are transmitted or inherited through one sex and sterilise, kill or change the sex of the other offer the most significant potential for control (Bax and Thresher, 2009). As females are predominantly the agents of damage (*via* biting or ovipositing), and generally determine the effective population size, most approaches have focused on releasing benign males that produce either male progeny or none at all (Rendón et al., 2004). Elimination of females ensures long-term suppression and immediately reduces the associated damage caused by biting or egg-laying.

GPM technologies for insect population suppression currently under development seek to improve on older systems by spreading female-targeting genetic loads through a population or converting female progeny into functional males. These newer technologies also make use of contemporary molecular biology tools—particularly those involved in gene editing such as CRISPR/Cas9. However, what is common to all is that they exploit knowledge of the sex determination pathways of the target species, thus exemplifying the importance of incorporating fundamental biological principles to underpin applied science in GPM (Leftwich et al., 2016; Leftwich et al., 2021).

Here we first introduce the fundamentals of insect sex determination systems, focusing on species of interest to GPM. We detail which components are conserved and which show more rapid evolution, what types of primary signals have evolved and in which "direction" they push downstream cascades (e.g., towards maleness or femaleness). We then provide a framework for understanding how sex determination systems have been used to develop insect population suppression tools. We describe three "generations" of genetic engineering technologies with related components or goals. First-generation systems are genetically engineered analogues of the classical Sterile Insect Technique. Second-generation systems are under development and are made possible by the development of the CRISPR/Cas9 platform. Finally, we discuss the challenges inherent in developing "third-generation" control technologies that seek to achieve the goal of sex-conversion by manipulating master regulators of sex determination.

INSECT SEX DETERMINATION SYSTEMS

Sex determination systems can be described as a cascade in the form of a pyramid. There is an initial primary signal ("master regulator"—top of the pyramid) that initiates a limited series of intermediary elements (middle) that then result in diverse downstream sexual differentiation and development (base of the pyramid). In insects, the genes at the base of the pyramid tend to be highly conserved, while the elements at the top show marked diversity, both in identity and function (Adolfi et al., 2021; Hopkins and Kopp, 2021). The hypothesis is that these basal genes—which generally consist of transcription factors—represent ancient mechanisms of sex determination [e.g., *doublesex* is shared with some non-dipteran arthropods (Price et al., 2015; Wexler et al., 2019)]. At the same time, the primary signal can evolve rapidly, even within a taxonomic order. The differences in conservation between the basal and intermediate elements of the sex determination pathway are shown in **Figure 1**.

Amongst the downstream basal elements, the most well conserved is *doublesex (dsx)* (Saccone et al., 2002; Price et al., 2015; Verhulst and van de Zande, 2015; Wexler et al., 2019). Ubiquitous amongst insects, *dsx* is the "central nexus" between sex determination and sexual differentiation cascades (Verhulst and van de Zande, 2015). It functions as a transcription factor activating or repressing thousands of downstream genes which cause female or male somatic differentiation. Its role in this regulation (male or female biasing) is determined by whether it exists in a male or female "form" as a protein. In most cases, this is determined by sex-specific alternative splicing of the initial *dsx* pre-mRNA—itself determined by intermediary regulators between the primary signal and *dsx*. "Male" *dsx* typically represents the constitutive splicing isoform; while female-specific *dsx* isoforms require the splice enhancing factor transformer (*tra*) (**Figure 2**). However, even within this most conserved member of the insect sex determination cascade, some variation does exist. For example, there are significant differences in the number and "style" of exon skipping between different insect species (Verhulst and van de Zande, 2015). For example, in lepidopterans, the constitutive dsx isoform is female with male determining factors required to shift splicing towards a male form (Lee et al., 2015; Xu et al., 2017; Visser et al., 2021). Further, in at least two species of termites, *dsx* has evolved towards male-only expression rather than sex-alternate splicing (Wexler et al., 2019; Miyazaki et al., 2021).

An intermediary element that exists directly above *dsx* is *transformer (tra)*. Although not as highly conserved as *dsx* (Verhulst et al., 2010b), *tra* homologues have been identified in a variety of insect orders, e.g., Coleopterans [*Tribolium castaneum* (Shukla and Palli, 2012)], Hymenopterans [*Apis mellifera* (Gempe et al., 2009), *Nasonia vitripennis* (Verhulst et al., 2010a)] and dipterans [*Drosophila melanogaster* (Sosnowski et al., 1989; Inoue et al., 1990), *Musca domestica* (Inoue and Hiroyoshi, 1986) and a number of Tephritid fruitflies (Pane et al., 2002; Lagos et al., 2007; Ruiz et al., 2007)]. In these groups, sex-specific splicing of *tra* leads to a "functional" female-specific or "non-functional" male-specific protein. Interestingly,

40

FIGURE 1 | The variety of sex determination systems employed by pest insects. The more upstream elements of the sex determination pathway such as the primary signal vary widely across pest insect species. Several species have XX/XY chromosome structures although they do not utilise them in the same way. Ceratitis and Aedes employ masculinizing elements carried on the Y chromosome whereas Drosophila uses X counting to determine sex.

tra has often been found to self-regulate its splicing (except in *D. melanogaster*) acting as a positive self-regulatory element to increase its own expression (Pane et al., 2002; Gempe et al., 2009; Salvemini et al., 2009; Verhulst et al., 2010a; Hediger et al., 2010). Functional *tra* acts as a splicing enhancer, binding *dsx* pre-mRNA and promoting the inclusion of female-specific exons in the final transcript.

Although *tra* is an essential gene in the sex determination pathway of many dipterans and hymenopterans; in other species, there may be, as yet, no identified homologue, as is the case for *Aedes aegypti* (Nene et al., 2007). In these cases, there may be functional quasi-equivalents for *transformer*. For example, in silkmoth (*Bombyx mori*) males, P-element somatic inhibitor and IGF-II mRNA binding proteins interact to form a complex which binds *dsx* pre-mRNA. This complex inhibits internal splice site junctions, excluding female-specific exons to produce "male" form dsx mRNA (Suzuki et al., 2010). This example illustrates that while there may be orthogonal splicing factor/s, analogous to *tra*, the role the new factor/s plays may be very different (promoting male-form, rather than female-form *dsx* alternate splicing). The *transformer-2* gene (*tra2*) is also involved in the sex determination pathway of many insects. It is often an additional factor that forms an essential part of the splicing enhancer complex, which helps sustain and regulate the splicing of *tra* (Salvemini et al., 2009). It is, however, not a homologue of *transformer* itself. *Tra-2*, unlike *tra*, has also been shown to have both expression and function in males (Salvemini et al., 2009).

Above *tra* (or other intermediary elements) in the sex determination pathway lies the primary signal or master regulator underpinning the sexual determination cascade. The identity and function of these master regulators vary enormously between species even within the same order due to a high turnover rate of the primary signaler at this level (Gempe and Beye, 2011). For example, in four dipteran species, the mosquitos *Anopheles gambiae*, *Ae. aegypti*, the Mediterranean fruitfly *Ceratitis capitata*, and the house fly *M. domestica*, the master regulators of sex determination are evolutionarily unrelated [*Yob, Nix, MoY*, and *Mdmd* respectively (Hall et al., 2015; Krzywinska et al., 2016; Sharma et al., 2017; Meccariello et al., 2019)]. While the exact mechanisms by which these primary signals act remains largely unknown, the mosquito species assessed so far (including those listed above) do not appear to possess a *tra* homologue (Nene et al., 2007). In contrast, the sex-specific splicing of *tra* is integral to the sex determination cascade in *C. capitata and M. domestica*, strongly suggesting divergent functions in regulating intermediary elements between the top and bottom levels of the pyramid in mosquitos and other diptera (Saccone et al., 2011). In *D. melanogaster*, sex is determined by an X-chromosome counting mechanism. The expression ratios of specific X-linked (*sis-a, sis-b, sis-c* and *run*) and autosomal genes determine the correct expression of an autoregulatory-splicing female-determining gene (*sex-lethal*) (Cline, 1993; Kaiser and Bachtrog, 2010). A 1:1 X: A ratio (implying two X chromosomes) leads to functional sex-lethal expression and the female sex determination cascade (Baker and Ridge, 1980; Sánchez and Nöthiger, 1982; Parkhurst et al., 1990). Although *tra* plays a crucial intermediary role in *D. melanogaster* (as in *C. capitata*), the sex-determining role of *sex-lethal* appears to be Drosophilid-specific (Meise et al., 1998). In Lepidoptera, primary signals can vary, as both Z-chromosome counting and dominant primary signals exist in different species (Traut et al., 2007). In *B. mori*, a

41

FIGURE 2 | Sex-specific splicing patterns of doublesex. The conserved elements of the sex determination pathway that result in the alternative splicing of the pre-mRNA dsx. Depending on the upstream signalling pathways TRA is either present as a functional protein (in female development) or a non-functional protein (in male development). TRA acts as a splicing enhancer to promote the recruitment of the splicing machinery to the weak splice acceptor prior to exon 4 of the dsx pre-mRNA (Shukla and Nagaraju, 2010). This allows for the retention of exon 4 in the mRNA resulting in the female version of dsx. Doublesex exon numbers vary among species with the use of splicing to retain or remove sex-specific exons remaining constant, Ceratitis capitata pathway illustrated in this figure.

dominant female-linked (*feminizer*) piRNA system encoded on the female-specific W chromosome silences Z-linked genes that would otherwise initial male sex-determination (Hasselmann et al., 2008; Katsuma et al., 2018); this then directs *dsx* splicing. In Hymenoptera, sexual fate is effectively regulated by the presence or absence of a paternal genome. E.g., in the honeybee, *A. mellifera*, it is determined by heterozygosity at a single locus the complementary sex determiner (csd) gene (Gempe et al., 2009), single sex alleles within an organism result in male development (homozygous/hemizygous) and mismatched sex alleles develop into females (heterozygous). While in the haplodiploid wasp *N. vitripennis* the sex determination gene *wasp overruler of masculinization* (*wom*) is only transcribed from the paternally provided genome (Zou et al., 2020). In both systems, *tra* and *dsx* are still employed as intermediate and basal elements (Cho et al., 2007; Zou et al., 2020).

The genomic location of the master regulator and, specifically, whether it exists on a heteromorphic sex chromosome is a further important aspect of insect sex determination systems in this review. In many cases, the evolution of distinct sex-

chromosomes necessitates a mechanism for equalising the expression of shared, sex-linked genes between sexes (dosage compensation). In *D. melanogaster*, the absence of *sex-lethal* in males initiates hypertranscription from the single X-chromosome to make up for two X's in females. Inactivation of *sex-lethal* in females leads to deadly X-chromosome hypertranscription due to carrying two X chromosomes. This "coupled" sex determination and dosage compensation has implications for manipulating these systems for genetic pest control. For example, it is a significant challenge to aim to alter sexual fate without simultaneously "programming" the dosage compensation pathway. As a result, in many cases where the master regulator has been identified and ectopically expressed in females (e.g., *Yob* in *An. gambiae; Guy1* in *Anopheles stephensi*), the result is the death of the XX individuals rather than their conversion to fertile males (Criscione et al., 2016; Krzywinska and Krzywinski, 2018; Qi et al., 2019). This represents a high hurdle if the most potent application of manipulating sex determination for GPM suppression systems is to convert a population to a single-sex rather than selectively kill off females.

Fortunately, neither fully differentiated sex chromosomes, nor coupled-dosage compensation pathways where heterogametic sex chromosomes exist, appear universal amongst insects that are of concern to human health or agriculture. For example, *Aedes albopictus* (and other culicine mosquitos) do not possess heterogametic sex chromosomes but rather a small "Male-determining" locus on chromosome 1 (Hall et al., 2014; Gomulski et al., 2018). Transgenic expression of the male-determining gene from within this locus (*Nix*) in transgenic *Ae. albopictus* was sufficient to convert females into functioning males (Lutrat et al., 2022). Similarly, in *C. capitata*, transient ectopic expression of *MoY* is enough to convert karyotypic females to functional males, suggesting either a lack of dosage compensation or an "un-coupled" version in this species (Meccariello et al., 2019).

Understanding the nature of the sex determination pathway that has evolved in a pest of concern, and its possible interaction with dosage compensation, provides some potential routes for manipulating a species for genetic pest management. If the goal is female to male sex *conversion*, then the upper levels of the pyramid will likely need to be manipulated to ensure complete sex conversion and the viability/fertility of converted individuals. However, this goal may be difficult or practically impossible for species with lethal dosage compensation. If the goal is simply to kill one of the sexes, then lower, more conserved levels of the pyramid can be targeted, and this may also be beneficial in transferring efficient designs between related pests. Further exploration of the fundamental basis of sex determination mechanisms is, therefore, essential.

FIRST-GENERATION TECHNOLOGIES: AN IMPROVEMENT ON THE PAST

First-generation transgenic GPM systems are genetically engineered analogues of the classical sterile insect technique (SIT). In the SIT, the mass release of irradiated (sterilised) males results in a lower proportion of fertilised females in the field due to mating with the sterile males instead of the fertile, wildtype males. SIT is most efficient if only males are released (Rendón et al., 2004). Preventing the introduction into the population of females that can damage fruit crops or transmit disease is an additional advantage. However, male-only releases require an efficient mechanism for sex sorting. For this purpose, genetic sexing strains (GSS- **Box 1**) were developed that differentiated between males and females with selective markers such as pupal colour or conditional lethality when exposed to high temperatures (Rendon, 1996).

First-generation transgenic systems sought to improve these technologies by creating analogues of GSSs that could also be used as population suppression measures in the field. The most widely adopted of these was the Release of Insects carrying a Dominant Lethal technology (RIDL) (Thomas et al., 2000), but also see (Schetelig and Handler, 2012) (Ogaugwu et al., 2013; Schetelig et al., 2016). The basis behind RIDL is the genetic modification of

a pest to carry a deleterious/lethal gene whose expression can be turned off (repressed) during rearing, but which, when inherited by the progeny of released insects, will result in lethality for individuals in the field. As with SIT, mass releases of RIDL insects can thus suppress a target population by continually killing off field-born individuals before they can reproduce. Female-specific RIDL (fsRIDL) and genetic sexing strains have been developed in many species by combining this repressible lethality with sex-alternatively spliced introns from basal genes within the sex determination pathway (Fu et al., 2007; Schetelig and Handler, 2012; Ogaugwu et al., 2013; Tan et al., 2013), specifically *dsx* and *tra*. The pre-mRNA of these two genes is spliced differently between males and females—leading to the sex-specific inclusion or exclusion of exonic sequences. Sex-specific transgene expression can be designed by including the sequences responsible for sex-specific splicing (introns) embedded within components integral to the repressible-lethal system. As such, functional fsRIDL proteins are produced in one sex (usually females) and not the other (in the same manner as *tra* and sex-specific dsx proteins). Released fsRIDL individuals are homozygous fertile males that produce heterozygous male-only offspring when mating to wildtype females following release. These heterozygous fsRIDL males can then produce wildtype males and females as well as heterozygous fsRIDL males, resulting in a steadily diluting suppressive effect without continued releases. Female-specific RIDL lines have been developed in many insect pest species, including tephritid fruitflies (Fu et al., 2007; Ant et al., 2012), blow flies (Yan and Scott, 2020), screwworms (Concha et al., 2016), moths (Morrison et al., 2012; Jin et al., 2013) and mosquitoes (Phuc et al., 2007; Collado, 2013). Proof-of-principle demonstrations have also been made in beetles (Gregory, 2015). Caged and open-release trials have demonstrated that repeated releases of fsRIDL males can cause the rapid suppression of target populations (Harris et al., 2011; Leftwich et al., 2014; Carvalho et al., 2015; Shelton et al., 2020).

For these first-generation technologies, it is not necessary to know the precise means by which the sex-specific processing of *dsx* or *tra* components are regulated (i.e., the upstream elements of the pyramid which act upon them). All that is required is a basic understanding of the arrangement of the chosen gene and final mRNA differences between sexes. This conservation of function is a distinct advantage for adapting transgenic constructs, with minimal changes, across multiple species (Tan et al., 2013). A further advantage of using highly conserved basal elements of the pyramid is that the splicing signals which regulate their sex-specific splicing are often shared between closely related species (Ant et al., 2012). For example, an fsRIDL construct built using intronic sequences from pink bollworm (*Pectinophora gossypiella*) *dsx* functioned just as effectively in that species as it did in silkworm (*B. mori*) and diamondback moth (*Plutella xylostella*) (Jin et al., 2013; Tan et al., 2013). The limitation of using these downstream elements of the pyramid is that these systems are generally limited to *killing* females rather than their conversion to males. This, coupled with the self-limiting nature of these technologies, makes them far less potent than "second-generation" technologies (next section).

43

releases, leading for example, to the development of confinable split-drive technologies that also safeguard against accidental release (Li et al., 2020; López Del Amo et al., 2020).

HGDs designed for population suppression are inserted into (and therefore disrupt) a gene of essential function. If that essential gene is haplosufficient, the HGD causes a deleterious recessive phenotype. Heterozygous individuals are viable and contribute to the spread of the drive, but when the drive reaches a high allele frequency in the population, more and more non-viable homozygous individuals are produced, leading to a dramatic reduction in the reproductive capacity of the population (Deredec et al., 2011). The most efficient designs target viability in only one sex (usually females), allowing the drive to spread efficiently within the other sex, regardless of frequency.

As with any pest elimination strategy, selection for resistance to a gene drive is a concern. The most commonly cited mode of resistance is mutations in the target cut sites "resistance alleles," which prevent further recognition by the Cas9/sgRNA complex and therefore targeting by the drive. If resistance alleles do not severely disrupt the gene's coding potential (e.g., synonymous mutations, or small in-frame deletions), they may be rapidly selected for in a drive carrying population as the fitness differential between the resistance mutation and the drive is expected to be large (Champer et al., 2017; Carrami et al., 2018). Several strategies can mitigate this, including restricting expression of the gene drive to the germline (Hammond et al., 2021) targeting highly conserved gene sites, hoping that this indicates low tolerance for mutated alleles. In at least one example in *A. gambiae,* this appears to have been achieved by targeting the intron 4 - exon 5 boundary (splice junction) of *dsx* (Kyrou et al., 2018). Typically, this fifth exon is included in female, but not in male, *dsx* mRNA (female-specific exon). However, when the splice junction is disrupted, the 5th exon is instead excluded (skipped) in both sexes. Homozygous deletions at this junction incapacitated female sexual development leading to intersex and sterile females as these individuals could not produce functional dsxF (lacked the 5th exon). Male development and fertility, however, was unaffected as this 5th exon is canonically excluded from the final mRNA transcript in that sex. Within this study, this produced a highly effective drive, spreading female non-viability through a caged population, leading to a rapid population crash (Kyrou et al., 2018).

Much like earlier "first-generation" systems, the use of highly conserved, downstream basal elements of the sex determination pathway as essential components of a suppression drive produced reliable, predictable and effective mechanisms of generating female non-viability (in this case sterility, rather than lethality, as with first-generation technologies). The highly conserved and well-understood role of *dsx* in the downstream regulation of female sexual development means that it is highly likely that similar suppression drives could be developed in a range of other insect pest species. However, there are differences between the number and "style" of exon skipping which occurs between different insect species, which would require consideration.

While no functional resistant alleles were observed in this study, it is possible that at larger-scale releases, pre-existing or *de novo* alleles might eventually occur (Bier, 2021). Including multiple guide RNAs designed against numerous sequences at the target loci, also known as "multiplexing," is one frequently

discussed mitigation against this (Carrami et al., 2018). Pre-existing sequence variations or failed homing attempts must inhibit all target sequences simultaneously to inhibit the drive and are therefore less likely to generate functional resistant mutants (Marshall et al., 2017; Champer et al., 2018, 2020b; Oberhofer et al., 2018; Champer S. E. et al., 2020). However, the small target site of the exon-splice junction in *dsx* means that multiplexing guide RNAs would be difficult to engineer for this gene and would likely be an issue in using homologous targets in other species. Complementary alternatives, including combinate X-shredder drives, have been demonstrated (Simoni et al., 2020).

Precision-Guided Sterile Insect Technique

One alternative approach to using CRISPR/Cas9 to develop HEGs is to improve "first-generation" technologies with precise gene editing. Coupling the precision of CRISPR/Cas9 gene editing to enhance SIT has been proposed and developed as pgSIT (precision-guided sterile insect technique). This alternative approach to using CRISPR/Cas9 creates sterile males and kills or incapacitates females by targeting both a male fertility gene such as *beta2-tubulin* and elements of the sex determination pathway such as *dsx, tra* or *sxl*. RNA guides targeting *sxl* and *beta2-tubulin* coupled with a *Cas9* under a germline-specific promoter killed female embryos and produced sterile, male-only progeny in *Drosophila*. Targeted knock-out of *dsx and tra* resulted in intersex females (Kandul et al., 2019). Newer developments include a temperature inducible true-breeding strain that eliminates the requirement of maintenance and sexing of two independent parental strains (Cas9 and gRNA) (Kandul et al., 2021) Like suppression gene drives, the downstream, basal elements of the sex determination pathway are a reliable target for female non-viability. Versions of pgSIT have also been developed in mosquitoes (*Ae. aegypti*), and unlike a gene drive, this approach is self-limiting and is not predicted to persist or spread in the environment (Li et al., 2021).

THIRD-GENERATION TECHNOLOGIES: LOOKING TO THE FUTURE

Whereas first- and second-generation technologies seek to manipulate or disrupt basal elements of the sex determination pyramid in order to reduce the fitness of or kill females, future "third-generation" technologies may be designed to manipulate the master regulators of sexual fate, to affect full sex conversion. If attached to an efficient gene drive system, such a technology would spread through a target population causing a growing wave of sex distortion. This is theoretically more efficient than a second-generation system that kills or incapacitates one sex as homozygotes, because all inheritors, regardless of their genotype, continue to spread the system. This increased efficiency could potentially allow for a dramatic reduction in the number and size of releases required for population control. Depending on the efficiency of sex conversion, this could enable threshold-dependent gene drives to be used, previously discounted for population suppression because of their intolerance to high fitness loads (Leftwich et al., 2018; Champer et al., 2020a).

44

Many aspects of the theory underpinning such third-generation technologies pre-date second-generation strategies (Lyttle, 1991). However, they have proven challenging to enable in practical terms. Part of the reason is the non-conserved nature of the upstream components that need to be manipulated. This requires deep and specific knowledge of master regulators and their web of interactions with downstream elements for each specific pest species to be targeted. Even then, the unpredictable/inflexible nature of "coupled" sex determination and dosage compensation systems may make such a design unachievable in some species. Hence for third generation systems, the transfer of efficient gene drive designs between pests may not always be possible. In a similar vein, if sex-specific components of fitness are sex-linked it may be the case that efficient sex conversion can be achieved, but the sexual competitiveness of converted individuals is diminished.

The mechanisms for enacting sex conversion will vary greatly depending on whether the targeted gene(s) are the master regulators of sex determination (and whether these initiate a male or female cascade), or those genes directly downstream (*tra* or *tra-2*). In an XX/XY system, sex conversion through expression of a master regulator is likely to produce a dominant effect. This is highly likely to affect the dynamics of a gene drive. For example, the effects on population suppression would be seen much earlier than one where conversion is enacted through disruption of a recessive switch e.g., *tra* or *tra-2* (Hoshijima et al., 1991; Pane et al., 2002; Sarno et al., 2010). For systems where the presence of the master regulator determines femaleness [for example, the ZZ/ZW systems common to Lepidoptera (Suzuki et al., 2010)], maleness could be achieved by inactivating the master regulator or making the element below it resistant to its activity [see (Sakai et al., 2016)].

In the next section, we discuss evolutionary and empirical manipulation studies of dosage compensation and sex-linked fitness traits and outline the hurdles these may pose to engineering efficient third-generation technologies.

Dosage Compensation

In heterogametic sex chromosome systems, the loss of recombination between the dissimilar chromosomes leads to multiple evolutionary processes acting to reduce the size of the sex-limited chromosome, including mutation accumulation and gene loss (Bachtrog, 2013; Bachtrog et al., 2019). This can lead to a monoallelic state for the heterogametic sex, in which a single functional allele is present for multiple genes on the single copy of the X or Z chromosome, and the homogametic sex retains two functional copies. This imbalance of alleles between the sexes is often hypothesised to require dosage compensation mechanisms to restore a balanced state of gene expression: classically, this was thought to occur across the entirety of the X or Z chromosome (Ohno, 1967). If dosage compensation occurs across the entirety of sex chromosomes in a target pest species, it could prove challenging for the design of third-generation technologies, particularly if the dosage compensation pathway is downstream of the master regulator (i.e., the pathways are "coupled"). This is because, while such a system would ectopically express a sex determination master regulator, it would not alter the sex chromosome complement of

an individual. If the two pathways are "coupled", that individual (say, a female) would enact the dosage compensation pathway of the opposite sex (a male), despite having a "female" sex chromosome complement. This would lead to lethal misexpression of sex-linked genes and death, rather than conversion to the opposite sex.

Fortunately for GPM engineers, there is growing evidence from evolutionary studies of an alternative model of gene-by-gene dosage compensation. This alternative model states that only a minority of loci may be dosage-sensitive, specifically genes with particularly high expression levels, or those that have evolved through recent gene duplication. This may have a low correlation with levels of observed sex chromosome divergence (Furman et al., 2020). Where global dosage compensation is primarily observed is in XY systems, and could be driven by the stronger sexual selection and greater reproductive variance in males, this is predicted to result in slower evolution of Z than with X dosage compensation (Mullon et al., 2015). This could mean that sex chromosome dosage compensation may be less of a challenge in ZZ/ZW systems such as Lepidoptera (Gu et al., 2017).

In reality, the nature of dosage compensation appears to vary widely, and exceptions to "general" rules seem to be increasingly common. For example, *A. gambiae* has an XY heterogametic sex-determination system, with a single gene, *Yob,* identified as a Y-linked maleness factor (Krzywinska et al., 2016). The expression of *Yob* begins around 2 hours into embryonic development and precedes that of sex-specific splicing of *dsx* by about 6 hours. Ectopic expression of *Yob* has been confirmed to produce male splice-form *dsx* but leads to female embryonic death while leaving male development unaltered (Krzywinska and Krzywinski, 2018). This pattern of female lethality in the presence of *Yob* can be explained by gene overexpression by both X chromosomes as a result of misapplied dosage compensation leading to female death. Similar experiments in the mediterranean fruit fly (*C.capitata*), which also has an XY heterogametic sex-determination system, have also identified a Y-linked single gene determinant of maleness, *MoY*. Here knockdown of *MoY* was demonstrated to be sufficient to produce total loss of male-specific *tra* mRNA in embryos and complete XY feminisation. Conversely, ectopic expression of *MoY* produced no change in male development and partial or full masculinisation of XX flies (Meccariello et al., 2019). These XX pseudomales were also fertile, demonstrating that there are no genes *essential* to male fertility located on the Y chromosome in medfly. RNAi knockdown of *tra* in several other tephritid fruitflies and *M. domestica* have also produced female to male sex reversal, producing fertile converted males, indicating this approach may be possible in a number of pest insects (Dübendorfer et al., 2002; Pane et al., 2002; Lagos et al., 2007; Concha and Scott, 2009; Hediger et al., 2010; Schetelig et al., 2012) We note though the genetic factors that may influence the outcome of sperm competition have not yet been studied in these systems. In both *An. gambiae* and *C. capitata*; closely related species (*Anopheles arabiensis* with *Yob*; *Bactrocera oleae* with *MoY*), appear to be responsive to their respective male determining signals. However, the fast-evolving nature of the primary sex determination regulators means that these are likely to be restricted to closely related species with either direct homology to these genes or conserved downstream

interactions. The disparate responses of female death vs. female-male sex conversion between these two species were entirely unpredictable, and while the role of dosage compensation in this is speculative (and does not preclude that a dosage compensation mechanism exists that is uncoupled from sex determination); evidence to-date indicates these different fates for alteration of sexual development could be the result of just a handful of dosage-sensitive genes on the X chromosome of *An. gambiae* while none are present on the X chromosome of the medfly.

Essential Male Genes

One prediction of the evolution towards heterogametic sex chromosome systems is the accumulation of sex-specific fitness-enhancing genes on the sex-specific region, often linked to the master regulator through lack of recombination (Mank et al., 2014). As with dosage compensation, this arrangement may prove a hurdle for third-generation sex-conversion systems if these fitness-enhancing genes are not included alongside the master regulator. The Y chromosome of *D. melanogaster* contains male fertility factors. However, it contains only 16 protein-coding genes, and not all are essential for male fertility (Kaiser and Bachtrog, 2010; Zhang et al., 2020). Other examples of essential genes in males can be seen with the engineered manipulation of the male master regulator in *Ae. aegypti* and *Ae. albopictus*. These two mosquitos do not possess hetermorphic sex chromosomes, only a small, non-recombining male-determining region (M-locus) on the short arm of chromosome 1, an otherwise homomorphic autosome (Hall et al., 2014). A single gene *Nix* [a putative recent duplication of *tra2* (Gomulski et al., 2018)], has been identified as the master male determining gene in these species (Liu et al., 2019; Aryan et al., 2020). In *Ae. aegypti*, stable transgenic expression of *Nix* was sufficient to produce sex conversion of genotypic females into males. Dosage compensation in a species without a heterogametic sex would be unlikely, and the observed sex conversion over female death was in line with this prediction. However, while *Nix*

was sufficient for determining male sexual fate, the resultant pseudomales were incapable of flight as they lacked another gene *myo-sex*, also present in the M-locus, which is required for proper development of flight muscles in males (Aryan et al., 2020). Conversely, when analogous experiments were conducted in *Ae. albopictus* it was found that converted pseudomales were not only viable but capable of flight (Lutrat et al., 2022). Interestingly, despite evidence of an M-linked *myo-sex* gene, converted pseudomales could still express comparable levels of *myo-sex* transcripts to wildtype males. These results indicated at least one duplicate copy of *myo-sex* exists which is not M-linked in this species. These converted pseudomales, displayed reduced competitiveness compared to wildtype males, suggesting the possibility that the M-locus in this species may harbour other, as yet unknown, male fitness-enhancing genes. However, this is difficult to disentangle from the adverse effects of transgenesis (including ubiquitous marker gene expression, disruption of essential genes at the insertion sites or incomplete masculinization). These findings highlight that, even in species without apparent dosage compensation or heteromorphic sex chromosomes, efficient sex conversion technologies may prove more challenging to engineer than simply transgenically expressing the master regulator. Additionally, the significant differences between these two closely related species suggest that substantial fundamental research will be required to underpin the development of these technologies in novel pests.

CONCLUSION

Manipulating sex determination pathways for genetic pest management has many potential applications. Previous technologies have used the highly conserved "basal" elements of *dsx* and *tra* common to almost all insect species to produce reliable mechanisms of biasing sex ratios with the release of modified males carrying factors to generate female sterility or death. Newer

technological developments, including homing gene drives, demonstrate these basal elements continue to be predictable and reliable targets for control. Looking forward, development of genetic editing techniques to manipulate "master regulators" of sexual fate have the potential to improve the performance of a wide variety of genetic control methods. However, this approach has potential challenges—different species may exhibit sex-linked genes that are vital for viability or sexual fertility or have strong dosage compensation. However, this is a vibrant field of research and much experimental work is ongoing in a range of different pest species. While it is likely that the application of sex conversion for pest control will inevitably be applied on a case-by-case basis, active investigations on a number of fronts are likely to improve our understanding of the basic biology and evolution of sex determination, as well as expand our genetic toolbox for applied pest management.

AUTHOR CONTRIBUTIONS

AS, TH-S, and PL contributed to the conception of the manuscript. AS and TH-S wrote the first draft of the manuscript. PL and TC wrote sections of the manuscript and all authors contributed to editing and revisions, and approved the submitted version.

REFERENCES

Adolfi, M. C., Herpin, A., and Schartl, M. (2021). The Replaceable Master of Sex Determination: Bottom-Up Hypothesis Revisited. *Phil. Trans. R. Soc. B* 376, 20200090. doi:10.1098/rstb.2020.0090

Alcalay, Y., Fuchs, S., Galizi, R., Bernardini, F., Haghighat-Khah, R. E., Rusch, D. B., et al. (2021). The Potential for a Released Autosomal X-Shredder Becoming a Driving-Y Chromosome and Invasively Suppressing Wild Populations of Malaria Mosquitoes. *Front. Bioeng. Biotechnol.* 9. doi:10.3389/fbioe.2021.752253

Alphey, L. (2014). Genetic Control of Mosquitoes. *Annu. Rev. Entomol.* 59, 205–224. doi:10.1146/annurev-ento-011613-162002

Ant, T., Koukidou, M., Rempoulakis, P., Gong, H.-F., Economopoulos, A., Vontas, J., et al. (2012). Control of the Olive Fruit Fly Using Genetics-Enhanced Sterile Insect Technique. *BMC Biol.* 10, 51. doi:10.1186/1741-7007-10-51

Aryan, A., Anderson, M. A. E., Biedler, J. K., Qi, Y., Overcash, J. M., Naumenko, A. N., et al. (2020). Nix Alone Is Sufficient to Convert Female *Aedes aegypti* into Fertile Males and Myo-Sex Is Needed for Male Flight. *Proc. Natl. Acad. Sci. U.S.A.* 117, 17702–17709. doi:10.1073/pnas.2001132117

Atyame, C. M., Pasteur, N., Dumas, E., Tortosa, P., Tantely, M. L., Pocquet, N., et al. (2011). Cytoplasmic Incompatibility as a Means of Controlling *Culex pipiens* Quinquefasciatus Mosquito in the Islands of the South-Western Indian Ocean. *PLoS Negl. Trop. Dis.* 5, e1440. doi:10.1371/journal.pntd.0001440

Bachtrog, D., Mahajan, S., and Bracewell, R. (2019). Massive Gene Amplification on a Recently Formed Drosophila Y Chromosome. *Nat. Ecol. Evol.* 3, 1587–1597. doi:10.1038/s41559-019-1009-9

Bachtrog, D. (2013). Y-chromosome Evolution: Emerging Insights into Processes of Y-Chromosome Degeneration. *Nat. Rev. Genet.* 14, 113–124. doi:10.1038/nrg3366

Baker, B. S., and Ridge, K. A. (1980). Sex and the Single Cell. I. On the Action of Major Loci Affecting Sex Determination in *Drosophila melanogaster*. *Genetics* 94, 383–423. doi:10.1093/genetics/94.2.383

Bax, N. J., and Thresher, R. E. (2009). Ecological, Behavioral, and Genetic Factors Influencing the Recombinant Control of Invasive Pests. *Ecol. Appl.* 19, 873–888. doi:10.1890/07-1588.1

Bier, E. (2021). Gene Drives Gaining Speed. *Nat. Rev. Genet.* 23, 5–22. doi:10.1038/s41576-021-00386-0

Burt, A., and Crisanti, A. (2018). Gene Drive: Evolved and Synthetic. *ACS Chem. Biol.* 13, 343–346. doi:10.1021/acschembio.7b01031

Burt, A., and Deredec, A. (2018). Self-limiting Population Genetic Control with Sex-Linked Genome Editors. *Proc. R. Soc. B* 285, 20180776. doi:10.1098/rspb.2018.0776

Burt, A. (2003). Site-specific Selfish Genes as Tools for the Control and Genetic Engineering of Natural Populations. *Proc. R. Soc. Lond. B* 270, 921–928. doi:10.1098/rspb.2002.2319

Carrami, E. M., Eckermann, K. N., Ahmed, H. M. M., Sánchez C., H. M., Dippel, S., Marshall, J. M., et al. (2018). Consequences of Resistance Evolution in a Cas9-Based Sex Conversion-Suppression Gene Drive for Insect Pest Management. *Proc. Natl. Acad. Sci. U.S.A.* 115, 6189–6194. doi:10.1073/pnas.1713825115

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

Champer, J., Kim, I. K., Champer, S. E., Clark, A. G., and Messer, P. W. (2020a). Performance Analysis of Novel Toxin-Antidote CRISPR Gene Drive Systems. *BMC Biol.* 18, 27. doi:10.1186/s12915-020-0761-2

Champer, J., Liu, J., Oh, S. Y., Reeves, R., Luthra, A., Oakes, N., et al. (2018). Reducing Resistance Allele Formation in CRISPR Gene Drive. *Proc. Natl. Acad. Sci. U.S.A.* 115, 5522–5527. doi:10.1073/pnas.1720354115

Champer, J., Reeves, R., Oh, S. Y., Liu, C., Liu, J., Clark, A. G., et al. (2017). Novel CRISPR/Cas9 Gene Drive Constructs Reveal Insights into Mechanisms of Resistance Allele Formation and Drive Efficiency in Genetically Diverse Populations. *PLoS Genet.* 13, e1006796. doi:10.1371/journal.pgen.1006796

Champer, J., Yang, E., Lee, E., Liu, J., Clark, A. G., and Messer, P. W. (2020b). A CRISPR Homing Gene Drive Targeting a Haplolethal Gene Removes Resistance Alleles and Successfully Spreads through a Cage Population. *Proc. Natl. Acad. Sci. U.S.A.* 117, 24377–24383. doi:10.1073/pnas.2004373117

Champer, S. E., Oh, S. Y., Liu, C., Wen, Z., Clark, A. G., Messer, P. W., et al. (2020). Computational and Experimental Performance of CRISPR Homing Gene Drive Strategies with Multiplexed gRNAs. *Sci. Adv.* 6, eaaz0525. doi:10.1126/sciadv.aaz0525

Cho, S., Huang, Z. Y., and Zhang, J. (2007). Sex-specific Splicing of the Honeybee Doublesex Gene Reveals 300 Million Years of Evolution at the Bottom of the Insect Sex-Determination Pathway. *Genetics* 177, 1733–1741. doi:10.1534/genetics.107.078980

Cline, T. W. (1993). The Drosophila Sex Determination Signal: How Do Flies Count to Two? *Trends Genet.* 9, 385–390. doi:10.1016/0168-9525(93)90138-8

Collado, A. (2013). *Development of Strains and Procedures for Genetic Control of Aedes aegypti (Diptera: Culicidae)*. PhD Thesis. United Kingdom: University of Oxford

47

Concha, C., Palavesam, A., Guerrero, F. D., Sagel, A., Li, F., Osborne, J. A., et al. (2016). A Transgenic Male-Only Strain of the New World Screwworm for an Improved Control Program Using the Sterile Insect Technique. *BMC Biol.* 14, 72. doi:10.1186/s12915-016-0296-8

Concha, C., and Scott, M. J. (2009). Sexual Development in *Lucilia cuprina* (Diptera, Calliphoridae) Is Controlled by the Transformer Gene. *Genetics* 182, 785–798. doi:10.1534/genetics.109.100982

Criscione, F., Qi, Y., and Tu, Z. (2016). GUY1 Confers Complete Female Lethality and Is a Strong Candidate for a Male-Determining Factor in *Anopheles stephensi.* *eLife* 5. doi:10.7554/eLife.19281

De Barro, P. J., Murphy, B., Jansen, C. C., and Murray, J. (2011). The Proposed Release of the Yellow Fever Mosquito, *Aedes aegypti* Containing a Naturally Occurring Strain of Wolbachia Pipientis, a Question of Regulatory Responsibility. *J. Verbr. Leb.* 6, 33–40. doi:10.1007/s00003-011-0671-x

Deredec, A., Godfray, H. C. J., and Burt, A. (2011). Requirements for Effective Malaria Control with Homing Endonuclease Genes. *Proc. Natl. Acad. Sci. U.S.A.* 108, E874–E880. doi:10.1073/pnas.1110717108

Deutsch, C. A., Tewksbury, J. J., Tigchelaar, M., Battisti, D. S., Merrill, S. C., Huey, R. B., et al. (2018). Increase in Crop Losses to Insect Pests in a Warming Climate. *Science* 361, 916–919. doi:10.1126/science.aat3466

Dübendorfer, A., Hediger, M., Burghardt, G., and Bopp, D. (2002). *Musca domestica,* a Window on the Evolution of Sex-Determining Mechanisms in Insects. *Int. J. Dev. Biol.* 46, 75–79.

Fasulo, B., Meccariello, A., Morgan, M., Borufka, C., Papathanos, P. A., and Windbichler, N. (2020). A Fly Model Establishes Distinct Mechanisms for Synthetic CRISPR/Cas9 Sex Distorters. *PLoS Genet.* 16, e1008647. doi:10.1371/journal.pgen.1008647

Fu, G., Condon, K. C., Epton, M. J., Gong, P., Jin, L., Condon, G. C., et al. (2007). Female-specific Insect Lethality Engineered Using Alternative Splicing. *Nat. Biotechnol.* 25, 353–357. doi:10.1038/nbt1283

Furman, B. L. S., Metzger, D. C. H., Darolti, I., Wright, A. E., Sandkam, B. A., Almeida, P., et al. (2020). Sex Chromosome Evolution: So Many Exceptions to the Rules. *Genome Biol. Evol.* 12, 750–763. doi:10.1093/gbe/evaa081

Galizi, R., Doyle, L. A., Menichelli, M., Bernardini, F., Deredec, A., Burt, A., et al. (2014). A Synthetic Sex Ratio Distortion System for the Control of the Human Malaria Mosquito. *Nat. Commun.* 5, 3977. doi:10.1038/ncomms4977

Galizi, R., Hammond, A., Kyrou, K., Taxiarchi, C., Bernardini, F., O'Loughlin, S. M., et al. (2016). A CRISPR-Cas9 Sex-Ratio Distortion System for Genetic Control. *Sci. Rep.* 6, 31139. doi:10.1038/srep31139

Gamez, S., Chaverra-Rodriguez, D., Buchman, A., Kandul, N. P., Mendez-Sanchez, S. C., Bennett, J. B., et al. (2021). Exploiting a Y Chromosome-Linked Cas9 for Sex Selection and Gene Drive. *Nat. Commun.* 12, 7202. doi:10.1038/s41467-021-27333-1

Gantz, V. M., and Bier, E. (2015). The Mutagenic Chain Reaction: A Method for Converting Heterozygous to Homozygous Mutations. *Science* 348, 442–444. doi:10.1126/science.aaa5945

Gantz, V. M., Jasinskiene, N., Tatarenkova, O., Fazekas, A., Macias, V. M., Bier, E., et al. (2015). Highly Efficient Cas9-Mediated Gene Drive for Population Modification of the Malaria Vector Mosquito *Anopheles stephensi.* *Proc. Natl. Acad. Sci. U.S.A.* 112, E6736–E6743. doi:10.1073/pnas.1521077112

Gempe, T., and Beye, M. (2011). Function and Evolution of Sex Determination Mechanisms, Genes and Pathways in Insects. *Bioessays* 33, 52–60. doi:10.1002/bies.201000043

Gempe, T., Hasselmann, M., Schiøtt, M., Hause, G., Otte, M., and Beye, M. (2009). Sex Determination in Honeybees: Two Separate Mechanisms Induce and Maintain the Female Pathway. *PLoS Biol.* 7, e1000222. doi:10.1371/journal.pbio.1000222

Gomez-Zavaglia, A., Mejuto, J. C., and Simal-Gandara, J. (2020). Mitigation of Emerging Implications of Climate Change on Food Production Systems. *Food Res. Int.* 134, 109256. doi:10.1016/j.foodres.2020.109256

Gomulski, L. M., Mariconti, M., Di Cosimo, A., Scolari, F., Manni, M., Savini, G., et al. (2018). The Nix Locus on the Male-specific Homologue of Chromosome 1 in *Aedes albopictus* Is a Strong Candidate for a Male-Determining Factor. *Parasites Vectors* 11, 647. doi:10.1186/s13071-018-3215-8

Gregory, M. (2015). *The Genetic Control of Agricultural Pests (Plutella Xylostella and* Tribolium castaneum*).* PhD Thesis. UK: University of Oxford.

Gu, L., Walters, J. R., and Knipple, D. C. (2017). Conserved Patterns of Sex Chromosome Dosage Compensation in the Lepidoptera (WZ/ZZ): Insights

from a Moth Neo-Z Chromosome. *Genome Biol. Evol.* 9, 802–816. doi:10.1093/gbe/evx039

Hall, A. B., Basu, S., Jiang, X., Qi, Y., Timoshevskiy, V. A., Biedler, J. K., et al. (2015). A Male-Determining Factor in the Mosquito *Aedes aegypti.* *Science* 348, 1268–1270. doi:10.1126/science.aaa2850

Hall, A. B., Timoshevskiy, V. A., Sharakhova, M. V., Jiang, X., Basu, S., Anderson, M. A. E., et al. (2014). Insights into the Preservation of the Homomorphic Sex-Determining Chromosome of *Aedes aegypti* from the Discovery of a Male-Biased Gene Tightly Linked to the M-Locus. *Genome Biol. Evol.* 6, 179–191. doi:10.1093/gbe/evu002

Hammond, A., Galizi, R., Kyrou, K., Simoni, A., Siniscalchi, C., Katsanos, D., et al. (2016). A CRISPR-Cas9 Gene Drive System Targeting Female Reproduction in the Malaria Mosquito Vector *Anopheles gambiae.* *Nat. Biotechnol.* 34, 78–83. doi:10.1038/nbt.3439

Hammond, A., Karlsson, X., Morianou, I., Kyrou, K., Beaghton, A., Gribble, M., et al. (2021). Regulating the Expression of Gene Drives Is Key to Increasing Their Invasive Potential and the Mitigation of Resistance. *PLoS Genet.* 17, e1009321. doi:10.1371/journal.pgen.1009321

Harris, A. F., Nimmo, D., McKemey, A. R., Kelly, N., Scaife, S., Donnelly, C. A., et al. (2011). Field Performance of Engineered Male Mosquitoes. *Nat. Biotechnol.* 29, 1034–1037. doi:10.1038/nbt.2019

Hasselmann, M., Gempe, T., Schiøtt, M., Nunes-Silva, C. G., Otte, M., and Beye, M. (2008). Evidence for the Evolutionary Nascence of a Novel Sex Determination Pathway in Honeybees. *Nature* 454, 519–522. doi:10.1038/nature07052

Hawkins, N. J., Bass, C., Dixon, A., and Neve, P. (2018). The Evolutionary Origins of Pesticide Resistance. *Biol. Rev.* 94, 135–155. doi:10.1111/brv.12440

Hediger, M., Henggeler, C., Meier, N., Perez, R., Saccone, G., and Bopp, D. (2010). Molecular Characterization of the Key Switch F Provides a Basis for Understanding the Rapid Divergence of the Sex-Determining Pathway in the Housefly. *Genetics* 184, 155–170. doi:10.1534/genetics.109.109249

Hopkins, B. R., and Kopp, A. (2021). Evolution of Sexual Development and Sexual Dimorphism in Insects. *Curr. Opin. Genet. Dev.* 69, 129–139. doi:10.1016/j.gde.2021.02.011

Hoshijima, K., Inoue, K., Higuchi, I., Sakamoto, H., and Shimura, Y. (1991). Control of Doublesex Alternative Splicing by Transformer and Transformer-2 in Drosophila. *Science* 252, 833–836. doi:10.1126/science.1902987

Inoue, H., and Hiroyoshi, T. (1986). A MATERNAL-EFFECT SEX-TRANSFORMATION MUTANT OF THE HOUSEFLY, MUSCA DOMESTICA L. *MUSCA Domest. L. Genet.* 112, 469–482. doi:10.1093/genetics/112.3.469

Inoue, K., Hoshijima, K., Sakamoto, H., and Shimura, Y. (1990). Binding of the Drosophila Sex-Lethal Gene Product to the Alternative Splice Site of Transformer Primary Transcript. *Nature* 344, 461–463. doi:10.1038/344461a0

Jin, L., Walker, A. S., Fu, G., Harvey-Samuel, T., Dafa'alla, T., Miles, A., et al. (2013). Engineered Female-specific Lethality for Control of Pest Lepidoptera. *ACS Synth. Biol.* 2, 160–166. doi:10.1021/sb300123m

Kaiser, V. B., and Bachtrog, D. (2010). Evolution of Sex Chromosomes in Insects. *Annu. Rev. Genet.* 44, 91–112. doi:10.1146/annurev-genet-102209-163600

Kandul, N. P., Liu, J., and Akbari, O. S. (2021). Temperature-Inducible Precision-Guided Sterile Insect Technique. *CRISPR J.* 4, 822–835. doi:10.1089/crispr.2021.0077

Kandul, N. P., Liu, J., Sanchez C., H. M., Wu, S. L., Marshall, J. M., and Akbari, O. S. (2019). Transforming Insect Population Control with Precision Guided Sterile Males with Demonstration in Flies. *Nat. Commun.* 10, 84. doi:10.1038/s41467-018-07964-7

Katsuma, S., Kiuchi, T., Kawamoto, M., Fujimoto, T., and Sahara, K. (2018). Unique Sex Determination System in the Silkworm, *Bombyx mori*: Current Status and beyond. *Proc. Jpn. Acad. Ser. B Phys. Biol. Sci.* 94, 205–216. doi:10.2183/pjab.94.014

Klassen, W. (2005). "Area-Wide Integrated Pest Management and the Sterile Insect Technique," in *Sterile Insect Technique. Principles and Practice in Area-wide Integrated Pest Management.* Editors G. Franz, V. A. Dyck, J. Hendrichs, and A. S. Robinson (Vienna, Austria: Springer), 427.

Krzywinska, E., Dennison, N. J., Lycett, G. J., and Krzywinski, J. (2016). A Maleness Gene in the Malaria Mosquito *Anopheles gambiae.* *Science* 353, 67–69. doi:10.1126/science.aaf5605

Krzywinska, E., and Krzywinski, J. (2018). Effects of Stable Ectopic Expression of the Primary Sex Determination Gene Yob in the Mosquito *Anopheles gambiae.* *Parasites Vectors* 11, 648. doi:10.1186/s13071-018-3211-z

48

Kyrou, K., Hammond, A. M., Galizi, R., Kranjc, N., Burt, A., Beaghton, A. K., et al. (2018). A CRISPR-Cas9 Gene Drive Targeting Doublesex Causes Complete Population Suppression in Caged *Anopheles gambiae* Mosquitoes. *Nat. Biotechnol.* 36, 1062–1066. doi:10.1038/nbt.4245

Lagos, D., Koukidou, M., Savakis, C., and Komitopoulou, K. (2007). The Transformer Gene in Bactrocera Oleae: the Genetic Switch that Determines its Sex Fate. *Insect Mol. Biol.* 16, 221–230. doi:10.1111/j.1365-2583.2006.00717.x

Lee, J., Kiuchi, T., Kawamoto, M., Shimada, T., and Katsuma, S. (2015). Identification and Functional Analysis of aMasculinizerorthologue inTrilocha varians(Lepidoptera: Bombycidae). *Insect Mol. Biol.* 24, 561–569. doi:10.1111/imb.12181

Leftwich, P. T., Bolton, M., and Chapman, T. (2016). Evolutionary Biology and Genetic Techniques for Insect Control. *Evol. Appl.* 9, 212–230. doi:10.1111/eva.12280

Leftwich, P. T., Edgington, M. P., Harvey-Samuel, T., Carabajal Paladino, L. Z., Norman, V. C., and Alphey, L. (2018). Recent Advances in Threshold-dependent Gene Drives for Mosquitoes. *Biochem. Soc. Trans.* 46, 1203–1212. doi:10.1042/BST20180076

Leftwich, P. T., Koukidou, M., Rempoulakis, P., Gong, H.-F., Zacharopoulou, A., Fu, G., et al. (2014). Genetic Elimination of Field-Cage Populations of Mediterranean Fruit Flies. *Proc. R. Soc. B* 281, 20141372. doi:10.1098/rspb.2014.1372

Leftwich, P. T., Spurgin, L. G., Harvey-Samuel, T., Thomas, C. J. E., Paladino, L. C., Edgington, M. P., et al. (2021). Genetic Pest Management and the Background Genetics of Release Strains. *Phil. Trans. R. Soc. B* 376, 20190805. doi:10.1098/rstb.2019.0805

Li, M., Yang, T., Bui, M., Gamez, S., Wise, T., Kandul, N. P., et al. (2021). Suppressing Mosquito Populations with Precision Guided Sterile Males. *Nat. Commun.* 12, 5374. doi:10.1038/s41467-021-25421-w

Li, M., Yang, T., Kandul, N. P., Bui, M., Gamez, S., Raban, R., et al. (2020). Development of a Confinable Gene Drive System in the Human Disease Vector *Aedes aegypti*. *eLife* 9. doi:10.7554/eLife.51701

Liu, P., Jin, B., Li, X., Zhao, Y., Gu, J., Biedler, J. K., et al. (2020). Nix Is a Male-Determining Factor in the Asian Tiger Mosquito *Aedes albopictus*. *Insect Biochem. Mol. Biol.* 118, 103311. doi:10.1016/j.ibmb.2019.103311

López Del Amo, V., Bishop, A. L., Sánchez C., H. M., Bennett, J. B., Feng, X., Marshall, J. M., et al. (2020). A Transcomplementing Gene Drive Provides a Flexible Platform for Laboratory Investigation and Potential Field Deployment. *Nat. Commun.* 11, 352. doi:10.1038/s41467-019-13977-7

Lutrat, C., Olmo, R. P., Baldet, T., Bouyer, J., and Marois, E. (2022). Transgenic Expression of Nix Converts Genetic Females into Males and Allows Automated Sex Sorting in *Aedes albopictus*. *Commun. Biol.* 5, 210. doi:10.1038/s42003-022-03165-7

Lyttle, T. W. (1991). Segregation Distorters. *Annu. Rev. Genet.* 25, 511–581. doi:10.1146/annurev.ge.25.120191.002455

Mank, J. E., Hosken, D. J., and Wedell, N. (2014). Conflict on the Sex Chromosomes: Cause, Effect, and Complexity. *Cold Spring Harb. Perspect. Biol.* a017715. doi:10.1101/cshperspect.a017715

Marshall, J. M., Buchman, A., Sánchez C., H. M., and Akbari, O. S. (2017). Overcoming Evolved Resistance to Population-Suppressing Homing-Based Gene Drives. *Sci. Rep.* 7, 3776. doi:10.1038/s41598-017-02744-7

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

Meccariello, A., Krsticevic, F., Colonna, R., Del Corsano, G., Fasulo, B., Papathanos, P. A., et al. (2021). Engineered Sex Ratio Distortion by X-Shredding in the Global Agricultural Pest Ceratitis Capitata. *BMC Biol.* 19, 78. doi:10.1186/s12915-021-01010-7

Meccariello, A., Salvemini, M., Primo, P., Hall, B., Koskinioti, P., Dalíková, M., et al. (2019). Maleness-on-the-Y (MoY) Orchestrates Male Sex Determination in Major Agricultural Fruit Fly Pests. *Science* 365, 1457–1460. doi:10.1126/science.aax1318

Meise, M., Hilfiker-Kleiner, D., Dübendorfer, A., Brunner, C., Nöthiger, R., and Bopp, D. (1998). Sex-lethal, the Master Sex-Determining Gene in Drosophila, Is Not Sex-Specifically Regulated in *Musca domestica*. *Development* 125, 1487–1494. doi:10.1242/dev.125.8.1487

Miyazaki, S., Fujiwara, K., Kai, K., Masuoka, Y., Gotoh, H., Niimi, T., et al. (2021). Evolutionary Transition of Doublesex Regulation from Sex-specific Splicing to Male-specific Transcription in Termites. *Sci. Rep.* 11, 15992. doi:10.1038/s41598-021-95423-7

Morrison, N. I., Simmons, G. S., Fu, G., O'Connell, S., Walker, A. S., Dafa'alla, T., et al. (2012). Engineered Repressible Lethality for Controlling the Pink Bollworm, a Lepidopteran Pest of Cotton. *PLoS ONE* 7, e50922. doi:10.1371/journal.pone.0050922

Mullon, C., Wright, A. E., Reuter, M., Pomiankowski, A., and Mank, J. E. (2015). Evolution of Dosage Compensation under Sexual Selection Differs between X and Z Chromosomes. *Nat. Commun.* 6, 7720. doi:10.1038/ncomms8720

Nene, V., Wortman, J. R., Lawson, D., Haas, B., Kodira, C., Tu, Z., et al. (2007). Genome Sequence of *Aedes aegypti* , a Major Arbovirus Vector. *Science* 316, 1718–1723. doi:10.1126/science.1138878

Nguyen, T. N. M., Choo, A., and Baxter, S. W. (2021). Lessons from Drosophila: Engineering Genetic Sexing Strains with Temperature-Sensitive Lethality for Sterile Insect Technique Applications. *Insects* 12, 243. doi:10.3390/insects12030243

Oberhofer, G., Ivy, T., and Hay, B. A. (2018). Behavior of Homing Endonuclease Gene Drives Targeting Genes Required for Viability or Female Fertility with Multiplexed Guide RNAs. *Proc. Natl. Acad. Sci. U.S.A.* 115, E9343–E9352. doi:10.1073/pnas.1805278115

Ogaugwu, C. E., Schetelig, M. F., and Wimmer, E. A. (2013). Transgenic Sexing System for Ceratitis Capitata (Diptera: Tephritidae) Based on Female-specific Embryonic Lethality. *Insect Biochem. Mol. Biol.* 43, 1–8. doi:10.1016/j.ibmb.2012.10.010

Ohno, S. (1967). *Sex Chromosomes and Sex-Linked Genes*. Berlin, Heidelberg: Springer Berlin Heidelberg. doi:10.1007/978-3-642-88178-7

Pane, A., Salvemini, M., Bovi, P. D., Polito, C., and Saccone, G. (2002). Thetransformergene inCeratitis Capitataprovides a Genetic Basis for Selecting and Remembering the Sexual Fate. *Development* 129, 3715–3725. doi:10.1242/dev.129.15.3715

Parkhurst, S. M., Bopp, D., and Ish-Horowicz, D. (1990). X:A Ratio, the Primary Sex-Determining Signal in Drosophila, Is Transduced by Helix-Loop-Helix Proteins. *Cell* 63, 1179–1191. doi:10.1016/0092-8674(90)90414-a

Phuc, H. K., Andreasen, M. H., Burton, R. S., Vass, C., Epton, M. J., Pape, G., et al. (2007). Late-acting Dominant Lethal Genetic Systems and Mosquito Control. *BMC Biol.* 5, 11. doi:10.1186/1741-7007-5-11

Price, D. C., Egizi, A., and Fonseca, D. M. (2015). The Ubiquity and Ancestry of Insect Doublesex. *Sci. Rep.* 5, 13068. doi:10.1038/srep13068

Qi, Y., Wu, Y., Saunders, R., Chen, X.-G., Mao, C., Biedler, J. K., et al. (2019). Guy1, a Y-Linked Embryonic Signal, Regulates Dosage Compensation in *Anopheles stephensi* by Increasing X Gene Expression. *eLife* 8. doi:10.7554/eLife.43570

Raban, R. R., Marshall, J. M., and Akbari, O. S. (2020). Progress towards Engineering Gene Drives for Population Control. *J. Exp. Biol.* 223. doi:10.1242/jeb.208181

Rendon, P. A. (1996). *Development and Evaluation of a Temperature Sensitive Lethal (TSL) Genetic Sexing Strain of the Mediterranean Fruit Fly*. United Kingdom: The University of Manchester.

Rendón, P., McInnis, D., Lance, D., and Stewart, J. (2004). Medfly (Diptera: Tephritidae) Genetic Sexing: Large-Scale Field Comparison of Males-Only and Bisexual Sterile Fly Releases in Guatemala. *ec* 97, 1547–1553. doi:10.1603/0022-0493-97.5.1547

Robinson, A. S. (2002). Genetic Sexing Strains in Medfly, Ceratitis Capitata, Sterile Insect Technique Programmes. *Genetica* 116, 5–13. doi:10.1023/a:1020951407069

Ruiz, M. F., Milano, A., Salvemini, M., Eirín-López, J. M., Perondini, A. L. P., Selivon, D., et al. (2007). The Gene Transformer of Anastrepha Fruit Flies (Diptera, Tephritidae) and its Evolution in Insects. *PLoS ONE* 2, e1239. doi:10.1371/journal.pone.0001239

Saccone, G., Pane, A., and Polito, L. C. (2002). Sex Determination in Flies, Fruitflies and Butterflies. *Genetica* 116, 15–23. doi:10.1023/a:1020903523907

Saccone, G., Salvemini, M., and Polito, L. C. (2011). The Transformer Gene of Ceratitis Capitata: a Paradigm for a Conserved Epigenetic Master Regulator of Sex Determination in Insects. *Genetica* 139, 99–111. doi:10.1007/s10709-010-9503-7

Sakai, H., Sumitani, M., Chikami, Y., Yahata, K., Uchino, K., Kiuchi, T., et al. (2016). Transgenic Expression of the piRNA-Resistant Masculinizer Gene Induces Female-specific Lethality and Partial Female-To-Male Sex Reversal

49

in the Silkworm, *Bombyx mori. PLoS Genet.* 12, e1006203. doi:10.1371/journal. pgen.1006203

Salvemini, M., Robertson, M., Aronson, B., Atkinson, P., Polito, L. C., and Saccone, G. (2009). Ceratitis Capitata Transformer-2 Gene Is Required to Establish and Maintain the Autoregulation of Cctra, the Master Gene for Female Sex Determination. *Int. J. Dev. Biol.* 53, 109–120. doi:10.1387/ijdb.082681ms

Sánchez, L., and Nöthiger, R. (1982). Clonal Analysis Ofsex-Lethal, a Gene Needed for Female Sexual Development inDrosophila Melanogaster. *Wilhelm Roux' Arch.* 191, 211–214. doi:10.1007/BF00848339

Sarno, F., Ruiz, M. F., Eirín-López, J. M., Perondini, A. L., Selivon, D., and Sánchez, L. (2010). The Gene Transformer-2 of Anastrephafruit Flies (Diptera, Tephritidae) and its Evolution in Insects. *BMC Evol. Biol.* 10, 140. doi:10. 1186/1471-2148-10-140

Schetelig, M. F., and Handler, A. M. (2012). A Transgenic Embryonic Sexing System for Anastrepha Suspensa (Diptera: Tephritidae). *Insect Biochem. Mol. Biol.* 42, 790–795. doi:10.1016/j.ibmb.2012.07.007

Schetelig, M. F., Milano, A., Saccone, G., and Handler, A. M. (2012). Male Only Progeny in Anastrepha Suspensa by RNAi-Induced Sex Reversion of Chromosomal Females. *Insect Biochem. Mol. Biol.* 42, 51–57. doi:10.1016/j. ibmb.2011.10.007

Schetelig, M. F., Targovska, A., Meza, J. S., Bourtzis, K., and Handler, A. M. (2016). Tetracycline-suppressible Female Lethality and Sterility in the Mexican Fruit fly,Anastrepha Ludens. *Insect Mol. Biol.* 25, 500–508. doi:10.1111/imb.12238

Sharma, A., Heinze, S. D., Wu, Y., Kohlbrenner, T., Morilla, I., Brunner, C., et al. (2017). Male Sex in Houseflies Is Determined by Mdmd , a Paralog of the Generic Splice Factor Gene CWC22. *Science* 356, 642–645. doi:10.1126/science.aam5498

Shelton, A. M., Long, S. J., Walker, A. S., Bolton, M., Collins, H. L., Revuelta, L., et al. (2020). First Field Release of a Genetically Engineered, Self-Limiting Agricultural Pest Insect: Evaluating its Potential for Future Crop Protection. *Front. Bioeng. Biotechnol.* 7. doi:10.3389/fbioe.2019.00482

Shukla, J. N., and Palli, S. R. (2012). Sex Determination in Beetles: Production of All Male Progeny by Parental RNAi Knockdown of Transformer. *Sci. Rep.* 2, 602. doi:10.1038/srep00602

Simoni, A., Hammond, A. M., Beaghton, A. K., Galizi, R., Taxiarchi, C., Kyrou, K., et al. (2020). A Male-Biased Sex-Distorter Gene Drive for the Human Malaria Vector *Anopheles gambiae. Nat. Biotechnol.* 38, 1054–1060. doi:10.1038/ s41587-020-0508-1

Sosnowski, B. A., Belote, J. M., and McKeown, M. (1989). Sex-specific Alternative Splicing of RNA from the Transformer Gene Results from Sequence-dependent Splice Site Blockage. *Cell* 58, 449–459. doi:10.1016/0092-8674(89)90426-1

Sultana, S., Baumgartner, J. B., Dominiak, B. C., Royer, J. E., and Beaumont, L. J. (2020). Impacts of Climate Change on High Priority Fruit Fly Species in Australia. *PLoS ONE* 15, e0213820. doi:10.1371/journal.pone.0213820

Suzuki, M. G., Imanishi, S., Dohmae, N., Asanuma, M., and Matsumoto, S. (2010). Identification of a Male-specific RNA Binding Protein that Regulates Sex-specific Splicing of Bmdsx by Increasing RNA Binding Activity of BmPSI. *Mol. Cell. Biol.* 30, 5776–5786. doi:10.1128/MCB.00444-10

Tan, A., Fu, G., Jin, L., Guo, Q., Li, Z., Niu, B., et al. (2013). Transgene-based, Female-specific Lethality System for Genetic Sexing of the Silkworm, *Bombyx mori. Proc. Natl. Acad. Sci. U.S.A.* 110, 6766–6770. doi:10.1073/pnas.1221700110

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

Traut, W., Sahara, K., and Marec, F. (2007). Sex Chromosomes and Sex Determination in Lepidoptera. *Sex. Dev.* 1, 332–346. doi:10.1159/000111765

Verhulst, E. C., Beukeboom, L. W., and van de Zande, L. (2010a). Maternal Control of Haplodiploid Sex Determination in the Wasp Nasonia. *Science* 328, 620–623. doi:10.1126/science.1185805

Verhulst, E. C., van de Zande, L., and Beukeboom, L. W. (2010b). Insect Sex Determination: it All Evolves Around Transformer. *Curr. Opin. Genet. Dev.* 20, 376–383. doi:10.1016/j.gde.2010.05.001

Verhulst, E. C., and van de Zande, L. (2015). Double Nexus--Doublesex Is the Connecting Element in Sex Determination. *Briefings Funct. Genomics* 14, 396–406. doi:10.1093/bfgp/elv005

Visser, S., Voleníková, A., Nguyen, P., Verhulst, E. C., and Marec, F. (2021). A Conserved Role of the Duplicated Masculinizer Gene in Sex Determination of the Mediterranean Flour Moth, Ephestia Kuehniella. *PLoS Genet.* 17, e1009420. doi:10.1371/journal.pgen.1009420

Wexler, J., Delaney, E. K., Belles, X., Schal, C., Wada-Katsumata, A., Amicucci, M. J., et al. (2019). Hemimetabolous Insects Elucidate the Origin of Sexual Development via Alternative Splicing. *eLife* 8. doi:10.7554/eLife.47490

Windbichler, N., Papathanos, P. A., Catteruccia, F., Ranson, H., Burt, A., and Crisanti, A. (2007). Homing Endonuclease Mediated Gene Targeting in *Anopheles gambiae* Cells and Embryos. *Nucleic Acids Res.* 35, 5922–5933. doi:10.1093/nar/gkm632

Xu, J., Chen, S., Zeng, B., James, A. A., Tan, A., and Huang, Y. (2017). *Bombyx mori* P-Element Somatic Inhibitor (BmPSI) Is a Key Auxiliary Factor for Silkworm Male Sex Determination. *PLoS Genet.* 13, e1006576. doi:10.1371/journal.pgen. 1006576

Yan, Y., and Scott, M. J. (2020). Building a Transgenic Sexing Strain for Genetic Control of the Australian Sheep Blow Fly *Lucilia cuprina* Using Two Lethal Effectors. *BMC Genet.* 21, 141. doi:10.1186/s12863-020-00947-y

Zhang, J., Luo, J., Chen, J., Dai, J., and Montell, C. (2020). The Role of Y Chromosome Genes in Male Fertility in *Drosophila melanogaster. Genetics* 215, 623–633. doi:10.1534/genetics.120.303324

Zheng, X., Zhang, D., Li, Y., Yang, C., Wu, Y., Liang, X., et al. (2019). Incompatible and Sterile Insect Techniques Combined Eliminate Mosquitoes. *Nature* 572, 56–61. doi:10.1038/s41586-019-1407-9

Zou, Y., Geuverink, E., Beukeboom, L. W., Verhulst, E. C., and van de Zande, L. (2020). A Chimeric Gene Paternally Instructs Female Sex Determination in the Haplodiploid Wasp Nasonia. *Science* 370, 1115–1118. doi:10.1126/science. abb8949

50

2 Chapter Two- Engineering controlled Cas9 expression in *Ceratitis capitata*

2.1 Abstract

The Mediterranean fruit fly (medfly, *Ceratitis capitata*) is an agricultural pest of global significance and a model system of growing importance. It has been studied in the context of research into pest control strategies, sex determination mechanisms and evolutionary ecology. However, research into this model, especially in terms of the development of strains for pest control is currently severely hampered by the lack of a genetic toolbox. For example, key tools and reagents for the development of genetic sorting strains and gene drives are sorely needed. The aim of the experiments described in this chapter was to address an important gap in this area by constructing and characterising transgenic lines of medfly that express Cas9. This is a valuable goal because stable Cas9 expression is necessary for the latest genetic pest control strategies and for studies into gene function in general. I designed lines using endogenous promotors as these result in more stable Cas9 expression, with germline specific promotors being favoured to allow for precise timing of expression. Two different promotor-Cas9 expressing lines were established: *mei-W68*-Cas9 and *bgcn*-Cas9 (*Benign gonial cell neoplasm)* and I describe the selection, testing and assembly of these promotors. I quantified the Cas9 expression patterns of the resulting lines by using qPCR across several tissue types. Both lines showed Cas9 expression and expression was higher overall in *mei-W68*-driven lines across all tissue types. However, further validation through qPCR with more stable reference genes would be needed to confirm if this is a true difference. None of the lines showed evidence of a fitness cost to carrying the transgene through the initial fitness study that was conducted. The lines generated provide an important expansion of the genetic toolbox of genetic modification for medfly, as they represent the first described stable Cas9 expressing lines under the control of endogenous germline specific promotors.

2.2 Introduction

Ceratitis capitata (commonly known as medfly) is a pest species of global agricultural importance. With the warming climate, the invasive potential of medfly is predicted to increase (Gutierrez et al. 2021; Sultana et al. 2020; De Meyer et al. 2007). Due to its increasing prevalence as an invasive pest, further methods of pest control are needed. Methods of genetic control are some of the most promising strategies, following the success of sterile insect technique (SIT) for localised eradication. The development of genetic sexing strains (GSS) (Robinson 2002) further improved the efficiency of SIT-based eradication methods and they have been successfully used for population suppression (Hendrichs et al. 2002). Despite the success seen with SIT there is a constant need for improvement in genetic control strategies. SIT is costly and relies upon mass rearing of males and continuous releases. Although genetic sexing strains have sped up the sex-sorting process there are still challenges with the often low competitiveness of irradiated males (Virginio et al. 2017) resulting in high release numbers being needed to outcompete wildtype males, which must be reared and maintained under very specific, often costly, conditions (Caceres 2002).

Strategies based upon CRISPR/Cas9 represent among the most promising new methods for genetic control in medfly. Not only can this system be used to develop new transgenics for improved genetic sexing strains (Nguyen et al. 2021) but also to develop more durable methods of control such as gene drives (McFarlane et al. 2018). Endogenous promotor-driven Cas9 had been identified as a necessity for gene drive designs in medfly (Carrami et al. 2018). Despite this, when beginning this research there were no published examples of Cas9 expressing lines under a native promotor in medfly.

In previous research where genetic modification of medfly was needed, methods based upon RNA interference (RNAi) (Saccone et al. 2007) or the co-injection of Cas9 either as protein or a Cas9-expressing plasmid were used (Aumann et al. 2018; Primo et al. 2020; Meccariello et al. 2017). Establishing a line with stable Cas9 expression is desirable. For guide validation, stable Cas9 lines are the most reliable for confirming efficacy (Kouranova et al. 2016) as they provide a constant expression level and can be used effectively for comparison between guides. Cas9-expressing lines also provide a continuous expression of Cas9 that persists throughout generations as opposed to the transient expression achieved by co-injection of Cas9.

Genetic technologies have a variety of requirements with respect to Cas9 promotors. The timing of Cas9 expression is an important factor to consider when designing lines. Using an endogenous promotor to drive the expression of Cas9 has been shown to result in higher editing efficiencies and lower rates of chimerism than observed with non-endogenous promotors (Wang et al. 2015; Mao et al. 2016; Lei et al. 2021; Feng et al. 2021). Expression in the germline is a necessity for Cas9 promotors as they drive the early expression in germ cells which is needed to established stable mutations as opposed to chimeras. Cas9 expression in somatic tissue causes cell specific mutations, resulting in chimerism. The promotors from germline specific genes are often used to drive Cas9 expression, as these exhibit high expression in the germline and reduced expression in other tissue types.

Depending upon the intended use of the Cas9 promotor, expression outside of the germline may or may not be tolerated. For use in homing drives and drive types in which resistance allele formation is an issue, Cas9 expression must be tightly controlled and germline-restricted. This is because, outside of the germline, homology directed repair is much less likely to occur than non-homologous end joining, and the latter often results in the formation of resistance alleles (Raban et al. 2020). However, complete restricted expression within the germline may be rare, and even if expression is highest in the germline, some somatic expression may often still be present. Somatic expression and deposition of Cas9 often results in fitness costs, with somatic expression of Cas9 having the strongest deleterious effects (Port et al. 2014; Hammond et al. 2021; Hammond et al. 2016).

Cas9 deposition into eggs is also relatively common and has been observed in studies based on several candidate germline-expressed gene promoters. This can represent a problem for some Cas9-based strategies because Cas9 is a relatively stable protein. In homing drives, Cas9 deposition into the eggs often does not increase drive inheritance, rather, it results in an increase in resistance allele formation (Champer et al. 2017; Champer et al. 2018; Gantz et al. 2015). However, when used for toxin/antidote systems, Cas9 deposition and even somatic expression can be tolerated (Champer et al. 2021). This is because the constructs developed don't rely on the homing pathway, meaning that resistance allele formation can be slowed through the use of essential genes. High levels of Cas9 deposition can be desirable when testing guide RNAs through injection as this is more likely to have a dominant knockout effect in the first generation, making phenotypic studies easier to complete. Toxin/antidote systems can utilise Cas9 deposition to achieve the accumulation of broken alleles, therefore driving the need for the rescue drive carrying element in the population as it becomes the antidote

is more likely to be required (Champer et al. 2020). Cas9 deposition could also be used in the development of precision-guided SIT (pgSIT) (M. Li et al. 2021) as these strategies could utilise a dominant knockout to target a male reproductive factor, avoiding the fitness damaging irradiation traditionally used in SIT.

I focused in the work described here on germline promotors, as they have a wider range of applications in Cas9-based strategies than do ubiquitous native promotors. There are several candidates for germline promotors which have been tested in other model organisms. The transferability of promotors between species has proven difficult (Du et al. 2023). Promotors have been shown to have efficient levels of cutting in one species, yet, when the same native gene is used in another species, similar results are often not achieved. Although it may frustrate control efforts whenever a functional Cas9 promotor is found in one species that does not transfer to another, it does mean that promotors which have not previously been successful may often work in a new species.

Functional Cas9 expression can be affected by other factors in addition to the promotor used. For example, the 3' untranslated region (3' UTR) has been shown to have an effect on transgene expression. The use of non-native 3' UTR such as *Autographa californica* nucleopolyhedrovirus (AcNPV) *p10* gene (Smith et al. 1983) or simian virus 40 (SV40) (Van Oers et al. 1999) to increase transgene expression is common. However, the use of native 3'UTRs can result in lower levels of mRNA degradation (Garneau et al. 2007) and, from studies in nematodes, may be more critical in achieving gene expression in germline tissues than the promotor region itself (Merritt et al. 2008). The insertion site of the Cas9 construct can also have an impact on the expression levels of Cas9. Insertion into a transcriptionally inactive section of the genome is likely to result in downregulation of the Cas9. The use of *piggybac* (X. Li et al. 2013) elements results in a random insertion location of transgenic constructs and this must be considered if there is evidence of an insertion event having lower expression than that of other lines.

This study aimed to further expand the genetic toolbox for *Ceratitis capitata* with the production of stable Cas9 lines under germline promotors. A well annotated genome is often a necessity in selecting potential Cas9 promotors. Fortunately, medfly does have a fully annotated genome, which allows for the expansion of genetic tools in this model organism (Papanicolaou et al. 2016). Several potential germline promotors were identified in this chapter, with the expression across tissues measured for each. Genes with appropriate expression for use as a Cas9 promotor were selected with the regulatory region to be used

as a promotor and 5'UTR identified and extracted where possible. These were constructed with non-native 3'UTR and integrated into the medfly genome through *piggybac*-mediated, randomised genomic insertion. Once insertion was confirmed, the Cas9 expression across tissue types was determined through qPCR. These Cas9 expressing lines were tested for fitness costs in a small-scale fitness study to identify any strong fitness deficits in comparison to the wildtypes.

2.3 Methods

2.3.1 Fly Husbandry

2.3.1.1 Medfly stock lines

Three wildtype medfly stock lines were used throughout these experiments. 'Cepa Petapa' is a mass reared wildtype strain originally collected from the wild in Guatemala, Central America (Rendon, 1996). TOLIMAN is a wildtype strain from Guatemala, Central America originally collected in 1990. Both strains have been reared at Oxitec LTD (Milton Park, Abingdon) from 2004 and a sub-culture of each was established at the University of East Anglia (UEA) in 2010. The wild-type Benakeion strain (originally isolated in Athens, Greece) was provided by A. Meccariello (Imperial College, London) and maintained at UEA from 2022.

2.3.1.2 Medfly rearing and dissections for qPCR or end point RT-PCR/RACE

All wild type medfly cultures were reared in a controlled environment with temperature set at 25°C ±0.5°C, humidity at 60% ± 10% RH and on a 12 h light/dark schedule. Depending upon the demand for eggs from cultures, adult populations were kept in 1 of 3 different cage sizes: small cages (10cm x 10cm x 10cm) initiated with 50 pupae, medium cages (13cm x 13cm x 14cm) with 100 pupae and large cages (33cm x 33cm x 16cm) with 500 pupae. The emerging adult flies held in these cages were fed on an *ad libitum* diet of 3:1 sucrose:hydrolysed yeast mixture, changed every seven days. Water was supplied through the side of each cage via dental rolls soaked in RO water, which were renewed every 4 days.

All cages were designed with a mesh layer covering the majority of the surface area of one vertical wall of the cage to allow egg deposition. Water pots containing RO water were placed under the mesh side of each cage to collect eggs. Eggs laid through the cage mesh were collected after a period of no more than 24 hours of egg laying, by removing eggs with a Pasteur pipette and transferring up to two drops of the eggs contained in the water solution to filter paper. The filter papers with eggs were then placed into 1/3 pint milk bottles containing 100ml of ASG larval food (1L ASG food: 850ml RO water, 12.5g agar, 73.5g sucrose, 67g maize, 47.5g Brewer's Yeast, 2ml Propionic acid, 25ml Nipagin). 7 days post egg collection, bottles were prepared to allow the third instar larvae to exit and start pupation. To do this, 2 filter paper folded strips were placed in each bottle before laying the bottles down on a thin layer of sand within a pupation box (170mm x 130mm x 50mm). The box was

then sealed with a lid containing a mesh for a further 7-9 days. During this time third instar 'jumping' larvae exited the bottles and pupated within the sand. After this period, pupae were sieved from the sand through a standard metal sieve and transferred into petri dishes. From this the appropriate number of pupae required for each procedure were randomly selected and placed into a fresh cage or used in experiments.

To obtain flies for dissections, a large TOLIMAN cage containing adults was placed at 4°C for 30-60 minutes until all flies were unconscious. A subset of flies was then removed and placed into glass petri dishes kept on ice. The glass plate was kept free from condensation by adding a thin layer of agar. Flies used for qPCR were reared in small cages of 50 individuals. On day 7 post eclosion, these cages were placed at 4°C for 30-60 minutes until flies were unconscious, then transferred into Eppendorfs of up to 5 flies of the same line/sex and snap frozen in liquid nitrogen. These samples were then stored at -80°C and removed in batches for dissection. Batches were removed for dissection and kept on dry ice.

Prior to beginning dissections all equipment and surfaces were sterilised with 70% ethanol and Invitrogen RNaseZap™. All equipment in direct contact with flies was wiped down with ethanol between dissections and full sterilisation was effected by applying Invitrogen RnaseZap™ between different samples. Single flies were removed and placed in a petri dish of 100% ethanol for 1 minute to prevent floating in PBS. Flies were then placed on a dissecting plate (glass petri dish containing silicon) and held in place with dissecting pins before covering the sample with a drop of sterile Phosphate Buffered Saline (PBS). Fine tip forceps were then used to peel open the abdominal cavity and extract testes or ovaries. All excess tissue was removed and for dissections on live samples, the early germline cells were separated from the late by using a scalpel, with samples placed in an Eppendorf and covered in a thin layer of PBS. The remaining fly carcasses were kept for use as the somatic samples. Samples collected are shown in *Table 2.1*.

Table 2.1 Samples collected from dissections for each experiment. All samples were snap frozen in liquid nitrogen and stored at -80°C until RNA extraction.

	Experiment	
	End-point RT-PCR/RACE	**qPCR**
Live or Frozen Tissue	Live	Frozen
Separated into early and late germline?	Yes	No
Number of testes per sample	11 pairs (22 total)	9 pairs (18 total)
Number of ovaries per sample	9 pairs (18 total)	3 pairs (6 total)
Number of somatic tissue per sample	3	1

2.3.1.3   Egg sample collection for end point RT-PCR

A large TOLIMAN cage was set up with the purpose of allowing females to lay eggs for 1h intervals, and to allow single water drop egg collections to be taken, and stored in Eppendorfs. To collect eggs at different development stages these samples were maintained under standard temperature conditions before removing any excess water and were then snap frozen in liquid nitrogen at the different time points required. The egg development times tested for RNA expression post laying were 1h, 3h, 6h, 24h, 48h and 72h. As the collections were a unknown mix of developing embryos and unfertilised eggs they are referred to as post egg collection time. Due to high levels of desiccation the 72h samples were not carried forward.

2.3.2 Transgenic Line Establishment

2.3.2.1 Microinjection of medfly eggs to generate transgenic strains

Medfly microinjection was performed at the Imperial College London insectary by A. Meccariello. A further purification step was performed on midiprep samples for microinjection using Millipore Ultrafree®-MC GV Centrifugal Filters. The full midiprep sample was loaded to the column and centrifuged at 14,000 RPM for 4 minutes. 40µl of injection mix was used per microinjection session with a final concentration of 500ng/µl plasmid of interest, 200ng/µl [i]hyPBase transposase helper plasmid (Eckermann et al. 2018). This was made up to a final volume with injection buffer (Rubin and Spradling 1982). Embryos were collected after an egg laying period of 45 minutes, transferred to double sided sticky tape and allowed to dry for 6 minutes. Eggs were manually dechorionated using the tip of a blunted needle. Once the chorion was removed, eggs were transferred to slides coated in a thin layer of glue (double sided sticky tape dissolved in heptane) with the posterior of the egg facing outwards. These egg slides were desiccated by placing them in a petri dish containing calcium chloride for 6 minutes, and then covered with a thin layer of Halocarbon™ 700 Oil. Eggs on these slides were then injected using WPI Sutter Quartz needles (outer diameter (OD) = 1.00mm, inner diameter (ID) = 0.70mm) that had been drawn on a Sutter P-200/G laser needle puller. Needles were loaded with injection mix using an Eppendorf Microloader. Microinjections were done using an inverted stereo microscope with an Eppendorf microinjection set-up and compressed air-driven pump.

Post injection, eggs were left on slides until the larvae hatched (typically after 48h) at which point the larvae were placed onto larval food (30g paper, 30g sucrose, 30g yeast extract, 10ml cholesterol stock (5g cholesterol, 140 ml distilled water, 30ml 95% EtOH), 2ml HCl Stock (384ml distilled water, 66ml 37% HCl), 8.5ml Benzoic Stock (50g Benzoic acid, 300ml 95% EtOH, 150ml distilled water) made up to 400ml with distilled water). After 7 days larval culture bottles were laid down in pupation boxes with a thin layer of fine sand and allowed to pupate. Surviving pupae were then sent to UEA within three sealed layers.

2.3.2.2 Fluorescence screening for medfly transgenics

Transgenic lines were established and maintained by screening flies for fluorophore marker expression at each generation. When establishing transgenics, live adults were screened on the day of eclosion from the pupae. Flies were anaesthetised over ice for screening. A Leica

MZ7.5 dissection microscope was used for screening and was fitted with fluorescence filters to visualise DsRed2 (excitation wavelength 558nm, emission wavelength 583nm) or AmCyan (excitation wavelength 458nm, emission wavelength 485nm) depending on the marker being screened. Established transgenic lines were screened as pupae as the fluorophore expression was clearer at this stage.

2.3.3   Investigating fitness costs in transgenic lines

2.3.3.1   Measuring pupal development and eclosion rates

Egg collections were taken from the Benakeion wildtype and the *M68*.3 and *bgcn*.2C transgenic lines on three consecutive days. After 7 days of development, bottles were laid on sand in pupation boxes. Sand was sieved daily and pupal number recorded until 23 days post egg collection. 150 pupae were taken from the first egg collection, from the day 12 pupae produced by each line. Transgenic lines were screened to select 150 transgenic carrying pupae. Eclosions were monitored daily with sex of flies recorded. Three biological replicates were performed for each line.

2.3.3.2   Fitness tests

The first pupae to eclose from the wild type and transgenic lines were used to set up small cages of 25 males and 25 females in order to undertake survival fitness tests under both standard and food-stressed conditions. A stressed cage was provided with no food or water and deaths were monitored daily until all flies were dead. A standard cage was set up with regular food and water and deaths were recorded daily for one month, which allowed median death to be reached for all cages. Three biological replicates performed for each line for each stress treatment.

2.3.4   Molecular investigation of candidate gene expression

2.3.4.1   DNA Extraction

DNA extraction was performed on whole flies of both TOLIMAN and Cepa Petapa wildtype strains. Whole flies were frozen at -20°C for a minimum of 48 hours with 3 individuals per sample. Samples were removed from the freezer and placed onto dry ice. Samples were then homogenised with a sterilised, plastic pestle over dry ice. Qiagen DNeasy® Blood & Tissue Kit was used as per the manufacturer's instructions with the addition of an extra elution step

with 25µl molecular grade water after the original 50µl elution. The concentration of each sample was measured using a Thermo fisher Scientific Nanodrop™ 8000 Spectrophotometer.

2.3.4.2   Primer design for End-point RT-PCR

Six genes were selected as potential Cas9 promotors (*nanos, vasa, bgcn*, *ß2-tubulin, mei-W68, innexin-5*). Primer pairs were placed to ensure the resulting PCR product covered an intron/exon boundary to allow for differentiation of DNA versus RNA amplification. Three primer pairs were designed for each gene of interest. All primer pairs were tested on 0.5µl TOLIMAN gDNA (324ng/ul) with New England Labs® *OneTaq*® 2X Master Mix with Standard Buffer used for all PCR reactions. Cycling conditions were 3 minutes at 94°C, 28 cycles of 30 s at 94°C, 30 s at 55°C, 2 minutes at 72°C with final extension of 10 minutes at 72°C. The resulting PCR products were visualised using gel electrophoresis, and primer pairs to be used for the cDNA amplification were selected based on which pairs showed the strongest, correct sized bands on these gel results. All primers were synthesised by Integrated DNA Technologies and the sequences are available in (*Supplementary information 2.6.1.1*).

2.3.4.3   RNA Extraction

Prior to commencing RNA extractions, all equipment and worktops were sterilised with 70% ethanol and Invitrogen RNaseZap™. Consumables and equipment were then placed in a UV hood for 15 minutes to reduce the risk of RNase contamination. Samples for RNA extractions were removed from the freezer and placed onto dry ice. Samples were then homogenised with a sterilised, pre-chilled, plastic pestle over dry ice. The Invitrogen *mir*Vana™ miRNA Isolation Kit was used for RNA extractions as per the manufacturer's instructions. Samples were eluted in 20µl Invitrogen™ THE Ambion™ RNA Storage solution.

2.3.4.4   Concentration quantification

The concentration of all DNA and RNA samples was measured via a Thermo fisher Scientific Nanodrop™ 8000 Spectrophotometer in 1µl volume aliquots. Optical density was measured at 260/280 with the ideal value being 1.8 and at 260/230 with the ideal value being 2.0. Samples with too low a concentration were repeated where possible.

2.3.4.5   DNase treatment

DNase treatments were conducted to remove DNA from the RNA samples as per the manufacturer's instructions, using the Invitrogen Turbo DNA-*free*™ kit for qPCR samples or with the Invitrogen Ambion® DNA-*free*™ Kit for RT-PCR/RACE samples.

2.3.4.6   Reverse Transcriptase PCR (RT-PCR)

All DNased samples for end-point RT-PCR only were diluted to below 20ng/μl. The Thermo Scientific RevertAid H Minus First strand cDNA Synthesis Kit was used as per the manufacturer's instructions. All samples were diluted to the same concentration as the lowest concentration sample. Qiagen QuantiTect® Reverse Transcription Kit was then used for the RT-PCR reactions as per the manufacturer's instructions.

2.3.4.7   End-point RT-PCR

New England Labs® *OneTaq*® 2X Master Mix with Standard Buffer was used for all PCR reactions for *nanos* and *bgcn.* The PCR cycling conditions were 30 s at 94°C, 28 cycles of 30 s at 94°C, 30 s at 54°C, 1 minute at 68°C with a final extension of 10 minutes at 68°C. *mei-W68, innexin-5* and *vasa* PCRs were run with Thermo Scientific™ DreamTaq Green PCR Master Mix (2X) as per the manufacturer's instructions with a 20μl total reaction volume. The cycling conditions were 3 minutes at 95°C, 28 cycles of 30 s at 95°C, 30 s at 54°C, 1 minute at 72°C with a final extension of 10 minutes at 72°C. PCR products were all visualised with gel electrophoresis (see 2.3.4.1 below).

2.3.4.8   Rapid amplification of cDNA ends (RACE)

The Takara SMARTer® RACE 5'/3' Kit was used as per the manufacturer's instructions to generate RACE-Ready cDNA using the highest concentration RNA extraction sample (early ovaries) post DNase digestion. This cDNA was then used as the input for the RACE PCR reactions. Gene-specific primers for amplification of the 5' ends were designed following the manufacturer's guidance, for *bgcn*, *mei-W68*, *nanos*, *innexin* and *vasa* (*Supplementary information 2.6.1.2*). Three primers were recommended for each gene. However, only two could be designed for *innexin*, four primers were required for *mei-W68* and five for *vasa*. Primers were obtained from Integrated DNA Technologies. Reactions were run with both positive and negative controls, on medfly cDNA and mouse heart cDNA as a control. Cycling conditions were: 30 cycles of 30 s at 94°C, 30 s at 68°C, 4 minutes at 72°C. Nested PCR was

done with 5µl of PCR product diluted in 245µl molecular grade water, with the same cycling conditions when there was difficulty generating sufficient product in the first PCR reaction. Positive PCR products were identified with gel electrophoresis and prepared for sequencing.

2.3.5   General Molecular Techniques

2.3.5.1   Gel electrophoresis

Products to be visualised by gel electrophoresis that did not have any loading dye in PCR mastermix were combined with New England BioLabs® Purple Gel Loading Dye (6X) prior to loading. Agarose gels were made at 1.2% (w/v) with 1X TBE and stained with 3µl Ethidium Bromide per 100ml gel. Gels were run with appropriate ladders in each case to determine band sizes. The ladders used were New England BioLabs 1kb Plus DNA Ladder; Thermo Scientific™ GeneRuler™ 50bp DNA Ladder; Thermo Scientific™ GeneRuler™ 1kb DNA Ladder; Thermo Scientific™ GeneRuler™ 1kb Plus DNA Ladder. Gels were run at 90-120V for 60-90 minutes depending on product size.

2.3.5.2   PCR Purification

The Thermo Scientific™ GeneJET PCR Purification Kit was used for the purification of PCR products, as per the manufacturer's instructions with final elution in 25µl molecular grade water after a 1 minute incubation at room temperature.

2.3.5.3   Gel Extraction

Thermo Scientific™ GeneJET Gel Extraction Kit used for purification of excised bands from agarose gel. Used as per the manufacturer's instructions with final elution in 25µl molecular grade water after a 1 minute incubation at room temperature.

2.3.5.4   Sequence Confirmation of PCR products

Purified PCR samples were diluted to 10ng/µl and sent for sequencing using the Eurofins Genomics TubeSeq Sequencing service. Sequencing primers were sent with samples and are provided in *Supplementary information 2.6.1*. Sequence alignment of the resulting reads was performed using the Benchling sequence alignment tool (Benchling 2023).

2.3.5.5 DpnI Digestion

DpnI digestion (New England Biolabs®) was completed on PCR products for Gibson assembly, to remove the original methylated plasmid template. Digestions were performed as per the manufacturer's instructions.

2.3.6 Construct Building

2.3.6.1 Construct Design

All constructs were designed in Benchling using the Gibson assembly tool. Primers were designed manually or generated by Benchling using the pre-set parameters (Benchling 2023).

2.3.6.2 Gibson Assembly

2.3.6.2.1 Primer Design

All primers were designed to create a 20bp overlap region between parts to allow for assembly. All primers were obtained from Integrated DNA Technologies and all primer sequences are available in the supp info (*Supplementary information 2.6.1.3*).

2.3.6.2.2 Generating Gibson Assembly parts

Individual parts for the Gibson assemblies were generated via PCR with overhanging primers, using New England BioLabs® Q5® High fidelity 2X Master Mix to ensure sequence accuracy. Primer annealing temperatures were calculated using New England BioLabs® Tm Calculator and cycling conditions as outlined in manufacturer's instructions were used. Cycling conditions for all reactions were determined by manufacturer's instructions. Medfly gDNA or plasmid DNA were used as templates for all PCR reactions. All plasmids used can be found in *Supplementary information 2.6.2*. PCR products were purified, and their concentration measured by using the Nanodrop™.

2.3.6.2.3 Gibson Assembly Reaction

Invitrogen™ GeneArt™ Gibson Assembly HiFi Master Mix was used as per the manufacturer's instructions for the assemblies of *bgcn* and *mei-W68* Cas9 plasmid. After incubation, Gibson reactions were diluted in a ratio of 1:5 and used for transformation into *E. coli* competent cells.

2.3.6.3 Transformation of *E. coli* competent cells

Three competent cell types were used and the transformation protocols are summarised below:

Table 2.2 Transformation protocols of *E. coli* competent cells used for construct assembly. mei-W68 Cas9 was established in Takara Stellar cells and all other constructs were established in SURE2 cells.

	Takara Stellar™ Competent Cells	Invitrogen™ Subcloning Efficiency™ DH5α Competent Cells	Agilent SURE® 2 Supercompetent Cells
Volume used	4µl diluted Gibson : 100µl cells	2µl diluted Gibson : 50µl cells	5µl diluted Gibson : 50µl cells
Ice incubation	30 minutes		
Heat shock	60 s at 45°C	60 s at 42°C	30 s at 42°C
Ice incubation	2 minutes	5 minutes	2 minutes
Volume of SOC added	450µl	950µl	700µl
Incubation	1h at 37°C, 200rpm	1h at 37°C, 100rpm	1h at 37°C, 100rpm
Spin step	N/A	N/A	2 minutes at 4000rpm, pellet resuspended in 100µl SOC
Final plated volume onto LB Ampicillin (100µg/ml)	100µl		

Plate incubation	Overnight at 37°C

2.3.6.4 Colony PCR

Individual colonies were screened for inserts after Gibson assembly. Colonies of varying sizes were selected that did not overlap other colonies on the growth plate. A 10µl pipette tip was used to remove the edge of each screened colony, and the material was then resuspended in 10µl of molecular grade water. 1µl of this colony mix was then used as the template in a Thermo Scientific™ DreamTaq Green PCR Master Mix (2X) reaction with a total volume of 20µl. Primers were designed to cover insert boundaries where possible to confirm presence of both the insert and the backbone. All primers used are listed in *Supplementary information 2.6.1.3*.

2.3.6.5 Plasmid Miniprep

Positive clones were miniprepped to prepare them for sequencing. Overnight cultures were inoculated with a single positive colony into 4ml LB Ampicillin (100µg/ml) contained within a 15ml falcon tube to allow sufficient aeration. Cultures were incubated overnight at 37°C with constant agitation at 180rpm. After confirming growth, 2ml of the culture was prepared by using the Thermo Scientific™ GeneJET Plasmid Miniprep Kit, as per the manufacturer's instructions. The final elution step was performed with 30µl molecular grade water. The purity and concentration of all samples were checked with Nanodrop™ before being diluted to the correct concentration for sequencing.

2.3.6.6 Sequencing confirmation of plasmids

Plasmids were sent for Sanger sequencing with Eurofins Genomics TubeSeq Sequencing after dilution to 50ng/µl with sequencing primers which covered insert boundaries. Subsequently, I switched to employing whole plasmid sequencing. Whole plasmid sequencing was completed by diluting miniprepped plasmids to 30ng/µl and sending to SNPsaurus© Plasmidsaurus Whole Plasmid Sequencing (Nanopore based long-read) Service. Sequence alignment of returned reads was performed using the Benchling sequence alignment tool (Benchling 2023).

2.3.6.7 Plasmid Midiprep

After sequence confirmation, plasmids were prepared for microinjection using the Macherey-Nagel NucleoBond® Xtra Midi EF, Midi kit for endotoxin-free plasmid DNA kit. Cultures were revived from glycerol stocks onto a streak plate which was incubated overnight at 37°C. A single culture from this was used for inoculation of a 4ml LB Ampicillin (100µg/ml) starter culture and incubated for 6-8h at 37°C with constant agitation at 180rpm. After confirming growth in starter cultures, 400µl was used to inoculate 100ml LB Ampicillin (100µg/ml) to be incubated overnight at 37°C with constant agitation at 200rpm. This culture was then used as per the manufacturer's instructions for midiprep. The final pellet was resuspended in 50µl endotoxin-free molecular grade water.

2.3.6.8 Preparing glycerol stocks

Plasmids were maintained for long term storage as glycerol stocks. For the preparation of glycerol stocks, single colonies were selected to inoculate 4ml LB Ampicillin (100µg/ml). This was then incubated overnight at 37°C with constant agitation at 180rpm (typically, this was the culture that was also used for plasmid miniprep). From this, 500µl of culture was added to 500µl 50% glycerol and mixed well by inversion. Glycerol stocks were then stored at -80°C, and only the top layer was thawed when reviving them.

2.3.7 Molecular validation of transgenic lines

2.3.7.1 Real time quantitative PCR (qPCR)

qPCR was performed with the BioRad iTaq™ Universal SYBR Green Supermix, following the manufacturer's instructions and using 12.5ng total cDNA assuming a 1:1 conversion during first strand synthesis. Cas9 primer sequences were obtained from A. Meccariello (Imperial College, London) and *RPL19, RPE and Actin* primer sequences obtained from a study of housekeeping genes in medfly (Sagri et al. 2017). All primers were synthesised by Integrated DNA Technologies (*Supplementary information 2.6.1.4*). 3.3pmol of each primer was used per reaction to a final total reaction volume of 20µl. All samples were run with two reference genes (*RPE* and *RPL19*) on the same plate. One single identical sample was run on all plates to allow for interplate calibration. Three biological replicates of every sample were run with three technical replicates for every reaction. BioRad Multiplate™ 96-Well PCR Plates were loaded and sealed with BioRad Microseal 'B' PCR Plate Sealing Film. All reactions were run on the BioRad CFX Connect™ Real-Time PCR Detection System with the following cycling

conditions: 30s at 95°C, 39 cycles of 5s at 95°C, 30s at 60°C followed by a final 5s at 65°C and 5s at 95°C.

2.3.7.2   qPCR analysis

Technical replicates with a Ct value difference of greater than 0.5 were excluded from the dataset, with the closest two Ct values being retained. When no technical replicates were within 0.5 Ct of each other the whole sample was excluded. The mean expression level was calculated for each sample from all the technical replicates that had been retained in the dataset. Primer efficiency was calculated using Thermo Fisher qPCR Efficiency Calculator. Interplate calibration (IPC) was performed using the calibration sample run on every plate by dividing all calculated mean Ct values by the Ct value of the IPC on that plate and multiplying by the mean of the IPC on every plate (Hellemans et al. 2007). Sample types were randomised on each plate to minimise the risk of batch effects. *RPL19* was used as the sole reference gene as *RPE* had, on average, a higher Ct value leading to the exclusion of a greater number of samples due to the technical replicate threshold not being met. Relative expression was calculated using the Pfaffl equation $2^{-\Delta Ct}$ (Pfaffl 2001) with -ΔCt being the difference in the Ct values for Cas9 and RPL19.

2.4　Results

2.4.1　Selection of germline promotors to drive Cas9 expression

2.4.1.1　Identifying germline specific genes

A primary literature search was performed to identify germline specific genes that have been characterised in medfly. Previous reports of potentially useful candidates for driving Cas9 expression in germline tissue have highlighted *nanos* in *Drosophila* (Champer et al. 2016; 2018) and *Aedes aegypti* (Adelman et al. 2007). Expression characterisation of *nanos* has also been done in medfly (Ogaugwu and Wimmer 2013). *vasa* has similarly been used for driving Cas9 expression in *Drosophila* (Champer et al. 2018; 2016) and for driving transgene expression in *Anopheles gambiae* (Papathanos et al. 2009)*.* Based on these findings, *nanos* and *vasa* were identified as two of the most promising candidates for driving Cas9 expression due to their early, germline restricted expression in both male and female germline. However, neither *vasa* nor *nanos* have previously been used to drive Cas9 expression in medfly. Therefore, it was decided to test expression of an additional 3 germline specific genes. *Benign gonial cell neoplasm* (*bgcn*) had previously been tested as a (unsuccessful) promotor in a homing endonuclease gene drive system in *Drosophila melanogaster* (Chan et al. 2013). However, I retained *bgcn* as a candidate to explore here because I planned to test a longer promotor region than previous investigated, which could have a different and more useful expression pattern. *mei-W68* and *innexin-5* were included as additional candidates as they are both well-annotated, germline-specific genes found in medfly. Sequences for all candidate genes were obtained from GenBank as shown in *Table 2.3*.

Table 2.3 GenBank accession IDs for candidate genes used. The table lists the sequences used for design of primers for RT-PCR, sequencing alignment and for final promotor design.

Candidate gene	Accession ID
bgcn	LOC:101457599 NW_019378542 REGION: 3472104..3477014
innexin-5	LOC:101449162 NW_019376375 REGION: complement(2138041..2139705)
mei-W68	LOC: 101459815 NW_019378542 REGION: complement(3565992..3569179)
nanos	LOC: 101462214 KC595268
vasa	LOC: 101456741 NW_019376383 REGION: complement(197451..223995)

2.4.1.2   End point RT-PCR to determine expression patterns of candidate genes

To test the tissue specificity of expression for the candidate genes selected, RT-PCR was performed. Three primer sets were designed for each gene – each of these covered an intronic region to enable differentiation between DNA and RNA products. All primer sets were tested on medfly gDNA to select the most appropriate for use.
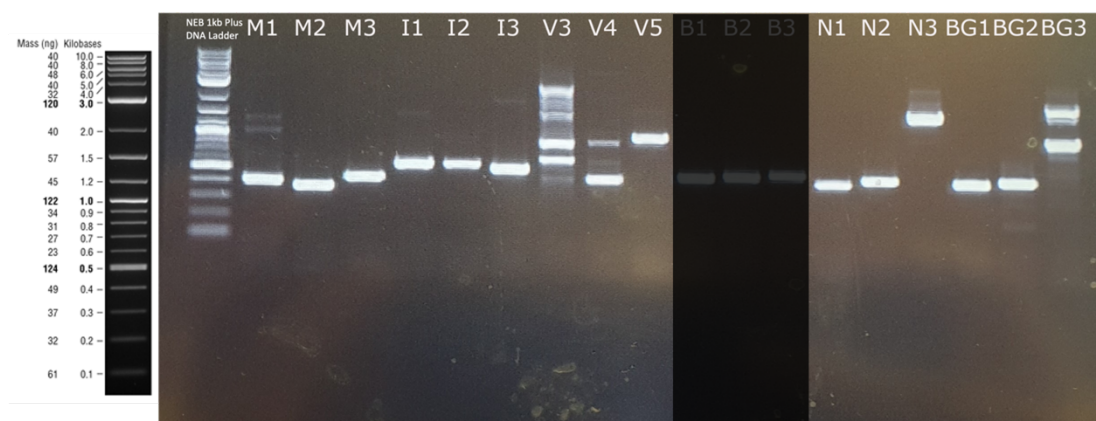
Figure 2.1 Gel electrophoresis of PCR products for multiple primer sets amplifying fragments of candidate genes from genomic DNA. Order of loading displayed in *Table 2.4*. The gel was run with NEB 1kb Plus DNA Ladder. The obscured lanes represent samples run for a germline specific promotor that was not intended for Cas9 expression.

Table 2.4 Primer sets for each candidate gene. Shown in the expected size of each product and if this was empirically confirmed. Primer sets in bold were selected for the expression study.

Primer Set	Predicted size (bp)	Size confirmed	Multiple banding
mei-W68 1 (M1)	414	Y	Y
***mei-W68* 2 (M2)**	**379**	**Y**	**N**
mei-W68 3 (M3)	446	Y	N
innexin-5 1 (I1)	579	Y	N
***innexin-5* 2 (I2)**	**576**	**Y**	**N**
innexin-5 3 (I3)	549	Y	N
vasa 3 (V3)	899	N	Y
vasa 4 (V4)	459	N	Y
***vasa* 5 (V5)**	**987**	**Y**	**N**
nanos 1 (N1)	449	Y	N
***nanos* 2 (N2)**	**498**	**Y**	**N**
nanos 3 (N3)	1645	Y	Y
bgcn 1 (BG1)	474	Y	N
***bgcn* 2 (BG2)**	**499**	**Y**	**N**
bgcn 3 (BG3)	1049	N	Y

The results from the PCR on gDNA shown in *Figure 2.1* and *Table 2.4* show that all candidate genes had a usable primer set (highlighted in bold) which showed no evidence of off target amplification.

To determine whether these genes have expression that is restricted to the germline, RNA was extracted from: male somatic tissue, early germline male testes (upper 1/3rd), late germline male testes (lower 2/3rd), female somatic tissue, early germline ovaries (upper 1/3rd), late germline ovaries (lower 2/3rd) and egg collections at several development times (0h, 1h, 3h, 6h, 24h and 48h). First strand synthesis was performed on the RNA extracted from all samples to generate cDNA. Although all samples were subjected to a Dnase digestion step, all primers were also designed to yield a different band size for the amplification of RNA vs DNA to ensure that mRNA expression could clearly be confirmed. PCR was run for every candidate gene with the chosen primer set in all tissue types. The products were visualised with gel electrophoresis (*Figure 2.2*).
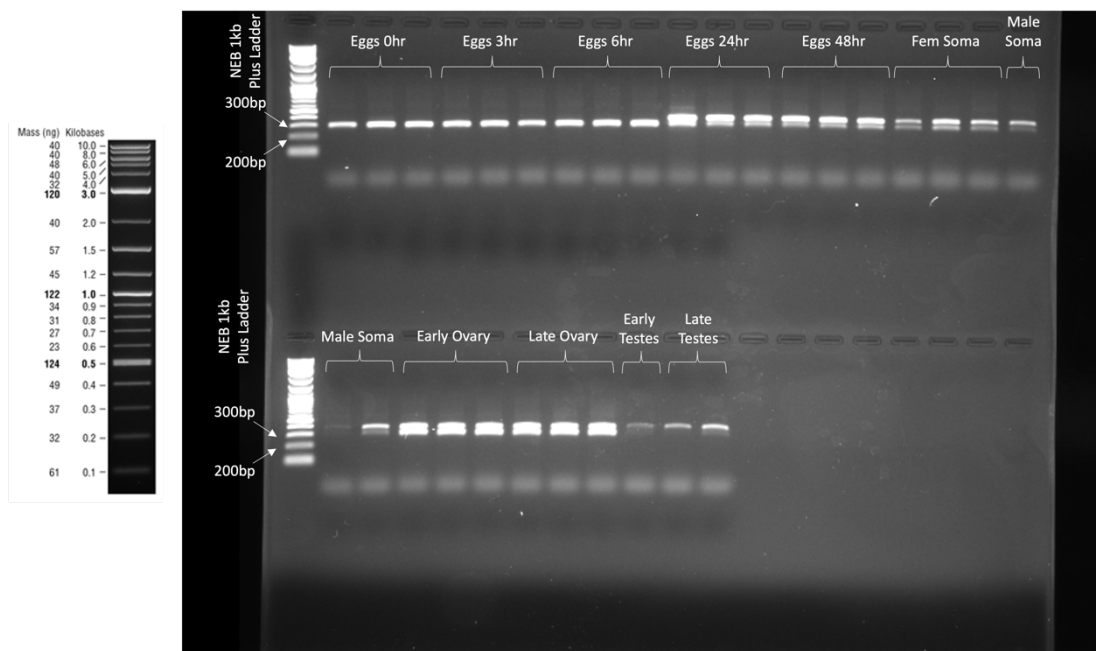
Figure 2.2 RNA expression pattern of *mei-W68*. End point RT-PCR results indicate in which tissue types *mei-W68* mRNA is expressed. The smaller sized band indicates the presence of spliced mRNA, whereas the larger band is unspliced mRNA. The gel was run with a NEB 1kb Plus Ladder with band sizes shown. Order of loading: Ladder, Eggs 0h post collection, Eggs 3h post collection, Eggs 6h post collection, Eggs 24h post collection, Eggs 48h post collection, Female Soma, Male Soma, Early Ovary, Late Ovary, Early Testes, Late Testes. Three biological replicates of every sample type were run except early testes for which only one pooled sample could be produced, and late testes with two pooled samples only.

The expression of *mei-W68* remained relatively constant, with expression shown in all tissue types (*Figure 2.2*). The difference between tissue types was evident in the levels of spliced vs unspliced mRNA. In early egg collection samples (<24h post collection), only spliced mRNA was present whereas in the testes, late embryos and somatic tissue there appeared to be

higher levels of unspliced mRNA. This expression across tissues, whether spliced or unspliced, means that the promotor of *mei-W68* would be suitable for driving Cas9 expression in these tissue types. Germline expression was evident, with ovary samples showing high levels of expression. Testes expression did not appear as strong. However, this does not rule it out as a potential promotor as the level of expression can only be fully confirmed only by additional qPCR testing and a lower level of testes expression could still be sufficient to achieve gene editing.
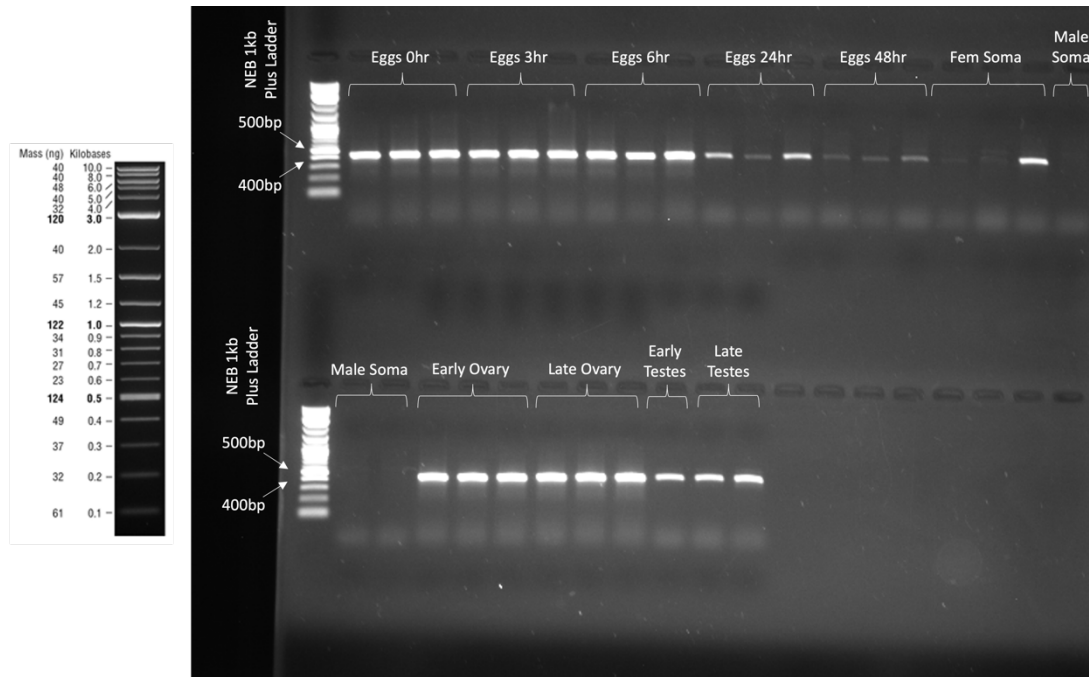
Figure 2.3 RNA expression pattern of *innexin-5.* End point RT-PCR results showing in which tissue types *innexin-5* mRNA is expressed. The gel was run with NEB 1kb Plus Ladder with band sizes shown. Order of loading: Ladder, Eggs 0h post collection, Eggs 3h post collection, Eggs 6h post collection, Eggs 24h post collection, Eggs 48h post collection,, Female Soma, Male Soma, Early Ovary, Late Ovary, Early Testes, Late Testes. Three biological replicates of every sample type were run, except early testes for which only one pooled sample could be produced and late testes for which there are two.

The expression pattern of *innexin* appeared to be ideal for use as a Cas9 promotor (*Figure 2.3*). Expression in the germline was consistent in male and female tissues, with high expression in early embryo cells and decreasing as the embryos mature. No expression was seen in the male soma, with weak expression in the female soma. The cause of this sex difference is unclear, as is the reason for the variability across the female soma biological replicates.
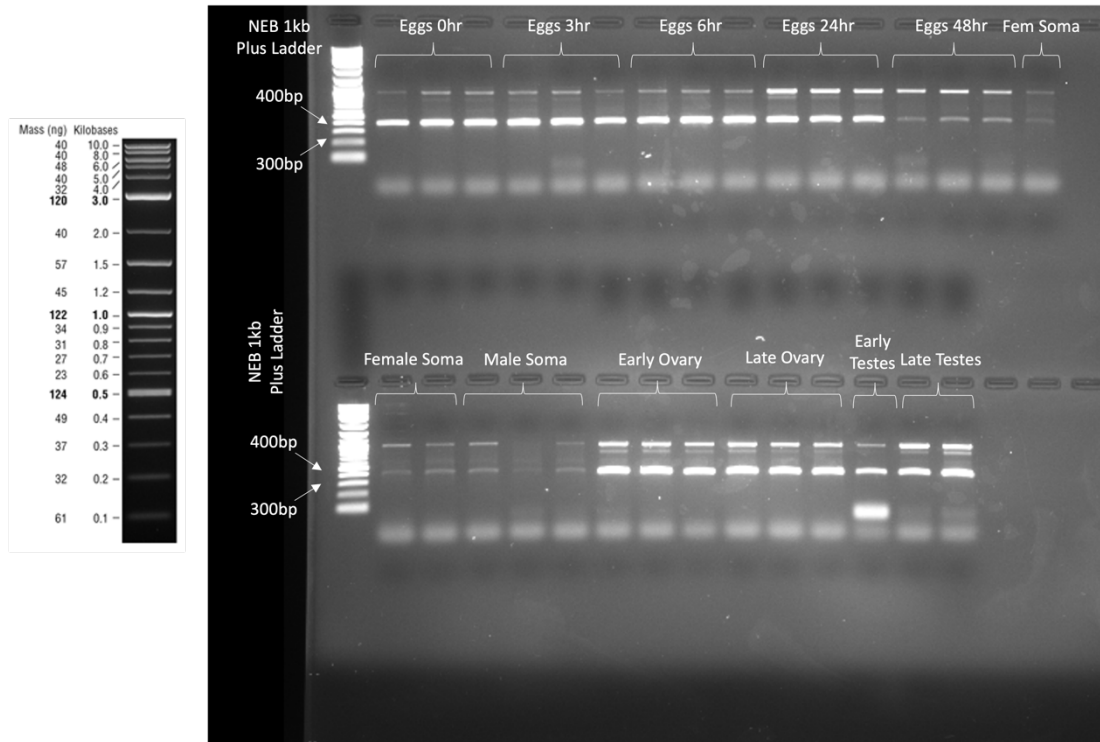
Figure 2.4 RNA expression pattern of *vasa.* End point RT-PCR results showing in which tissue types *vasa* mRNA is expressed. The smaller sized band represents spliced *vasa* mRNA whereas the larger bands are alternatively spliced/unspliced isoforms of *vasa* (size could not be accurately determined as the ladder has not separated). The gel was run with the NEB 1kb Plus Ladder with band sizes shown. Order of loading: Ladder, Eggs 0h post collection, Eggs 3h post collection, Eggs 6h post collection, Eggs 24h post collection, Eggs 48h post collection,, Female Soma, Male Soma, Early Ovary, Late Ovary, Early Testes, Late Testes. There were three biological replicates of every sample type, except early testes for which only one pooled sample could be produced and late testes for which there were two.

vasa appeared to be expressed across all tissue types, albeit with slightly weaker expression in somatic tissue. Multiple banding was seen across all samples (*Figure 2.4*) with the strongest band varying across tissue types. This was most likely caused by the alternative splicing of *vasa* which has several different isoforms annotated on the genome (*Supplementary information 2.6.5)*. Different isoforms could be preferentially expressed in different tissue types explaining this variance. For use as a Cas9 promotor, *vasa* does have the high levels of germline expression needed. However, its expression is not restricted to the germline.
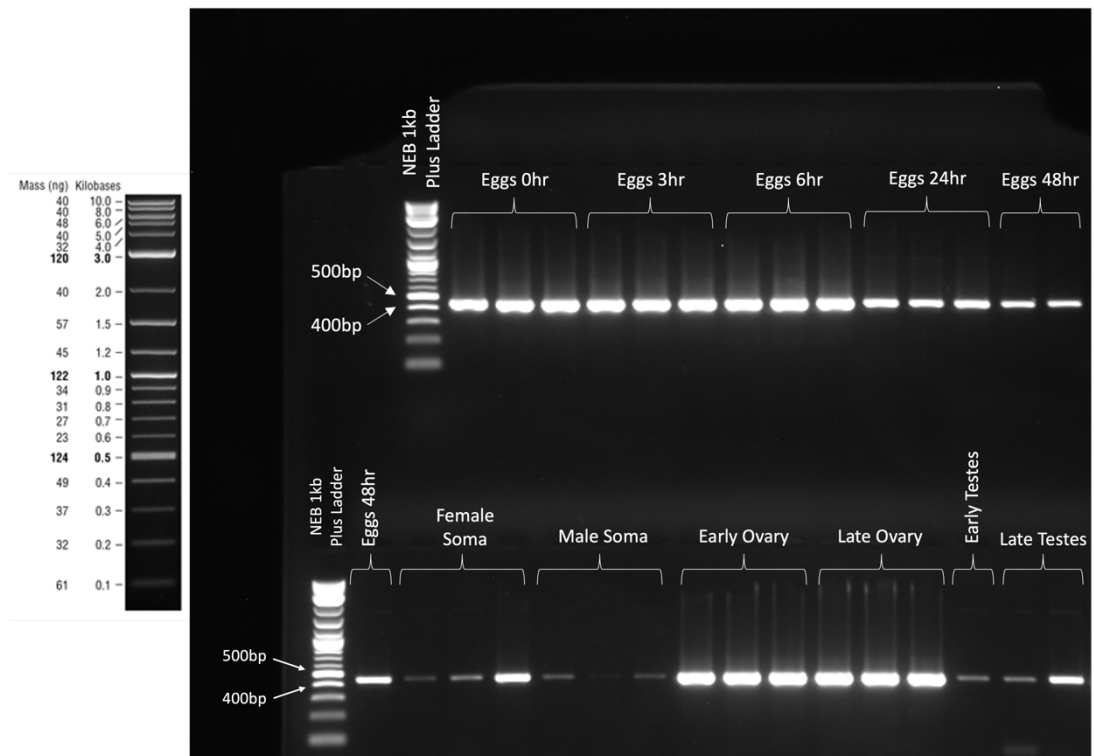
Figure 2.5 RNA expression pattern of *nanos.* End point RT-PCR results showing in which tissue types *nanos* mRNA is expressed. Ran with NEB 1kb Plus Ladder with band sizes shown. Order of loading; Ladder, Eggs 0h post collection, Eggs 3h post collection, Eggs 6h post collection, Eggs 24h post collection, Eggs 48h post collection,, Female Soma, Male Soma, Early Ovary, Late Ovary, Early Testes, Late Testes. There were three biological replicates of every sample type except early testes, for which only one pooled sample could be produced, and late testes for which there were two pooled samples.

nanos showed strong expression in the ovaries and early embryos (*Figure 2.5)* with a lower level of expression in the somatic tissue. Expression was present in the testes but not as clearly as was observed in the other tissue types. There were no problems with multiple banding in any of the tissue samples.
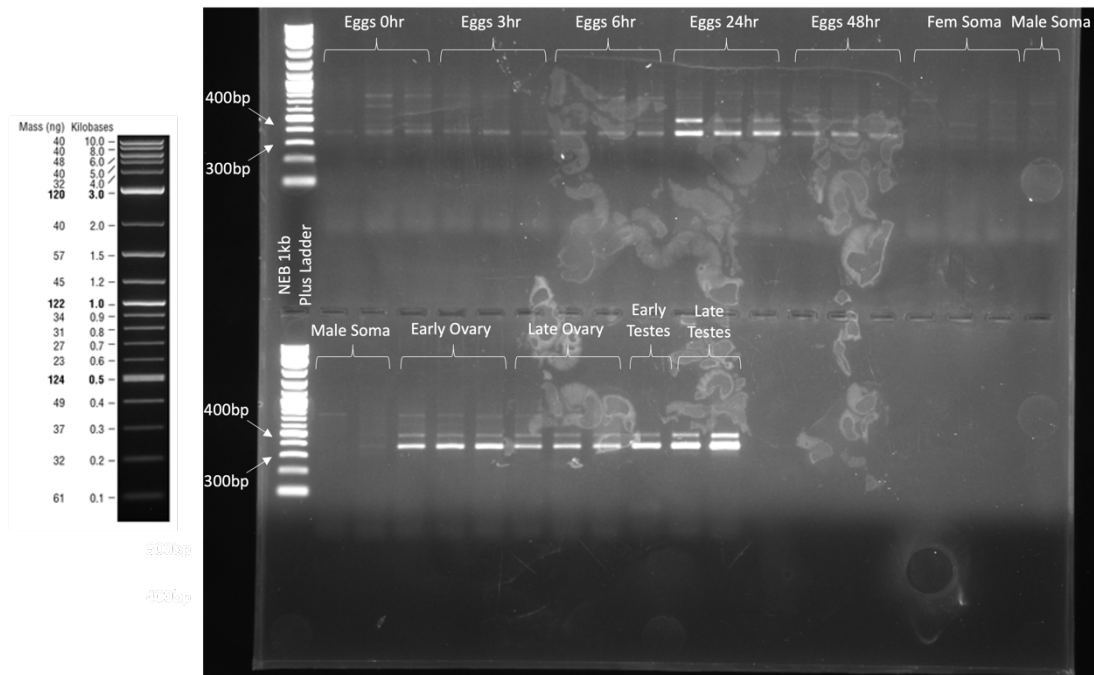
Figure 2.6 RNA expression pattern of *bgcn.* End point RT-PCR results showing in which tissue types *bgcn* mRNA is expressed. The band of smaller size represents spliced mRNA whereas the larger bands are predicted to be alternatively/unspliced mRNA. The gel was run with NEB 1kb Plus Ladder with band sizes shown. Order of loading; Ladder, Eggs 0h post collection, Eggs 3h post collection, Eggs 6h post collection, Eggs 24h post collection, Eggs 48h post collection, Female Soma, Male Soma, Early Ovary, Late Ovary, Early Testes, Late Testes. There were three biological replicates of every sample type except early testes, for which only one pooled sample could be produced, and late testes for which there were two pooled samples. The cloudy markings on the gel were due to an issue with burnt agarose on the visualiser and can be ignored.

bgcn showed good germline restriction of expression in comparison to other candidate genes (*Figure 2.6*)*.* Expression in the testes appeared as strong as it did in the ovaries, which was a pattern not observed across the other candidate genes. Somatic expression appeared very minimal for all isoforms of the mRNA. Expression appeared weak in early embryo cells, with bands becoming clearer at the 24-hour timepoint. This low expression in the early embryonic tissue could be beneficial for a Cas9 promotor where mosaicism caused by leakage is a concern.

In the vasa results (*Figure 2.4*) there was a much stronger band present in early in comparison to late testes. Overall, *vasa* had the highest number of different sized bands present, which could be interpreted as the presence of alternative splicing isoforms. The band primarily present in early testes is smaller than the expected for a fully spliced mRNA product and it was most evident in a sample that showed lower expression. Therefore, this band could indicate the presence of primer dimers, and is the correct size for that interpretation. The localisation of a primer dimer band in the sample with the lowest amount of expression is

also expected, and for *vasa* such a band would be expected in the somatic tissue rather than the testes, in which expression was relatively high.

Across several candidate genes, bands of corresponding size to gDNA amplification were also present. However, this was not consistently observed in the same samples as would be expected if it had been caused by gDNA contamination. To test for gDNA contamination PCR was performed on an intronic region in a 24-hour post egg collection sample and early ovary sample with a gDNA sample as control. No bands were present for either cDNA sample. Therefore, DNA contamination appears unlikely. Another possible cause of the larger bands is unspliced or differentially spliced mRNA. This could explain the difference between candidate genes, as splicing varies across genes and the multiple banding in several samples could represent different isoforms. The expression pattern of *vasa* had the most instances of multiple banding across all sample types (*Figure 2.4*) which further supports this conclusion, as *vasa* has many different isoforms derived from alternative splicing.

No candidate gene's expression was entirely restricted to the germline. For Cas9 expression complete restriction to the germline is not required, as leaky expression is primarily a concern when designing homing drives in which resistance allele formation is more common. Cas9 expression outside of the germline can be tolerated for functional studies or other drive types which do not rely on homing, and can sometimes even be beneficial due to the accumulation of broken alleles when using for toxin-antidote systems (Champer et al. 2021). *nanos* and *vasa* were identified as the most commonly used promotors for germline expression of transgenes (Champer et al. 2016; Champer et al. 2018; Adelman et al. 2007; Papathanos et al. 2009) in the literature search which informed this experiment. My results show that, whilst these genes did have strong expression in the germline they also displayed a lack of specificity, with expression being high throughout tissue types. The less commonly used *bgcn* and *innexin* genes displayed the necessary germline expression with more specificity, having minimal somatic expression. *bgcn* and *innexin* have both been tested as potential Cas9 promotors in other model organisms (*bgcn* in *Drosophila melanogaster* (Chan et al. 2013) and *innexin-4* (a different innexin gene to the *innexin-5* used in this work) in fall armyworm (Chen and Palli 2022)) yet my results here showed no somatic expression in any experiment.

Some candidate genes also displayed a difference in expression across male and female tissues. Expression in the ovaries was consistent throughout the candidate genes, whereas expression in the testes didn't appear as strongly in *mei-W68* or *nanos* when compared to the ovaries. A band was present in all testes samples for the candidate genes, meaning some

expression is present. However, without the use of qPCR, or a similar quantitative method, it is hard to judge exact levels of expression across the tissue types. From these results it was decided that all genes were expressed in the germline, and could be retained as candidates to use as promotors for Cas9 expression (next section).

2.4.2   Identification of the promotor region of candidate genes

2.4.2.1   Rapid Amplification of cDNA Ends (RACE)

To determine the 5' untranslated regions (UTR) of the promotors, a RACE reaction was carried out for all of the candidate genes retained from the initial gene expression tests described above. RACE-ready cDNA was generated from an early ovary sample as this had the strongest expression across all candidates. RACE PCR was run with the first primer set for all candidate genes and visualised on a 1.3% agarose gel (*Figure 2.7*). All primer sets used in this experiment available in *Supplementary information 2.6.1.2*.
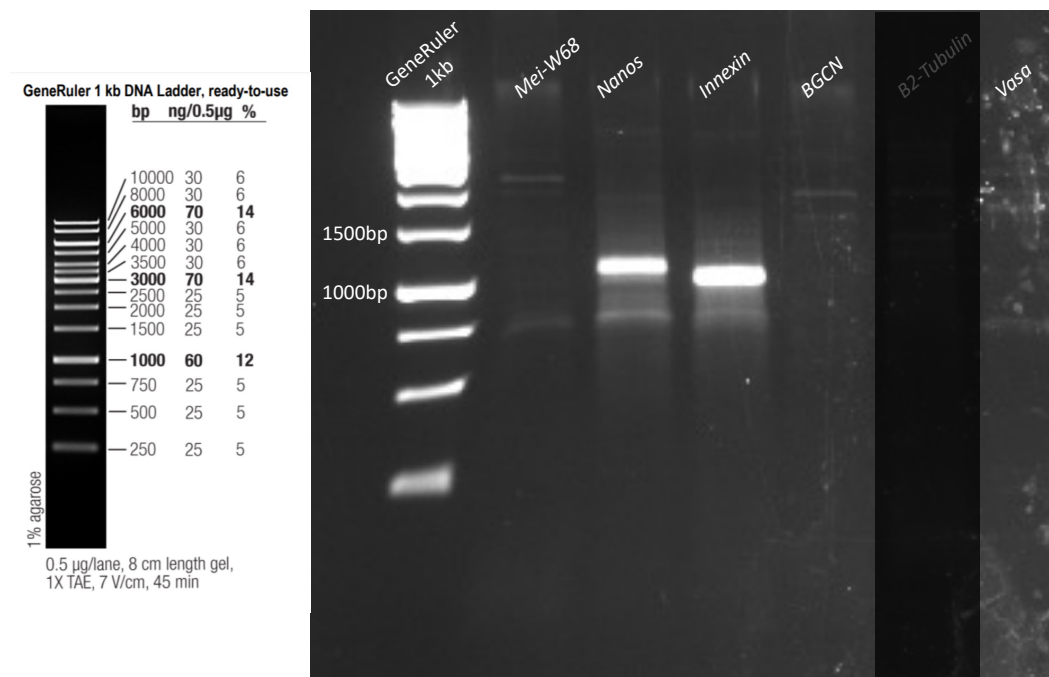
Figure 2.7 Gel electrophoresis of RACE PCR products of candidate genes run alongside the GeneRuler 1kb DNA Ladder. Order of loading: Ladder, *mei-W68, nanos, innexin, bgcn, vasa.* The blanked out lane labelled *B2-Tubulin* was a sample run for a different experiment and can be ignored. Strong bands were present for both *nanos* and *innexin* and could be extracted for sequencing. Weaker bands were present for all other samples, but none were strong enough to send products for sequencing.

Strong bands of the expected size were seen for both *nanos* and *innexin*. These products were purified and sent for sequencing. Some very weak bands were present for the other

candidate genes, none of which were considered sufficiently strong to be an unequivocal positive and to justify sending for sequencing. Nested PCR was carried out on the diluted PCR reactions of *mei-W68, bgcn* and *vasa* using the second primer set for all genes. This yielded no results, so all of these original bands were assumed to represent non-specific amplification. RACE PCR was repeated on RACE-ready cDNA using the second and third set for *bgcn* and *vasa* (*mei-W68* could be run only with its second primer set as no third primer set could be designed due to difficulty in finding a binding position). Products were visualised on a 1.2% gel (*Figure* 2.8).
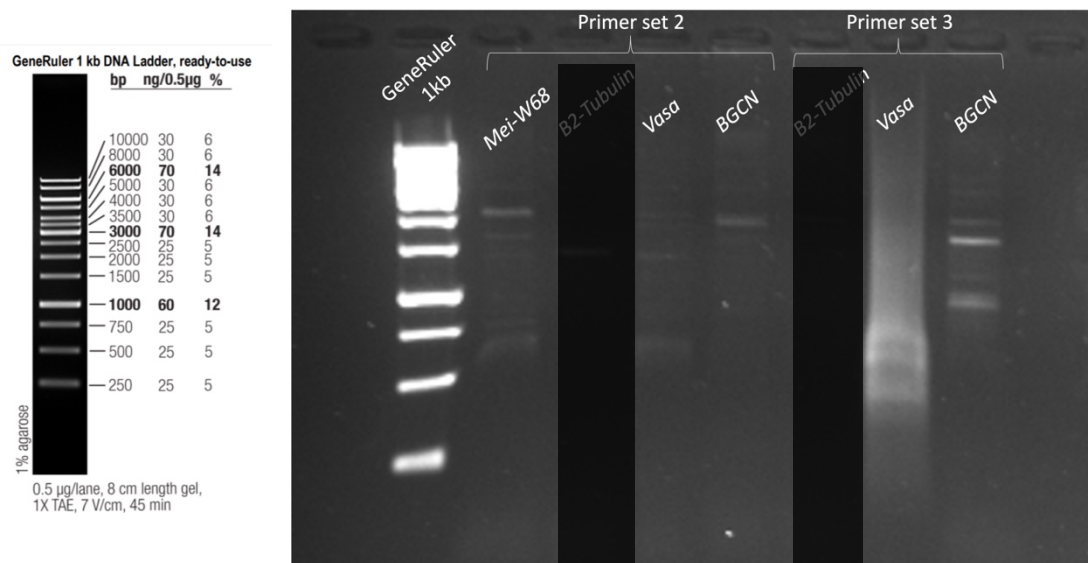
Figure *2.8* Gel electrophoresis of RACE PCR products of candidate genes run alongside a GeneRuler 1kb DNA Ladder. Order of loading; Ladder, *mei-W68* with primer set 2, *vasa* with primer set 2, *bgcn* with primer set 2, *vasa* with primer set 3, *bgcn* with primer set 3. Blanked out lane labelled *B2-Tubulin* was a sample for a different experiment and can be ignored. *bgcn* showed bands present for both primer sets, with the strongest band present for primer set 3. *mei-W68* showed a band present for primer set two. Weaker bands were present for all samples, but none were strong enough to send products for sequencing.

bgcn produced PCR products from both primer sets 2 and 3. However, neither band was deemed strong enough for purification. Primer set 3 did produce a stronger band than for primer set 2. Therefore, it was thought that with a more specific template the product amplification might be strengthened by using a nested PCR. Nested PCR with primer set 3 was run with the diluted primer set 2 PCR product in an attempt to do this (*Figure 2.9*). *mei-W68* primer set 2 produced a sufficiently strong band for use in a nested PCR. However, a third primer set was not immediately available for this candidate gene due to difficulties in placing a final primer that adhered to the recommendations for RACE. A new gene-specific primer was designed for nesting and ordered. None of the products for *vasa* were deemed

suitable for nesting or sequencing. Accordingly, 2 new gene specific primers were designed for *vasa* (primers 53 and 54 sequences in *Supplementary information* 2.6.1.2).
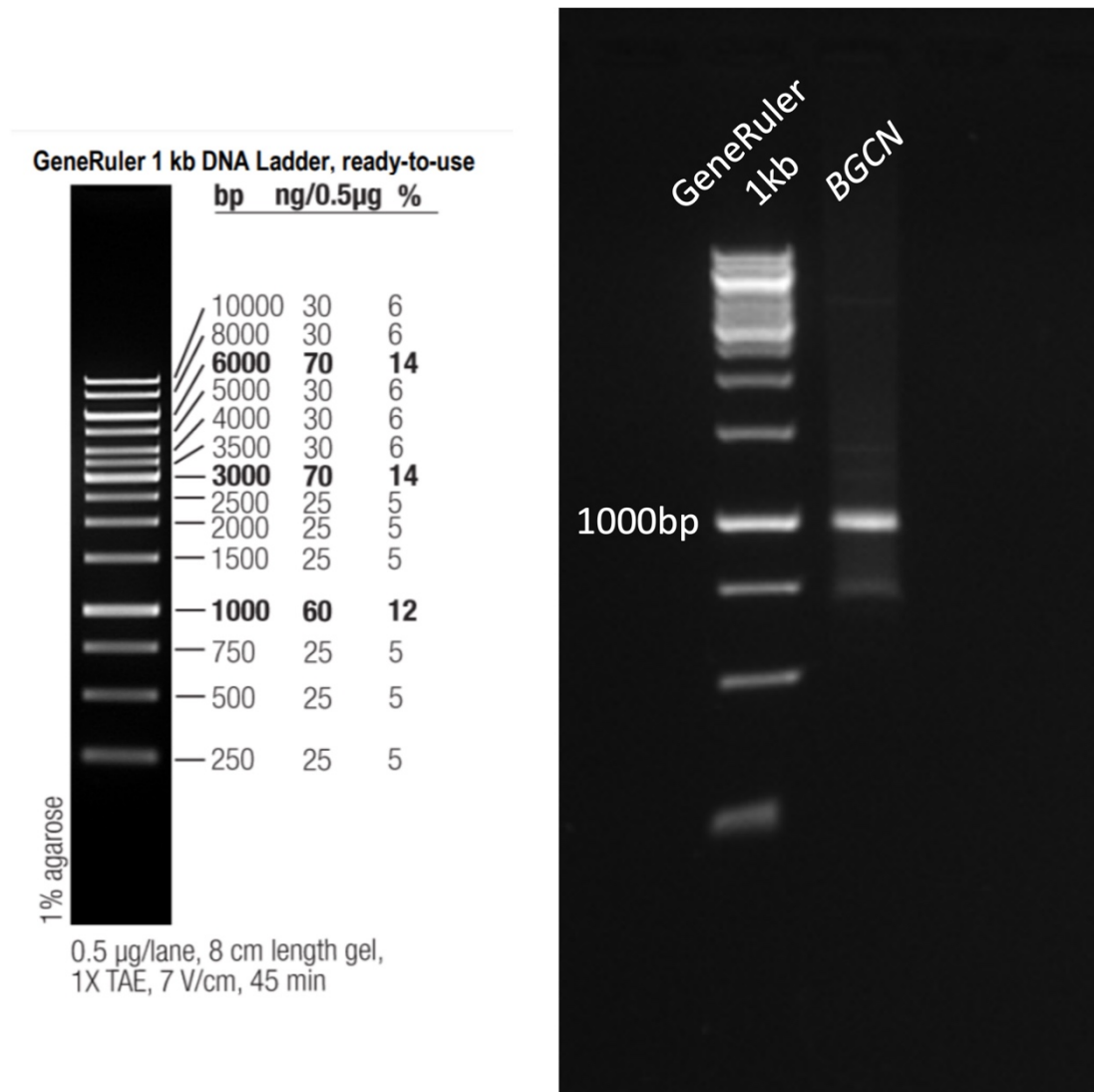
Figure 2.9 Nested RACE PCR of bgcn using PCR product from primer set 2 as the template to be amplified with primer set 3. Samples were run alongside the GeneRuler 1kb DNA Ladder. Order of loading: Ladder; *bgcn.* A band was present at the expected size of just over 1000bp. Sufficient product was produced for purification and subsequent sequencing.

The nested PCR was successful at amplifying a RACE product for *bgcn* (*Figure 2.9*) and this was then purified and prepared for sequencing. The newer primers for nesting of *mei-W68* were run with the product of *mei-W68* primer set 2, using the newly designed nesting primer (*mei-W68* primer set 4). The newly designed gene specific primers for *vasa* were used on RACE-ready cDNA as no products had been amplified to use for nesting. The resulting PCR products were visualised on a 1.2% gel (*Figure 2.10*).
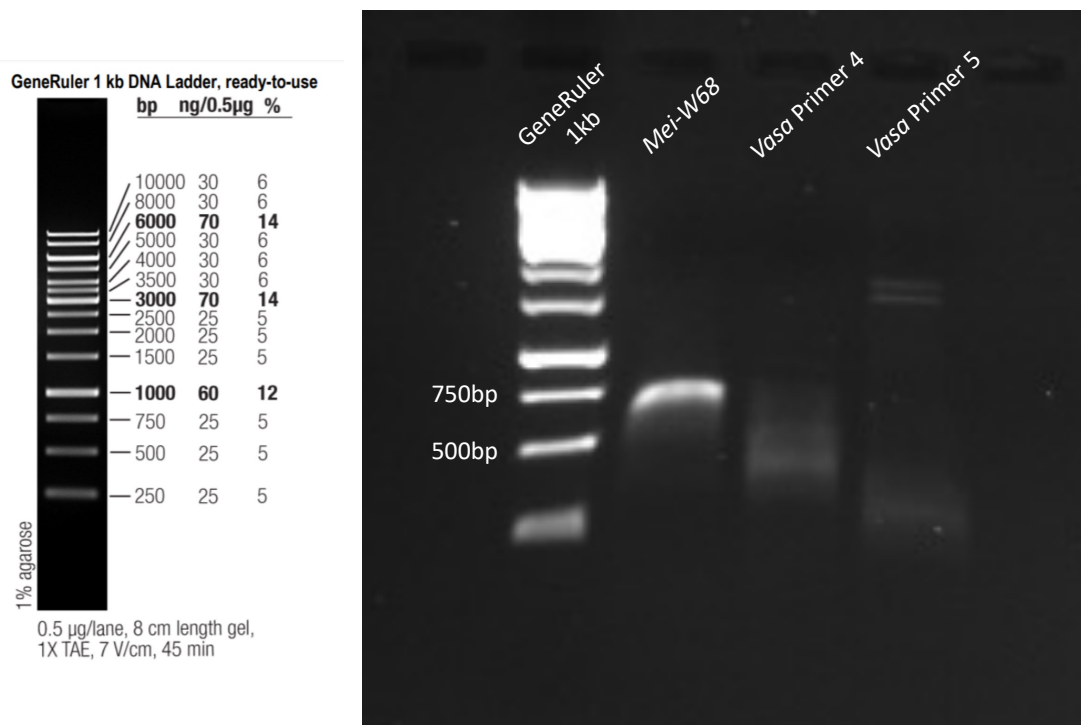
Figure 2.10 Nested RACE PCR of *mei-W68* using PCR product from primer set 2 as the template to be amplified with primer set 4. Race PCR of *vasa* was run with new primers 4 and 5 using RACE-ready cDNA as template. Samples were run alongside the GeneRuler 1kb DNA Ladder. Order of loading: Ladder, *mei-W68*, *vasa* primer 4, *vasa* primer 5. A band was present for *mei-W68* of the expected size (~ 750bp). No bands for *vasa* were of sufficient quality.

mei-W68 product was obtained through nested PCR (*Figure 2.10*). This was suitable for purification and sequencing alongside *innexin, bgcn* and *nanos*. A usable product could not be obtained for *vasa*. This gene has a very large potential promotor region (approximately 6kb) annotated on the genome, so a RACE sequence was needed to facilitate the use of a much smaller fragment. Given my troubleshooting of this *vasa* RACE reaction had not provided a resolution, I contacted a collaborator who had constructed a *vasa*-Cas9 plasmid for medfly, to discover whether they had tested a smaller promotor fragment. However, they had made the decision to use the full 6kb, which is significantly larger than the promotor size envisioned in the TARE designs for this project. Without the ability to confirm a shorter *vasa* promotor, the decision was made to focus on the other candidate genes.

2.4.2.2   Sequencing of RACE products and selection of promotor fragment

RACE products of *nanos, bgcn, innexin and mei-W68* purified and sent for sequencing with appropriate sequencing primers (*Supplementary information 2.6.1.2*). Sequencing results

were aligned to the medfly genome to identify the 5' UTR. All sequencing files aligned to genome in *Supplementary information 2.6.2*. Once the 5' UTR was identified for each candidate, approximately 2kb of the preceding genome sequence was selected as the promotor fragment, up to the start site of each gene.

2.4.2.2.1 nanos

The annotation of *nanos* on the medfly genome is detailed and the 5' UTR was identified. Hence the RACE sequence produced here could be used for sequence confirmation. The RACE sequence aligned well with the annotated 5' UTR confirming the genomic location. This allowed me to select a promotor region of 2.5kb which included 2kb of the 5' UTR. Using 2.5KB of *nanos* 5'UTR did result in the inclusion of the start of a mitochondrial import inner membrane subunit which is upstream of *nanos.* Only a small part of this was included, and it was not expected to result in any transcription of this off-target gene (shown in *Supplementary information 2.6.2)*. The sequence of identified *nanos* fragment to be extracted from the genome can be found in *Supplementary information 2.6.2* with the primers to be used for gDNA amplification in *Supplementary information 2.6.1.3*.

2.4.2.2.2 innexin

innexin had the mRNA regions annotated on the genome, which aligned well with the RACE read. A 2340bp region including the 5' UTR was identified as potential promotor and selected. This included the start site for an uncharacterised gene which was annotated upstream of *innexin*. Hence, taking only 2340bp (2kb of 5'UTR) ensured that only a small section of this off-target sequence was included, meaning that its transcription should not be possible. The sequence of identified *innexin* fragment to be extracted from the genome can be found in *Supplementary information 2.6.2* with the primers to be used for gDNA amplification in *Supplementary information 2.6.1.3*.

2.4.2.2.3 bgcn

bgcn had an annotated mRNA sequence on the genome. However, the RACE sequencing results were consistently shorter than that of the annotated region on the genome. Sequencing was attempted with 3 different primers, none of which achieved full coverage. This was interpreted as the 5' UTR beginning potentially slightly later than that annotated on the genome. This resulted in a difference of approximately 100bp, and little change in the final promotor region which included 2kb of 5'UTR. Upstream of the 5' UTR are two

annotated coding regions, with one as close as 500bp. To pull the 2kb upstream of the 5' UTR, both would be retained. Neither coding region was fully retained in the region selected as the promotor with the end of an actin related protein and the start of ubiquinone assembly factor 3 being the sections included. The sequence of identified *bgcn* fragment to be extracted from the genome can be found in *Supplementary information 2.6.2* with the primers to be used for gDNA amplification in *Supplementary information 2.6.1.3*.

2.4.2.2.4 *mei-W68*

mei-W68 had an annotated box A binding factor which aligned well with the RACE sequencing read, making the start of the 5' UTR identification simple. The end of the 5' UTR was slightly more difficult to identify, as the annotated start of the gene did not begin with a canonical (ATG) start codon. Due to this, the start of translation had to be predicted by locating the nearest start codon with an appropriate reading frame. The primer designed for sequencing *mei-W68* was misplaced, causing an issue with covering the entire 5' UTR. Intron 1 was not covered in the RACE sequence therefore could not be confirmed. Had the aim been to engineer the 5' UTR from sequencing, it would have been necessary to repeat this to confirm this region. However, as the whole 5' UTR is being included in promotor fragments, the confirmation was deemed unnecessary. When selecting the genomic region to identify as the promotor, the entirety of the coding region for an ADP-ribodylation factor-like protein 6 was included in the 2.5kb selected (2kb of 5'UTR). This could not be avoided as it is too close to the 5' UTR to omit. The sequence of identified *mei-W68* fragment to be extracted from the genome can be found in *Supplementary information 2.6.2* with the primers to be used for gDNA amplification in *Supplementary information 2.6.1.3*.

2.4.2.3 Extracting the promotor region from gDNA

Promotor regions were extracted from TOLIMAN gDNA using a high-fidelity polymerase (NEB Q5®) using the following primers: *mei-W68* amplified with primers 132 and 133; *bgcn* amplified with primers 134 and 135; *nanos* amplified with primers 136 and 137; *innexin* amplified with primers 138 and 139 (*Supplementary information 2.6.1.3)*. All primers had appropriate overhangs for Gibson assembly. Promotor parts were obtained for *mei-W68* and *nanos* on the first attempt. *bgcn* had only a weak band present. Hence to generate sufficient DNA for purification the PCR was redone using the PCR product from the previous run as template DNA. This generated sufficient DNA for purification. *innexin* had no band present after the first PCR or the subsequent PCR run at a lower temperature. Flanking primers 83

and 84 had been designed to create a product which could be used for a nested PCR. However, this PCR was run but no product was observed. *innexin* was then dropped at this point as troubleshooting attempts had not resolved the difficulties. *mei-W68*, *bgcn* and *nanos* products were purified, and their concentrations quantified using a nanodrop.

Table 2.5 Concentrations of promotor parts for Gibson Assembly.

Part	260/280	260/230	Concentration (ng/µl)
bgcn	1.78	1.36	160.5
nanos	1.56	1.10	32.88
mei-W68	1.69	0.45	12.60

The quantification results shown in *Table 2.5* indicate that only *bgcn* had a sufficiently high concentration to be confidently used in the subsequent Gibson assembly. *nanos* and *mei-W68* PCR repeated using the purified product as template which generated final concentrations of *nanos* at 126.3ng/µl and *mei-W68* at 154.6ng/µl.

2.4.3   Cas9 plasmid design and construction

2.4.3.1   Plasmid components

2.4.3.1.1   Development of identifying fluorescent marker

To identify transgenics a marker gene must be contained within the inserted construct to create an easy to detect phenotype only expressed in those individuals that carry the insertion. The fluorophore DsRed2 (Nishizawa et al. 2006) was used as the marker in all Cas9 plasmids. DsRed2 expression was driven by the commonly used insect expression promotor Hr5-IE1 (Jarvis et al. 1996) with the *Drosophila* native K10 acting as the 3'UTR and terminator. A nuclear localisation signal (NLS) flanked the DsRed2, which allowed for a clearer phenotype by restricting DsRed2 expression to the nucleus. This creates a distinct spotting pattern on the pupal surface in transgenic individuals that is more easily distinguishable from the autofluorescence that is observed in all individuals. The entire DsRed2 marker region, including the 5' and 3' UTR sequences, was taken from the plasmid AGG1103 (*Supplementary*

information 2.6.2) kindly provided by T. Harvey-Samuel (project collaborator Arthropods genetic group at the Pirbright Institute). PCR was completed using primers 96 and 107 (*Supplementary information 2.6.1.3)* with the final concentration after DpnI digestion of 52.65ng/µl.

2.4.3.1.2 Cas9 Protein

The Cas9 coding sequence used was the human codon-optimised *S. pyogenes* Cas9 (hSpCas9) flanked by nuclear localisation signals. This was extracted from plasmid AGG1583 originally synthesised at the Pirbright Institute Arthropod Genetics Group. The PCR was completed using primers 91 and 109 (*Supplementary information 2.6.1.3*), with the final concentration after DpnI digestion and repeat PCR on the product being 201.1ng/µl. The 3' UTR for read termination was *Autographa californica multiple nucleopolyhedrovirus* (AcMNPV) P10 taken from plasmid AGG1522 synthesised at the Pirbright Institute. The PCR was completed using primers 106 and 108 (*Supplementary information 2.6.1.3*), with the final concentration after DpnI digestion and repeat PCR on product being 125.5ng/µl.

2.4.3.1.3 Plasmid backbone

The parts of the plasmid outlined above were all to be inserted into the medfly genome. The *piggyBac* transposition system (X. Li et al. 2013) was proposed to be used for genome integration. Two *piggyBac* arms flanked the region for insertion, as taken from plasmid AGG1583. On the outside of these arms are the elements needed for plasmid propagation in *E. coli.* The standard pUC origin of replication was used with an AmpR sequence coding for Ampicillin resistance. As this was in the plasmid containing the *piggyBac* arms, this whole section could be amplified as a continuous, single fragment. The PCR was completed using primers 124 and 125 (*Supplementary information 2.6.1.3*), with the final concentration after DpnI digestion being 48ng/µl.

2.4.3.2 Final plasmid designs

Plasmids were to be assembled with Gibson assembly for scarless assembly as parts are designed with complementary overhangs. PCR was used to amplify specific parts and was completed by using overhang primers which produced compatible ends between the amplified fragments. Three plasmids were constructed as below. All of the plasmids used the same backbone, but varied in terms of the promotor regions. The assembly process of plasmids is described in *2.4.3.3.*
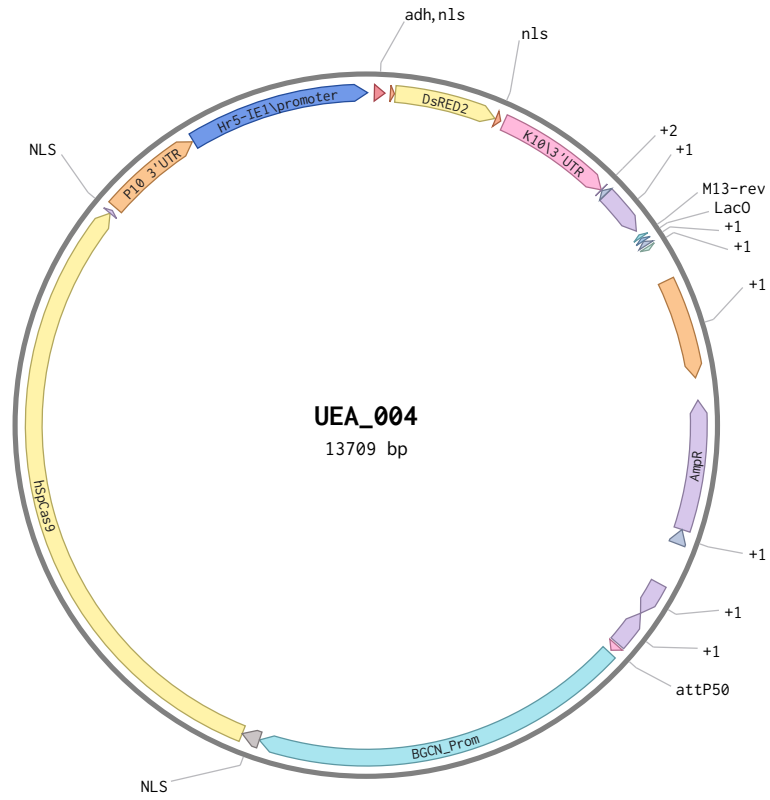
Figure 2.11 Final plasmid design for *bgcn*-Cas9 Plasmid. The full sequence can be found in the *Supplementary information* 2.6.2.
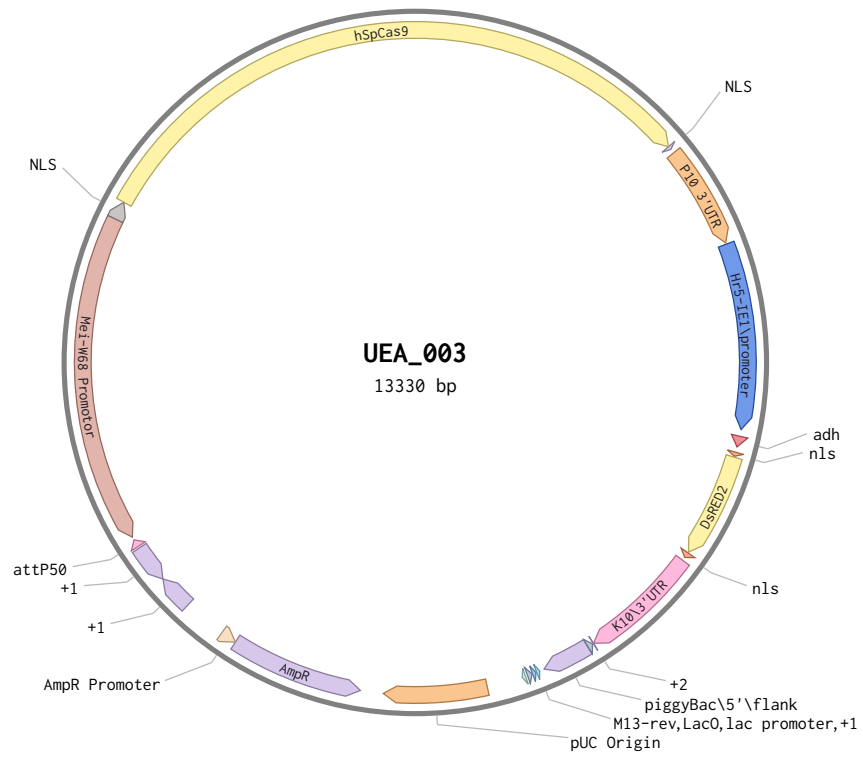
Figure 2.12 Final plasmid design for *mei-W68*-Cas9 plasmid. The full sequence can be found in the *Supplementary information* 2.6.2.
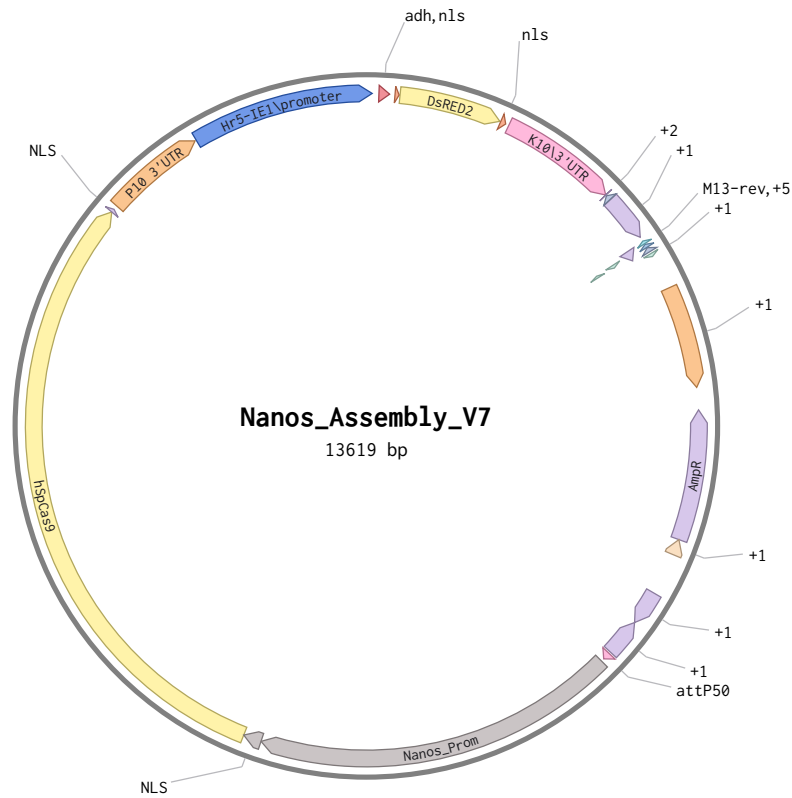
Figure 2.13 Final plasmid design for *nanos*-Cas9 plasmid. The full sequence can be found in the *Supplementary information* 2.6.2.

2.4.3.3   Plasmid construction

All plasmid assemblies were attempted simultaneously using the GeneArt Gibson Assembly, with the Cas9 part used as the vector for assembly. Post Gibson transformation was completed into Takara Stellar™ competent *E. coli* cells, and colonies were present for all reactions. Ten colonies were selected from each of the *nanos*-Cas9 and *mei-W68*-Cas9 plates. The *bgcn*-Cas9 reaction yielded only seven colonies which were all selected for colony PCR. Four colony PCR reactions were performed which covered every insert junction to ensure insertion of all parts. The presence of a positive band for all reactions allowed me to select which bands could be subjected to sequencing. *mei-W68* visualised on a single agarose gel shown in *Figure 2.14*, with reactions 1-4 for all colonies being shown.
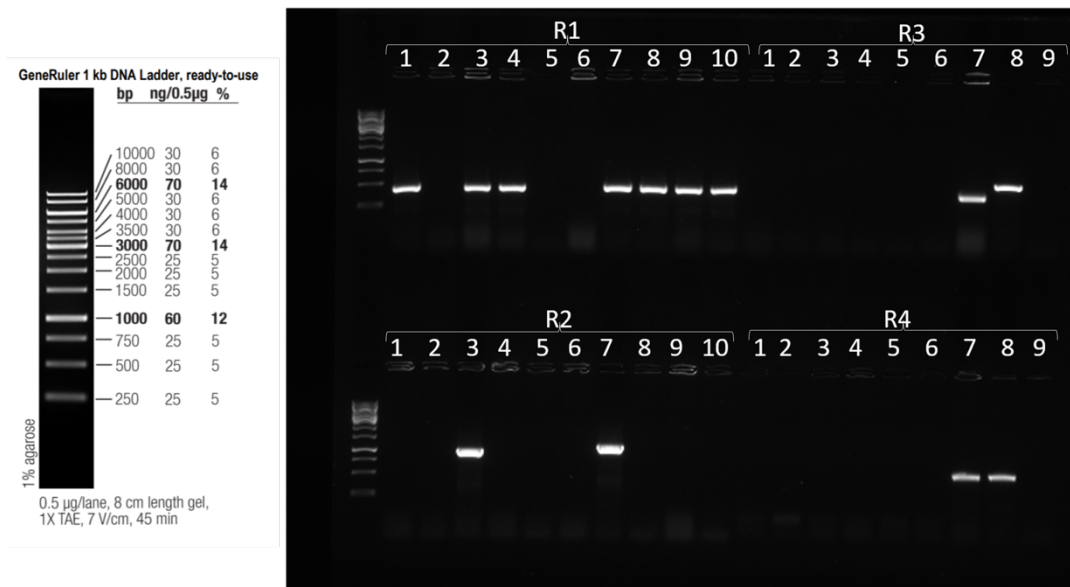
Figure 2.14 Colony PCR of *mei-W68*-Cas9 transformants. The gel was run alongside the GeneRuler 1kb DNA Ladder. Order of loading upper section, ladder, reaction one (colonies 1-10), reaction 3 (colonies 1-9). Order of loading lower section; ladder, reaction two (colonies 1-10), reaction 4 (colonies 1-9). Four reactions were performed, each covering a different insertion junction. Reaction one used primers 156 and 157, reaction two used primers 162 and 163, reaction 3 used primers 150 and 161, reaction 4 used primers 151 and 158. All primer sequences available in *Supplementary information* 2.6.1.3. Colony number is displayed and was loaded in same order for each reaction. Colony 10 reaction 3 and 4 is absent due to lack of available lanes on gel. A band had to be present for the same colony in all 4 reactions to be selected for sequencing. Colony 7 was the only positive colony with product for all 4 reactions. The product for reaction 3 was slightly smaller than its expected size (approximately 100bp smaller than expected).

The *mei-W68* colony PCR results showed only one fully assembled construct. Colony 7 had a product across all four reactions. The band size for reaction 3 was slightly smaller than that expected for a fully assembled plasmid, with the correct size band being visible for colony 8 for comparison. This colony PCR reaction covered the *mei-W68*/Cas9 junction. As this was the only colony with all parts present, it was selected for sequencing to determine the problem with assembly. Short read sequencing was used with primers 150, 151, 156, 157, 158, 161, 162 and 163 which covered all junctions (*Supplementary information 6.2.1.3*).

Colony PCR of *nanos* and *bgcn* were visualised together with reactions one and two being shown in *Figure 2.15* and reactions 3 and 4 in *Figure 2.16*.
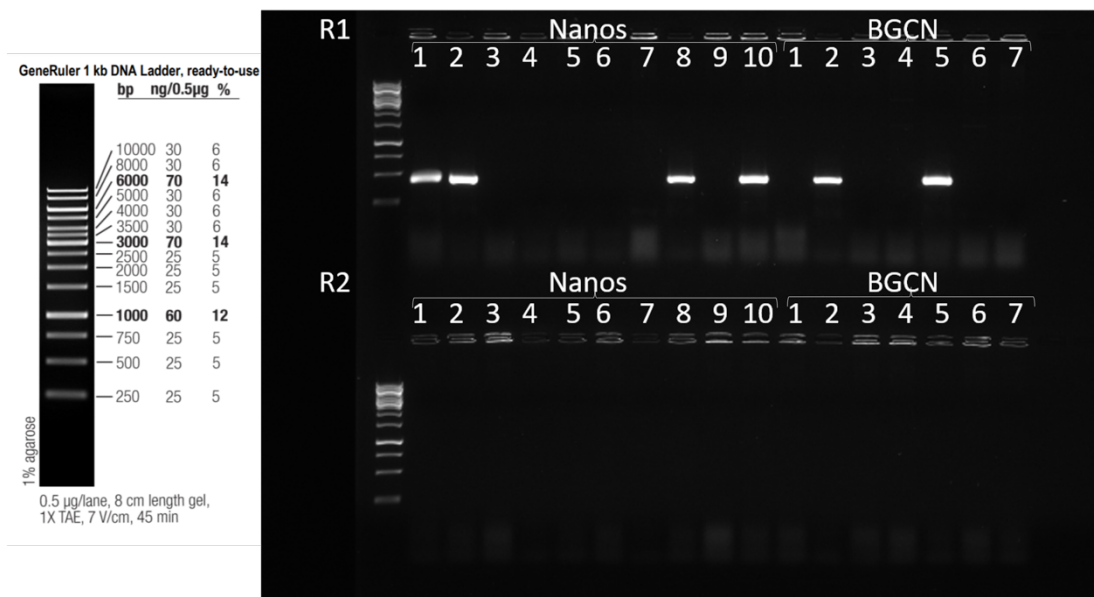
Figure 2.15 Colony PCR reactions one and two for *nanos-Cas9* plasmid and *bgcn-Cas9* plasmid. Gel is run alongside the GeneRuler 1kb DNA Ladder. Order of loading: Upper section; ladder; reaction one (*nanos-Cas9* colonies 1-10); reaction one (*bgcn-Cas9* colonies 1-7). Lower section; ladder; reaction two (*nanos-Cas9* colonies 1-10); reaction two (*bgcn-Cas9* colonies 1-7). Two reactions were performed, each covering a different insertion junction. Colony number is displayed and loaded in same order for each reaction. Reaction one used primers 156 and 157. Reaction two used primers 162 and 163. All primer sequences available in *Supplementary information 2.6.1.3*.
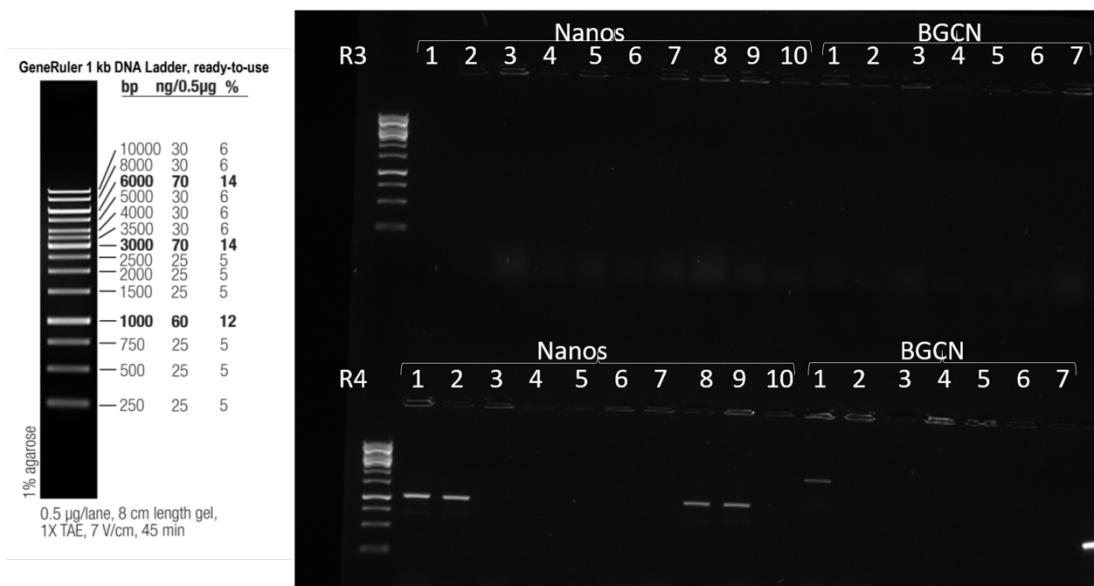
Figure 2.16 Colony PCR reactions three and four for *nanos-Cas9* and *bgcn-Cas9*. Gel is run alongside the GeneRuler 1kb DNA Ladder. Order of loading: Upper section; ladder; reaction three (*nanos-Cas9* colonies 1-10); reaction three (*bgcn* colonies 1-7). Lower section; ladder; reaction four (*nanos-Cas9* colonies 1-10); reaction four (*bgcn-Cas9* colonies 1-7). Two reactions were performed, each covering a different insertion junction. Colony number is

displayed and loaded in same order for each reaction. Reaction three used primers 161 and 160 (*bgcn-Cas9*) or 152 (*nanos-Cas9*). Reaction four used primers 158 and 159 (*bgcn-Cas9*) or 153 (*nanos-Cas9*). All primer sequences available in *Supplementary information 2.6.1.3*.

nanos-Cas9 and *bgcn*-Cas9 produced no colonies with a product for all reactions, with reactions two and three being unsuccessful across all samples as no product was produced. This indicates that at least one of the assembly parts covered by these PCRs was not present therefore the primers could not bind resulting in no product. None of these colonies were assessed as strong candidates to be sent for validation by sequencing.

The sequencing results received for *mei-W68*-Cas9 colony 7 are shown in the *Supplementary information 2.6.2*. All junctions where two parts were annealed appeared correct, with full assembly except that of the *mei-W68* and Cas9. All sequencing reads showed the presence of both fragments by covering the junction where they connect meaning the sequencing read should be able to confirm the presence of two parts. These junctions are also where mutations are most likely to occur and therefore these reads will identify any issues around these sites. In the junction where the *Meiw-68* promotor and Cas9 part anneal, the Cas9 appeared to stop after approximately 100bp which matches with the results observed in the colony PCR. This means the Cas9 had most likely truncated, preventing it from forming a functional protein due to the significant portion of missing sequence. That the Cas9 fragment was the only concern in this assembly means that this plasmid retained its potential for future use in constructions, to minimise the number of parts going into the Gibson assembly reaction, which typically improves success. For the *mei-W68* assembly, all parts except the Cas9 had successfully been assembled. The Cas9 part and *mei-W68* backbone were obtained through PCR. A universal backbone fragment was obtained through PCR which contained the P10, marker and backbone parts fully assembled using primer 125 and 108 (*Supplementary information 2.6.1.3*). This produced two bands, and gel excision and purification was used to ensure the presence of only the larger band, with the final concentration being 38.09ng/μl.

To understand to what extent the Cas9 had truncated after assembly, a PCR was performed to check the size of the Cas9 fragment in *mei-W68*-Cas9 colony 7. This showed a product of 600bp as opposed to the expected 4284bp of the full fragment. During part generation of the Cas9 fragment, there was a weak, smaller band present which I suggest could be interfering with the assembly. The Cas9 part generation PCR was repeated and the correct sized band was gel excised to ensure the smaller part could not interfere with the reaction. The final concentration of this new Cas9 fragment was 19.65ng/μl.

The concentration of the gel excised fragments was lower than optimal for use in Gibson assembly. In an attempt to generate more product, the excised fragments were used as template in a repeat of the same PCRs with a higher reaction volume, which could then be loaded to the gel giving more product to purify. This procedure yielded final concentrations of 28.81ng/µl of Cas9 and 32.66ng/µl of the backbone. This was still not sufficient to allow for the use of the recommended 0.08pmoles in the Gibson reactions and I used 0.05pmoles in the reactions instead. Transformed into Takara Stellar™ competent cells and colony PCR was performed on the resulting colonies. As *mei-W68* was a two part assembly, only two colony PCRs needed to be completed. *nanos* and *bgcn* were three part assemblies, making 3 colony PCRs necessary. The results of the colony PCRs were visualised using gel electrophoresis with the results summarised in *Table 2.6*.

Table 2.6 Colony PCR results of transformants. Black indicates no correct sized product, green indicates correctly sized product amplified and detected on the gel.

Construct	Colony	Reaction 1	Reaction 2	Reaction 3
mei-W68-Cas9	1	green	green	
	2	green	green	
	3	black	black	
	4	black	green	
	5	black	black	
	6	green	green	
	7	black	black	
	8	green	green	
	9	green	green	
	10	green	green	
nanos-Cas9	1	black	black	black
	2	black	green	green
	3	green	black	black
	4	black	black	black
	5	green	green	green
	6	green	black	black
	7	green	green	black
	8	black	black	black
	9	green	green	black
	10	green	green	green
bgcn-Cas9	1	green	green	green
	2	green	green	green
	3	green	green	green
	4	green	green	green
	5	black	green	green
	6	green	green	green
	7	green	green	green
	8	green	green	green
	9	green	green	green
	10	green	black	black

All assemblies had colonies with all bands present, those highlighted in bold in _Table 2.6_ were sent for short read sequencing of all junctions (_Supplementary information 2.6.2_). _mei-W68-_Cas9 reads all aligned well, colony 10 was chosen to carry forward, as colony 9 had a single base pair deletion at the end of the promotor region. The _nanos_-Cas9 constructs were not suitable for further progression, due to a common issue across both colonies, of Cas9 truncation. _bgcn_-Cas9 results indicated a problem with the promotor fragment in which the sequencing reads would break down after covering a few base pairs of the promotor region. Size confirmation of the promotor of _bgcn_-Cas9 and the Cas9 of _nanos_-Cas9 was completed. Sequencing results of _mei-W68_-Cas9 did not show any evidence of Cas9 truncation. However, not all of the Cas9 region was sequenced. Therefore, the size confirmation of _mei-W68_-Cas9 colony 10 Cas9 part was also completed. Products were visualised by gel electrophoresis (_Figure 2.17_).
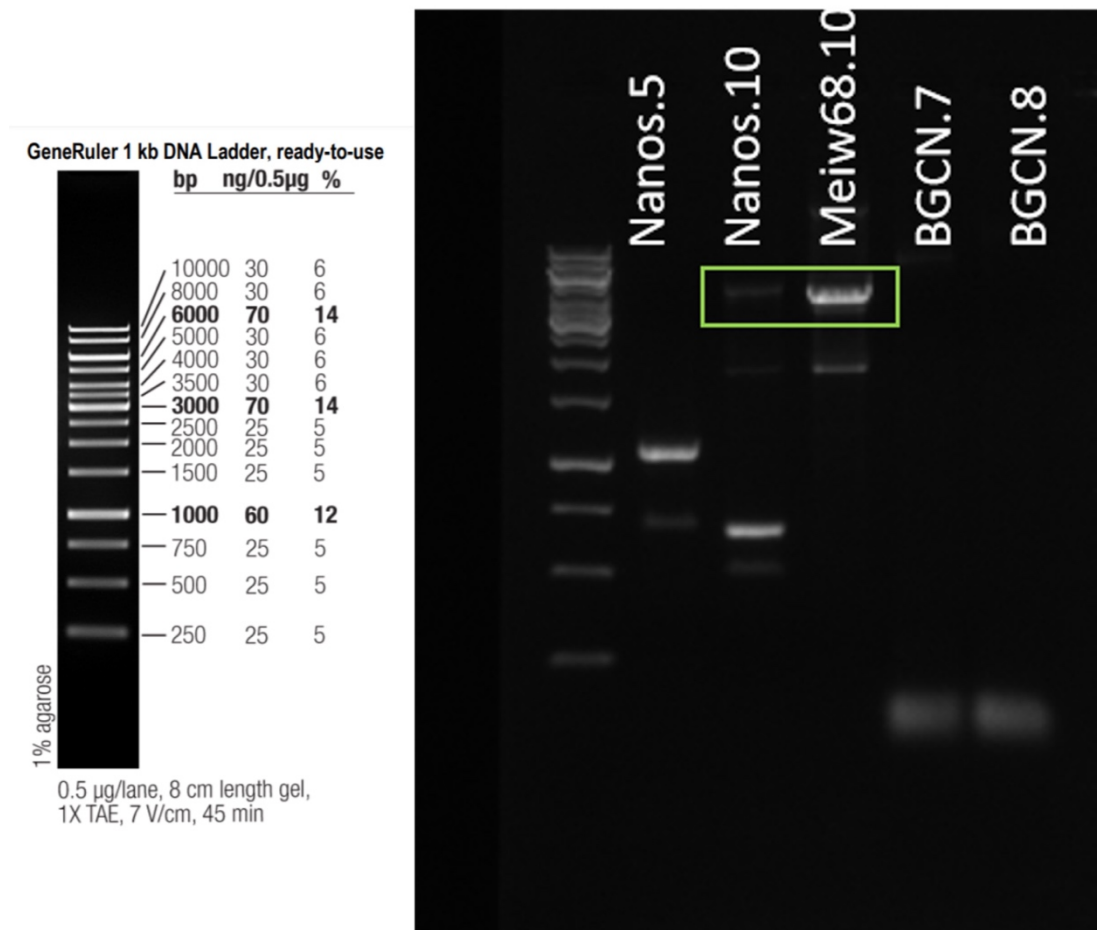
Figure 2.17 Size confirmation of parts of assembled plasmids. Bands indicated in the green box are of the expected size. Products were run alongside the GeneRuler 1kb DNA Ladder. Order of loading; ladder; _nanos_-Cas9 colony 5 Cas9 confirmation; _nanos_-Cas9 colony 10 Cas9

The PCR results confirmed the truncation of Cas9 in *nanos*-Cas9. A weak band was visible at the correct size for *nanos*-Cas9 colony 10, with the much smaller product present at a much higher concentration. A band of the correct size was seen for *mei-W68*-Cas9, providing confidence that truncation was not an issue in this construct, which was then taken forward for midiprep. No bands were produced for the promotor fragment of *bgcn*-Cas9, confirming that this had not been inserted. To further reduce the number of parts for Gibson assembly, *mei-W68*-Cas9 colony 10 was used to provide a backbone for assembly of *bgcn*-Cas9 by creating a fragment which covered the entire plasmid except the promotor region, with primers 91 and 125 (*Supplementary information 2.6.1.3*). A backbone for *nanos*-Cas9 can be produced using *nanos*-Cas9 colony 5 with primers 108 and 137 (*Supplementary information 2.6.1.3*) to generate a fragment containing all parts except the Cas9 insert. This strategy allows for a two-part assembly for both constructs, reducing the difficulty of assembly. To reduce the risk of recombination, all Gibson reactions from this point onwards were transformed into low recombination Agilent SURE®2 Supercompetent Cells. *bgcn*-Cas9 assembly only produced five colonies, 3 of which appeared correctly assembled after colony as shown by the PCR results (*Figure 2.18*).
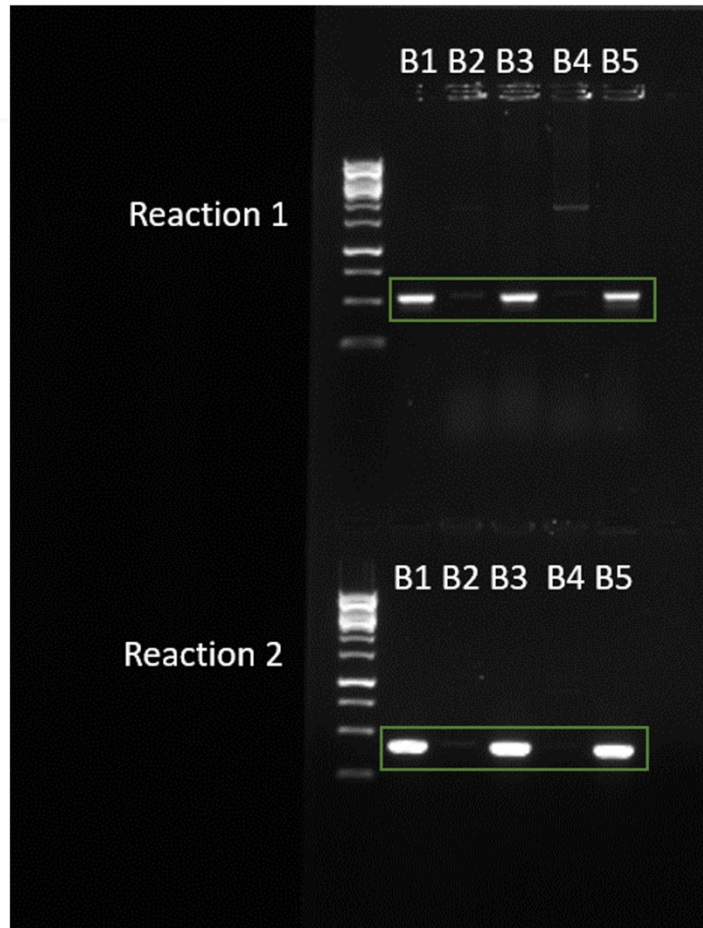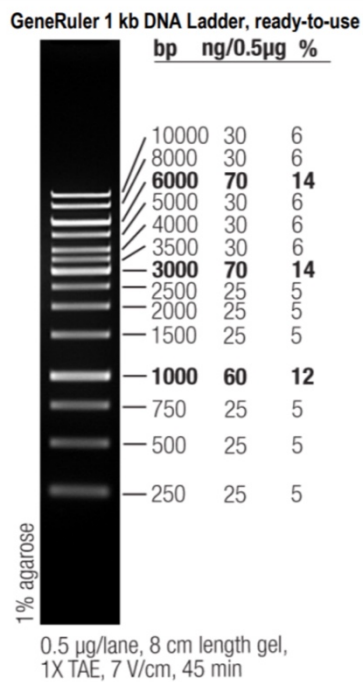
Figure 2.18 Colony PCR of *bgcn*-Cas9 from two part Gibson assembly. Bands indicated in the green box are of the expected size. The gel was run alongside the GeneRuler 1kb DNA Ladder. Order of loading: upper section; ladder; reaction one (colonies 1-5); lower section; ladder; reaction two (colonies 1-5). Colonies 1, 3 and 5 all appear correctly assembled and were sent for sequencing.

Sequencing results of *bgcn*-Cas9 colony 3 (*Supplementary information 2.6.2*) aligned well, revealing no issues with assembly. This construct was taken forward for midiprep. All transformation attempts with *nanos* resulted in truncated Cas9. Gibson assembly was reattempted in the same configuration that had been successful for *bgcn*, inserting *nanos* into the promotor free backbone generated from *mei-W68*-Cas9. No colonies were obtained from several Gibson assembly attempts to insert *nanos* promotor into *mei-W68*-Cas9 backbone. With all troubleshooting attempts failing from this point, only the *mei-W68* and *bgcn* Cas9 constructs were prepared for microinjection.

2.4.4 Establishment of transgenic lines

Microinjection survivors were received as pupae. Due to logistical and time constraints caused by covid-related issues, the pupae received were set up as inbreeding crosses. *mei-*

W68-Cas9 (UEA_003) pupae were divided equally between four small cages. *bgcn*-Cas9 (UEA_004) had slightly higher survivor numbers to set up 5 cages. Offspring from these cages screened as adults with 7 transgenics found in total. All UEA_004 transgenics came from the same cage (cage 2) and were therefore designated bgcn.2 with differing letters to identify the different crosses set up, the two UEA_003 transgenics came from different cages (cages 3 and 4) and were designated M68.3 and M68.4. The following crosses were set up:

bgcn.2 A - 2 transgenic males x 5 wildtype Benakeion females

bgcn.2 B - 2 transgenic females x 2 wildtype Benakeion males

bgcn.2 C - 1 transgenic female x 2 wildtype Benakeion males

M68.3 - 1 transgenic female x 2 wildtype Benakeion male

M68.4 - 1 transgenic male x 6 wildtype Benakeion females

All offspring form these cages were screened to check number of transgenic individuals. With the use of inbreeding cages, multiple insertion events within a single line are a possibility which had to be checked. All lines had the expected 50% transgenic offspring except those from cage M68.4 of which approximately 75% of screened offspring were transgenic. This could be the result of multiple insertions, though further investigation was not possible at that time due to covid restrictions. The presence of multiple insertions would not negatively affect the experiments envisioned for these lines and could still be tested for at a later date if needed.

Stable transgenic lines were established from all five cages. Lines were maintained as heterozygotes for several generations before being homozygote enriched for ease of maintenance. An example of a stable transgenic pupa detected at the screening stage is shown in *Figure 2.19*.
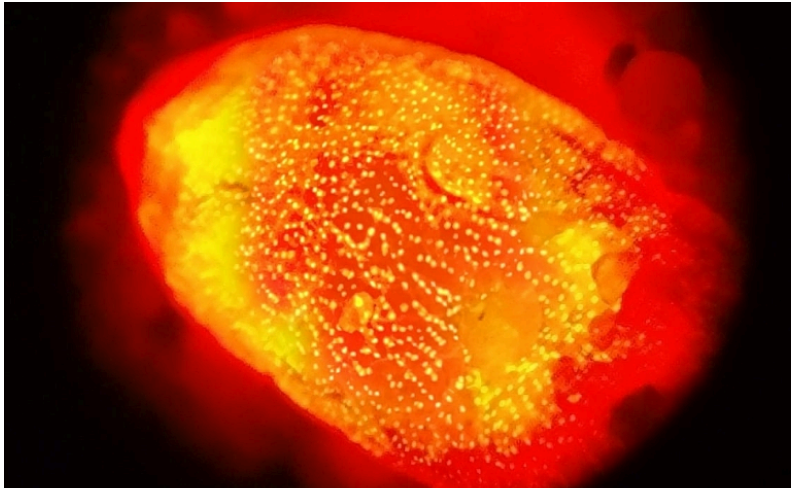
Figure 2.19 Stable transgenic pupa expressing DsRed marker under fluorescent microscope. The spotting pattern observed is due to the use of nuclear localisation signals isolating DsRED expression to the nucleus.

2.4.5   Testing the fitness of transgenic lines

2.4.5.1   Pupal Development

Pupal development was measured for two of the established transgenic lines (M68.3 and bgcn.2C) in comparison to the Benakeion wildtype. Pupal number was measured daily from 7 days post egg collection up 26 days post egg-collection (*Figure 2.20*).
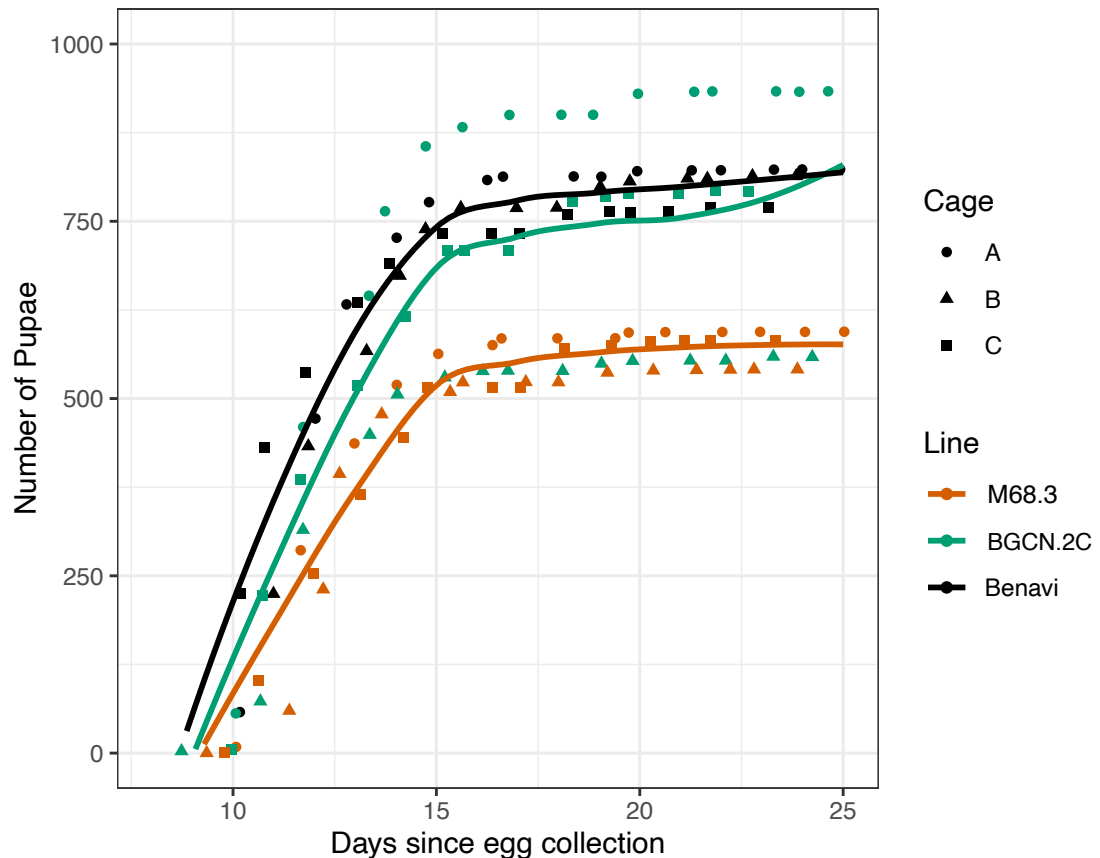
Figure 2.20 Number of pupae produced in the transgenic lines. The number of pupae emerging from each line was determined over a period of 26 days. Three different lines were measured with three biological replicates of each (transgenic lines M68.3 and bgcn.2C and the Benavi wildtype). Biological replicates comprised egg collections from 3 consecutive days. A line of best fit has been used for each line, averaging across biological replicates. Pupal number for bgcn.2C was very variable between biological replicates.

All lines showed a similar rate of pupal emergence, with most pupae emerging by day 15. Transgenic line bgcn.2C showed a similar final number of pupae number to that of the wildtype, suggesting no significant fitness costs in terms of the number and speed of pupae emerging. However, there three biological replicates and pupal survival varied widely between them, with one replicate being similar to M68.3, one similar to the wildtype and the third replicate even higher than the wildtype. Transgenic line M68.3 did appear to have consistently lower pupal number emergence, and to have slower development times to pupae, with very little variance across biological replicates, indicating the existence of some fitness costs. An ANOVA was used to test for differences in the final number of pupae by line. The model showed no overall evidence for a significant difference in final pupal number between lines ($F_{2,6}$ = 3.63, p = 0.093). However, within this, post hoc comparisons between lines showed evidence for a significantly lower number of pupae being produced by line M68.3 (t_6= -2.52, p=0.045) with an average of 229 fewer pupae than for the wildtype. There

was no evidence for a difference in the number of pupae produced by line bgcn.2C in comparison to the wild type (t_6 = -0.44, p = 0.673). P-values were computed using a Wald t-distribution approximation.

2.4.5.2 Pupae survivability

To test whether the transgenic lines had reduced eclosion rates from pupae, 150 pupae were selected from each line and followed through to eclosion. There was no significant difference detected between any of the lines in terms of eclosion rates of each transgenic line compared to wildtype (($F_{2,6}$ = 0.87, p = 0.467), analysed using ANOVA.

2.4.5.3 Testing the fecundity of Cas9 lines

To test the fecundity of the transgenic lines in comparison to the wild type, 50 flies (typical small cage set up with 50:50 sex distribution) were allowed to lay for 1h and the number of eggs produced in this time was then counted. Three biological replicates were of each line were set up and the findings are summarised in *Figure 2.21*.
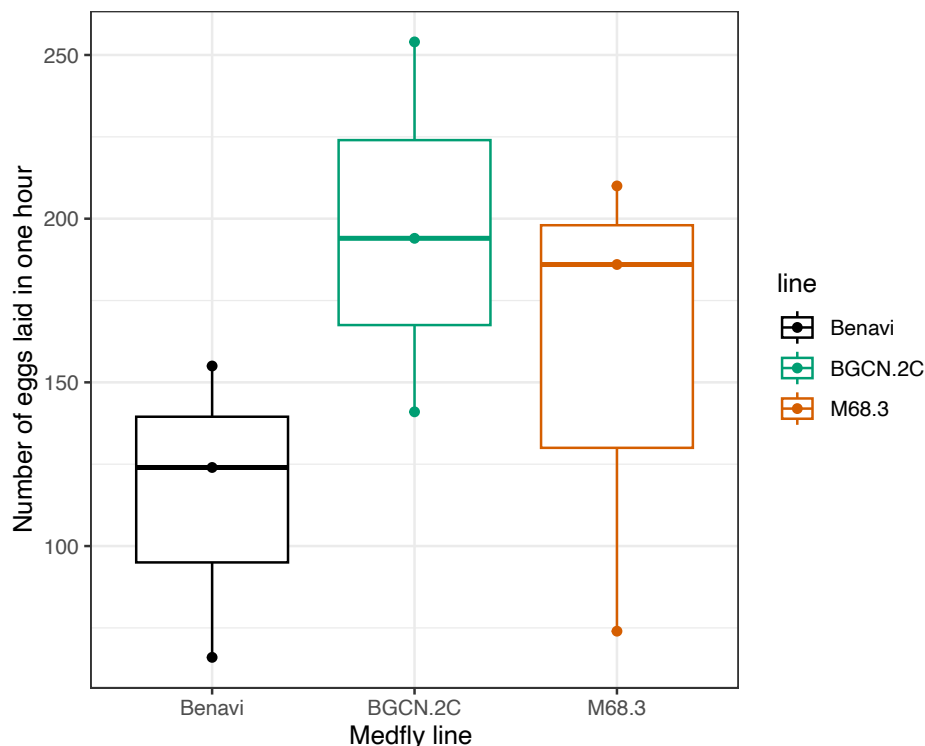
Figure 2.21 Number of eggs laid in one hour across the different lines (transgenic lines 00.3. and bgcn.2C and the Benavi wildtype). Egg collections were taken at day 7 post eclosion. Both transgenic lines produced a larger number of eggs than the wildtype line, giving no indication of a fecundity cost of carrying the transgene. The line within the boxplot represents the median value with the upper and lower bounds representing the first and 3rd quartile

respectively. The whiskers indicate the upper and lower values obtained with the three data points represented by the points.

There was no indication of any significant fecundity cost to carrying the transgene (*Figure 2.21*) and the transgenic lines appeared to have an even slightly higher level of egg laying than the wildtype (ANOVA; $F_{2,6} = 1.42$, p = 0.313).

2.4.5.4 Fitness tests

Stressed and unstressed adult survival tests were completed on cages of 50 flies with 3 biological replicates for each line. In stressed survival tests, no food or water was provided and deaths per day were measured until all flies were dead. In the unstressed survival test, due to the time taken for all flies to die, flies were monitored for approximately one month which allowed median deaths to be reached. There was no evidence of a loss in fitness in terms of adult survival in the transgenic lines as shown in *Figure 2.22* and the transgeneic lines survived as well as or even better than, the wild type.

This was supported by an analysis using a Cox Proportional-Hazards Model. In the stressed test conditions significant differences in survival were detected (Wald $\chi^2_2 = 47.2$, n =450, p-value <0.01). There was no significant difference detected in a comparison between Benavi and bgcn.2C (z = -1.310, p = 0.19) and a significant difference between Benavi and M68.3 (z = -6.492, p <0.01) indicating higher survival in M68.3. In the non-stressed test conditions, there was no evidence for a statistically significant difference between any of the lines (Cox Proportional-Hazards Model, Wald Wald $\chi^2_2 = 1.54$, 2, n =312, p-value = 0.5; no significant difference detected between either line and Benavi, with line M68.3 having a z of -1.242 with a p-value of 0.214, and line bgcn.2C a z of -0.558 with a p-value of 0.556).

A)
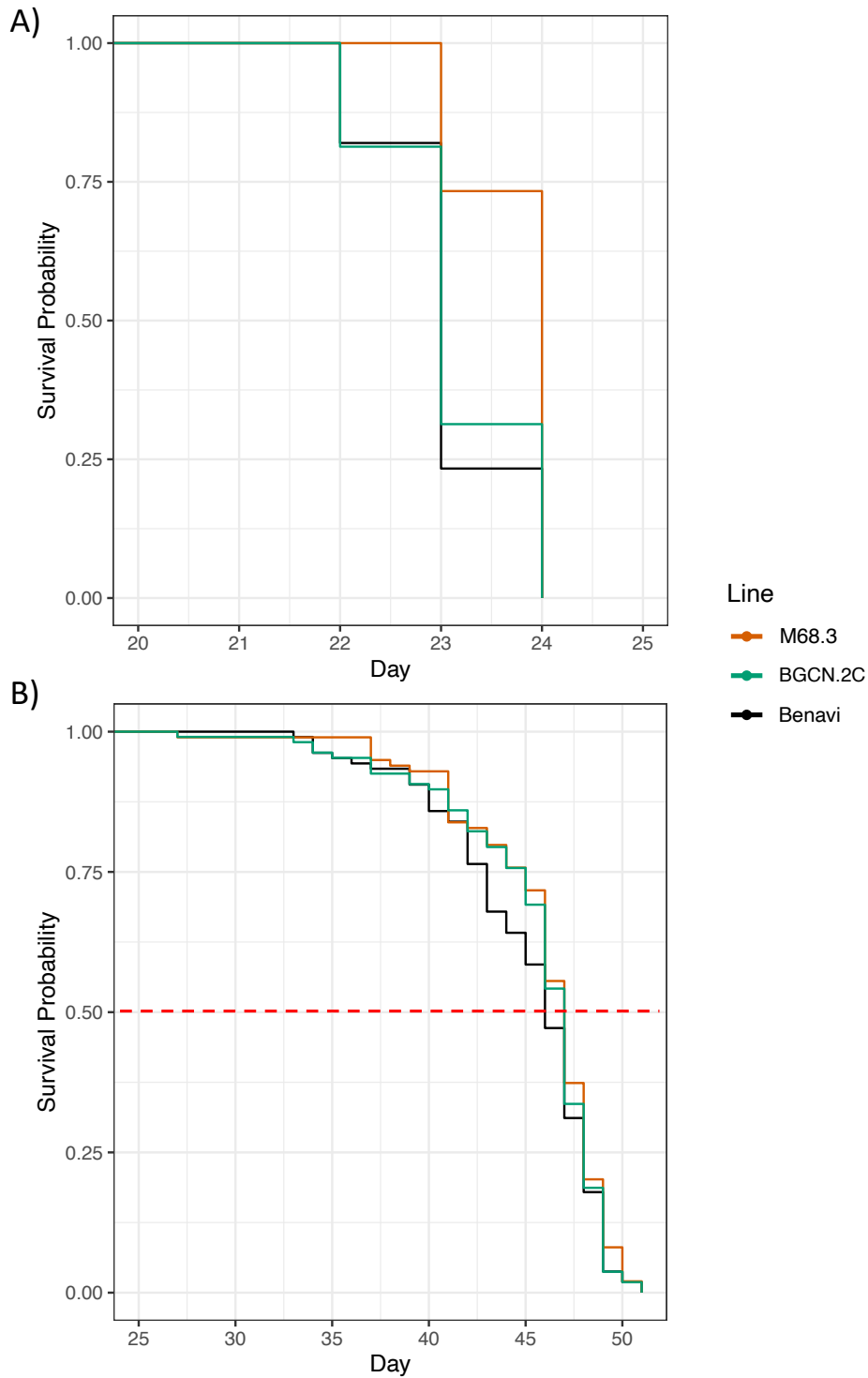
Figure 2.22 Survival test of 50 medfly in both stressed and unstressed conditions from day of egg collection with eclosion from pupae on day 21. A) Stressed adult survival fitness test with no access to food or water for eclosed flies with deaths recorded daily until all flies were dead (three biological replicates per line). B) Unstressed adult survival fitness test with deaths recorded daily. The median death day is indicated with the horizontal dashed red line. There was no evidence for a loss of survival in the transgenic lines in comparison to the control, under either non stressed or stressed conditions.

2.4.6 Validation of Cas9 expression with Real Time Quantitative PCR

2.4.6.1 Establishing reaction conditions

Standard curves were generated for two reference genes (*RPE* and *RPL19*) and Cas9 primers to determine primer efficiency. Cas9 and *RPE* results were generated from the ovary tissue, *RPL19* results were generated from testis tissue. Linear regression lines were plotted with the slope of this line used to calculate the primer efficiency (*Figure 2.23*).
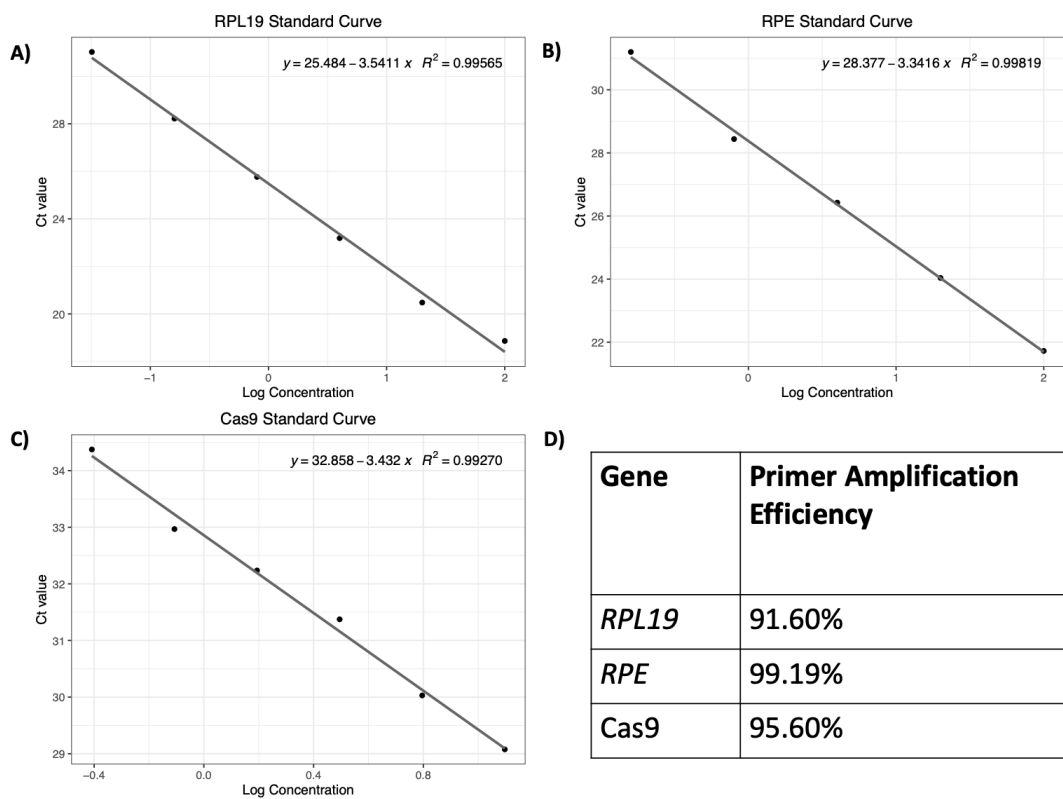
Figure 2.23 Standard curves of all primer sets. Ct value plotted against the log concentration. Samples with a Ct value greater than 34 were excluded due to poor read quality. Primer efficiency was calculated from the slope of the curve. A) Standard curve of *RPL19*. B) Standard curve of *RPE*. C) Standard curve of Cas9. D) Primer amplification efficiencies of all primers as calculated form the slope of the curve.

All primers had amplification efficiencies within the appropriate range to be used for qPCR (90%-110%). To obtain an appropriate standard curve for Cas9, the serial dilution factor used was reduced from a 5X dilution to a 2X dilution factor. This allowed for a higher number of reads to be obtained before the sample became too dilute to be detected before cycle number 34.

2.4.6.2 Testing of reference genes

To compare expression across tissue types, a reference gene which has stable expression across all tissue types to be tested (male soma, female soma, testes and ovaries) was needed. Expression levels across the tissue types were obtained for both reference genes (*Figure 2.24*).

RPE Expression

A)

RPL19 Expression

B)
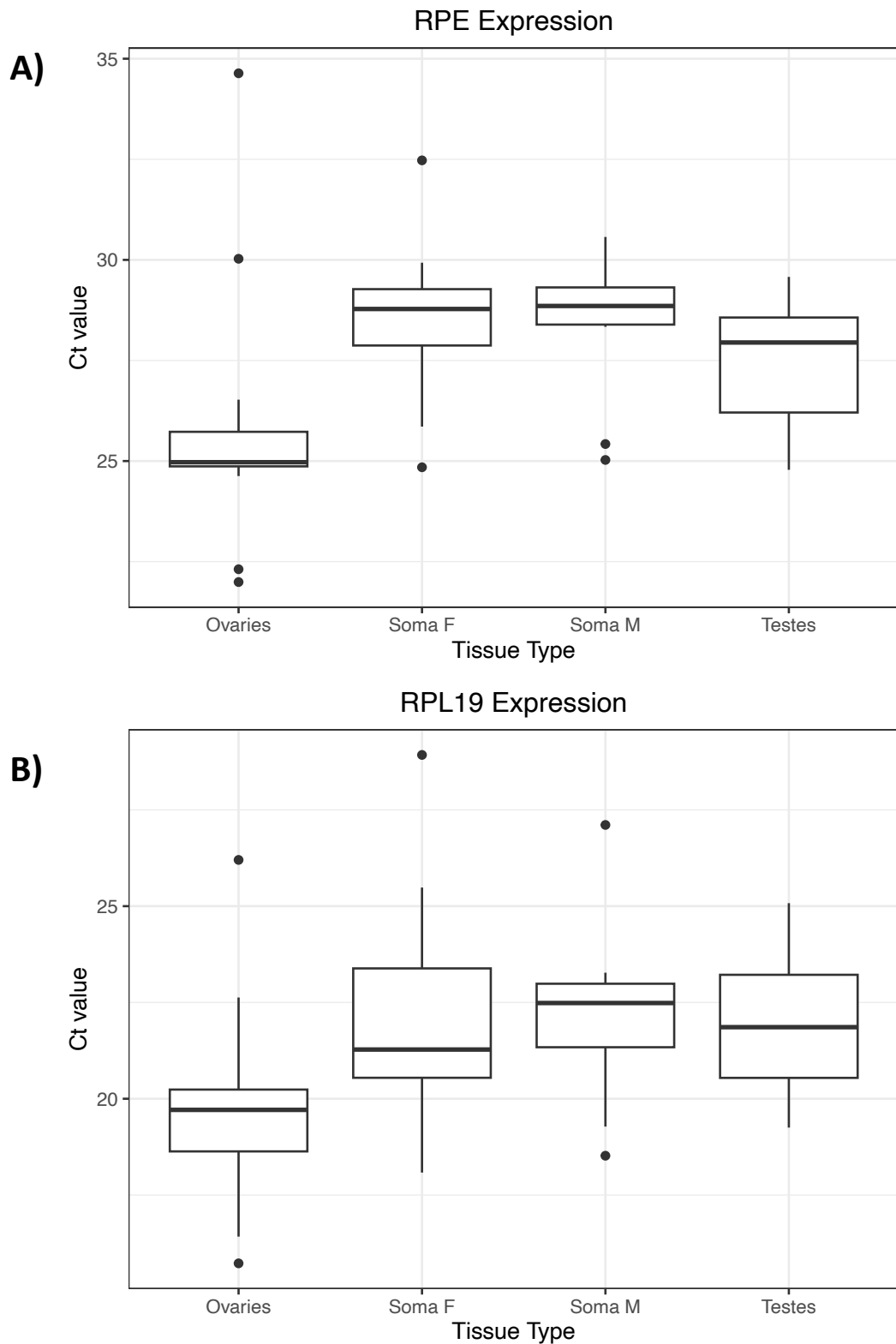
Figure 2.24 Ct values of reference gene expression as measured by qPCR, across all the tissue types tested. Technical replicates were averaged and all values standardised by using an interplate calibrator. A) The Ct values of *RPE* across tissue types. B) The Ct values of *RPL19* across tissue types. To allow for comparison between tissue types, a reference gene with consistent expression across tissue types is needed. In both *RPE* and *RPL19* expression in the ovaries was higher than that in the other tissue types.

Expression in the ovaries of both reference genes appeared higher than that in other tissue types. One-way ANOVA results (*Supplementary Information 2.6.3.1*) confirmed this and showed a significant difference in expression across tissue types (p <0.01). The consistency of the ovary samples, having the highest level of expression across reference genes, could indicate that the ovaries do have higher levels of cDNA which would make the reference genes a good control. The ovary samples did have the highest concentration of RNA after RNA extraction and although all samples were diluted to the same concentration for cDNA synthesis, samples of a higher starting concentration may be of better quality for cDNA synthesis. *RPE* had a lower level of expression than that of *RPL19* resulting in a greater number of samples needing to be excluded because the Ct value was too high, or because of divergence between the technical replicates (of > 0.5 cycles). Therefore, *RPL19* was used as the single reference gene owing to its increased stability of expression across tissues.

2.4.6.3   Expression of Cas9

The expression of four Cas9 transgenic lines was measured (for lines bgcn.2A, bgcn.2C, M68.3, M68.4) across tissue types. Cas9 expression was also measured in the Benavi wildtype to check for off-target amplification. No wildtype sample showed any Cas9 amplification. Relative expression was calculated with *RPL19* as the reference gene. Initially, there were 3 biological replicates of each tissue type per line. However, some biological replicates were excluded if not all 3 technical replicates were within 0.5 cycles times, or if Ct values were above 34. Replicates excluded included two bgcn.2C ovary samples, one M68.4 female somatic sample, one bgcn.2A male somatic sample and one M68.4 male somatic sample. All remaining samples relative expression shown in *Figure 2.25*.

A)

B)
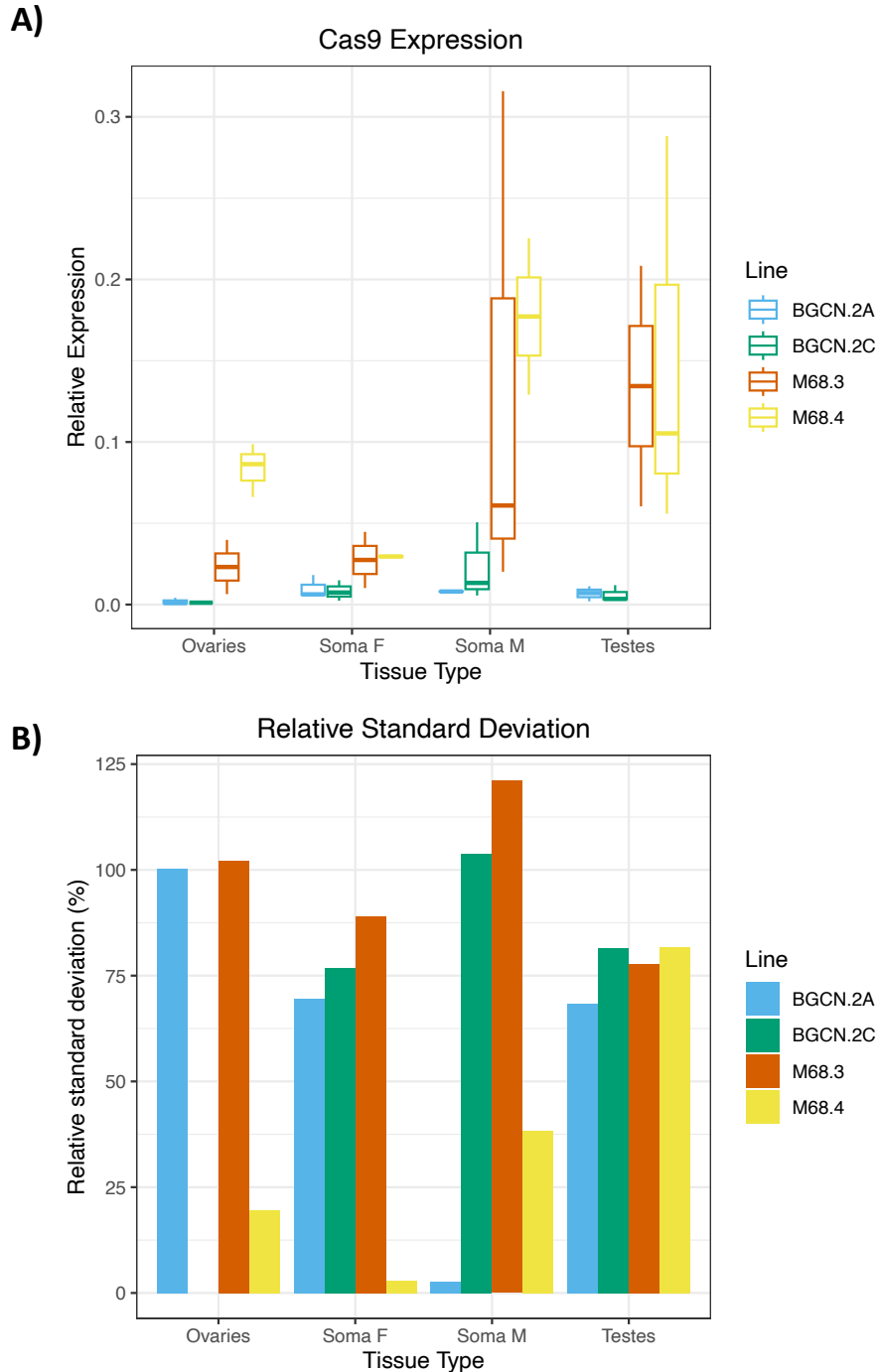
Figure 2.25 A) Relative expression of Cas9 across different tissue types in four established transgenic lines. Comparison between tissue types was challenging as one reference gene showed near to consistent expression across tissues. M68 lines consistently showed higher expression across all tissue types than did bgcn lines. Sample exclusion due to reads not meeting the criteria for use meant not all tissue types within lines had 3 biological replicates however at least one usable sample was obtained for each condition. B) Relative standard deviation of Cas9 relative expression, expressed as a percentage of the mean (coefficient of variation). The M68.2C ovary sample had only one biological replicate. Therefore, the standard deviation could not be calculated for that sample. M68.4 Female Soma, bgcn.2A male soma and M68.4 male soma only had two replicates.

Across all tissue types, Cas9 expression was consistently higher in M68 lines. There was a variability among biological replicates in some tissue types, which increased as the overall relative expression level increased. When the standard deviation was expressed as a percentage of the mean (*Figure 2.25* B*)* the variation across samples appeared to be fairly consistent. Relative expression in the *bgcn*-Cas9 lines was present but appeared consistently lower than for the *mei-W68*-Cas9 lines. Lower observed levels of Cas9 expression could be caused by not only a weaker Cas9 promotor but could also be the result of more tightly controlled expression. As the germline cells were not split into early and late cells, the lack of expression could be due to the Cas9 being active in a much smaller number of cells than that of a promotor with leakier expression. All lines showed expression in the somatic tissue, this was expected for *mei-W68*-Cas9, which displayed a lack of germline restriction in original RT-PCR. *bgcn*-Cas9 did not show somatic expression in the original RT-PCR experiment, yet expression remained higher in somatic tissue than germline tissue.

2.5   Discussion

A range of effective Cas9 promotors are a necessary tool in the germline editing of medfly. Previous studies in medfly have relied on the co-injection of Cas9 protein and mRNA for gene editing (Meccariello et al. 2017; Aumann et al. 2018; Primo et al. 2020; Aumann et al. 2020), as there were no effective Cas9 expressing transgenics in this species. Endogenous expression of Cas9 has proved a highly effective source of Cas9 nuclease for gene editing in a range of other species, and is therefore highly desirable for medfly (Champer et al. 2017; Kandul et al. 2020; Gantz et al. 2015; Hammond et al. 2016). Characterising the expression Cas9 transgenes is essential for balancing effective editing against potential fitness costs. If Cas9 cutting occurs too late in embryonic development, total gene knockout is much less likely to be achieved, with mosaicism obscuring the expression of a knockout phenotype. Promotors specific to the germline are highly desirable for the development of HDR-based gene drives. However, in order to maximise homing rates and minimise the formation of NHEJ generated resistance alleles, less tissue-restricted Cas9 expression is required, and this could have a broad range of functions. Indeed, heavy maternal loading of Cas9, while unsuitable for HDR-gene drives, is highly desirable for alternative gene drive strategies such as those based on Toxin-Antidotes, or for dominant sex conversion and male sterility in applications such as pgSIT (precision guided sterile insect technique). Other applications include the ability to set up crosses with transgenic sgRNA expressing lines or inject embryos with sgRNAs to produce highly penetrant gene knockouts.

I analysed the expression patterns of several candidate promotor genes, which provided useful information concerning the challenges of finding a germline specific gene with no somatic expression. *nanos* and *vasa* (or homologues of these genes) are two of the most commonly used Cas9 promotors to drive germline expression (Hammond et al. 2016; Champer et al. 2017; Feng et al. 2021). However, somatic expression, albeit lower than that of the germline, was seen in both promoters. Complete restriction to the germline has traditionally been difficult to achieve with germline promotors (Du et al. 2023). *nanos* has been recorded to have lower somatic expression than *vasa* but results in higher levels of maternal deposition and there is a high level of variability in promotor function between species (Carballar-Lejarazú et al. 2020). The only genes in this study which showed very low levels of somatic expression were *bgcn* and *innexin. innexin* had not previously been tested as a Cas9 promotor and *bgcn* had only been used once, though unsuccessfully (Chan et al. 2013). Thus, the genes identified as the most promising promotors in my research work, were

not those predicted to be most promising from my initial literature search, showing the value of the systematic tests of existing and new candidates described here.

The decision to split germline tissues into early and late was taken with the aim of capturing expression differences between the two. Expression in early germline cells is desirable for designs which utilise homology directed repair which occurs at much higher levels in early germline cells. However, in none of the candidate gene tested did I observe a difference in expression between early and late ovaries. Early and late testes expression occasionally appeared to differ, as is evident in the strength of the gel bands observed. However, given the difficulties in extracting sufficient RNA from early testes samples due to the small amount of available tissue, this is more likely to be due to differences in sample quality.

Identifying the promotor regions of the candidate genes was relatively simple since the genome was well annotated and the RACE results aligned well with these. *vasa* was the promotor which the RACE was most useful, due to its 6kb 5' UTR and multiple splice variants. However, a RACE product was not obtained from *vasa,* which could be due to the multiple variants causing multiple start sites. *vasa* requires further investigation, as taking the entire 6kb would have produced a construct much larger than optimal for subsequent *piggybac* insertion. The other identifiable issue was that by including 2kb upstream of the identified 5' UTR, sections of other genes were often included in these regions. However, none of them covered the entirety of another gene, so the likelihood for these off-target genes being transcribed is low.

The amplification of the identified promotor regions from the genome was achievable with the exception of *innexin-5.* A product could not be produced with either the original primers or the flanking primers designed for nesting. This could be due to sequence complexity or poor primer binding but would require more time to troubleshoot. As three candidate promotors were still available, assembly was conducted with these.

The assembly of *bgcn* and *mei-W68* only needed a small amount of troubleshooting. *nanos* continually truncated Cas9 when inserted. It did not consistently truncate in the same place and this was not an issue in either of the other constructs. Switching to low recombination SURE2 cells did not help solve this issue and I ultimately had to drop further work with two of the most commonly used Cas9 promotors - *nanos* and *vasa.* Despite *innexin-5* not having been used as a Cas9 promotor to date, the original expression profile I derived for it showed that it was one of the most promising, with the lowest level of somatic expression.

Starting with five candidate genes, two constructs were finalised for microinjection, both of which were successfully integrated into the genome. Although a variety of cas9 promotors was the original aim, even having a single Cas9 germline promotor would be a highly valuable addition to the medfly Cas9 toolbox. As these constructs had been integrated by *piggybac* insertion, there was some variability between flies carrying the same construct. The establishment of these lines did allow for the potential of multiple insertion events. One of the *mei-W68* lines M68.4 had a higher than expected number of transgenic F1 offspring. This could indicate that there were multiple insertions. The other line established from *mei-W68* M68.3 did not show this pattern and was established from a different cage of microinjection survivors giving two separate insertion events for comparison. This was useful to allow for comparison of how the genomic location of the transgenic elements can affect the expression of Cas9.

The lines were maintained as heterozygotes for several generations. However, for logistical reasons the lines were then enriched through inbreeding to increase the percentage of transgenic alleles within the population. This meant that to generate flies for experiments, a cross to wildtype was done, to generate heterozygote test individuals. This could represent a problem in lines with more than one insertion event (which may have been the case for the M68.4 line). However, it would be possible detect multiple insertions if, following two backcrosses to wildtype, the offspring were more than 50% transgenic. Multiple insertions in a single line may limit comparisons of insertion site effects. As the primary question was to determine which candidate gene provides the most effective Cas9 promotor, it is the difference between the promotors that should be the primary focus. As the insertion is random, knowing if the promotor works regardless of the insert location or the number of inserts will show the effect on variability.

To determine whether the transgenic lines displayed any fitness cost to carrying the transgene a small-scale fitness study was completed. Fitness costs from transgenes can be caused by a variety of issues. When insertion location is random, the transgene can be inserted into a transcriptionally active region which can inhibit the function of a native gene (Scolari et al. 2011). The accumulation of transgenic proteins such as marker elements can also have a negative effect on fitness (Marrelli et al. 2006). This guided my decision to use the lines which showed the highest production of RFP. One line of each promotor type was tested (M68.3 and bgcn.2C) and these were chosen based on which lines expressed the strongest marker phenotype. None of the fitness tests indicated a decrease in fitness of

transgenic lines in comparison to the wild type, with some even showing higher indices of fitness in the transgenic lines. However, it should be noted that this was a small-scale study of fitness and there are many other ways fitness could be affected in addition to pupal development, egg laying rate and survival. For example, mating competitiveness is often compromised in transgenic lines and thus additional tests should be completed in the future (Dilani et al. 2022; Harvey-Samuel et al. 2014).

To determine the expression of Cas9, qPCR was completed on four of the lines (M68.3, M68.4, BGCN.2A and BGCN.2C). BGCN.2B was excluded as it displayed the least distinguishable marker expression, and it was deemed unnecessary to test three lines of what could be the same insertion. Seven tissue types were considered for testing. However, this was limited due to both technical capabilities and for logistical reasons. Germline tissue was to be split into early and late cells to identify how tightly controlled expression is. Separation of tissue into early and late was challenging once each tissue had been snap frozen. This may have introduced some variation in dissection consistency. This, in combination with seeing limited differences in the expression patterns between early and late germline in the original expression study, determined that the separation of germline tissue into early and late was not necessary.

Designing a qPCR experiment with multiple lines and tissue types to allow comparisons between tissue types required a reference gene whose expression was extremely stable across all tissue types. Three reference genes were originally tested *RPL19*, RPE and *actin3.* *Actin3* was immediately excluded as it showed very limited amplification and a standard curve could not be obtained. *RPL19* and *RPE* both functioned well with primer efficiencies being in acceptable range however the expression across tissue types in both was not at a consistent level. Expression in the ovaries was higher in both reference genes than all other tissue types as seen in Figure *2.24*. Thus, this reference gene was not consistently expressed across tissue types. The expression in the ovaries was higher across both reference genes implying that this might not be due to higher expression of these genes in the ovaries, rather, that the ovary samples have an overall higher level of cDNA. If so, the reference genes would both be suitable as the Cas9 expression would need to be standardised to account for this higher level of cDNA. After RNA extraction, ovary samples consistently had the highest concentration of all the tissue types and although all samples were diluted to the same concentration, samples at an original higher concentration are more likely to be higher quality RNA samples for cDNA synthesis.

Ideally both reference genes would be used for this experiment and all samples would be run with all three primer sets (Cas9, *RPL19* and RPE). RPE had lower expression across samples than that of *RPL19* which often led to high Ct values for RPE. Some had to be excluded as they were above the 34 cycle threshold while others (although below 34) had too much variability between technical replicates as variability between technical replicates tends to increase as the Ct value increases (Ruiz-Villalba et al. 2021). For RPE to be used for calculating the final delta values, several samples would have been excluded as there would not be a value for RPE. *RPL19* was therefore the only reference gene used for calculating the expression of Cas9. *RPL19* is a commonly used reference gene for qPCR in medfly and among the best ranked (Sagri et al. 2017). Studies often consider expression stability in a single tissue type, as opposed to identifying reference genes which can be used across tissue types. Finding a gene with equal expression across tissue types can be difficult resulting in an inability to confidently compare delta values between different tissues. *RPL* genes were identified as the most stable across insect tissue types in several studies (Lü et al. 2018) however in this work clear variation across tissue types was seen. The use of a single reference gene makes it difficult to confirm if this is variation in expression between tissue types or if it is due to higher starting RNA concentration in these tissue types. This could lead to incorrect assumptions being made about tissue specific expression levels. Further validation should be completed on reference genes to be used for qPCR in *Ceratitis*, with particular focus on identifying reference genes with stable expression across tissue types.

In the qPCR results all tissues are mapped onto the same graph despite the uncertainty of whether we can make a direct comparison between tissue types as it still allows comparison between the lines. Three biological replicates were run for each tissue in every line however not all of these produced usable results (exclusion due to technical replicates being too far apart or over the 34-cycle cut off). A minimum of one reading was obtained for each tissue in every line so a value could be plotted for each however this meant statistical analysis could not be ran. There was not enough cDNA of each sample for the qPCR to be repeated and redoing the RT-PCR could introduce further variability. The most important information was whether the lines were producing Cas9, which they were, although the levels appeared very low in the *bgcn* lines. The same qPCRs were run on wildtype flies as controls to confirm no off-target binding in any tissue and none of these samples showed expression. This means that although the *bgcn* lines appear to have very low expression when compared to the mei-W68 lines, they are still expressing Cas9. Whether the *mei-W68* lines showing higher Cas9 expression will make them more effective is unknown. *bgcn* lines may have more tightly

controlled expression which results in their overall expression appearing lower, if this was the case *bgcn* would be the more suitable Cas9 promotor. It is promising that all lines are expressing Cas9 and can be used for crosses to guide lines to confirm their cutting ability.

To confidently compare across tissue types, I first considered whether such comparisons were legitimate by comparing the level of control *RPL19* expression. *RPL19* was consistent across the testes and somatic tissue allowing me to directly compare expression levels with relative confidence. Hence, the observed difference in expression between male and female somatic tissue in *Figure 2.25* was surprising, as neither *mei-W68* or *bgcn* are reported to have sex specific expression patterns, and this was also not evident in the original end point RT-PCR expression studies. Male tissue type also appeared to show greatest variation between biological replicates. To determine if this was truly greater variation, the relative standard deviation was also calculated. This shows that the variation appearing larger in *mei-W68* than *bgcn* samples seemed primarily due to *mei-W68* samples having a higher relative expression. Sex differences in expression could be caused by insertion site of the transgene if insertion occurred in a region which is upregulated in males. However, the consistency across lines of males showing higher expression would make this explanation unlikely as the insertion site should vary when founded from separate insertion events.

When comparing expression between germline and somatic tissue, the original hope was for minimal expression in somatic tissue and higher expression in the germline. Expression is equal if not higher than germline expression in most lines, the only line in which germline expression is higher was M68.4, with ovary expression being higher than that of the female soma. Given the uncertainty on how comparable the ovaries are to other tissue types due to the reference gene variability, we cannot be sure this is a true reflection of how expression in these two tissues compares. The comparison of the somatic tissue to the testes can be made more confidently – and these showed similar levels of expression. Therefore, Cas9 expression is unlikely to be restricted to the germline in either of these lines. Depending upon the intended purpose of a Cas9 expressing line, restriction to the germline is not always a necessity. All lines show expression of Cas9, to test their effectiveness as promotors they would all be crossed to guide lines to investigate their cutting rates.

At the beginning of this experiment there were very few native genes being used as promotors for Cas9 in medfly. In the work descried here, I designed two further Cas9 promotors for use in *Ceratitis capitata*. Both promotors have demonstratable levels of Cas9 expression and will be further validated through the crossing to guide lines. Once cutting

efficiencies have been determined, these promotors can be used to drive Cas9 expression in genetic pest control methods utilising CRISPR/Cas9 technology. Several other interesting potential promotors have been identified and assembly of these could be continued if further germline promotors are needed. My research has increased the number of available germline-specific Cas9 promotors in the toolbox for genetic modification in medfly.

2.6   Supplementary Information

2.6.1   Primers

2.6.1.1   Primers used for genomic sequencing

Primer Number	Primer Name	Sequence
1	EX1_M_F1	tgtcagttatccgcacgcagaa
2	EX1_M_R1	agcagtcgatcatgtcaccgct
3	EX1_M_F2	gcattttcgcatcaggcaaaggc
4	EX1_M_R2	atgcacggtagagagacgttgt
5	EX1_M_F3	gcagtcttgcacacaacaccatgg
6	EX1_M_R3	tggatcagcatcaaaaagtgca
7	EX1_I_F1	cggctgtgaaaccgctttccaa
8	EX1_I_R1	tccctccggaacaagagcatcg
9	EX1_I_F2	tctgcgttactatcaatgggtca
10	EX1_I_R2	ttgcgaaccacccggtccaaaa
11	EX1_I_F3	gcgatgctcttgttccggaggg
12	EX1_I_R3	tggacgcaattgagtgctgatca
13	EX1_V_F4	cgcggcttaagaaaaatgcggc
14	EX1_V_R4	tgtctctcacttcgagcaccac
15	EX1_V_F5	agcacggcaacatgtgtcacct
16	EX1_V_R5	tcgacgcaacaacgtatgccga
17	EX1_B_F1	agctggacaatgtggcaaccaga
18	EX1_B_R1	gttcccatgccggaacccgtac
19	EX1_B_F2	gccacaggcacttactatggcg
20	EX1_B_R2	tcggatacctttggggagggca
21	EX1_B_F3	tgagaagcggtaaaaccttaagt
22	EX1_B_R3	tggcccaattgttacccgctcc
23	EX1_N_F2	agagatatgctacttcccgaggg
24	EX1_N_R2	gccaacaaagcgaacattttaaggg
25	EX1_N_F1	tcgtacctgttcggtgaaatct
26	EX1_N_R1	acaggcacgttcatgtccagca
27	EX1_V_F3	gatcgaaatggtggtggcggct

28	EX1_V_R3	gaacctgtttgagcacatgccattaaatctc
29	EX1_BG_F1	acgccatatctcacgctattgcca
30	EX1_BG_R1	cgtcggttgcgcaccaaatcaa
31	EX1_BG_F2	ggcgatggaaggcacacctgag
32	EX1_BG_R2	acggtggtttggaaatggttgattgc
33	EX1_N_F3	tgcaacaataatggcaaccagagca
34	EX1_N_R3	ggctttacaaatcgggcaaatgtaag
35	EX1_BG_F3	taaaggcggcacttgtggcagg
36	EX1_BG_R3	tgctgccacagttctgcaagaa

2.6.1.2   Primers used for RACE

37	EX2_B2_3	gattacgccaagcttggggagggcaccaccgaaaaggtattca
38	EX2_bgcn_3	gattacgccaagcttcgcctcgtgtcggtagacgtcgctaaca
39	EX2_B2_1	gattacgccaagcttctcgccctcttcatcggcggtgg
40	EX2_bgcn_2	gattacgccaagcttgcactcaggtgtgccttccatcgcc
41	EX2_B2_2	gattacgccaagcttggccgccactgttaagtaccgtccatga
42	EX2_bgcn_1	gattacgccaagcttcgtgcgtcggttgcgcaccaaatcaa
43	EX2_Inn_1	gattacgccaagcttctcgcttagtttgcgaaccacccggtcc
44	EX2_M68_2	gattacgccaagctttgttgtgtgcaagactgcacagacgtcc
45	EX2_M68_1	gattacgccaagcttgcgccaccatgcgtatagcagtcgatc
46	EX2_vasa_2	gattacgccaagcttaccttcgcactcgatccggagcgcc
47	EX2_nanos_3	gattacgccaagcttgacaggcacgttcatgtccagcactcc
48	EX2_nanos_2	gattacgccaagcttggcgacggggccggtaagttcagcatac

49	EX2_nanos_1	gattacgccaagcttttggcggcagcagctgctgcggcaactg
50	EX2_vasa_3	gattacgccaagcttactttgtctctcacttcgagcaccacggc
51	EX2_Inn_2	gattacgccaagcttataagttccctccggaacaagagcatcgcc
52	EX2_vasa_1	gattacgccaagcttgcgcaatggcgtggaagagagatcggtg
53	RACE_vasa_5	actcccattaccaaacccacggccggc
54	RACE_vasa_4	cgtccgccacctctaccgcgcctact
55	Meiw68_RACE_4	cacaccgctgcagtatcaaggccgactcag
56	bgcn_RACE_Seq	gatcttcatggagtttgcct

2.6.1.3 Primers used for assembly

67	Meiw68_Flank FWD	tctgcgatgggcggcctagaaacaccacaattttaaaattc
68	Meiw68_Flank REV	ccatgctagcttgagcacgttttcgtcaaggtttacatac
69	Cas9_BB_F	ggccgcccatcgcagatctc
70	Cas9_BB_R	agcgattcgagttaacGCGCCTGACTCTAGAatacattga
71	P10_F	atcaaggtaagaatgagtatgaatcgtttttaaaataac
72	bgcn_10kb FWD	gatctgcgatgggcggcccgttgttgcattattcaattca
73	bgcn_10kb REV	gctagcttgagcacgtaatgtttaaaatttctcacaacttc
74	Cas9_Frag_F	acgtgctcaagctagcat
75	P10_R	tgtattctagagtcaggcgcgttaactcgaatcgctatcc
76	innexin_10kb FWD	gatctgcgatgggcggccatttatggcttgtggaatgaat
77	innexin_10kb REV	atgctagcttgagcacgtctaaaaataaaacgcacatg
78	nanos_10kb FWD	gagatctgcgatgggcggccaattattactacagctatcagtattaca
79	nanos_10kb REV	ccatgctagcttgagcacgtcgcatttttgttacagccta
80	Cas9_Frag_R	attttaaaaacgattcatactcattcttaccttgatcg
81	bgcn_Outer_Prim_F	acaaagttctgacgttgcag
82	bgcn_Outer_Prim_R	aatcactctcccgttgcaac

83	Inn_Outer_Prim_F	ctccgtaccaccagcaactt
84	Inn_Outer_Prim_R	gtgacatcggctcatgggta
85	MW68_Outer_Prim_F	tcggtcgtcttagcccgatt
86	MW68_Outer_Prim_R	ggaaaggctttaggcacttcg
87	nanos_Outer_Prim_F	gtccagcactccatcgctac
88	nanos_Outer_Prim_R	atccgggtcgttccggtata
89	V4_BB_Plas_FWD	Ctaccactggcgcgtttggggactagtataacttcgtataatgtatgctatacgaagt
90	V4_BB_Plas_REV	ggccgcccatcgcagatctc
91	V4_Cas9_FWD	atggactataaggaccacgacggaga
92	V4_Cas9_REV	ttattttaaaaacgattcatttttttcgtggccgccggcct
93	V4_P10_FWD	Aggccggcggccacgaaaaaatgaatcgtttttaaaataacaaatcaa
94	V4_P10_REV	tagaattctactcgtaaagcgttaactcgaatcgctatcc
95	V4_Rescue_Plasmid_FWD	ggatagcgattcgagttaacgctttacgagtagaattctacgcg
96	V4_Rescue_Plasmid_REV	tatacgaagttatactagtccccaaacgcgccagtggtag
98	V4_nanos_Assembly_New REV	gtagtctccgtcgtggtccttatagtccatcgcattttttgttacagccta
97	V4_nanos_Assembly_New FWD	ctgagatctgcgatgggcggccaattattactacagctatcagtattaca
100	V4_innexin_Assembly REV	agtctccgtcgtggtccttatagtccatatttatggcttgtggaatgaat
99	V4_innexin_Assembly FWD	agactgagatctgcgatgggcggccctaaaaataaaacgcacatgaaaat
102	V4_Meiw68_New_Assembly REV	agtctccgtcgtggtccttatagtccattttcgtcaaggtttacatacaa
101	V4_Meiw68_New_Assembly FWD	gactgagatctgcgatgggcggcctagaaacaccacaattttaaaattct

103	V4_bgcn_10kb FWD	gacagactgagatctgcgatgggcggcccgttgttgcatta ttcaattca
104	V4_bgcn_10kb_REV	tctccgtcgtggtccttatagtccataatgtttaaaatttctc acaacttca
105	V4_BB_Plas_FWD_2	gactagtataacttcgtataatgtatgctatacgaagt
106	V4_P10_REV_2	actcgtaaagcttgtcgaccgttaactcgaatcgctatcc
107	V4_Res_FWD_Mod_1	ggatagcgattcgagttaacggtcgacaagctttacgagta gaattctac
108	V5_P10_FWD	aggcaaaaaagaaaaagtaaatgaatcgttttttaaaataa caaatcaa
109	V5_Cas9_REV	ttattttaaaaacgattcatttactttttctttttttgcctggc
110	110_vasa_1_fwd	gtcgacagactgagatctggcgcgcccttatagcctatttcc acc
111	111_vasa_1_rev	gttttggcgcataatgtttatagtaagaatctttagtac
112	112_vasa_2_fwd	taaacattatgcgccaaaactcttaccaac
113	113_vasa_2_rev	tatagtccattttccgtggtctctctgaaaatag
114	114_vasa_3_fwd	gaaaaagtaatagcactaccacatctataaaaac
115	115_vasa_3_rev	ccccgggcctcgaggtcaggcgcgccgcagtcgagtcgtc atttg
116	116_vasa_hCas9_fwd	accacggaaaatggactataaggaccac
117	117_vasa_hCas9_rev	ggtagtgctattactttttctttttttgcctg
118	118_bgcn_1_fwd	gtcgacagactgagatctggcgcgcctattggttattaccg aggaaatattg
119	119_bgcn_hCas9_fwd	tttaaacattatggactataaggaccac
120	120_bgcn_hCas9_rev	attcaattctttactttttctttttttgcctg
121	121_bgcn_2_fwd	gaaaaagtaaagaattgaattgttttaaaaagattttg
122	122_bgcn_1_rev	tatagtccataatgtttaaaatttctcacaacttc
123	123_bgcn_2_rev	ccccgggcctcgaggtcaggcgcgcctctaaattcaaagat gaagtgaaatag
124	V6_Cas9_BB_F	gactagtataacttcgtataatgtatgc
125	V6_Cas9_BB_R	cagatctcagtctgtcgacc
126	Marker_NO_FWD	ggtcgacaagctttacgagtagaattct
127	Marker_NO_REV	cccaaacgcgccagtggtagta

132	Meiw68_10kb FWD	ggtcgacagactgagatctggactgctcgaaaatggaata a
133	Meiw68_10kb REV	tcgtggtccttatagtccatcctcagtattcttaagaaacac a
134	bgcn_10kb FWD	ggtcgacagactgagatctgcgttgttgcattattcaattca
135	bgcn_10kb REV	tcgtggtccttatagtccataatgtttaaaatttctcacaact tca
136	nanos_Assembly_V5 FWD	ggtcgacagactgagatctgcgcattttttgttacagccta
137	nanos_Assembly_V5 REV	tcgtggtccttatagtccataattattactacagctatcagta ttaca
138	innexin_Assembly FWD	ggtcgacagactgagatctgctaaaaataaaacgcacatg aaaat
139	innexin_Assembly REV	tcgtggtccttatagtccatatttatggcttgtggaatgaat
150	M68_Cas9_junc_FW D	gactgcttctgtgctctg
151	M68_BB_junc_REV	gtcatagcccaaccgtga
152	nanos_Cas9_junc_FW D	cacagtagagattttcgtgg
153	nanos_BB_junc_REV	cgtaagtccatgaatgcgt
154	Inn_Cas9_junc_FWD	tgcagcgctgctaattgt
155	Inn_BB_junc_REV	cctcaccgaacgaacgta
156	Mark_BB_junc_FWD	tacggttacaattcccagcc
157	BB_mark_junc_REV	gcttgtcaatgcggtaagtg
158	BB_ins_junc_FWD	tcggtctgtatatcgaggtt
159	bgcn_BB_junc_REV	ttgcggaagaaggagctt
160	bgcn_Cas9_junc_FW D	gagcatttgtcgggtgaa
161	Cas9_ins_REV	ctccgatcaggttcttcttg
162	Cas9_P10_junc_FWD	cctacaacaagcaccgggat
163	Mar_P10_junc_REV	cgatcgtgcgttacacgtag

2.6.1.4 Primers used for qPCR

331	qRT-PCR-Cas9 FWD	cagattcgcctggatgacca
332	qRT-PCR_Cas9_RVS	atccgctcgatgaagctctg
333	RPE FWD Primer	gatgaaagtactctccgaac
334	RPE REV Primer	tgtaacacagcattcatctc
335	RPL19 FWD Primer	aacaaacgtgtactgatgg
336	RPL19 RVS Primer	cacgtactttatgtcgtctg

2.6.2 Plasmids, sequences and parts

All sequences and aligned reads can be obtained at https://benchling.com/siddall/f_/sqcsnseT-alex-siddall-thesis-supplementary-info/

2.6.3 Statistical tests

2.6.3.1 Reference Gene ANOVA Summary

RPE

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
qPCR by tissue	3	89.07	29.689	7.533	0.000358
Residuals	44	173.42	3.941		

RPL19

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
qPCR by tissue	3	59.51	19.838	5.014	0.00392
Residuals	53	209.71	3.957		

2.6.4 Scripts and data

All data used to generate figures and complete analysis have been uploaded to a GitHub repository with the R scripts used. This can be found at https://github.com/A-Sidd96/Thesis_scripts/tree/f62f5eb13cd423987733d29ece1d3eeae2224ef5. The original raw data can be found at https://ueanorwich-my.sharepoint.com/:f:/g/personal/gvy19usu_uea_ac_uk/EmJ_Dv-5nolNigVfT5N2BrYBqSwLs06Bkh0HmDBw3VAaEg?e=cEoTlk.

2.6.5 *vasa* alternative isoforms
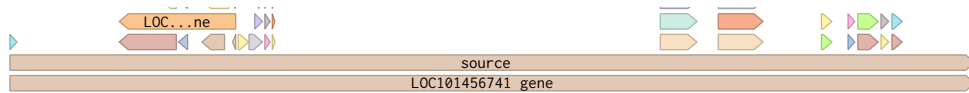
Figure 2.26 Annotated *vasa* alternate transcripts. Generated on Benchling.

2.7 References

Adelman, Zach N., Nijole Jasinskiene, Sedef Onal, Jennifer Juhn, Aurora Ashikyan, Michael Salampessy, Todd MacCauley, and Anthony A. James. 2007. '*nanos* Gene Control DNA Mediates Developmentally Regulated Transposition in the Yellow Fever Mosquito *Aedes aegypti*'. *Proceedings of the National Academy of Sciences* 104 (24): 9970–75. https://doi.org/10.1073/pnas.0701515104.

Aumann, Roswitha A., Irina Häcker, and Marc F. Schetelig. 2020. 'Female-to-Male Sex Conversion in *Ceratitis capitata* by CRISPR/Cas9 HDR-Induced Point Mutations in the Sex Determination Gene Transformer-2'. *Scientific Reports* 10 (1): 18611. https://doi.org/10.1038/s41598-020-75572-x.

Aumann, Roswitha A., Marc F. Schetelig, and Irina Häcker. 2018. 'Highly Efficient Genome Editing by Homology-Directed Repair Using Cas9 Protein in *Ceratitis capitata*'. *Insect Biochemistry and Molecular Biology* 101 (October): 85–93. https://doi.org/10.1016/j.ibmb.2018.08.004.

Benchling. 2023. 'Biological Software'. Benchling.com.

Caceres, C. 2002. 'Mass Rearing of Temperature Sensitive Genetic Sexing Strains in the Mediterranean Fruit Fly'. *Genetica* 116 (1): 107–16. https://doi.org/10.1023/A:1020967810703.

Carballar-Lejarazú, Rebeca, Christian Ogaugwu, Taylor Tushar, Adam Kelsey, Thai Binh Pham, Jazmin Murphy, Hanno Schmidt, Yoosook Lee, Gregory C. Lanzaro, and Anthony A. James. 2020. 'Next-Generation Gene Drive for Population Modification of the Malaria Vector Mosquito, *Anopheles gambiae*'. *Proceedings of the National Academy of Sciences* 117 (37): 22805–14. https://doi.org/10.1073/pnas.2010214117.

Carrami, Eli M., Kolja N. Eckermann, Hassan M. M. Ahmed, Héctor M. Sánchez C., Stefan Dippel, John M. Marshall, and Ernst A. Wimmer. 2018. 'Consequences of Resistance Evolution in a Cas9-Based Sex Conversion-Suppression Gene Drive for Insect Pest Management'. *Proceedings of the National Academy of Sciences* 115 (24): 6189–94. https://doi.org/10.1073/pnas.1713825115.

Champer, Jackson, Anna Buchman, and Omar S Akbari. 2016. 'Cheating Evolution: Engineering Gene Drives to Manipulate the Fate of Wild Populations'. *Nat. Rev. Genet.* 17 (3): 146–59. https://doi.org/10.1038/nrg.2015.34.

Champer, Jackson, Samuel E Champer, Isabel K Kim, Andrew G Clark, and Philipp W Messer. 2021. 'Design and Analysis of CRISPR-Based Underdominance Toxin-Antidote Gene Drives'. *Evol. Appl.* 14 (4): 1052–69. https://doi.org/10.1111/eva.13180.

Champer, Jackson, Esther Lee, Emily Yang, Chen Liu, Andrew G Clark, and Philipp W Messer. 2020. 'A Toxin-Antidote CRISPR Gene Drive System for Regional Population Modification'. *Nat. Commun.* 11 (1): 1082. https://doi.org/10.1038/s41467-020-14960-3.

Champer, Jackson, Jingxian Liu, Suh Yeon Oh, Riona Reeves, Anisha Luthra, Nathan Oakes, Andrew G. Clark, and Philipp W. Messer. 2018. 'Reducing Resistance Allele

Formation in CRISPR Gene Drive'. *Proceedings of the National Academy of Sciences* 115 (21): 5522–27. https://doi.org/10.1073/pnas.1720354115.

Champer, Jackson, Riona Reeves, Suh Yeon Oh, Chen Liu, Jingxian Liu, Andrew G. Clark, and Philipp W. Messer. 2017. 'Novel CRISPR/Cas9 Gene Drive Constructs Reveal Insights into Mechanisms of Resistance Allele Formation and Drive Efficiency in Genetically Diverse Populations'. Edited by Harmit S. Malik. *PLOS Genetics* 13 (7): e1006796. https://doi.org/10.1371/journal.pgen.1006796.

Chan, Yuk-Sang, David S. Huen, Ruth Glauert, Eleanor Whiteway, and Steven Russell. 2013. 'Optimising Homing Endonuclease Gene Drive Performance in a Semi-Refractory Species: The *Drosophila melanogaster* Experience'. Edited by Shree Ram Singh. *PLoS ONE* 8 (1): e54130. https://doi.org/10.1371/journal.pone.0054130.

Chen, Xien, and Subba Reddy Palli. 2022. 'Development of Multiple Transgenic CRISPR/Cas9 Methods for Genome Editing in the Fall Armyworm, *Spodoptera frugiperda*'. *Journal of Pest Science*, July. https://doi.org/10.1007/s10340-022-01546-9.

Dilani, Padukka Vidanelage Desha, Ranil Samantha Dassanayake, Brij Kishore Tyagi, and Yasanthi Illika Nilmini Silva Gunawardene. 2022. 'The Impact of Transgenesis on Mosquito Fitness: A Review'. *Frontiers in Insect Science 2* (September): 957570. https://doi.org/10.3389/finsc.2022.957570.

De Meyer, M., M. P. Robertson, A. T. Peterson, and M. W. Mansell. 2007. 'Ecological Niches and Potential Geographical Distributions of Mediterranean Fruit Fly (*Ceratitis capita*) and Natal Fruit Fly (*Ceratitis rosa*)'. *Journal of Biogeography* 0 (0): 070821084123001-??? https://doi.org/10.1111/j.1365-2699.2007.01769.x.

Du, Jie, Weizhe Chen, Xihua Jia, Xuejiao Xu, Emily Yang, Ruizhi Zhou, Yuqi Zhang, Matt Metzloff, Philipp W. Messer, and Jackson Champer. 2023. 'New Germline Cas9 Promoters Show Improved Performance for Homing Gene Drive'. Preprint. Genetics. https://doi.org/10.1101/2023.07.16.549205.

Eckermann, Kolja N., Hassan M.M. Ahmed, Mohammad KaramiNejadRanjbar, Stefan Dippel, Christian E. Ogaugwu, Peter Kitzmann, Musa D. Isah, and Ernst A. Wimmer. 2018. 'Hyperactive piggyBac Transposase Improves Transformation Efficiency in Diverse Insect Species'. *Insect Biochemistry and Molecular Biology* 98 (July): 16–24. https://doi.org/10.1016/j.ibmb.2018.04.001.

Feng, Xuechun, Víctor López Del Amo, Enzo Mameli, Megan Lee, Alena L. Bishop, Norbert Perrimon, and Valentino M. Gantz. 2021. 'Optimized CRISPR Tools and Site-Directed Transgenesis towards Gene Drive Development in *Culex quinquefasciatus* Mosquitoes'. *Nature Communications* 12 (1): 2960. https://doi.org/10.1038/s41467-021-23239-0.

Gantz, Valentino M, Nijole Jasinskiene, Olga Tatarenkova, Aniko Fazekas, Vanessa M Macias, Ethan Bier, and Anthony A James. 2015. 'Highly Efficient Cas9-Mediated Gene Drive for Population Modification of the Malaria Vector Mosquito *Anopheles stephensi*'. *Proc. Natl. Acad. Sci. U. S. A.* 112 (49): E6736-43. https://doi.org/10.1073/pnas.1521077112.

Garneau, Nicole L., Jeffrey Wilusz, and Carol J. Wilusz. 2007. 'The Highways and Byways of mRNA Decay'. *Nature Reviews Molecular Cell Biology* 8 (2): 113–26. https://doi.org/10.1038/nrm2104.

Gutierrez, Andrew Paul, Luigi Ponti, Markus Neteler, David Maxwell Suckling, and José Ricardo Cure. 2021. 'Invasive Potential of Tropical Fruit Flies in Temperate Regions under Climate Change'. *Communications Biology* 4 (1): 1141. https://doi.org/10.1038/s42003-021-02599-9.

Hammond, Andrew, Roberto Galizi, Kyros Kyrou, Alekos Simoni, Carla Siniscalchi, Dimitris Katsanos, Matthew Gribble, et al. 2016. 'A CRISPR-Cas9 Gene Drive System Targeting Female Reproduction in the Malaria Mosquito Vector *Anopheles gambiae*'. *Nat. Biotechnol.* 34 (1): 78–83. https://doi.org/10.1038/nbt.3439.

Hammond, Andrew, Xenia Karlsson, Ioanna Morianou, Kyros Kyrou, Andrea Beaghton, Matthew Gribble, Nace Kranjc, et al. 2021. 'Regulating the Expression of Gene Drives Is Key to Increasing Their Invasive Potential and the Mitigation of Resistance'. Edited by Owain Rhys Edwards. *PLOS Genetics* 17 (1): e1009321. https://doi.org/10.1371/journal.pgen.1009321.

Harvey-Samuel, Tim, Thomas Ant, Hongfei Gong, Neil I. Morrison, and Luke Alphey. 2014. 'Population-level Effects of Fitness Costs Associated with Repressible Female-lethal Transgene Insertions in Two Pest Insects'. *Evolutionary Applications* 7 (5): 597–606. https://doi.org/10.1111/eva.12159.

Hellemans, Jan, Geert Mortier, Anne De Paepe, Frank Speleman, and Jo Vandesompele. 2007. 'qBase Relative Quantification Framework and Software for Management and Automated Analysis of Real-Time Quantitative PCR Data'. *Genome Biology* 8 (2): R19. https://doi.org/10.1186/gb-2007-8-2-r19.

Hendrichs, J, A.S. Robinson, J. P. Cayol, and W. Enkerlin. 2002. 'MEDFLY AREAWIDE STERILE INSECT TECHNIQUE PROGRAMMES FOR PREVENTION, SUPPRESSION OR ERADICATION: THE IMPORTANCE OF MATING BEHAVIOR STUDIES'. *Florida Entomologist* 85 (1): 1–13. https://doi.org/10.1653/0015-4040(2002)085[0001:MASITP]2.0.CO;2.

Jarvis, Donald L., Carla Weinkauf, and Linda A. Guarino. 1996. 'Immediate-Early Baculovirus Vectors for Foreign Gene Expression in Transformed or Infected Insect Cells'. *Protein Expression and Purification* 8 (2): 191–203. https://doi.org/10.1006/prep.1996.0092.

Kandul, Nikolay P, Junru Liu, Anna Buchman, Valentino M Gantz, Ethan Bier, and Omar S Akbari. 2020. 'Assessment of a Split Homing Based Gene Drive for Efficient Knockout of Multiple Genes'. *G3 Genes|Genomes|Genetics* 10 (2): 827–37. https://doi.org/10.1534/g3.119.400985.

Kouranova, Evguenia, Kevin Forbes, Guojun Zhao, Joe Warren, Angela Bartels, Yumei Wu, and Xiaoxia Cui. 2016. 'CRISPRs for Optimal Targeting: Delivery of CRISPR Components as DNA, RNA, and Protein into Cultured Cells and Single-Cell Embryos'. *Human Gene Therapy* 27 (6): 464–75. https://doi.org/10.1089/hum.2016.009.

Lei, Jianfeng, Peihong Dai, Jiyang Li, Mi Yang, Xiuqin Li, Wanqi Zhang, Guantong Zhou, WangzhenGuo, and Xiaodong Liu. 2021. 'Tissue-Specific CRISPR/Cas9 System of Cotton Pollen with GhPLIMP2b and GhMYB24 Promoters'. *Journal of Plant Biology* 64 (1): 13–21. https://doi.org/10.1007/s12374-020-09272-4.

Li, Ming, Ting Yang, Michelle Bui, Stephanie Gamez, Tyler Wise, Nikolay P. Kandul, Junru Liu, et al. 2021. 'Suppressing Mosquito Populations with Precision Guided Sterile Males'. *Nature Communications* 12 (1): 5374. https://doi.org/10.1038/s41467-021-25421-w.

Li, Xianghong, Erin R. Burnight, Ashley L. Cooney, Nirav Malani, Troy Brady, Jeffry D. Sander, Janice Staber, et al. 2013. '*piggyBac* Transposase Tools for Genome Engineering'. *Proceedings of the National Academy of Sciences* 110 (25). https://doi.org/10.1073/pnas.1305987110.

Lü, Jing, Chunxiao Yang, Youjun Zhang, and Huipeng Pan. 2018. 'Selection of Reference Genes for the Normalization of RT-qPCR Data in Gene Expression Studies in Insects: A Systematic Review'. *Frontiers in Physiology* 9 (November): 1560. https://doi.org/10.3389/fphys.2018.01560.

Mao, Yanfei, Zhengjing Zhang, Zhengyan Feng, Pengliang Wei, Hui Zhang, José Ramón Botella, and Jian-Kang Zhu. 2016. 'Development of Germ-Line-Specific CRISPR-Cas9 Systems to Improve the Production of Heritable Gene Modifications in *Arabidopsis*'. *Plant Biotechnology Journal* 14 (2): 519–32. https://doi.org/10.1111/pbi.12468.

Marrelli, Mauro T., Cristina K. Moreira, David Kelly, Luke Alphey, and Marcelo Jacobs-Lorena. 2006. 'Mosquito Transgenesis: What Is the Fitness Cost?' *Trends in Parasitology* 22 (5): 197–202. https://doi.org/10.1016/j.pt.2006.03.004.

McFarlane, Gus R, C Bruce A Whitelaw, and Simon G Lillico. 2018. 'CRISPR-Based Gene Drives for Pest Control'. *Trends Biotechnol.* 36 (2): 130–33. https://doi.org/10.1016/j.tibtech.2017.10.001.

Meccariello, Angela, Simona Maria Monti, Alessandra Romanelli, Rita Colonna, Pasquale Primo, Maria Grazia Inghilterra, Giuseppe Del Corsano, et al. 2017. 'Highly Efficient DNA-Free Gene Disruption in the Agricultural Pest Ceratitis Capitata by CRISPR-Cas9 Ribonucleoprotein Complexes'. *Scientific Reports* 7 (1): 10061. https://doi.org/10.1038/s41598-017-10347-5.

Merritt, Christopher, Dominique Rasoloson, Darae Ko, and Geraldine Seydoux. 2008. '3' UTRs Are the Primary Regulators of Gene Expression in the *C. Elegans* Germline'. *Current Biology* 18 (19): 1476–82. https://doi.org/10.1016/j.cub.2008.08.013.

Nguyen, Thu N. M., Amanda Choo, and Simon W. Baxter. 2021. 'Lessons from Drosophila: Engineering Genetic Sexing Strains with Temperature-Sensitive Lethality for Sterile Insect Technique Applications'. *Insects* 12 (3): 243. https://doi.org/10.3390/insects12030243.

Nishizawa, Keito, Yoichi Kita, Masahiko Kitayama, and Masao Ishimoto. 2006. 'A Red Fluorescent Protein, DsRed2, as a Visual Reporter for Transient Expression and Stable Transformation in Soybean'. *Plant Cell Reports* 25 (12): 1355–61. https://doi.org/10.1007/s00299-006-0210-x.

Ogaugwu, Christian E., and Ernst A. Wimmer. 2013. 'Molecular Cloning and Expression of nanos in the Mediterranean Fruit Fly, *Ceratitis capitata* (Diptera: Tephritidae)'. *Gene Expression Patterns* 13 (5–6): 183–88. https://doi.org/10.1016/j.gep.2013.03.002.

Papanicolaou, Alexie, Marc F. Schetelig, Peter Arensburger, Peter W. Atkinson, Joshua B. Benoit, Kostas Bourtzis, Pedro Castañera, et al. 2016. 'The Whole Genome Sequence of the Mediterranean Fruit Fly, *Ceratitis capitata* (Wiedemann), Reveals Insights into the Biology and Adaptive Evolution of a Highly Invasive Pest Species'. *Genome Biology* 17 (1): 192. https://doi.org/10.1186/s13059-016-1049-2.

Papathanos, Philippos A, Nikolai Windbichler, Miriam Menichelli, Austin Burt, and Andrea Crisanti. 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): 65. https://doi.org/10.1186/1471-2199-10-65.

Pfaffl, M. W. 2001. 'A New Mathematical Model for Relative Quantification in Real-Time RT-PCR'. *Nucleic Acids Research* 29 (9): 45e–445. https://doi.org/10.1093/nar/29.9.e45.

Port, Fillip, Hui-Min Chen, Tzumin Lee, and Simon L. Bullock. 2014. 'Optimized CRISPR/Cas Tools for Efficient Germline and Somatic Genome Engineering in *Drosophila*'. *Proceedings of the National Academy of Sciences* 111 (29). https://doi.org/10.1073/pnas.1405500111.

Primo, Pasquale, Angela Meccariello, Maria Grazia Inghilterra, Andrea Gravina, Giuseppe Del Corsano, Gennaro Volpe, Germano Sollazzo, et al. 2020. 'Targeting the Autosomal *Ceratitis capitata Transformer* Gene Using Cas9 or dCas9 to Masculinize XX Individuals without Inducing Mutations'. *BMC Genetics* 21 (S2): 150. https://doi.org/10.1186/s12863-020-00941-4.

Raban, Robyn R., John M. Marshall, and Omar S. Akbari. 2020. 'Progress towards Engineering Gene Drives for Population Control'. Edited by Michael H. Dickinson, Leslie B. Vosshall, and Julian A. T. Dow. *Journal of Experimental Biology* 223 (Suppl_1): jeb208181. https://doi.org/10.1242/jeb.208181.

Rendon Arana, Pedro Alfonso., The University of Manchester (United Kingdom), and Department not provided. 1996. 'Development and Evaluation of a Temperature Sensitive Lethal (TSL) Genetic Sexing Strain of the Mediterranean Fruit Fly, *Ceratitis capitata* (Wied.)'.

Robinson, A S. 2002. 'Genetic Sexing Strains in Medfly, *Ceratitis capitata*, Sterile Insect Technique Programmes'. *Genetica* 116 (1): 5–13. https://doi.org/10.1023/a:1020951407069.

Rubin, Gerald M., and Allan C. Spradling. 1982. 'Genetic Transformation of *Drosophila* with Transposable Element Vectors'. *Science* 218 (4570): 348–53. https://doi.org/10.1126/science.6289436.

Ruiz-Villalba, Adrián, Jan M. Ruijter, and Maurice J. B. Van Den Hoff. 2021. 'Use and Misuse of Cq in qPCR Data Analysis and Reporting'. *Life* 11 (6): 496. https://doi.org/10.3390/life11060496.

Saccone, G, A Pane, A De Simone, M Salvemini, A Milano, L Annunziata, U Mauro, and L C Polito. 2007. 'New Sexing Strains for Mediterranean Fruit Fly *Ceratitis capitata*: Transforming Females into Males'. *Area-Wide Control of Insect Pests*, 95–102. https://doi.org/10.1007/978-1-4020-6059-5_7.

Sagri, Efthimia, Panagiota Koskinioti, Maria-Eleni Gregoriou, Konstantina T. Tsoumani, Yiannis C. Bassiakos, and Kostas D. Mathiopoulos. 2017. 'Housekeeping in Tephritid Insects: The Best Gene Choice for Expression Analyses in the Medfly and the Olive Fly'. *Scientific Reports* 7 (1): 45634. https://doi.org/10.1038/srep45634.

Scolari, F., P. Siciliano, P. Gabrieli, L. M. Gomulski, A. Bonomi, G. Gasperi, and A. R. Malacrida. 2011. 'Safe and Fit Genetically Modified Insects for Pest Control: From Lab to Field Applications'. *Genetica* 139 (1): 41–52. https://doi.org/10.1007/s10709-010-9483-7.

Smith, Gale E., Just M. Vlak, and Max D. Summers. 1983. 'Physical Analysis of *Autographa Californica* Nuclear Polyhedrosis Virus Transcripts for Polyhedrin and 10,000-Molecular-Weight Protein'. *Journal of Virology* 45 (1): 215–25. https://doi.org/10.1128/jvi.45.1.215-225.1983.

Sultana, Sabira, John B. Baumgartner, Bernard C. Dominiak, Jane E. Royer, and Linda J. Beaumont. 2020. 'Impacts of Climate Change on High Priority Fruit Fly Species in Australia'. Edited by Nikos T. Papadopoulos. *PLOS ONE* 15 (2): e0213820. https://doi.org/10.1371/journal.pone.0213820.

Van Oers, Monique M., Just M. Vlak, Harry O. Voorma, and Adri A. M. Thomas. 1999. 'Role of the 3' Untranslated Region of Baculovirus P10 mRNA in High-Level Expression of Foreign Genes'. *Journal of General Virology* 80 (8): 2253–62. https://doi.org/10.1099/0022-1317-80-8-2253.

Virginio, Jair F., Maylen Gómez, Aline M. Pinto, Gessyca G. Aniely, Beatriz J. Paranhos, Carlos A.T. Gava, Carlos Cáceres, and Júlio M.M. Walder. 2017. 'Male Sexual Competitiveness of Two *Ceratitis capitata* Strains, Tsl Vienna 8 and OX3864A Transgenics, in Field Cage Conditions'. *Entomologia Experimentalis et Applicata* 164 (3): 318–26. https://doi.org/10.1111/eea.12615.

Wang, Zhi-Ping, Hui-Li Xing, Li Dong, Hai-Yan Zhang, Chun-Yan Han, Xue-Chen Wang, and Qi-Jun Chen. 2015. 'Egg Cell-Specific Promoter-Controlled CRISPR/Cas9 Efficiently Generates Homozygous Mutants for Multiple Target Genes in *Arabidopsis* in a Single Generation'. *Genome Biology* 16 (1): 144. https://doi.org/10.1186/s13059-015-0715-0.

3 Chapter Three – Designing Cas9 promotors for homing drives

3.1 Abstract

Ceratitis capitata (medfly) is an important pest species of global agricultural significance. Current control strategies include the use of pesticides and the release or sterile irradiated males in regions where medfly have been detected. Improved methods are needed for this important pest species as no current strategy provides a sustained level of control. Since the advent of CRISPR/Cas9, homing drives have become one of the most promising methods of genetic pest control. This is because they can result in the rapid spread of a transgene throughout a population while ensuring their own persistence, even if they incur a fitness cost. Prior to the initiation of the research described in this thesis, there was no completed Cas9 homing drive for *Ceratitis,* though the current level of understanding of the medfly genome had advanced to make this a possibility. One of the hurdles in constructing homing drives is the necessity for tightly controlled expression of Cas9. This requires the identification of the regulatory elements of endogenous, germline specific genes, although the success of these as Cas9 promotors is reported to vary across species. Here, I developed two Cas9 promotors in medfly using the regulatory elements of *vasa* and *bgcn (benign gonial cell neoplasm),* for use in a gene drive strategy. After successful integration of *vasa* into a homing drive, my *vasa* driver was shown to be the most effective of three promotors tested. This driver was then subsequently integrated by the Meccariello research team at Imperial College into the first homing drives constructed in *Ceratitis*. High levels of homing were observed, proving that homing is a viable strategy in *Ceratitis*. However, further optimisation of the drive is needed due to high levels of somatic expression and deposition. The resulting paper describing the characteristics of those gene drives is appended.

3.2   Introduction

The applied potential of homing drives was first proposed over 20 years ago (Burt 2003). However, it wasn't until the advent of CRISPR/Cas9 gene editing tools that these designs became a more promising prospect in the field of pest control. Homing drives utilising CRSIPR to trigger homology directed repair were quickly proposed (Esvelt et al. 2014; Gantz and Bier 2015), with potential applications including the prevention of pathogen transmission in vectors or population suppression (Champer et al. 2016). The ability to spread throughout a population even with an associated fitness cost at relatively low introduction frequency makes homing gene drives ideal for use in pest control. CRISPR homing drives bias inheritance by cutting the wildtype allele and repairing the site through homology directed repair, thus ensuring the drive is present on *both* alleles, as shown in *Figure 3.1*.
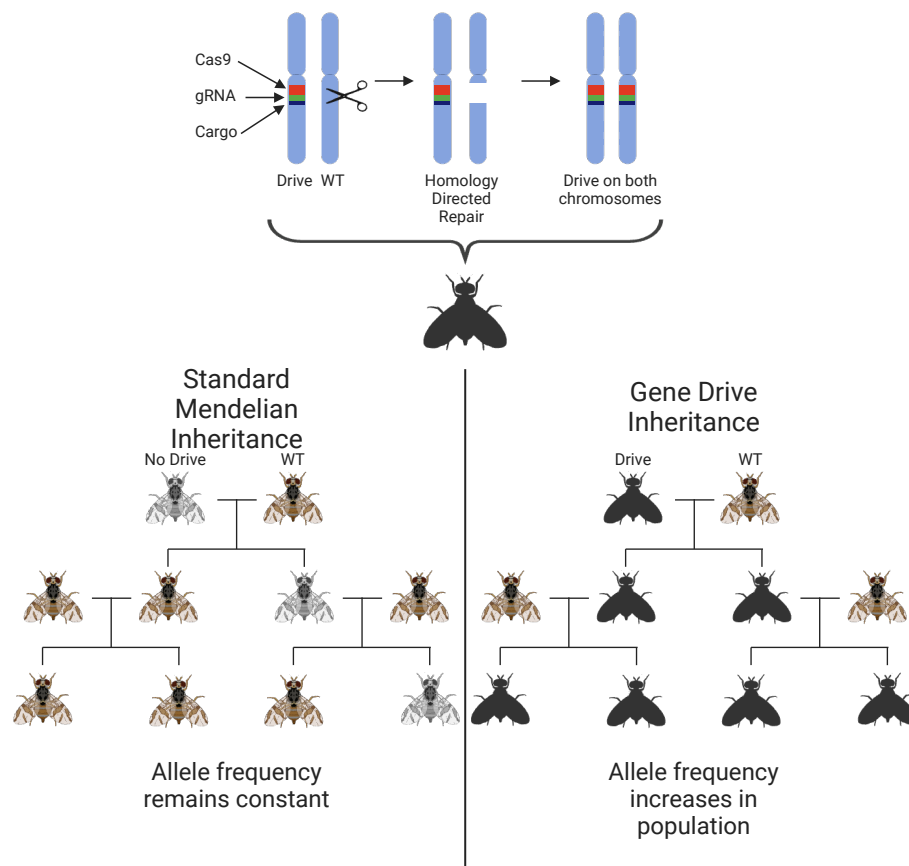
Figure 3.1 The genetic principle and biased inheritance of a homing gene drive. Homing drives ensure and drive their own transmission to the next generation through homology-directed repair. The Cas9 element of the construct, coupled with specific guides cut the wild type chromosome at a precise genomic location triggering a double-strand break. This break is then repaired using homology-directed repair with the remaining chromosome being used as a template. The drive construct is therefore present on *both* chromosome copies ensuring 100% offspring inheritance. A cargo element is also coupled with the drive elements which can spread a desired trait through a population. This biases the usual 50% inheritance

Homing drives have since been developed in a variety of insect pest species. The first CRISPR homing drive was developed as a proof of principle in *Drosophila melanogaster* (Gantz and Bier 2015) and has since then been developed in several pests, including several species of mosquito including *Anopheles stephensi* and *Anopholes gambiae* (Hammond et al. 2016; Gantz et al. 2015; Carballar-Lejarazú et al. 2020; Kyrou et al. 2018; Verkuijl et al. 2022) and *Drosophila suzukii* (Yadav et al. 2023).

The development of homing drives requires an endogenous toolbox for genome editing. In medfly these tools are now being progressed (Meccariello et al. 2021). However, further improvements are still needed to create efficient homing drives. The expression of Cas9 must be specific and tightly controlled to the germline for use in homing drives. This is due to the reliance on homology directed repair which is only favoured over non-homologous end joining (NHEJ) in certain stages of development (Raban et al. 2020). NHEJ is the primary cause of resistance allele formation in homing drives (Unckless et al. 2017). Therefore, to prevent this, Cas9 expression should be avoided during the developmental periods when DNA repair by NHEJ is favoured. Somatic or embryonic expression and maternal or paternal depositions of ribonucleoprotein complexes into the embryos are the primary challenges when attempting to prevent NHEJ (Champer et al. 2018). Therefore, to avoid expression in these tissues, genes that are restricted to the germline are typically used to drive Cas9 expression.

Cas9 expression from endogenous germline promotors achieve varying levels of success in both driving Cas9 and restricting expression to the germline. The germline expressing genes that have been tested so far have varying expression patterns, with some showing higher levels of expression than others, although this is often at the expense of higher levels of embryonic deposition. The level of expression in the soma also varies across native genes controlled by endogenous germline promotors and when those promoters are used to drive the expression of Cas9 (Du et al. 2023). In order to achieve a tighter level of expression control a native 3'UTR can be used. The use of a native 3' UTR has been shown to more closely mimic native gene expression (Ebron and Shukla 2016) as this may have an essential role in the localisation of mRNA and translational regulation (Gavis et al. 1996). This could reduce the amount of somatic Cas9 expression by more closely mimicking the pattern of native germline specific gene expression.

There has been extensive work done on Cas9 promotors for homing drives in other species. However, prior to this work, this had not yet been attempted in medfly. Germline promotors for Cas9 have been tested in several insect species (Champer et al. 2017; Feng et al. 2021; Adolfi et al. 2020; Verkuijl et al. 2022; Anderson et al. 2023) though none are reported to show complete restriction of expression to the germline. The high levels of variation in promotor activity across species must also be considered (Du et al. 2023). It is often the case that a promotor that has worked well in one species may not display the same activity in another. This highlights the need to test a variety of promotors in *Ceratitis* to discover the most effective one for use in a homing drive and to understand more about how phylogenetic distance relates to conservation of function.

In this work, two Cas9 promotors were designed and produced. *vasa* and *bgcn* (*benign gonial cell neoplasm*) were selected as potential promotors from the candidate promotors identified in Chapter two. The design for *bgcn* was relatively simple, as the 5'UTR was short and easily identifiable. *vasa* was more complex given that it has several splicing isoforms and intronic regions, which when removed in previous studies have resulted in sex-limited expression (Papathanos et al. 2009). *vasa* was tested against two externally produced (Angela Meccariello, Imperial College, London) Cas9 promotors (*nanos* and *zpg (zero population growth)*)) and shown to be the most effective for use in a homing drive of those tested. This was then integrated into the first constructed gene drive in *Ceratitis* by collaborators (Meccariello et al. 2023).

3.3 Methods

3.3.1 Fly Husbandry

3.3.1.1 Medfly stock lines

Two wildtype medfly stock lines were used throughout these experiments. TOLIMAN is a wildtype strain from Guatemala, Central America originally collected in 1990 and reared at Oxitec LTD (Milton Park, Abingdon) from 2004 with a sub-culture established at the University of East Anglia (UEA) in 2010. The wild-type Benakeion strain (originally isolated in Athens, Greece) in which the transgenic line was established and maintained by A. Meccariello (Imperial College, London).

3.3.1.2 Medfly rearing

All wildtype medfly cultures were reared in a controlled environment with temperature set at 25°C ±0.5°C, humidity at 60% ± 10% RH and on a 12h light/dark schedule. Depending upon the demand for eggs from cultures, adult populations were kept in 1 of 3 different cage sizes: small cages (10cm x 10cm x 10cm) initiated with 50 pupae, medium cages (13cm x 13cm x 14cm) with 100 pupae and large cages (33cm x 33cm x 16cm) with 500 pupae. The emerging adult flies held in these cages were fed on an *ad libitum* diet of 3:1 sucrose:hydrolysed yeast mixture, changed every seven days. Water was supplied through the side of each cage via dental rolls soaked in RO water, which were renewed every 4 days.

All cages were designed with a mesh layer covering the majority of the surface area of one vertical wall of the cage to allow egg deposition. Water pots containing RO water were placed under the mesh side of each cage to collect eggs. Eggs laid through the cage mesh were collected after a period of no more than 24h of egg laying, by removing eggs with a Pasteur pipette and transferring up to two drops of the eggs contained in the water solution to filter paper. The filter papers with eggs were then placed into 1/3 pint milk bottles containing 100ml of ASG larval food (1L ASG food: 850ml RO water, 12.5g agar, 73.5g sucrose, 67g maize, 47.5g Brewer's Yeast, 2ml Propionic acid, 25ml Nipagin). 7 days post egg collection, bottles were prepared to allow the third instar larvae to exit and start pupation. To do this, 2 filter paper folded strips were placed in each bottle before laying the bottles down on a thin layer of sand within a pupation box (170mm x 130mm x 50mm). The box was then sealed with a lid containing a mesh for a further 7-9 days. During this time third instar 'jumping' larvae

exited the bottles and pupated within the sand. After this period, pupae were sieved from the sand through a standard metal sieve and transferred into petri dishes. From this the appropriate number of pupae required for each procedure were randomly selected and placed into a fresh cage or used in experiments.

3.3.2 Transgenic Line Establishment

3.3.2.1 Microinjection of medfly eggs to generate transgenic strains

Medfly microinjection and transgenic line establishment was performed at the Imperial College London insectary by A. Meccariello. 40µl of injection mix was used per microinjection session with a final concentration of 500ng/µl plasmid of interest, 200ng/µl [i]hyPBase transposase helper plasmid (Eckermann et al. 2018). This was made up to a final volume with injection buffer (Rubin and Spradling 1982). Embryos were collected after an egg laying period of 45 mins, transferred to double sided sticky tape and allowed to dry for 6 mins. Eggs were manually dechorionated using the tip of a blunted needle. Once the chorion was removed, eggs were transferred to slides coated in a thin layer of glue (double sided sticky tape dissolved in heptane) with the posterior of the egg facing outwards. These egg slides were desiccated by placing them in a petri dish containing calcium chloride for 6 minutes, and then covered with a thin layer of Halocarbon™ 700 Oil. Eggs on these slides were then injected using WPI Sutter Quartz needles (outer diameter (OD) = 1.00mm, inner diameter (ID) = 0.70mm) that had been drawn on a Sutter P-200/G laser needle puller. Needles were loaded with injection mix using an Eppendorf Microloader. Microinjections were performed using an inverted stereo microscope with an Eppendorf microinjection set-up and compressed air-driven pump.

Post injection, eggs were left on slides until the larvae hatched (typically after 48h) at which point the larvae were placed onto larval food (30g paper, 30g sucrose, 30g yeast extract, 10ml cholesterol stock (5g cholesterol, 140 ml distilled water, 30ml 95% EtOH), 2ml HCl Stock (384ml distilled water, 66ml 37% HCl), 8.5ml Benzoic Stock (50g Benzoic acid, 300ml 95% EtOH, 150ml distilled water) made up to 400ml with distilled water). After 7 days larval culture bottles were laid down in pupation boxes with a thin layer of fine sand and allowed to pupate. Transgenic line establishment and subsequent testing was completed by A. Meccariello.

3.3.2.2 Validation of transgenic lines

Cas9 expressing transgenic lines were tested for Cas9 expression through Real-time qPCR completed by A.Meccariello with the following protocol (Meccariello et al. 2023). Adult flies were dissected to obtain male and female somatic tissue, testes and ovaries. The samples for each group were collected in biological triplicates and were homogenized using Ambion® TRIzol reagent. RNA extraction was performed using the protocol adapted from Chomczynski and Mackey (1995) and quantified using Nanodrop™ One Spectrophotmeter. cDNA was synthesised from the total obtained RNA using Thermo Fisher Scientific™ Maxima H Minus First Strand cDNA synthesis Kit with dsDNase as per manufacturer instructions. qPCR was performed on the Applied Biosystems® 7500 Fast Real-Time PCR System using the Applied Biosystems® Fast SYBRTM Green Master Mix with cycling conditions as follows: 20s at 95°C, 40 cycles of 3s at 95°C and 30s at 60°C, followed by 15s at 95°C, 1 min at 60°C, 15s at 95°C and 15s at 60°C. Amplifications were carried out in a solution containing 10μL Thermo Scientific 2X Fast SYBR™ Green Master Mix , 2μL first-stranded cDNA (diluted 1:10) and 800nM of each primer (*Supplementary information 3.6.1*). To check reproducibility, each assay was performed with technical duplicates for each of the three biological samples. The Cas9 expression levels were normalized to the housekeeping gene *RPL19* and expression to wild-type samples, and data analysis was performed using the PCR package for R.

3.3.3 Molecular techniques

3.3.3.1 DNA Extraction

DNA extraction was performed on whole flies of both TOLIMAN and Cepa Petapa wildtype strains. Whole flies were frozen at -20°C for a minimum of 48h with 3 individuals per sample. Samples were removed from the freezer and placed onto dry ice. Samples were then homogenised with a sterilised, plastic pestle over dry ice. Qiagen DNeasy® Blood & Tissue Kit was used as per the manufacturer's instructions with the addition of an extra elution step with 25μl molecular grade water after the original 50μl elution. The concentration of each sample was measured using a Thermo Fisher Scientific Nanodrop™ 8000 Spectrophotometer.

3.3.4   Construct design and building

3.3.4.1   Construct Design

All constructs were designed in Benchling using the Gibson assembly tool. Primers were designed manually or generated by Benchling (Benchling 2023).

3.3.4.2   DNA Synthesis

All synthesised genetic parts were designed in Benchling (Benchling 2023) and ordered from Azenta Life Sciences GENEWIZ® service in a pUC-GW-Amp plasmid backbone.

3.3.4.3   Plasmid Midiprep

After sequence confirmation, plasmids were prepared for microinjection using the Macherey-Nagel NucleoBond® Xtra Midi EF, Midi kit for endotoxin-free plasmid DNA kit. Cultures were revived from glycerol stocks onto a streak plate which was incubated overnight at 37°C. A single culture from this was used for inoculation of a 4ml LB Ampicillin (100µg/ml) starter culture and incubated for 6-8h at 37°C with constant agitation at 180rpm. After confirming growth in starter cultures, 400µl was used to inoculate 100ml LB Ampicillin (100µg/ml) to be incubated overnight at 37°C with constant agitation at 200rpm. This culture was then used as per the manufacturer's instructions for midiprep. The final pellet was resuspended in 50µl endotoxin-free molecular grade water.

3.3.4.4   Midiprep Purification

A further purification step was performed on midiprep samples for microinjection using Millipore Ultrafree®-MC GV Centrifugal Filters. The full midiprep sample was loaded to the column and centrifuged at 14,000 RPM for 4 minutes.

3.3.4.5   Preparing glycerol stocks

For the preparation of glycerol stocks for long term plasmid storage, single colonies were selected to inoculate 4ml LB Ampicillin (100µg/ml). This was then incubated overnight at 37°C with constant agitation at 180rpm (typically, this was the culture that was also used for plasmid miniprep). From this, 500µl of culture was added to 500µl 50% glycerol and mixed well by inversion. Glycerol stocks were then stored at -80°C, and only the top layer was thawed when reviving them.

3.4   Results

3.4.1   Cas9 promotor design

Two of the candidate genes from Chapter 2 were selected as potential Cas9 promotors for use in a homing drive. As homing drives require more tightly controlled expression both the native 5' and 3' UTR were use in preference to the native 5' UTR and a viral 3' UTR. *bgcn* and *vasa* were selected as the most promising candidates based on their usage in other species and expression patterns obtained (Chapter 2). The 3' UTR was simple to identify and select for both *vasa* and bgcn, with the 1kb immediately following the STOP codon used in both cases. The design of *bgcn* was also relatively simple as there was no evidence of large intronic regions, so the entire promotor region could be used, which was confirmed through RACE sequencing in previous work (Chapter 2). The design for the *vasa* 5' UTR was more complex as there are several *vasa* isoforms resulting in over 6kb 5' UTR of *vasa* including intronic sequence.

The full length *vasa* sequence was analysed with the Berkeley Drosophila Genome Project splice checker (Reese et al. 1997) and the identified putative splice sites were annotated, most of which aligned well with the genome annotations already present. The start sites of these variants span approximately 7kb, which is too large to be used as a promotor as larger transgenic constructs incur a higher fitness cost to carrying and expressing them. Within this, there is a 5kb intronic region which contained no variant splice sites, start sites, identified transcription factors or binding sites, this allowed for the *vasa* promotor region to be truncated into two sections by removing this intronic sequence while retaining as many 5' UTR regions as possible. This was done to retain isoforms, as it is difficult to predict which isoforms are the most prevalent in the germline. The entire construct was flanked by AscI digestion sites which to allow for insertion into the homing construct. The final design of both these constructs can be found in *Supplementary information 3.6.2* with the final parts of this used in assembled plasmids shown in *Figure 3.2* and *Figure 3.3*.

3.4.2   Assembly of Cas9 constructs

The Cas9 constructs were designed to be put together via Gibson assembly. *bgcn* had a simpler design requiring the amplification of only two sections from the genome, a single 5' UTR and a single 3' UTR which would flank the cas9. *vasa* required the amplification of a single 3' UTR but the 5'UTR consisted of two fragments to be amplified from the genome and

annealed to each other. Amplification through PCR from the genomic DNA was completed. However, after several assembly attempts, no fully assembled constructs were obtained. Due to time constraints, the parts were synthesised externally (Azenta Life Sciences GENEWIZ®) with restriction sites that would allow for insertion into the backbone designed by Meccariello. The final plasmid designs are shown in *Figure 3.2* and *Figure 3.3* with construct parts indicated in green.

Figure 3.2 Plasmid map of UEA_001 (*vasa*_Cas9) in a standard pUC synthesis backbone (indicated in orange). Full plasmid sequence can be obtained in *Supplementary information 3.6.2*.
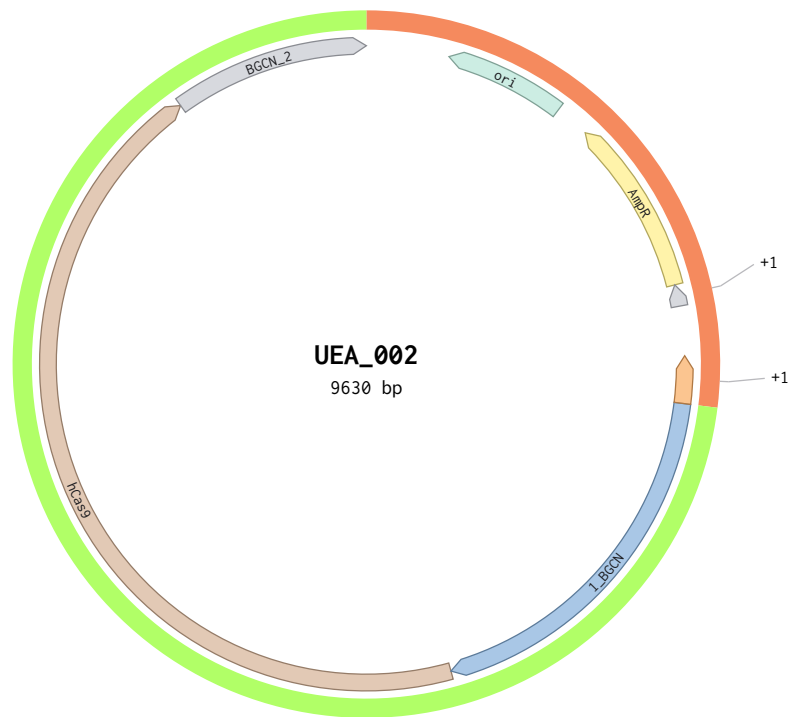
Figure 3.3 Plasmid map of UEA_002 (*bgcn*_Cas9) in a standard pUC synthesis backbone (indicated in orange). Full plasmid sequence can be obtained in *Supplementary information 3.6.2*.

These plasmids were sent to Meccariello for genome insertion through AscI restriction digest and ligation into the backbone used for the envisioned homing drive (Meccariello et al. 2023). The design strategy for Cas9 part insertion is shown in *Figure 3.4*.
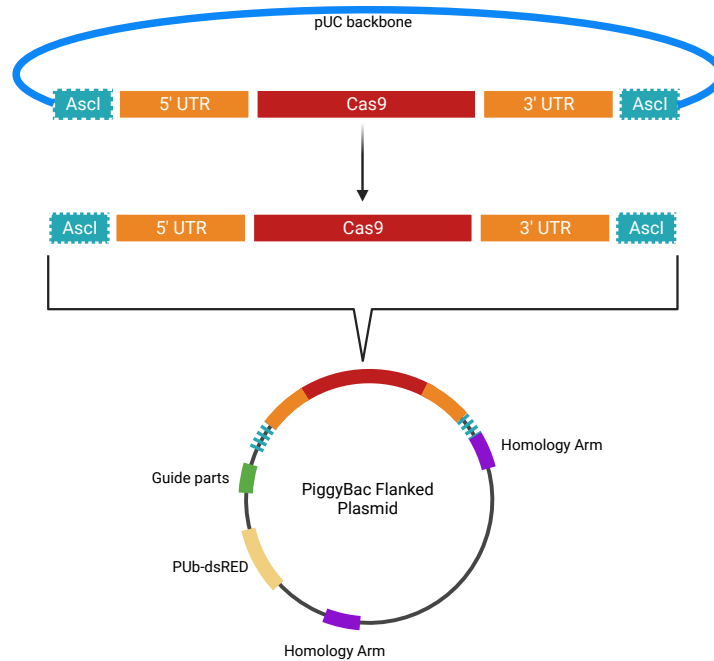
Figure 3.4 Strategy for the plug and play Cas9 part insertion which will allow for the insertion of multiple Cas9 promotors through AscI (restriction sites highlighted in blue) restriction digest and ligation.

3.4.3 Validation of Cas9 expressing lines through RT-qPCR

The *vasa* construct was successfully integrated into the medfly genome alongside two other Cas9 homing constructs designed by A. Meccariello (*zpg* and *nanos*). *bgcn* was not successfully integrated on the first attempt and we did not proceed further with it. A guide targeting *white eye* and the necessary guide scaffolding were also part of this construct. The levels of Cas9 expression across four tissue types produced by these three constructs were tested through real-time qPCR with *RPL19* used as refence gene. The expression was measured across four tissue types (male somatic, female somatic, ovaries and testes). The relative expression of Cas9 produced by each promotor is shown in *Figure 3.5*.
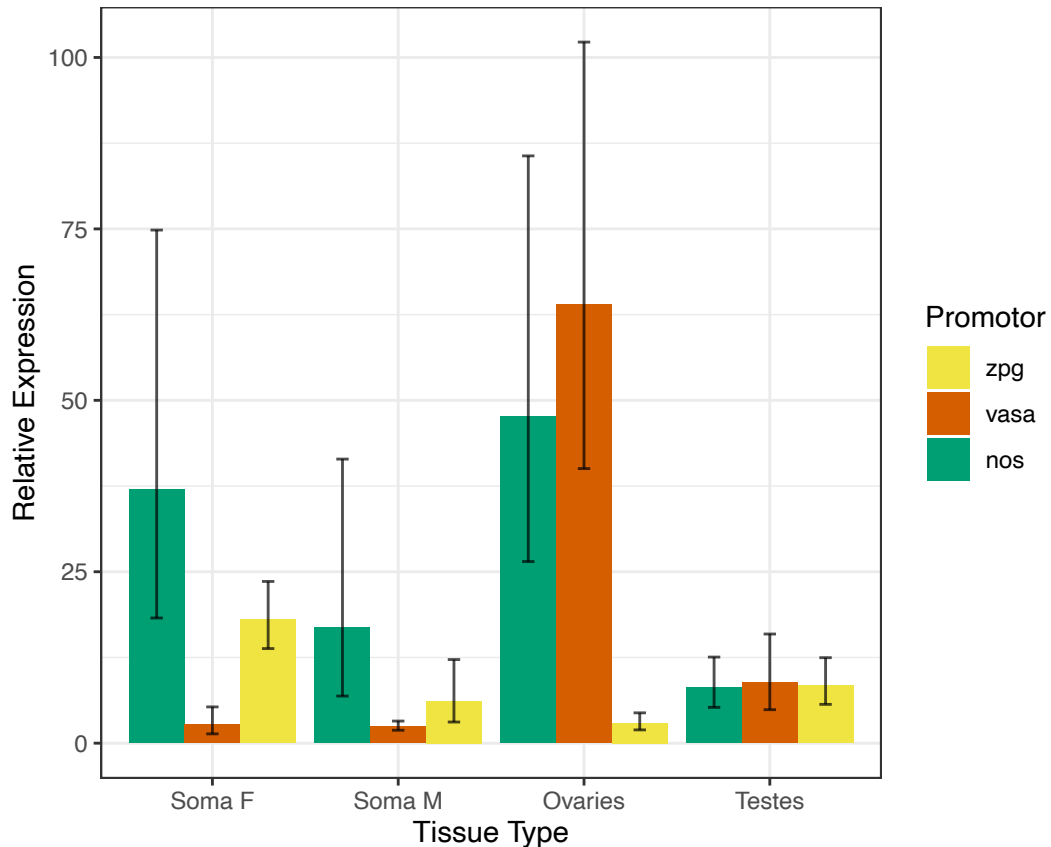
Figure 3.5 qPCR relative expression of Cas9 from each promotor across different tissue types. Error bars represent the two technical replicates of each biological sample. The three tested promotors include *zpg* and *nanos* (*nos*) designed and constructed by A. Meccariello and *vasa* designed and constructed in this Chapter. Low expression in somatic tissue and high expression in the germline is desirable for the gene drive strategy. *vasa* showed the lowest somatic and highest germline expression and was selected as the promotor to be used in the final homing drive design. *RPL19* was used as the reference gene for standardisation. Tissue types tested are female somatic (Soma F), male somatic (Soma M), ovaries and testes.

High expression in the germline tissues with low expression in the somatic tissue is necessary for a Cas9 promotor to be used in a homing drive. The *vasa* construct showed the most promising expression with both the highest level of expression in both the ovaries and testes in addition to the lowest expression in somatic tissues. This promotor was selected for subsequent use in the homing drive constructed by A. Meccariello and outlined in Meccariello et al. (2023).

3.4.4   Use in homing drives

From the preliminary qPCR data obtained, the *vasa* promotor was the most promising for use in a homing gene drive. All three promotors were integrated into a homing drive targeting the *white-eye* gene to determine cutting efficiencies of each promotor. All constructs caused super-Mendelian inheritance however only *vasa* showed balanced levels of drive in both

males and females, whereas both *nanos* and *zpg* were only able to drive from transgenic females. Therefore, my *vasa*-Cas9 construct was selected for use in the final homing drive design, targeting the sex determination gene *tra*. *Tra* disruption was seen in this homing drives, demonstrating that homing is a viable strategy in medfly. However, leaky Cas9 expression did not make this a viable homing drive for sex conversion. The full results of the homing drive experiments are presented in (Meccariello et al. 2023).

3.5   Discussion

The need for an expanded toolbox for Cas9 genome editing in medfly is clear in the context of the construction of functional homing drives. Cas9 promotors require a high degree of specificity for use in homing drives, with expression outside of the germline being deleterious because it results in mosaicism and increases the chance of resistance allele formation through NHEJ (Unckless, Clark, and Messer 2017). Homing gene drives reply upon the use of homology directed repair (HDR) rather than NHEJ to repair cuts in DNA. However, HDR is favoured over NHEJ only in specific cell developmental stages (Raban, Marshall, and Akbari 2020). For use in homing drives expression must be minimised in somatic tissue and embryos to prevent the formation of resistance alleles (Champer et al. 2018). Here, we designed endogenous germline promotors to restrict expression to the germline and tested which was most effective at inducing homing.

Two Cas9 promotors were designed in this work, *vasa* and *bgcn*. *vasa* has been used extensively as a Cas9 promotor (Champer et al. 2017; Feng et al. 2021; Adolfi et al. 2020). *bgcn* has been used less commonly although it has been tested in *Aedes* though low cutting rates were observed (Verkuijl et al. 2022). *nanos* was previously identified as superior to *vasa* for use as a Cas9 promotor due to lower somatic expression levels in *Drosophila melanogaster* (Champer et al. 2018) and *Anopheles gambiae* (Kyrou et al. 2018). However, in the expression data obtained here, *vasa* had demonstrably lower levels of somatic expression in both male and female tissue than was observed in the other promotors tested including *nanos*. This supports the need for species specific research into promotor activity to resolve this unexplained variation between species (Du et al. 2023). Even with transgene expression mimicking that seen in the gene from which the regulatory regions were taken, this can still present problems due to protein persistence and deposition. Cas9 deposition into embryos is often an issue with Cas9 promotors and can lead to the rapid accumulation of resistance alleles (Champer et al. 2018). This is of particular concern when high levels of Cas9 are produced as this can lead to higher levels of persistence (Champer et al. 2017).

High Cas9 expression in the germline is desirable though it could also result in an increase in deposition if there is more protein around to persist. Previous designs I have constructed used a non-endogenous 3-UTR which is suitable for most uses as it upregulates transgene expression. However, high levels of expression although desirable in homing drives, is a secondary need to confining the expression to the germline therefore the designs in this work used native 3' UTR. The use of a native 3' UTR has been shown to increase protein stability

(Tangphatsornruang et al. 2011) and more closely mimic native gene expression (Ebron and Shukla 2016). The 3' UTR has been shown to have an essential role in localisation in *Drosophila* (Gavis et al. 1996). This further supports the decision to include the native 3' UTR when designing Cas9 regulatory elements for homing drives as it could allow for expression which more closely mimics that of the native gene. An investigation into Cas9 transgene localisation of *nanos* and *vasa* driven Cas9 in comparison to native *nanos* and *vasa* expression showed localisation, within the ovaries, of Cas9 that was not the same as the endogenous pattern in *Anopheles* with Cas9 transcripts accumulating in the outer follicle cytoplasm due to the absence of a *nanos* signal (Terradas et al. 2022). The regulatory elements used in these studies did not perfectly match that of the native promotors (the removal of introns) which I hypothesised might have affected localisation patterns. This evidence does however raise questions about how the editing in the *vasa* 5' UTR will affect localisation if any non-intronic sequence was inadvertently removed.

The removal of intronic sequences in 5' of the *vasa* construct was risky, but necessary to reduce the sequence to a length suitable for use in a transgene. Previous removal of the intronic region of the *vasa* 5'UTR has resulted in undesirable sex-limited expression in *Anopheles spp* (Papathanos et al. 2009). The intronic removal done here was conservative and efforts were made to retain all isoform splice sites. There was no evidence of single sex expression with the construct I made. However, it is possible that the removed intronic regions could have an unknown role in localisation.

vasa was identified as the most promising promotor for use in a homing drive for medfly from those tested. It demonstrated the most balanced level of drive between sexes. However, after integration into homing drives it did exhibit some limitations. High levels of somatic expression and deposition resulted in female fitness costs, which could promote the accumulation of resistance alleles. This highlights the current limitations of qPCR for determining tissue-specific expression levels. As discussed in Chapter 2, the use of a single reference gene can make it difficult to distinguish if differences in expression between tissue types are true differences or if they are affected by a difference in reference gene expression between tissue types.

Despite this, the results show the first attempt at homing in *Ceratitis* and promising results were obtained, such as high efficiency of homing and the ability to manipulate *tra* for conversion. Further optimisation efforts will continue. The constructed *bgcn*-Cas9 promotor

can now be tested, as the expression patterns of this gene were among the most promising for use as a Cas9 promotor. The *bgcn* construct will be continued with and tested with the *white eye* guide to determine if this may be a more suitable promotor. Using alternative germline-specific genes which have not yet been tested may provide a superior Cas9 promotor. There still remains a wealth of germline specific genes which have not been tested so work in this field has several potential avenues to explore to find an effective Cas9 promotor for homing drives. This work highlights how continuing to expand the toolbox for genetic modification can contribute to future pest control strategies.

3.6   References

Adolfi, Adriana, Valentino M. Gantz, Nijole Jasinskiene, Hsu-Feng Lee, Kristy Hwang, Gerard Terradas, Emily A. Bulger, et al. 2020. 'Efficient Population Modification Gene-Drive Rescue System in the Malaria Mosquito *Anopheles stephensi*'. *Nature Communications* 11 (1): 5553. https://doi.org/10.1038/s41467-020-19426-0.

Anderson, Michelle A. E., Estela Gonzalez, Joshua X. D. Ang, Lewis Shackleford, Katherine Nevard, Sebald A. N. Verkuijl, Matthew P. Edgington, Tim Harvey-Samuel, and Luke Alphey. 2023. 'Closing the Gap to Effective Gene Drive in *Aedes aegypti* by Exploiting Germline Regulatory Elements'. *Nature Communications* 14 (1): 338. https://doi.org/10.1038/s41467-023-36029-7.

Benchling. 2023. 'Biological Software'. Benchling.com.

Burt, Austin. 2003. 'Site-Specific Selfish Genes as Tools for the Control and Genetic Engineering of Natural Populations'. *Proc. Biol. Sci.* 270 (1518): 921–28. https://doi.org/10.1098/rspb.2002.2319.

Carballar-Lejarazú, Rebeca, Christian Ogaugwu, Taylor Tushar, Adam Kelsey, Thai Binh Pham, Jazmin Murphy, Hanno Schmidt, Yoosook Lee, Gregory C. Lanzaro, and Anthony A. James. 2020. 'Next-Generation Gene Drive for Population Modification of the Malaria Vector Mosquito, *Anopheles gambiae*'. *Proceedings of the National Academy of Sciences* 117 (37): 22805–14. https://doi.org/10.1073/pnas.2010214117.

Champer, Jackson, Anna Buchman, and Omar S Akbari. 2016. 'Cheating Evolution: Engineering Gene Drives to Manipulate the Fate of Wild Populations'. *Nat. Rev. Genet.* 17 (3): 146–59. https://doi.org/10.1038/nrg.2015.34.

Champer, Jackson, Jingxian Liu, Suh Yeon Oh, Riona Reeves, Anisha Luthra, Nathan Oakes, Andrew G Clark, and Philipp W Messer. 2018. 'Reducing Resistance Allele Formation in CRISPR Gene Drive'. *Proc. Natl. Acad. Sci. U. S. A.* 115 (21): 5522–27. https://doi.org/10.1073/pnas.1720354115.

Champer, Jackson, Riona Reeves, Suh Yeon Oh, Chen Liu, Jingxian Liu, Andrew G. Clark, and Philipp W. Messer. 2017. 'Novel CRISPR/Cas9 Gene Drive Constructs Reveal Insights into Mechanisms of Resistance Allele Formation and Drive Efficiency in Genetically Diverse Populations'. Edited by Harmit S. Malik. *PLOS Genetics* 13 (7): e1006796. https://doi.org/10.1371/journal.pgen.1006796.

Du, Jie, Weizhe Chen, Xihua Jia, Xuejiao Xu, Emily Yang, Ruizhi Zhou, Yuqi Zhang, Matt Metzloff, Philipp W. Messer, and Jackson Champer. 2023. 'New Germline Cas9 Promoters Show Improved Performance for Homing Gene Drive'. Preprint. Genetics. https://doi.org/10.1101/2023.07.16.549205.

Ebron, Jey Sabith, and Girish C. Shukla. 2016. 'Molecular Characterization of a Novel Androgen Receptor Transgene Responsive to MicroRNA Mediated Post-Transcriptional Control Exerted via 3'-Untranslated Region: Characterization of a Novel Androgen Receptor Reporter'. *The Prostate* 76 (9): 834–44. https://doi.org/10.1002/pros.23174.

Eckermann, Kolja N., Hassan M.M. Ahmed, Mohammad KaramiNejadRanjbar, Stefan Dippel, Christian E. Ogaugwu, Peter Kitzmann, Musa D. Isah, and Ernst A. Wimmer. 2018. 'Hyperactive piggyBac Transposase Improves Transformation Efficiency in Diverse Insect Species'. *Insect Biochemistry and Molecular Biology* 98 (July): 16–24. https://doi.org/10.1016/j.ibmb.2018.04.001.

Esvelt, Kevin M, Andrea L Smidler, Flaminia Catteruccia, and George M Church. 2014. 'Concerning RNA-Guided Gene Drives for the Alteration of Wild Populations'. *Elife* 3 (July). https://doi.org/10.7554/eLife.03401.

Feng, Xuechun, Víctor López Del Amo, Enzo Mameli, Megan Lee, Alena L. Bishop, Norbert Perrimon, and Valentino M. Gantz. 2021. 'Optimized CRISPR Tools and Site-Directed Transgenesis towards Gene Drive Development in *Culex quinquefasciatus* Mosquitoes'. *Nature Communications* 12 (1): 2960. https://doi.org/10.1038/s41467-021-23239-0.

Gantz, Valentino M, and Ethan Bier. 2015. 'Genome Editing. The Mutagenic Chain Reaction: A Method for Converting Heterozygous to Homozygous Mutations'. *Science* 348 (6233): 442–44. https://doi.org/10.1126/science.aaa5945.

Gantz, Valentino M, Nijole Jasinskiene, Olga Tatarenkova, Aniko Fazekas, Vanessa M Macias, Ethan Bier, and Anthony A James. 2015. 'Highly Efficient Cas9-Mediated Gene Drive for Population Modification of the Malaria Vector Mosquito *Anopheles stephensi*'. *Proc. Natl. Acad. Sci. U. S. A.* 112 (49): E6736-43. https://doi.org/10.1073/pnas.1521077112.

Gavis, Elizabeth R., Daniel Curtis, and Ruth Lehmann. 1996. 'Identification Of cis-Acting Sequences That Control nanos RNA Localization'. *Developmental Biology* 176 (1): 36–50. https://doi.org/10.1006/dbio.1996.9996.

Hammond, Andrew, Roberto Galizi, Kyros Kyrou, Alekos Simoni, Carla Siniscalchi, Dimitris Katsanos, Matthew Gribble, et al. 2016. 'A CRISPR-Cas9 Gene Drive System Targeting Female Reproduction in the Malaria Mosquito Vector *Anopheles gambiae*'. *Nat. Biotechnol.* 34 (1): 78–83. https://doi.org/10.1038/nbt.3439.

Kyrou, Kyros, Andrew M Hammond, Roberto Galizi, Nace Kranjc, Austin Burt, Andrea K Beaghton, Tony Nolan, and Andrea Crisanti. 2018. 'A CRISPR–Cas9 Gene Drive Targeting Doublesex Causes Complete Population Suppression in Caged *Anopheles gambiae* Mosquitoes'. *Nature Biotechnology* 36 (11): 1062–66. https://doi.org/10.1038/nbt.4245.

Meccariello, Angela, Shibo Hou, Serafima Davydova, James Fawcett, Alexandra Siddall, Philip T. Leftwich, Flavia Krsticevic, Philippos Aris Papathanos, and Nikolai Windbichler. 2023. 'Gene Drive and Genetic Sex Conversion in the Global Agricultural Pest *Ceratitis capitata*'. Preprint. Synthetic Biology. https://doi.org/10.1101/2023.08.16.553191.

Meccariello, Angela, Flavia Krsticevic, Rita Colonna, Giuseppe Del Corsano, Barbara Fasulo, Philippos Aris Papathanos, and Nikolai Windbichler. 2021. 'Engineered Sex Ratio Distortion by X-Shredding in the Global Agricultural Pest *Ceratitis capitata*'. *BMC Biology* 19 (1): 78. https://doi.org/10.1186/s12915-021-01010-7.

Papathanos, Philippos A, Nikolai Windbichler, Miriam Menichelli, Austin Burt, and Andrea Crisanti. 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): 65. https://doi.org/10.1186/1471-2199-10-65.

Raban, Robyn R., John M. Marshall, and Omar S. Akbari. 2020. 'Progress towards Engineering Gene Drives for Population Control'. Edited by Michael H. Dickinson, Leslie B. Vosshall, and Julian A. T. Dow. *Journal of Experimental Biology* 223 (Suppl_1): jeb208181. https://doi.org/10.1242/jeb.208181.

Reese, Martin G., Frank H. Eeckman, David Kulp, and David Haussler. 1997. 'Improved Splice Site Detection in Genie'. *Journal of Computational Biology* 4 (3): 311–23. https://doi.org/10.1089/cmb.1997.4.311.

Rubin, Gerald M., and Allan C. Spradling. 1982. 'Genetic Transformation of *Drosophila* with Transposable Element Vectors'. *Science* 218 (4570): 348–53. https://doi.org/10.1126/science.6289436.

Siddall, Alex, Tim Harvey-Samuel, Tracey Chapman, and Philip T. Leftwich. 2022. 'Manipulating Insect Sex Determination Pathways for Genetic Pest Management: Opportunities and Challenges'. *Frontiers in Bioengineering and Biotechnology* 10 (June): 867851. https://doi.org/10.3389/fbioe.2022.867851.

Tangphatsornruang, Sithichoke, Ian Birch-Machin, Christine A. Newell, and John C. Gray. 2011. 'The Effect of Different 3' Untranslated Regions on the Accumulation and Stability of Transcripts of a Gfp Transgene in Chloroplasts of Transplastomic Tobacco'. *Plant Molecular Biology* 76 (3–5): 385–96. https://doi.org/10.1007/s11103-010-9689-1.

Terradas, Gerard, Anita Hermann, Anthony A James, William McGinnis, and Ethan Bier. 2022. 'High-Resolution *in Situ* Analysis of Cas9 Germline Transcript Distributions in Gene-Drive *Anopheles* Mosquitoes'. Edited by K McKim. *G3 Genes|Genomes|Genetics* 12 (1): jkab369. https://doi.org/10.1093/g3journal/jkab369.

Unckless, Robert L, Andrew G Clark, and Philipp W Messer. 2017. 'Evolution of Resistance Against CRISPR/Cas9 Gene Drive'. *Genetics* 205 (2): 827–41. https://doi.org/10.1534/genetics.116.197285.

Verkuijl, Sebald A. N., Estela Gonzalez, Ming Li, Joshua X. D. Ang, Nikolay P. Kandul, Michelle A. E. Anderson, Omar S. Akbari, Michael B. Bonsall, and Luke Alphey. 2022. 'A CRISPR Endonuclease Gene Drive Reveals Distinct Mechanisms of Inheritance Bias'. *Nature Communications* 13 (1): 7145. https://doi.org/10.1038/s41467-022-34739-y.

Yadav, Amarish K., Cole Butler, Akihiko Yamamoto, Anandrao A. Patil, Alun L. Lloyd, and Maxwell J. Scott. 2023. 'CRISPR/Cas9-Based Split Homing Gene Drive Targeting *Doublesex* for Population Suppression of the Global Fruit Pest *Drosophila suzukii*'. *Proceedings of the National Academy of Sciences* 120 (25): e2301525120. https://doi.org/10.1073/pnas.2301525120.

3.7 Appendices

3.7.1 Preprint homing drive paper – accepted in Nature Communications

Gene drive and genetic sex conversion in the global agricultural pest *Ceratitis capitata*

Angela Meccariello[1], Shibo Hou[1], Serafima Davydova[1], James Fawcett[1], Alexandra Siddall[2], Philip T. Leftwich[2], Flavia Krsticevic[3], Philippos Aris Papathanos[3], Nikolai Windbichler[1].

1. Department of Life Sciences, Imperial College London, London SW7 2AZ, UK.
2. School of Biological Sciences, University of East Anglia, Norwich NR4 7TJ, United Kingdom
3. Department of Entomology, Robert H. Smith Faculty of Agriculture, Food and Environment, Hebrew University of Jerusalem, Rehovot, Israel

Abstract

Homing-based gene drives are novel interventions promising the area-wide, species-specific genetic control of harmful insect populations. Here we characterise a first set of gene drives in a tephritid agricultural pest species, the Mediterranean fruit fly *Ceratitis capitata* (medfly). Our results show that the medfly is highly amenable to homing-based gene drive strategies. By targeting the medfly *transformer* gene, we also demonstrate two different mechanisms by which CRISPR-Cas9 gene drive can be coupled to sex conversion, whereby genetic females are transformed into fertile and harmless XX males. Given this unique malleability of sex determination, we modelled gene drive interventions that couple sex conversion and female sterility and found that such approaches could be effective and tolerant of resistant allele selection in the target population. Our results open the door for developing gene drive strains for the population suppression of the medfly and related tephritid pests by co-targeting female reproduction and shifting the reproductive sex ratio towards males. They demonstrate the untapped potential for gene drives to tackle agricultural pests in an environmentally friendly and economical way.

1

Introduction

Homing gene drives were originally proposed 20 years ago[1] are now under development in multiple insect species of medical, agricultural or ecological importance[2–13]. Suppressive homing gene drives are designed to reduce the population size of harmful pest organisms by, in most cases, targeting genes essential for insect fitness using a homing CRISPR endonuclease expressed in the germline. As a result, the homing construct, strategically placed within the target gene thereby disrupting it, is transmitted at rates exceeding Mendelian inheritance (gene drive). In most insect species, the number and productivity of females determines population size, and often crop damage or disease transmission occurs solely through the activities of females[14]. As such, approaches that interfere with female development or shift the sex ratio towards males are among the most promising strategies being explored[15]. No gene drive has ever been tested in the field, however for medically relevant mosquito species like the malaria mosquito *A. gambiae*, these technologies have reached a stage where such gene drives can reliably eliminate mosquito populations in the laboratory[16–18]. Consequently, efforts are now being made to establish the technical, social and regulatory frameworks for field testing and potential deployment.

Tephritid fruit flies, such as the Mediterranean fruit fly (medfly) *Ceratitis capitata*, represent a large group of economically important agricultural pests[19] that are exceedingly well suited for expanding the development of suppressive gene drives beyond mosquitoes: First, they are among the most important agricultural fruit fly pests affecting global food production[20]. Second, many species are non-endemic in affected areas, exacerbating their impact; for example, the medfly in the Americas[21–23]. The current gold-standard for Tephritid population suppression worldwide is genetic control, specifically the Sterile Insect Technique (SIT)[24]. Although transgenic approaches have been developed[25], classic SIT, based on genetic sexing strains and pupal irradiation, has seen the most widespread adoption and success for the last 30 years. The success of these SIT programs was built on decades of research and development that can now be used to fertilise the rapid development of gene drives in these species. This includes not only a deep understanding of species-specific traits that could now be leveraged, but also operational know-how on mass-rearing, release strategies and post-release monitoring.

Recently, CRISPR/Cas9 based genome editing has been established in a number of important Tephritid species. In the medfly specifically, an endogenous toolbox for CRISPR genome editing was established, including regulatory elements for germline and somatic Cas9 and sgRNA expression, fluorescent markers and endogenous markers to score gene editing efficiency markers[26]. This progress has coincided with the completion of functional genetic studies mapping the sex-determination pathway, leading to the identification of the Y-chromosome linked male-determing factor *MoY*, that overwrites default female development[27,28]. Interestingly, unlike mosquitoes or *Drosophila*, the pathway is unusually malleable when perturbed. For example, the generation of XX fertile males and XY fertile females is possible[28]. This opens exciting opportunities for genetic control through true genetic sex conversion, an impossibility in many other insects due to the role of the Y chromosome for male fertility, or dosage compensation of X-chromosome linked genes[29,30].

Here we sought to test a first set of gene drive constructs in the medfly targeting the known phenotypic marker gene *white-eye* using different regulatory elements to evaluate their performance in the medfly germline. We also sought to target the medfly sex determination

pathway to explore the possibility of generating sex conversion gene drives in this and related Tephritid species.

Results

Evaluating CRISPR-Cas9 induced homing in the medfly germline

We first established homing-capable gene drive constructs within the well-studied *white-eye* gene[31] (*white-eye*, GeneID_101458180) to test whether the medfly germlines are generally amenable to homologous recombination-based repair (**Figure 1**). For this purpose, we placed Cas9 under the transcriptional control of three conserved regulatory elements that have been used extensively in other insect species. These included orthologs of the *nanos*[32] (GeneID_101451248), *vasa* (GeneID_101456741) and *zpg* (GeneID_101449162) genes, from which we extracted both 5' and 3' regulatory regions. *Vasa* has at least five predicted splicing isoforms. To capture all possible 5'UTR variants and the endogenous promoter sequence, we removed 5951bp of intronic sequence between transcript variant X2 and X3, retaining only the splice junctions while preserving 978bp of 5'UTR and 1.5kb of 5' putative promoter sequence. The final gene drive constructs (**Figure S1, Supplementary File 1**) also included the *pUb-DsRed* marker gene and an expression cassette for a previously characterised gRNA targeting exon three of the *white-eye* gene[26,33]. All constructs contained approximately 1kb homology arms flanking the gRNA site, for insertion by homology based knockin, directly within the target site of *white-eye*, thereby disrupting it. The plasmids were injected into wild type (Benakion) embryos and transgenic F1 individuals displaying DsRed fluorescence were used to establish each transgenic strain by backcrossing.

To determine the rate of inheritance bias, we first crossed transgenic females or transgenic males to either *white-eye* homozygous mutant (**Figure 1**, upper panels) or wild type medflies (**Figure 1**, lower panels) and scored progeny for the DsRed marker. For all crosses involving transgenic females, we observed significant super-Mendelian inheritance above 70% (**Figure 1B**), from which *nanos* induced the highest mean level of gene drive ($\bar{x}$=85.6%). Activity in the male germline generally resulted in lower transmission bias, compared to the female germline for each construct. The highest transmission bias in males was driven by the vasa promoter, with no observable transmission bias for constructs featuring the *nanos* regulatory regions driving Cas9. For the *zpg* strain, the cross to *white-eye* mutant females ($\bar{x}$=57.5%, X^2=22.1, p<0.001) revealed significant levels of super-Mendelian transmission but not to wild type females ($\bar{x}$=51.8%, X^2=0.71, p=0.3989). Overall, the *vasa* construct showed balanced levels of drive in both sexes achieving a level of 78.1% and 70.7% mean transmission in the cross to wild type males and females, respectively. These results indicate that the medfly is generally amenable to homing-based gene drives in both sexes, and that gene drive performance could be further optimised in males by regulating Cas9 expression levels and timing.

Evaluating CRISPR-Cas9 activity and parental embryonic deposition in the medfly germline

To further correlate homing rates in these crosses with levels of Cas9 activity and to measure rates of parental CRISPR-Cas9 deposition, we analysed the DsRed negative progeny i.e. non-carriers of the drive element (**Figure 1C**) for eye colour phenotypes caused by disruption of the *white-eye* gene. Crosses to *white-eye* mutant medflies revealed that for transgenic

3

mothers on average 74.7% (*nanos*), 35.4% (*vasa*) and 49.5% (*zpg*) of chromosomes, that had not participated in gene conversion via homing (i.e DsRed-), carried disrupted target sites. These events are likely the result of non-homologous end joining repair following cleavage in the germline giving rise to white eyed progeny (**Figure 1C**, upper panel). Male transgenic crosses yielded mean mutation rates of 5.5% (*nanos*), 36.2% (*vasa*) and 34.8% (*zpg*) indicating that lower Cas9 levels in testes may explain the lower rates of homing observed for *nanos* and *zpg*. Overall, the large fraction of wild type, dark eyed (red eyed) individuals, indicative of unmodified chromosomes, in these experiments suggested that the Cas9 expression levels achievable with these promoters are likely limiting. No cases of somatic mosaicism with regards to eye phenotype were detected in the crosses to the *white-eye* mutants, as expected.

To evaluate parental embryonic deposition and post-zygotic activity of Cas9 we compared eye phenotypes in crosses of transgenic males and females to wild type individuals (**Figure 1C**, lower panel). In the cross of transgenic females to wild type males, activity of Cas9 acting on the paternal unmodified copy of *white-eye* is necessary for white eyed progeny to arise. In line with this expectation, we observed reduced levels of mutant progeny and some evidence of somatic mosaicism in the offspring of *nanos* (42.2% white eyed progeny) and *zpg* (7.5% white eyed progeny) drive females (**Figure S2A**). This trend did not hold for the *vasa* construct yielding 87.1% mutant progeny, suggesting very high levels of maternal deposition in this set of experiments. Finally, we observed no white eyed progeny when male transgenics were crossed to wild type females, except again for the *vasa* crosses where sporadic occurrence ($\bar{x}$=0.049%) of white eyed progeny suggests the possibility of paternal carryover of Cas9. The observed trends were broadly confirmed when we scored the eye phenotype of the DsRed positive fraction of the progeny (**Figure S2B**) although this set of experiments allows no distinction between somatic Cas9 expression and Cas9 carryover.

Generation of a gene drive targeting the medfly transformer gene

We next sought to exploit these insights to target the *C. capitata transformer* gene (*Cctra*, GeneID_101456163) with an in-locus homing gene drive. Since the *vasa* promoter offered the best overall activity in the male and female germlines, we used it to establish a gene drive construct located directly within the *Cctra* locus (*tra* drive), thereby disrupting it. The construct expresses a single gRNA targeting exon one of *tra* (**Figure S1**). Following the establishment of this strain from a G0 founder male we did not observe the occurrence of transgenic females in the progeny, instead observing males and intersex individuals showing either complete or partial "patchy" expression of the DsRed marker (**Figure S3A**). Over multiple generations, we crossed transgenic males carrying the *tra* drive to wild type females (**Figure 2A**). The overall rate of transmission was 83.1%, a higher rate of gene drive than observed from within the *white-eye* locus (**Figure 2B**). When we analysed the DsRed negative progeny for sexual phenotypes, we found a significant proportion of intersex individuals (20.0%) in addition to males (44.7%) and females (35.1%) (**Figure 2C** and **Figure S3B**). Strikingly, among the DsRed positive progeny no transgenic females were obtained. The progeny consisted exclusively of intersexes (35.3%) and males (representing 64.6% of the DsRed negative and 61.3% of the total progeny). To better understand the observed phenotypes, we carried out molecular analysis of the *tra* gene by Sanger sequencing, as well as karyotyping of DsRed positive male and intersex individuals utilising sex-chromosome specific PCR primers on genomic DNA. This analysis (**Supplementary Table 1**) revealed that

4

most individuals inheriting the *tra* drive showed, with regards to the amplicon from the maternally inherited chromosome, mixed sequencing traces starting near and centred around the predicted Cas9 cleavage site, and thus were genetic mosaics with regards to the *tra* target site. This suggests that leaky somatic Cas9 expression and somatic cleavage of *tra* resulted in sex conversion of females to XX intersex and XX males, which most likely accounts for the observed sex bias. No clear correlation was observed between the genotypes and karyotypes and the fluorescence pattern (complete or patchy) observed.

We next characterised the fertility of XX males and intersex individuals arising from crosses of the *tra* drive. By karyotyping 20 transgenic and 10 non-transgenic males that had been crossed to wildtype females individually we identified three transgenic XX male and two non-transgenic XX male individuals. All successfully sired offspring, with the XX non-transgenic males giving rise to all-female progeny (**Supplementary Table 2**). This finding corroborates previous studies, suggesting that XX males are viable and fertile[34]. By contrast, none of the crosses of XX intersex individuals yielded progeny (**Supplementary Table 3**). Despite some limitations of the *tra* drive construct, our transgenic approach demonstrates how gene drive and true sex conversion of females into fertile XX males could be achieved by a single construct in *C. capitata*.

Generation of a sex conversion gene drive disrupting the maternal provision of transformer

In the medfly, maternal *tra* gene product initiates a regulatory loop necessary for female development[27]. Disrupting *tra* expression in the ovary and inheritance via the female germline could therefore be a component of an attractive genetic control strategy based on sex conversion. To explore this strategy, we generated another strain where we supplemented the previously characterised gene drive construct targeting *white-eye* (utilising the *vasa* promoter) with an additional gRNA targeting transformer (*white* + *tra* drive) *in trans* (**Figure 2D**). In this strategy the expression of two gRNAs was designed to both bias inheritance of the transgene via homing into the *white-eye* locus and, at the same time, disrupt the *transformer* gene in the female germline by inducing disruptive indels. For this set of experiments, we were able to obtain transgenic individuals of both sexes, which were then crossed to wild type medflies. We then analysed DsRed fluorescence, as well as sexual and eye colour phenotypes in the resulting progeny.

First, we recorded transmission rates of 72.4% and 61.0% in female and male transgenic crosses, respectively (**Figure 2E**), rates of gene drive that were in line with the pilot experiments featuring the *white* drive with a single gRNA. Male transgenics gave rise to a weakly male-biased progeny (52.8%, X^2=9.2, p=0.002) with the sporadic occurrence of intersex individuals (**Figure 2F**). These results were consistent with our expectations of the recessive nature of the *Cctra* function, when mothers are wild types. By contrast, we observed a strong male bias in the progeny of transgenic mothers, with males and intersexes representing 80.3% and 7.3% of the total offspring, respectively (**Figure 2F**). These results suggest a strong maternal effect on the sex ratio of the progeny of *white* + *tra* females. With regards to the *white-eye* locus, the analysis of eye colour phenotypes in these crosses also confirmed a strong maternal effect, which led to progeny consisting almost exclusively of white-eyed individuals (**Figure S4**). By contrast, only a small fraction of mosaic (6.9%) or white eyed (0.79%) individuals was identified in the progeny of transgenic males, the majority of which were scored as DsRed positive (92%) offspring. Since targeting *tra* in the ovary could result in an unintended reduction in female fitness, we also analysed the number of eggs laid

5

by females (a measure of fecundity) and the number of eggs that hatched among those laid (a measure of fertility) of the *white + tra* drive and compared it to all other transgenic strains generated in this study (**Figure S5**). We found no significant differences in fertility of either male or female transgenic medflies compared to the wild type control. In summary, these results hint at a hitherto unexplored molecular mechanism, by which the reproductive sex ratio could be shifted towards males, whereby transgenic females would predominantly give rise to transgenic sons.

Modelling sex conversion gene drives

The *white + tra* drive as currently constituted does not alter female fertility as is the case in conventional suppressive gene drive designs that have been tested e.g. in mosquitoes. However, if it were to be retargeted towards a female fertility gene, it could constitute a powerful new intervention strategy, we term sterilising sex conversion (SSC). When SSC males are crossed to wild type females, super-Mendelian inheritance of the construct (thus inactivating the female fertility gene) and transmission of a disrupted *tra* allele occurs. However, in SSC females, in addition to the above, the provision of maternal *tra* would be disrupted leading to the generation of male-only XY and XXprogeny. As the SSC construct spreads, the progressive disruption of *tra* in the population would lead to an extreme male-bias, while targeting of female fertility genes would ensure that the remaining females are increasingly likely to be sterile. As the population becomes increasingly male biased, functional resistant alleles of the female fertility gene are more likely to be found in a phenotypic male, thereby limiting their relative selective advantage over non-functional alleles. To evaluate this possibility, we modelled this novel strategy using an agent-based discrete-generation gene drive model and compared it to canonical gene drives directly targeting either female fertility (**Figure 3A, E and I**) or directly targeting *transformer* function (**Figure 3B, F and J**) which have previously been modelled[29] (**Figure 3**). Briefly, we studied the effect of overall Cas9 activity i.e. the rate of cleavage in the germline and a proportional maternal effect) and the effect of the functional conservation of the target site, i.e. the rate at which non-functional R2 mutations are generated as opposed to R1 resistance mutations that maintain target gene function. The rate of homology directed repair (HDR) vs. non-homologous repair was the third dimension analysed in our model.

We find that the SSC strategy (**Figure 3C, G and K**) exhibits elevated tolerance for functional drive resistance. This is because SSC targets females in two synergistic ways: First, it can convert genetic daughters into fertile sons by cleaving the master female gene *transformer* in the germline, thereby abolishing maternal *tra* function, without which all progeny of SSC females develop as fertile males. The gRNA targeting *tra* is co-expressed from within a gene drive targeting female fertility. Since both gRNAs are always coinherited in the single SSC construct, its presence rarely affects the fertility of the individual harbouring it (homozygous carriers are fertile because they are also always male) and thus the spread of the drive results in rapid population suppression. For the same reason, SSC also sidesteps the issue of maternal deposition of Cas9 known to cause unwanted high rates of non-homologous end joining (NHEJ) in the embryo. Although maternal deposition of Cas9 in the embryos could generate R1 alleles at the female fertility target, these would not immediately be selected for, because the lack of maternal *tra* would cause such embryos to develop into males. SSC thus pre-empts and delays counterselection rendering it more robust against functional resistance at each target gene in our model.

6

Discussion

We describe the establishment of homing gene drives in a Tephritid agricultural pest species, the Mediterranean fruit fly *Ceratitis capitata*. Levels of homing recorded in the literature for different classes of target organisms vary dramatically with some species (e.g. *Anopheles* mosquitoes) showing near optimal levels, whereas for others (e.g. rodents) homing has not been shown to always occur at appreciable levels in the germline[13,35]. The levels of gene drive we observed in the medfly using the CRISPR-Cas9 system and employing a set of first generation of germline-specific promoter elements suggest that homing strategies are viable and attractive in this species and could be used to develop effective population suppression tools for this important pest. At the *white-eye* locus, homing in the female germline, was generally higher than in males, and the fraction of uncut chromosomes we observed in both sexes suggests that Cas9 activity levels, rather than a tendency towards non-homologous repair outcomes, was the rate-limiting factor for homing. Only the *vasa* promoter supported substantial levels of drive in both sexes. Improved regulatory elements should therefore be identified and tested as a next step to generate more effective drives.

We also generated a gene drive construct disrupting the *C. capitata transformer* locus. The medfly *tra* gene provides continuous female-specific Tra function in XX embryos and maintains cell-autonomous female identity. Indeed, it has been shown previously that interfering with *Cctra* expression in XX individuals can trigger complete sexual transformation of both the germline and somatic tissues in adult medflies, resulting in a fertile male XX phenotype[27]. The *tra* drive yielded high levels of homing in males and, unexpectedly, a male-biased progeny in crosses to wild type females. Given that our pilot experiments at the *white-eye* locus revealed a lack of evidence for substantial paternal carryover of Cas9, and since we observed at best a modest sex distortion in the DsRed negative fraction of the *tra* drive progeny, a significant male biased sex ratio as observed for of the DsRed positive progeny was unexpected. The maternal provision of *transformer* and inheritance of a maternal wild type *tra* allele in this cross should have resulted in an unbiased progeny. Our molecular analysis indicated that somatic activity and sex conversion of females accounted for this observation and this was most likely caused by the leaky expression of Cas9 in somatic tissues. The majority of XX individuals carrying the *tra* drive were found to be infertile intersexes, although a fraction was fully converted into fertile XX males. Since infertile XX intersex individuals represented a sizable fraction of the DsRed positive progeny of *tra* drive males, we concluded that this drive could not exhibit effective population-wide gene drive. Because no transgenic females could be obtained from it, the limitations which curtailed the usefulness of the *tra* drive as currently constituted also prevented us from exploring how the loss of maternal *tra* would affect sex determination.

Targeting the maternal provision of *transformer* is an alternative, novel strategy to effect sex conversion and interferes with the establishment of the autoregulatory loop whereby Tra controls the sex-specific splicing of its own pre-mRNA in females. The *white + tra* gene drive we have generated to test this idea indeed induced a substantial bias towards male progeny in the female, but not in the male transgenic crosses. The results are thus in line with our hypothesis that interfering with ovarian *tra* provision precludes normal female embryonic development in the progeny. However, our experiments also featured strong maternal deposition of Cas9. Further experiments will thus be necessary to evaluate to what degree these two different maternal effects contribute in determining the observed sex ratio in the

7

progeny and to what degree the loss of maternal *tra* results in male development irrespective of the genotype or karyotype of the developing embryo.

In Diptera evidence shows that in the families of Tephritidae (*Ceratitis*, *Anastrepha* and *Bactrocera*)[27,36–38], Muscidae (*Musca*)[39] and Calliphoridae (*Lucilia*)[40] maternal *transformer* (Tra_{mat}) and *transformer*2 ($Tra2_{mat}$) gene products initiate a positive feedback regulation of *tra* and enable female-specific splicing of the downstream genes *doublesex* and *fruitless*, unless a male determining factor (e.g. *MoY*, *Mdmd*)[28,41] is present. Furthermore, it has been shown that Tra_{mat} and $Tra2_{mat}$ accumulate in the nurse cells and the growing oocyte but not in the somatic parts of the ovary[42]. Since the *vasa* promoter too is predicted to be active in both of these cell types, it is likely that in our system the level of Tra_{mat} provided to the progeny is the compound result of cleavage and repair events in the 15 nurse cells and the oocyte. This dynamic process should also be explored in further experiments, since it is analogous to lethal/sterile mosaicism resulting from maternally deposited Cas9/gRNA, a process that can act dominantly if it occurs in a sufficient number of cells[9].

While a number of engineered sex ratio distortion systems and systems that interfere with sex determination have been previously described[26,43–45], our findings demonstrated the possibility of a true sex conversion gene drive that converts females into fertile XX males on the population level. For a gene drive homing directly into *tra* this would however require a tighter control of Cas9 expression at this locus than is currently the case in our *tra* drive strain, for example by making use of insulator elements. At the *white-eye* locus we did not find similar evidence for leaky somatic expression of Cas9, highlighting that ectopic Cas9 expression and its impact likely depend on the genomic context, the target gene with its spatiotemporal pattern of expression.

Our modelling explored coupling gene drive targeting female fertility with genetic sex conversion, a strategy we termed sterilising sex conversion (SSC), as a more attractive option. We show that, compared to a conventional female fertility drive, SSC is faster acting and also more tolerant towards the rates of CRISPR activity, homologous repair and the functional target site conservation achievable. Given the tools we have described in this manuscript, assembling such a system in the medfly should be possible in the near future.

Over the last 40 years there has been considerable progress in the development and integrated application of SIT against the medfly, as reflected by ongoing operational programs for eradication, prevention, and suppression as well as a dramatic increase in sterile fruit fly production capacity on multiple continents. There is however considerable scope for improving the efficiency and economics of medfly control and to develop control strategies for other Tephritid flies, for many of which currently few or no control options exist. As the first demonstration of effecting true and efficient genetic sex conversion in the female germline and the demonstration of gene drive in the medfly, our findings open the door to several novel control strategies, as well as genetic sexing mechanisms that could be combined with more classical approaches such as the sterile insect technique.

8

Materials & Methods

Ceratitis capitata rearing and maintenance

Wild type Benakeion and transgenic *C. capitata* strains were bred as described previously[26]. All transgenic strains were reared through backcrosses of male transgenic individuals with wild type Benakeion virgin females.

Generation of constructs

For sequence analysis we used the EGII-3.2.1 (GCA_905071925.1) medfly assembly. The annotated transformation vectors are provided in **Supplementary File 1**. Briefly, we used the gRNA (gRNA_white) targeting the *white-eye* gene, as previously described[33] to knock-in into the third exon of the gene (GeneID_101458180). All constructs were generated through a two-step Gibson assembly using the primers indicated (**Supplementary Table 4**). First, three fragments were cloned into the pUK21 backbone (Addgene #49787) which was digested using XhoI/XbaI to generate the intermediate plasmid pHA-white-dsRED. These included the left and right homology arms, 976 and 998 bp in size respectively, separated by a fragment amplified from piggyBac Cas9.w plasmid[26] containing an AscI restriction site, CcU6-gRNA_white and pUb-DsRed. pHA-white-dsRED was sequentially linearised with AscI for the second assembly in which a cassette was inserted containing Cas9, similarly amplified from the piggyBac Cas9.w plasmid. Three different endogenous germline promoters, *CcNanos*, *CcVasa* or *CcZpg*, were inserted upstream of Cas9, and their respective endogenous terminators were inserted downstream. The pVasa-w+tra_drive plasmid with an additional gRNA was assembled via linearisation of pVasa-w with FseI and the insertion of the CcU6-gRNA_tra fragment thereafter. This gRNA (gRNA_tra) (GTTGTTATTAAACGTAGATTTGG) was selected using CHOPCHOP v3[46] and targets the first female-specific exon of the *transformer* gene (*Cctra*, GeneID_101456163). Plasmid pGdtra, was similarly constructed using the pUK21 backbone linearised with XbaI/XhoI. First, the left and right homology arms were inserted, 937 and 907 bp in size respectively, separated by the EcoRV restriction site. The resulting plasmid was linearised with EcoRV and fragments amplified from the pVasa-w+tra_drive plasmid were inserted. These contained CcVasa-Cas9, CcU6-gRNA_tra and pUb-DsRed.

Generation and establishment of transgenic strains

Microinjections into wild type Benakeion strain embryos of the donor plasmids pNanos-w, pVasa-w and pZpg-w (250 ng/ml) were performed to generate the 3 *white* drive strains. To generate the *tra* drive and *white + tra* drive strains plasmids pGdtra and pVasa-w+tra_drive were used respectively (250 ng/ml). The injection master mix also included preassembled RNP complex of Cas9 protein (200 ng/ml) (PNA Bio)[33], and either gRNA_white (100 ng/ml) or gRNA_tra (100 ng/ml) obtained from Synthego. IInjected G_0 adults were reciprocally crossed in pools to wild type virgin flies. The resulting G_1 progeny was screened for DsRed at the adult stage, and fluorescent individuals were individually crossed to wild type virgin flies of the opposite sex. To confirm knock-in of the homing constructs into *white eye* and *transformer* target regions, gDNA from single flies was extracted using an adapted protocol from Holmes & Bonner[33,47]. PCRs were then carried out with primers binding to the homing construct and genomic sequence outside the homology arms (**Supplementary Table 4**).

Genotyping and karyotyping

Template gDNA from single flies was used for genotyping, karyotyping, gRNA-white target and gRNA-tra target PCRs (Supplementary Table 4). The genotyping to identify the drive allele and the wild type allele was performed with a multiplex PCR. For the *white* drive line the following primers were used; F-indel-w binding upstream to the cleavage site of *white-eye*

9

gene and For_dsRed binding in *white drive cassette* and reverse primer binding in the genome Rev_Genome-w. While for the *tra drive* line the following primers were used: For_Pub binding in *tra drive cassette*, F-indel-tra binding upstream to the cleavage site of *tra* gene and a reverse primer binding in the genome of the tra gene Rev_tra-screen. To detect indels at both the *we* and *tra* gene cleavage sites the following primers were used respectively: F-indel-w, R-indel-w; F-indel-tra and Rev-indel-tra. Karyotyping PCRs were conducted using RedTaq DNA Polymerase 2X Master Mix (VWR Life Science) with the following published primers: *CcYF*/*CcYR*[48]. The gRNA target amplicons were purified and Sanger sequenced.

Fecundity and fertility assays

Egg laying rates and egg hatching rates were assessed by crossing 10 transgenic males or intersexes with 20 wild type Benakeion virgin females in triplicates, alongside three control crosses of 10 wild type males and 20 wilt-type females. Upon maturation, eggs laid during a 5-hour period were collected on a black filter paper, images and counted immediately using ImageJ. After a further 4 days, unhatched eggs were counted again to determine the egg hatching rates. Statistical analysis was performed using an ANOVA (Dunnett's test).

Homing assays

Male and female transgenics were reciprocally mated with wild type Benakeion flies in standard crosses of 10 males and 20 females with the exception of the *tra* drive, where only male transgenics were available. The progeny of the *white* drive lines (*nanos*, *vasa* & *zpg* promoters) were screened for eye colour (red/mosaic/white) and fluorescence phenotype (DsRed+/DsRed-) for 10 consecutive generations. The progeny for the *white + tra* drive crosses were additionally screened for phenotypic sex characteristics (male/female/intersex), while the *tra* drive cross progeny were screened for fluorescence phenotype and phenotypic sex for three and five consecutive generations respectively. Eggs were collected at two timepoints, with an interval of 2 to 3 days and reared to adulthood. Adult flies were screened using the MVX-ZB10 Olympus with an RFP filter (excitation filter 530–560 nm, dichroic beam splitter 570 nm, barrier filter 585–670 nm). Statistical analysis was performed using the chi-square goodness-of-fit test.

Gene drive model

We employed the SMS agent-based gene drive model[49], where we considered populations of 1000 individuals with discrete generations. All gene drive designs were investigated at a starting allele frequency of 12.5% with populations consisting initially of 500 wild type females, 250 wild type males and 250 hemizygous drive males. The model considers fertility effects and target site resistance at each locus independently. Parameter sweeps were performed for the rate of Cas9 germline activity, the rate at which R1 (functional) versus R2 (non-functional) resistance alleles are formed at each target locus, as well as the HDR rate i.e. the rate at which cleavage leads to homologous (homing of the construct or of an R allele present at the other chromosome) versus non-homologous (R allele formation) repair. When Cas9 germline activity levels were varied we considered maternal Cas9 activity to vary proportionally. Each model is based on 360 simulations covering the parameter space indicated. The continuous rate of extinction was then calculated using a local polynomial regression model. The SMS model code and parameters are available at https://github.com/genome-traffic/SuperMendelianSandbox and R scripts for visualisation are available at https://github.com/genome-traffic/gRandTheftAutosome.

10

Figure Legends

Figure 1. Homing and Cas9 activity targeting *white-eye*. A. Genetic crossing scheme of transgenics crossed to *white-eye* (*w-*) mutant (upper panels) or wild type (lower panels) medflies and schematics of the gene drive constructs tested. **B.** Gene drive transmission as the percentage of DsRed positive progeny of female or male hemizygous transgenics. **C.** Eye pigmentation phenotypes of the DsRed negative fraction of the progeny of transgenic fathers or mothers. Shown is the percentage of white eyed and mosaic progeny obtained for each construct with red eyed progeny making up the remainder. The total number of individual offspring scored for male and female transgenic crosses were n=3644/n=2521 for *nanos*, n=5115/n=3622 for *vasa* and n=3770/n=3916 for *zpg* constructs respectively. Pooled crosses (10 males x 20 females) of *white-eye* mutant and wild type individuals were performed in 2 and 10 replicates respectively.

Figure 2. Sex converting gene drive in the medfly. A. Schematic showing the *tra* drive and the crossing scheme used to score the progeny. **B.** Transmission of the *tra* drive as the percentage of DsRed positive progeny of male hemizygous transgenics. No female transgenics were obtained. **C.** Overall percentage of DsRed positive and negative individuals and their sexual phenotypes in the progeny of *tra* drive males. **D.** Schematic showing the *white + tra* drive and the crossing scheme used to score the progeny. **E.** Transmission of the *white + tra* drive as the percentage of DsRed positive progeny of male and female hemizygous transgenics. **F.** Sexual phenotypes scored in the progeny of *white + tra* drive males or females crossed to the wild type. The total number of individual offspring scored for the *tra* drive was n=8361 from 5 replicate pooled crosses of 10 *tra* males to 20 females. The total number of individual offspring scored for 3 replicate pooled male and female transgenic crosses (10 males x 20 females) of the *white + tra* drive was n=2173 and n=3084 respectively.

Figure 3. Model of sex conversion gene drive strategies. A. A gene drive construct directly targeting a recessive female fertility gene. **B.** A gene drive construct directly targeting the medfly transformer gene. **C.** A sterilising sex conversion drive (SSC) homing into a recessive female fertility gene while also targeting the medfly transformer gene in *trans*. **D.** A gene drive construct directly targeting a recessive female fertility gene while targeting a second, separate female fertility gene in *trans*. Panels **E,F,G,H** explore the effect of the R2/R1 rate and cleavage rate on the mean likelihood of population extinction at a set HDR rate of 0.95 and a run duration of 20 generations. Panels **I,J,K,L** explore the effect of cleavage and HDR rates on the mean duration to extinction at a set R2/R1 rate of 0.99/0.01.

Figure S1. Gene drive constructs and target genes.

Figure S2. Eye pigmentation in the offspring of *white* drive carriers. A. Eye pigmentation phenotypes of exemplary wild type, *white-eye* mutant and mosaic individuals. **B.** Eye pigmentation phenotypes of the DsRed positive fraction of the progeny of transgenic fathers or mothers carrying the white drive constructs with Cas9 being driven by the *nanos*, *vasa* and *zpg* promoters. Shown is the percentage of white eyed and mosaic progeny obtained for each construct with red eyed progeny making up the remainder.

11

Figure S3. Fluorescence and sexual phenotypes of *tra* drive & *white* + *tra* drive individuals. **A.** Transmission and red fluorescence micrographs of *tra* drive (left panels) and *white* + *tra* drive (right panels) individuals. **B.** Sexual phenotypes of wild type (WT) and *tra* drive individuals (left panels) and close up images of genitalia (right panels).

Figure S4. Eye pigmentation phenotypes in the offspring of *white* + *tra* drive individuals. Eye pigmentation phenotypes of the DsRed positive and DsRed negative fractions of the progeny of transgenic fathers or mothers carrying the *white* + *tra* drive constructs. Shown is the percentage of red eyed, mosaic and white eyed progeny obtained.

Figure S5. Fertility and fecundity of transgenic medfly strains. A. Average egg output of pooled crosses (10 males x 20 females) of hemizygous transgenic medflies when crossed to wild type individuals compared to wild type control intercrosses. **B.** The larval hatching rate of the progeny of hemizygous transgenic and wild type individuals compared to wild type control intercrosses. Error bars indicate the standard deviation around the mean. The total number of embryos scored were n=1757 for the control, n=2025 for the tra drive, n=1858 for *white* + *tra* drive and n=2128/2235/1714 (*nanos*/*vasa*/*zpg*) for the *white* drive groups with a minimum of 3 replicate experiments performed per group.

Supplementary File 1. Annotated GenBank files of the transformation vectors.

Supplementary Table 1. Genotypes and karyotypes of the progeny of *tra* drive males crossed to wild type females.

Supplementary Table 2. Fertility of XX males.

Supplementary Table 3. Fertility of intersex individuals.

Supplementary Table 4. Primer database.

Acknowledgements

12

References

1. Burt, A. Site-specific selfish genes as tools for the control and genetic engineering of natural populations. *Proc. Biol. Sci.* **270**, 921–928 (2003).
2. Hammond, A. *et al.* A CRISPR-Cas9 gene drive system targeting female reproduction in the malaria mosquito vector Anopheles gambiae. *Nat. Biotechnol.* **34**, 78–83 (2016).
3. Anderson, M. A. E. *et al.* A multiplexed, confinable CRISPR/Cas9 gene drive propagates in caged Aedes aegypti populations. 2022.08.12.503466 Preprint at https://doi.org/10.1101/2022.08.12.503466 (2022).
4. Green, E. *et al.* A population modification gene drive targeting both Saglin and Lipophorin disables Plasmodium transmission in Anopheles mosquitoes. 2022.07.08.499187 Preprint at https://doi.org/10.1101/2022.07.08.499187 (2022).
5. Harvey-Samuel, T. *et al.* CRISPR-based gene drives generate super-Mendelian inheritance in the disease vector Culex quinquefasciatus. 2023.06.12.544656 Preprint at https://doi.org/10.1101/2023.06.12.544656 (2023).
6. Yadav, A. K. *et al.* CRISPR/Cas9-based split homing gene drive targeting doublesex for population suppression of the global fruit pest Drosophila suzukii. *Proc. Natl. Acad. Sci.* **120**, e2301525120 (2023).
7. Li, M. *et al.* Development of a confinable gene drive system in the human disease vector Aedes aegypti. *eLife* **9**, e51701 (2020).
8. Asad, M., Liu, D., Li, J., Chen, J. & Yang, G. Development of CRISPR/Cas9-Mediated Gene-Drive Construct Targeting the Phenotypic Gene in Plutella xylostella. *Front. Physiol.* **13**, (2022).
9. Adolfi, A. *et al.* Efficient population modification gene-drive rescue system in the malaria mosquito Anopheles stephensi. *Nat. Commun.* **11**, 5553 (2020).
10. Hoermann, A. *et al.* Gene drive mosquitoes can aid malaria elimination by retarding Plasmodium sporogonic development. *Sci. Adv.* **8**, eabo1733 (2022).
11. Gantz, V. M. *et al.* Highly efficient Cas9-mediated gene drive for population modification of the malaria vector mosquito Anopheles stephensi. *Proc. Natl. Acad. Sci.* **112**, E6736–E6743 (2015).
12. Carballar-Lejarazú, R. *et al.* Next-generation gene drive for population modification of the malaria vector mosquito, Anopheles gambiae. *Proc. Natl. Acad. Sci.* **117**, 22805–22814 (2020).
13. Xu, X. *et al.* Toward a CRISPR-Cas9-Based Gene Drive in the Diamondback Moth Plutella xylostella. *CRISPR J.* **5**, 224–236 (2022).
14. Bax, N. J. & Thresher, R. E. Ecological, behavioral, and genetic factors influencing the recombinant control of invasive pests. *Ecol. Appl.* **19**, 873–888 (2009).
15. Schliekelman, P., Ellner, S. & Gould, F. Pest Control by Genetic Manipulation of Sex Ratio. *J. Econ. Entomol.* **98**, 18–34 (2005).
16. Hammond, A. *et al.* Gene-drive suppression of mosquito populations in large cages as a bridge between lab and field. *Nat. Commun.* **12**, 4589 (2021).
17. Kyrou, K. *et al.* A CRISPR–Cas9 gene drive targeting doublesex causes complete population suppression in caged Anopheles gambiae mosquitoes. *Nat. Biotechnol.* **36**, 1062–1066 (2018).
18. Simoni, A. *et al.* A male-biased sex-distorter gene drive for the human malaria vector Anopheles gambiae. *Nat. Biotechnol.* **38**, 1054–1060 (2020).
19. White, I. M., Harris, M. M. E. & White, Ian M. *Fruit flies of economic significance; Their*

identification and bionomies. (CAB International, 1992).

20. *Trapping and the Detection, Control, and Regulation of Tephritid Fruit Flies: Lures, Area-Wide Programs, and Trade Implications*. (Springer Netherlands, 2014). doi:10.1007/978-94-017-9193-9.

21. Szyniszewska, A. M. & Tatem, A. J. Global Assessment of Seasonal Potential Distribution of Mediterranean Fruit Fly, Ceratitis capitata (Diptera: Tephritidae). *PLOS ONE* **9**, e111582 (2014).

22. Malacrida, A. R. *et al.* Globalization and fruitfly invasion and expansion: the medfly paradigm. *Genetica* **131**, 1–9 (2007).

23. Gutierrez, A. P., Ponti, L., Neteler, M., Suckling, D. M. & Cure, J. R. Invasive potential of tropical fruit flies in temperate regions under climate change. *Commun. Biol.* **4**, 1–14 (2021).

24. *Sterile Insect Technique: Principles and Practice in Area-Wide Integrated Pest Management*. (Springer Netherlands, 2005). doi:10.1007/1-4020-4051-2.

25. Leftwich, P. T. *et al.* Genetic elimination of field-cage populations of Mediterranean fruit flies. *Proc. Biol. Sci.* **281**, 20141372 (2014).

26. Meccariello, A. *et al.* Engineered sex ratio distortion by X-shredding in the global agricultural pest Ceratitis capitata. *BMC Biol.* **19**, 78 (2021).

27. Pane, A., Salvemini, M., Bovi, P. D., Polito, C. & Saccone, G. The *transformer* gene in *Ceratitis capitata* provides a genetic basis for selecting and remembering the sexual fate. *Development* **129**, 3715–3725 (2002).

28. Meccariello, A. *et al. Maleness-on-the-Y* (*MoY*) orchestrates male sex determination in major agricultural fruit fly pests. *Science* **365**, 1457–1460 (2019).

29. Carrami, E. M. *et al.* Consequences of resistance evolution in a Cas9-based sex conversion-suppression gene drive for insect pest management. *Proc. Natl. Acad. Sci.* **115**, 6189–6194 (2018).

30. Siddall, A., Harvey-Samuel, T., Chapman, T. & Leftwich, P. T. Manipulating Insect Sex Determination Pathways for Genetic Pest Management: Opportunities and Challenges. *Front. Bioeng. Biotechnol.* **10**, 867851 (2022).

31. Gomulski, L. M. *et al.* Genomic Organization and Characterization of the *white* Locus of the Mediterranean Fruitfly, *Ceratitis capitata*. *Genetics* **157**, 1245–1255 (2001).

32. Ogaugwu, C. E. & Wimmer, E. A. Molecular cloning and expression of nanos in the Mediterranean fruit fly, Ceratitis capitata (Diptera: Tephritidae). *Gene Expr. Patterns* **13**, 183–188 (2013).

33. Meccariello, A. *et al.* Highly efficient DNA-free gene disruption in the agricultural pest Ceratitis capitata by CRISPR-Cas9 ribonucleoprotein complexes. *Sci. Rep.* **7**, 10061 (2017).

34. Petrella Valeria, Marco Salvemini, Giuseppe Saccone. Digital differential expression analysis of early embryonic de novo assembled sexed transcriptomes of Ceratitis capitata, a world-wide agricultural pest, to identify new sex-specific and sex-biased genes. (2014).

35. Grunwald, H. A. *et al.* Super-Mendelian inheritance mediated by CRISPR–Cas9 in the female mouse germline. *Nature* **566**, 105–109 (2019).

36. Ruiz, M. F. *et al.* The Gene Transformer of Anastrepha Fruit Flies (Diptera, Tephritidae) and Its Evolution in Insects. *PLoS ONE* **2**, e1239 (2007).

37. Schetelig, M. F., Milano, A., Saccone, G. & Handler, A. M. Male only progeny in Anastrepha suspensa by RNAi-induced sex reversion of chromosomal females. *Insect*

14

Biochem. Mol. Biol. **42**, 51–57 (2012).

38. Liu, G., Wu, Q., Li, J., Zhang, G. & Wan, F. RNAi-Mediated Knock-Down of transformer and transformer 2 to Generate Male-Only Progeny in the Oriental Fruit Fly, Bactrocera dorsalis (Hendel). *PLOS ONE* **10**, e0128892 (2015).

39. Hediger, M. *et al.* Molecular Characterization of the Key Switch *F* Provides a Basis for Understanding the Rapid Divergence of the Sex-Determining Pathway in the Housefly. *Genetics* **184**, 155–170 (2010).

40. Li, F., Vensko, S. P., Belikoff, E. J. & Scott, M. J. Conservation and Sex-Specific Splicing of the transformer Gene in the Calliphorids Cochliomyia hominivorax, Cochliomyia macellaria and Lucilia sericata. *PLoS ONE* **8**, e56303 (2013).

41. Sharma, A. *et al.* Male sex in houseflies is determined by *Mdmd* , a paralog of the generic splice factor gene *CWC22*. *Science* **356**, 642–645 (2017).

42. Burghardt, G. *et al.* The transformer2 gene in Musca domestica is required for selecting and maintaining the female pathway of development. *Dev. Genes Evol.* **215**, 165–176 (2005).

43. Fasulo, B. *et al.* A fly model establishes distinct mechanisms for synthetic CRISPR/Cas9 sex distorters. *PLoS Genet.* **16**, e1008647 (2020).

44. Galizi, R. *et al.* A synthetic sex ratio distortion system for the control of the human malaria mosquito. *Nat. Commun.* **5**, 3977 (2014).

45. Galizi, R. *et al.* A CRISPR-Cas9 sex-ratio distortion system for genetic control. *Sci. Rep.* **6**, 31139 (2016).

46. Labun, K. *et al.* CHOPCHOP v3: expanding the CRISPR web toolbox beyond genome editing. *Nucleic Acids Res.* **47**, W171–W174 (2019).

47. Holmes, D. S. & Bonner, J. Preparation, molecular weight, base composition, and secondary structure of giant nuclear ribonucleic acid. *Biochemistry* **12**, 2330–2338 (1973).

48. Gabrieli, P. *et al.* Sex and the single embryo: early deveopment in the Mediterranean fruit fly, Ceratitis capitata. *BMC Dev. Biol.* **10**, 12 (2010).

49. Ellis, D. A. *et al.* Testing non-autonomous antimalarial gene drive effectors using self-eliminating drivers in the African mosquito vector Anopheles gambiae. *PLOS Genet.* **18**, e1010244 (2022).

Figure 1.
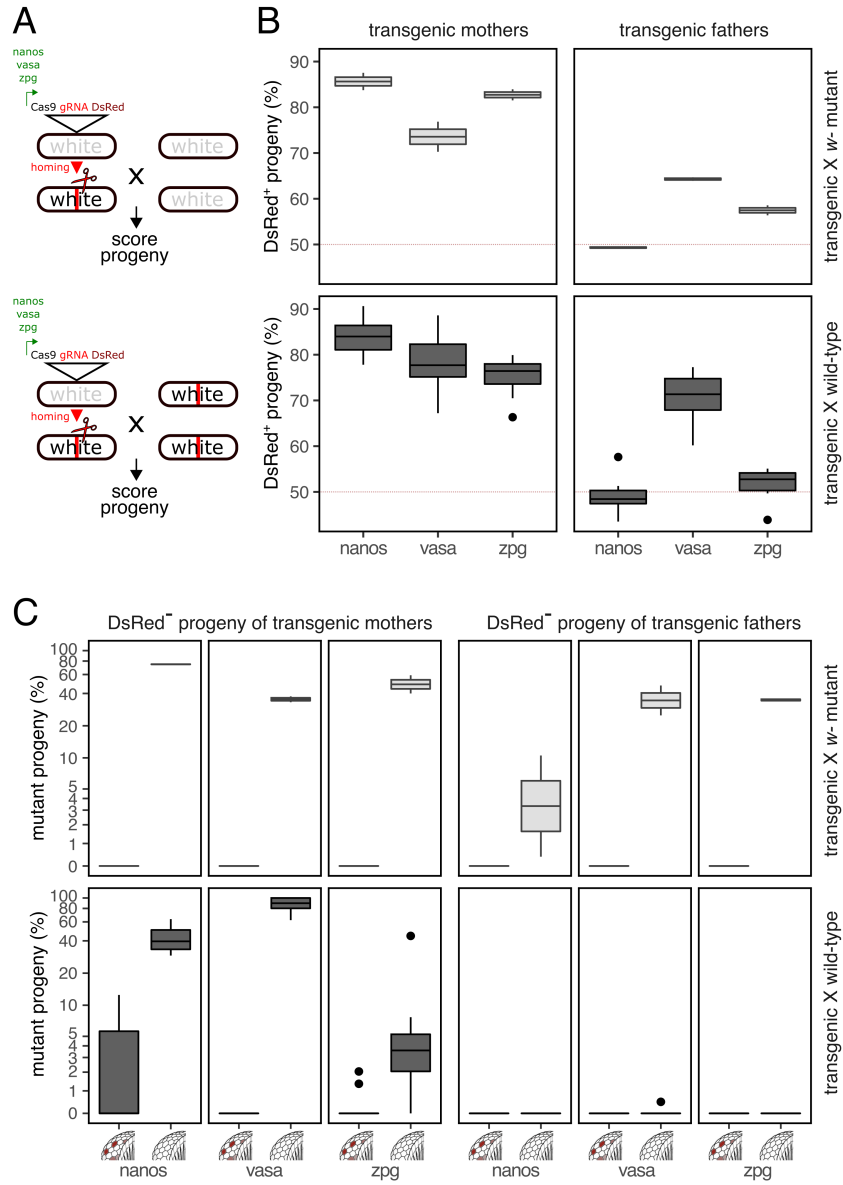
Figure 2.
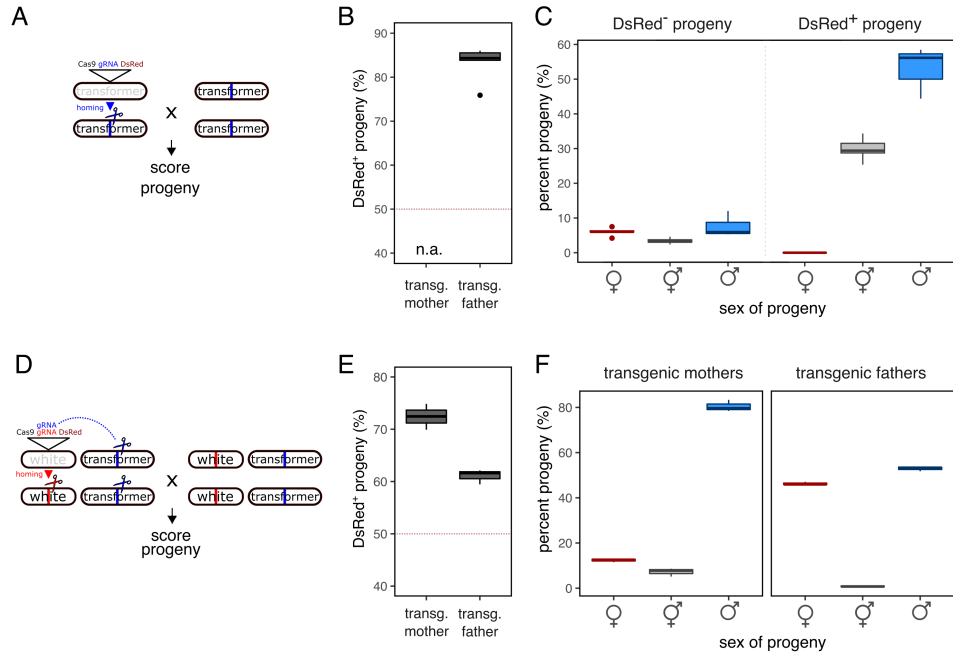
Figure 3.

Figure S1.

Figure S2.

A

wild-type   *white-eye* mutant   mosaic individual

B

Figure S3.

A

tra drive

complete   patchy

white + tra drive

transmission

fluorescence (RFP)

B

WT ♂ (XY male)

WT ♀ (XX female)

tra drive ♂ (XX male)

tra drive ⚥ (XX intersex)
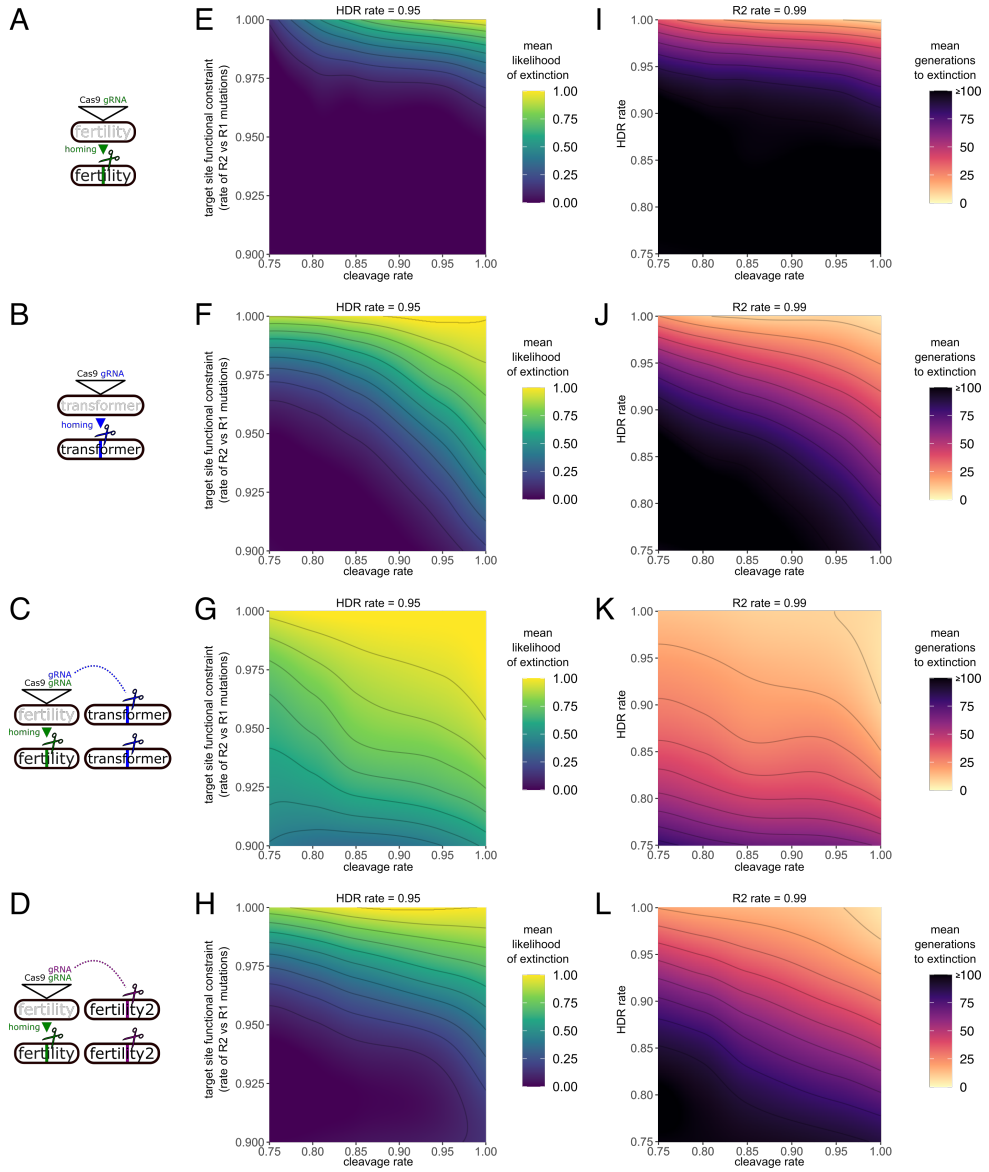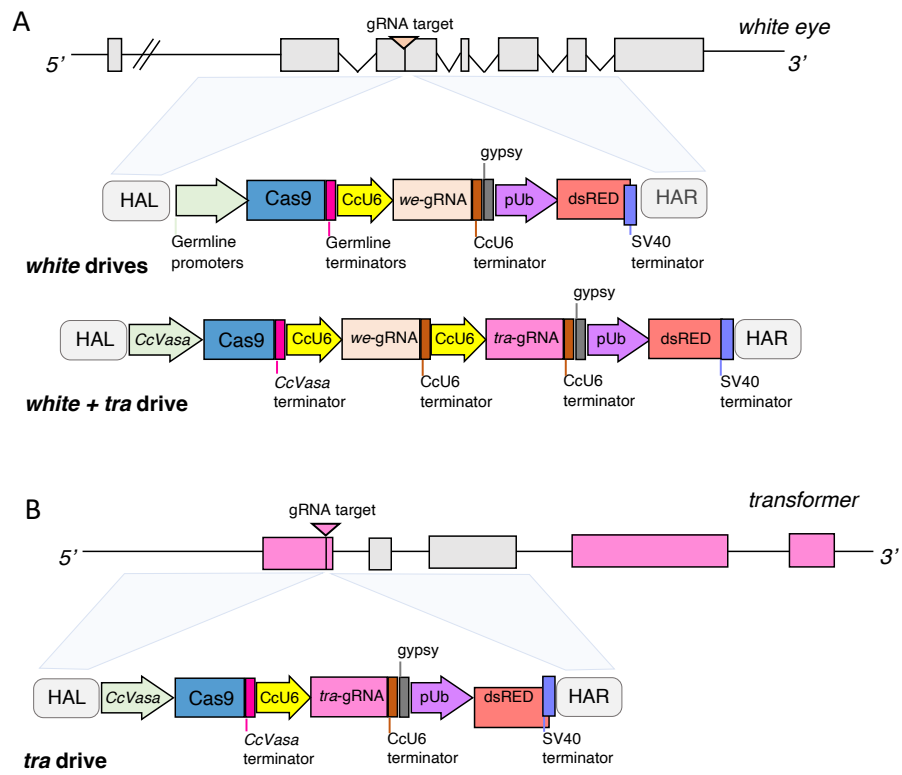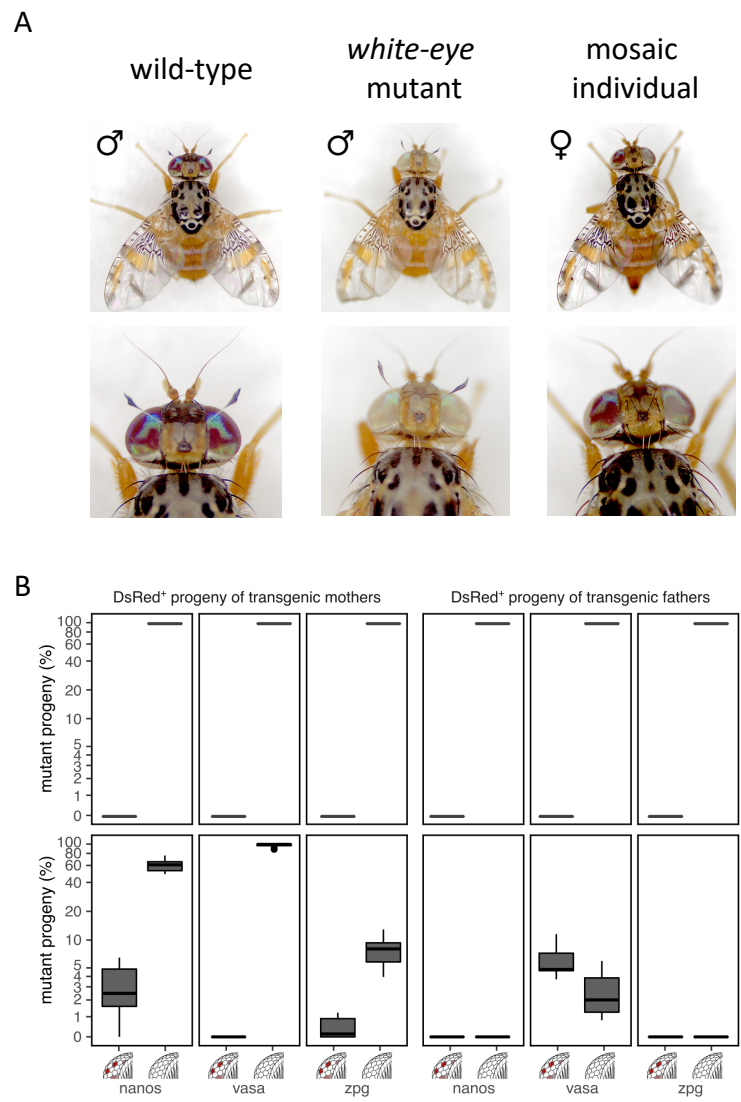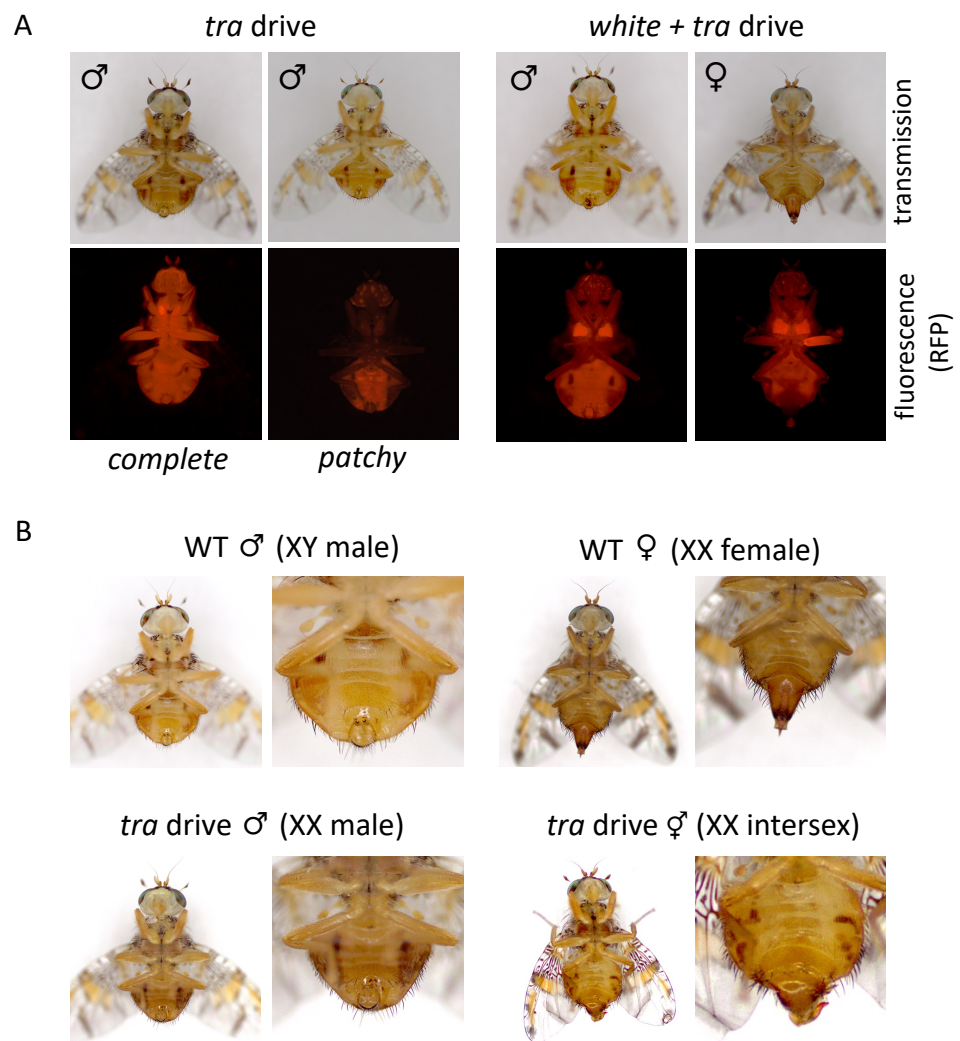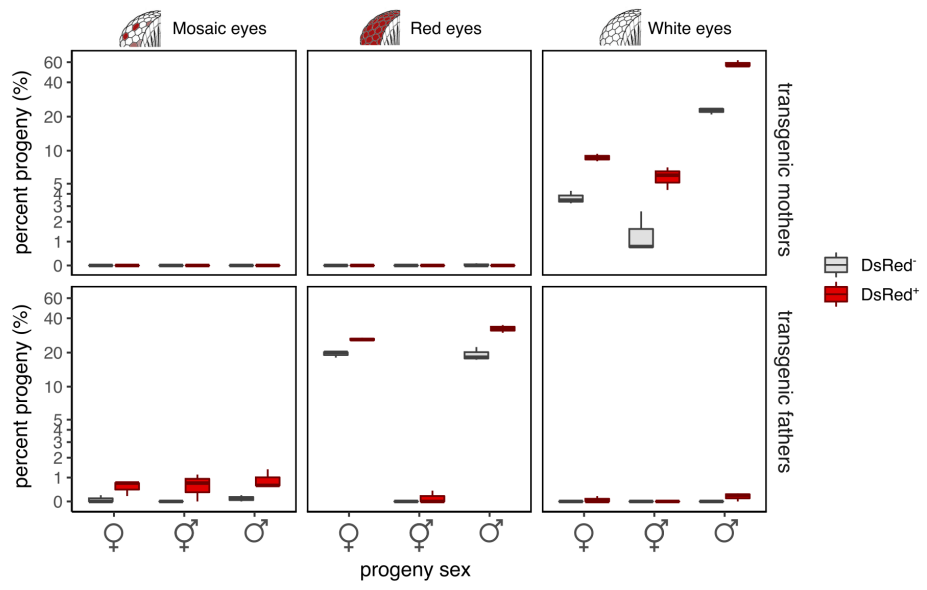
Figure S4.

Figure S5.
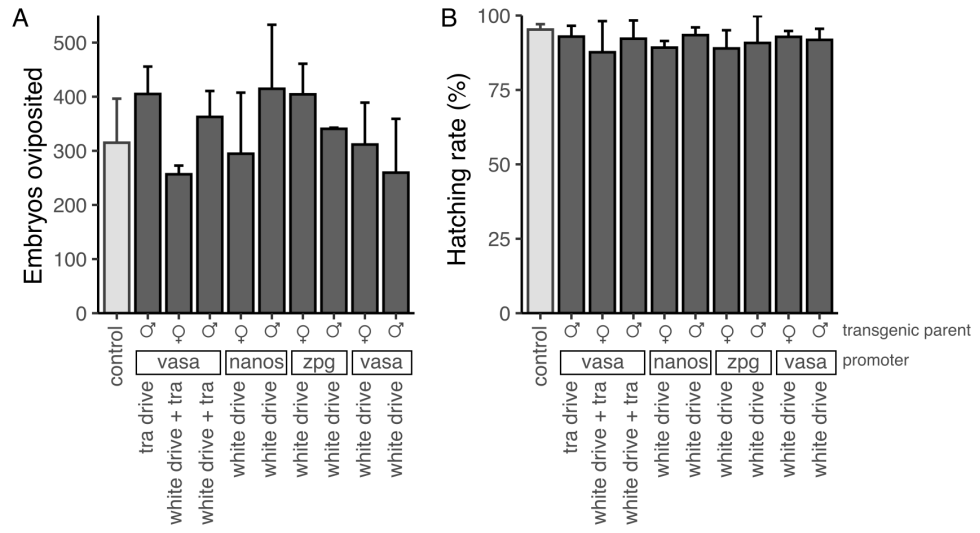
4 Chapter Four – Controlling the expression of *Maleness-on-the-Y* using the *ß2-tubulin* promotor

4.1 Abstract

Maleness-on-the-Y (MoY) is a recently discovered male determining factor in medfly. Manipulations of MoY can result in the generation of fertile XX pseudomales. This is of interest as this knowledge could be integrated into several pest control strategies such as gene drives and genetic sex sorting for mass rearing programmes. However, in order to use *MoY* in such control strategies, a deeper understanding of how the timing of *MoY* expression affects viability is required. By generating a transgenic construct to place *MoY* under the spermatogenesis-specific promotor *ß2-tubulin,* I tested whether restricting expression to a narrower temporal- and tissue-specific window would more closely mimic native expression of *MoY* than a constitutive promotor. The aim of this was to allow the establishment of a transgenic line with dominant, paternally-inherited expression of *MoY*, while minimising any lethality, by restricting expression to occur only within the process of spermatogenesis. The results showed that integration of the construct resulting in the expression of *MoY* resulted in lethality, with high levels of post injection embryonic death being observed. The results illustrated the critical importance of knowing the timing of the native expression timing of *MoY* in order to both realise its potential utility in genetic insect control, and to increase understanding the wider roles of this key sex determining factor.

4.2   Introduction

The medfly *Ceratitis capitata* is an agricultural pest of global significance, causing large scale crop loss through direct damage and expensive control strategies (White and Elson-Harris 1992). New routes for controlling this significant pest are urgently required. One promising route for genetic control is via the disruption of sex determination mechanisms thus, sex determination pathways in medfly have been investigated in detail due to their applied potential. Downstream elements of the sex determination pathway tend to be relatively well-conserved between species in comparison to upstream elements (Siddall et al. 2022; Hopkins and Kopp 2021; Adolfi et al. 2021) and have a history of applied use in insect transgenesis and sex sorting (Labbé et al. 2012). The sex determination genes *tra* and *doublesex* are well conserved downstream elements in the sex determination pathway, and their respective roles in sex determination have been identified in both medfly (Pane et al. 2002; Perrotta et al. 2023; Salvemini et al. 2009) and many other species. A large body of research has investigated how these genes influence sex determination. However, the genes upstream of *tra* in sex determination pathways remain relatively unknown in most insects. There is much more variability across species in the upstream 'master regulators' of sex determination. The evolution of the genes that encode such upstream elements is therefore predicted to be recent and to show high turnover (Gempe and Beye 2011). Because of this, upstream sex determination elements are much more challenging to identify in individual species. The targeting of upstream elements is still however an attractive option. This is due to the likelihood of obtaining viable offspring after manipulation. When downstream elements of the pathway are manipulated often this late interference in sex determination creates incomplete conversion, infertility and in some cases death. Targeting upstream elements of the pathway often yields offspring that incur lower fitness costs of the sex determination manipulations, as they are not receiving alternating sex signalling elements (Siddall et al. 2022).

In targeting the most upstream elements of the sex determination pathway, the driving of the male determining element (often carried on the Y chromosome) is the most obvious upstream regulator that can be manipulated. The development of a synthetic driving Y chromosome is one promising method of genetic pest management. The potential for such strategies was recognised over 40 years ago (Hamilton 1967) well before the technical ability to construct driving Ys was even possible. More recent designs for Y drives involve coupling them with an 'X-shredder' (Windbichler et al. 2008; Galizi et al. 2016) that is active during

spermatogenesis. X shredders target repetitive regions only found on the X chromosome and destroy the chromosome through multiple cuts during spermatogenesis. The aim of this is to induce the shredding of X-carrying male gametes and ensure the inheritance of only Y chromosomes (Beaghton et al. 2017; Galizi et al. 2016). The benefits of these types of designs are clear in terms of producing effective pest control. Unfortunately, there are significant technical barriers to the construction of such drives. When carried on an autosome they have no driving capabilities and will only be inherited by 50% of offspring. If X shredders are carried on the Y chromosome, they would bias their own inheritance as all viable offspring are Y carrying. Manipulating or achieving stable expression of genes or transgenes on Y chromosomes is much more difficult in comparison to autosomes due to meiotic sex chromosome inactivation, expression variation, unpredictability and rapid evolution of Y chromosomes (Hall et al. 2016; Gamez et al. 2021). Advances are resulting from an increased understanding of Y chromosome structure (Bernardini et al. 2014) and transgene expression from the Y chromosome has now been successfully achieved in *Drosophila melanogaster* (Gamez et al. 2021). Although this expression was weaker than for autosomal-linked sequences it was still sufficient to induce gene editing. However, overall, significant challenges remain in ensuring predictable levels of transgene expression from the Y, which makes the potential of synthetic driving Y chromosomes that do not rely on such expression all the more promising.

In general, sex manipulation strategies require knowledge of the key elements in sex determination pathways. The primary sex determining factor in medfly was originally suggested to be a masculinising element (M factor) over 20 years ago (Pane et al. 2002). The precise nature of this element was only recently identified, as the *Maleness-on-the-Y* (*MoY*) gene. *MoY* is the master gene determining male sexual development (Meccariello et al. 2019). It is located on the Y chromosome and its expression in XX individuals can generate fertile pseudomales (Meccariello et al. 2019). This indicates that, surprisingly, genes essential for male fertility in medfly are not Y linked. This is unusual, as in many species, XX pseudomales often suffer from reduced fitness through reduced fertility or impaired reproductive morphology (Lutrat et al. 2022; Aryan et al. 2020; Voelker and Kojima 1971). The fertility and viability observed in *MoY* exposed XX pseudomales in medfly makes *MoY* a particularly promising element in the field of genetic control, as its expression alone is sufficient to generate fertile pseudomales with as yet no clear fertility deficits identified.

The successful generation of pseudomales by manipulating *MoY* was achieved through the injection of linearised and plasmid *MoY* DNA into early embryos (Meccariello et al. 2019). Although this was a useful proof of principle of the action of *MoY*, for utility in genetic control, the *MoY* sequence would need to be integrated as a stable insertion, integrated into the medfly host DNA. To do this, an initial attempt was made to integrate *MoY* and its surrounding genetic regulatory elements onto an autosome, to see if this would result in sex conversion of females to XX phenotypic males. The results showed pervasive embryonic lethality post microinjection of the transgenic MoY constructs (Meccariello, pers comm.). This result could suggest two alternative explanations: (i) either timing of expression of *MoY* is key, with expression for extended periods of time being lethal or (ii) any overexpression of *MoY* is lethal. As MoY was integrated with its native regulatory regions we assumed that its genomic location allowed for a further level of control of expression. The sex conversion success achieved by expressing *MoY* transiently indicated that plasmid *MoY* DNA had been injected at the correct time point in embryogenesis during which expression of *MoY* is required, and that this procedure avoided persistence of MoY transcripts that appear to have lethal effects later in development. Therefore, we hypothesised that placing *MoY* under the control of a germline specific promotor might similarly restrict the expression to the necessary time window and rescue lethality effects, thus providing a potential additional tool for genetic control.

The candidate promoter fragment I tested was *ß2-tubulin*, a male testis-specific gene with expression restricted to spermatogenesis (Santel 2000). The *ß2-tubulin* gene promotor has been used extensively for transgene expression in several insects (Smith et al. 2007; Zimowska et al. 2009; Yan et al. 2021) and for sperm marking in medfly (Scolari et al. 2008). The envisioned use of a MoY construct driven by *ß2-tubulin* would allow the packaging of *MoY* into sperm cells which would be delivered to embryos. If *MoY* persisted beyond fertilisation this could trigger the male development pathway in the developing embryos before dissipating to avoid any lethality from continual/constitutive expression. The results could also be of interest in sperm biology in the context of testing whether the longevity of proteins in sperm cells can be extended to persist after fertilisation (Barroso et al. 2009; Castillo et al. 2018).

4.3 Methods

4.3.1 Fly Husbandry

4.3.1.1 Medfly stock lines

Three wildtype medfly stock lines were used throughout these experiments. 'Cepa Petapa' is a mass-reared wildtype strain originally collected from the wild in Guatemala, Central America (Rendon, 1996). TOLIMAN is a wildtype strain from Guatemala, Central America originally collected in 1990. Both strains have been reared at Oxitec LTD (Milton Park, Abingdon) from 2004 and a sub-culture of each was established at the University of East Anglia (UEA) in 2010. The wild-type Benakeion strain (originally isolated in Athens, Greece) was provided by A. Meccariello (Imperial College, London) and maintained at UEA from 2022.

4.3.1.2 Medfly rearing and dissections for end point RT-PCR

All wild type medfly cultures were reared in a controlled environment with temperature set at 25°C ±0.5°C, humidity at 60% ± 10% RH and on a 12h light/dark schedule. Depending upon the demand for eggs from cultures, adult populations were kept in 1 of 3 different cage sizes: small cages (10cm x 10cm x 10cm) initiated with 50 pupae, medium cages (13cm x 13cm x 14cm) with 100 pupae and large cages (33cm x 33cm x 16cm) with 500 pupae. The emerging adult flies held in these cages were fed on an *ad libitum* diet of 3:1 sucrose:hydrolysed yeast mixture, changed every seven days. Water was supplied through the side of each cage via dental rolls soaked in RO water, which were renewed every 4 days.

All cages were designed with a mesh layer covering the majority of the surface area of one vertical wall of the cage to allow egg deposition. Water pots containing RO water were placed under the mesh side of each cage to collect eggs. Eggs laid through the cage mesh were collected after a period of no more than 24h of egg laying, by removing eggs with a Pasteur pipette and transferring up to two drops of the eggs contained in the water solution to filter paper. The filter papers with eggs were then placed into 1/3 pint milk bottles containing 100ml of ASG larval food (1L ASG food: 850ml RO water, 12.5g agar, 73.5g sucrose, 67g maize, 47.5g Brewer's Yeast, 2ml Propionic acid, 25ml Nipagin). 7 days post egg collection, bottles were prepared to allow the third instar larvae to exit and start pupation. To do this, 2 filter paper folded strips were placed in each bottle before laying the bottles down on a thin layer of sand within a pupation box (170mm x 130mm x 50mm). The box was then sealed with a

lid containing a mesh for a further 7-9 days. During this time third instar 'jumping' larvae exited the bottles and pupated within the sand. After this period, pupae were sieved from the sand through a standard metal sieve and transferred into petri dishes. From this the appropriate number of pupae required for each procedure were randomly selected and placed into a fresh cage or used in experiments.

To obtain flies for dissections, a large TOLIMAN cage containing adults was placed at 4°C for 30-60 minutes until all flies were unconscious. A subset of flies was then removed and placed into glass petri dishes kept on ice. The glass plate was kept free from condensation by adding a thin layer of agar. Flies used for qPCR were reared in small cages of 50 individuals. On day 7 post eclosion, these cages were placed at 4°C for 30-60 minutes until flies were unconscious, then transferred into Eppendorfs of up to 5 flies of the same line/sex and snap frozen in liquid nitrogen. These samples were then stored at -80°C and removed in batches for dissection. Batches were removed for dissection and kept on dry ice.

Prior to beginning dissections all equipment and surfaces were sterilised with 70% ethanol and Invitrogen RNaseZap™. All equipment in direct contact with flies was wiped down with ethanol between dissections and full sterilisation was effected by applying Invitrogen RNaseZap™ between different samples. Single flies were removed and placed in a petri dish of 100% ethanol for 1 minute to prevent floating in PBS. Flies were then placed on a dissecting plate (glass petri dish containing silicon) and held in place with dissecting pins before covering the sample with a drop of sterile Phosphate Buffered Saline (PBS). Fine tipped forceps were then used to peel open the abdominal cavity and extract testes or ovaries. All excess tissue was removed and for dissections on live samples the early germline cells were separated from the late germline cells using a scalpel with samples placed in an Eppendorf and covered in a thin layer of PBS. The remaining fly carcasses were kept to be used for the somatic samples.

Table 4.1 Samples collected from dissections for each experiment. All samples were snap frozen in liquid nitrogen and stored at -80°C until RNA extraction.

Live or Frozen Tissue	Live
Separated into early and late germline?	Yes
Number of testes per sample	11 sets (22 total)
Number of ovaries per sample	9 sets (18 total)
Number of somatic tissues per sample	3

4.3.1.3   Egg sample collection for end point RT-PCR

A large TOLIMAN cage was set up with the purpose of allowing females to lay eggs for 1h intervals, and to allow single water drop egg collections to be taken and stored in Eppendorfs. To collect eggs at different development stages these samples were maintained under standard temperature conditions before removing any excess water and were then snap frozen in liquid nitrogen at the different time points required. The egg development times tested for RNA expression post laying were 1h, 3h, 6h, 24h, 48h and 72h. Due to high levels of desiccation the 72h samples were not carried forward.

4.3.2   Transgenic Line Establishment

4.3.2.1   Microinjection of medfly eggs to generate transgenic strains

Medfly microinjection was performed at the Imperial College London insectary by A. Meccariello. 40µl of injection mix was used per microinjection session with a final concentration of 500ng/µl plasmid of interest, 200ng/µl [i]hyPBase transposase helper plasmid (Eckermann et al. 2018). This was made up to a final volume with injection buffer (Rubin and Spradling 1982). Embryos were collected after an egg laying period of 45 minutes, transferred to double sided sticky tape and allowed to dry for 6 minutes. Eggs were manually dechorionated using the tip of a blunted needle. Once the chorion was removed, eggs were transferred to slides coated in a thin layer of glue (double sided sticky tape dissolved in heptane) with the posterior of the egg facing outwards. These egg slides were desiccated by placing them in a petri dish containing calcium chloride for 6 minutes, and then covered with a thin layer of Halocarbon™ 700 Oil. Eggs on these slides were then injected using WPI Sutter

Quartz needles (outer diameter (OD) = 1.00mm, inner diameter (ID) = 0.70mm) that had been drawn on a Sutter P-200/G laser needle puller. Needles were loaded with injection mix using an Eppendorf Microloader. Microinjections were done using an inverted stereo microscope with an Eppendorf microinjection set-up and compressed air-driven pump.

Post injection, eggs were left on slides until the larvae hatched (typically after 48h) at which point the larvae were placed onto larval food (30g paper, 30g sucrose, 30g yeast extract, 10ml cholesterol stock (5g cholesterol, 140 ml distilled water, 30ml 95% EtOH), 2ml HCl Stock (384ml distilled water, 66ml 37% HCl), 8.5ml Benzoic Stock (50g Benzoic acid, 300ml 95% EtOH, 150ml distilled water) made up to 400ml with distilled water). After 7 days larval culture bottles were laid down in pupation boxes with a thin layer of fine sand and allowed to pupate. Surviving pupae were then sent to UEA securely within three sealed layers.

4.3.2.2   Fluorescence screening for medfly transgenics

Transgenic lines were established and maintained by screening flies for fluorophore marker expression at each generation. When establishing transgenics, live adults were screened on the day of eclosion from the pupae. Flies were anaesthetised over ice for screening. A Leica MZ7.5 dissection microscope was used for screening and was fitted with fluorescence filters to visualise DsRed2 (excitation wavelength 558nm, emission wavelength 583nm) or AmCyan (excitation wavelength 458nm, emission wavelength 485nm) depending on the marker being screened.

4.3.3   Molecular investigation of *ß2-tubulin* RNA expression

4.3.3.1   DNA Extraction

DNA extraction was performed on whole flies of both TOLIMAN and Cepa Petapa wildtype strains. Whole flies were frozen at -20°C for a minimum of 48 hours with 3 individuals per sample. Samples were removed from the freezer and placed onto dry ice. Samples were then homogenised with a sterilised, plastic pestle over dry ice. Qiagen DNeasy® Blood & Tissue Kit was used as per the manufacturer's instructions with the addition of an extra elution step with 25µl molecular grade water after the original 50µl elution. The concentration of each sample was measured using a Thermo fisher Scientific Nanodrop™ 8000 Spectrophotometer.

4.3.3.2 Primer design for End-point RT-PCR

Primer pairs were placed to ensure the resulting PCR product covered an intron/exon boundary to allow for differentiation of DNA versus RNA amplification. 3 primer pairs were designed for *ß2-tubulin* amplification. All primer pairs were tested on 0.5µl TOLIMAN gDNA (324ng/ul) with New England Labs® *OneTaq*® 2X Master Mix with Standard Buffer used for all PCR reactions. Cycling conditions were 3 minutes at 94°C, 28 cycles of 30 s at 94°C, 30 s at 55°C, 2 minutes at 72°C with final extension of 10 minutes at 72°C. The resulting PCR products were visualised using gel electrophoresis, and primer pairs to be used for the cDNA amplification were selected based on which pairs showed the strongest, correct sized bands on these gel results. All primers were synthesised by Integrated DNA Technologies and the sequences are available in *Supplementary information 4.6.1*.

4.3.3.3 RNA Extraction

Prior to commencing RNA extractions, all equipment and worktops were sterilised with 70% ethanol and Invitrogen RNaseZap™. Consumables and equipment were then placed in a UV hood for 15 minutes to reduce the risk of RNase contamination. Samples for RNA extractions were removed from the freezer and placed onto dry ice. Samples were then homogenised with a sterilised, pre-chilled, plastic pestle over dry ice. The Invitrogen *mir*Vana™ miRNA Isolation Kit was used for RNA extractions as per the manufacturer's instructions. Samples were eluted in 20µl Invitrogen™ THE Ambion™ RNA Storage solution.

4.3.3.4 Concentration quantification

The concentration of all DNA and RNA samples was measured via a Thermo fisher Scientific Nanodrop™ 8000 Spectrophotometer in 1µl volume aliquots. Optical density was measured at 260/280 with the ideal value being 1.8 and at 260/230 with the ideal value being 2.0.

4.3.3.5 DNase treatment

DNase treatments were conducted to remove DNA from the RNA samples as per the manufacturer's instructions, using the Invitrogen Turbo DNA-*free*™ kit for qPCR samples or with the Invitrogen Ambion® DNA-*free*™ Kit for RT-PCR/RACE samples.

4.3.3.6   Reverse Transcriptase PCR (RT-PCR)

All DNased samples for end-point RT-PCR only were diluted to below 20ng/μl. The Thermo Scientific RevertAid H Minus First strand cDNA Synthesis Kit was used as per the manufacturer's instructions. All samples were diluted to the same concentration as the lowest concentration sample. Qiagen QuantiTect® Reverse Transcription Kit was then used for the RT-PCR reactions as per the manufacturer's instructions.

4.3.3.7   End-point RT-PCR

New England Labs® *OneTaq*® 2X Master Mix with Standard Buffer was used for all PCR reactions. The PCR cycling conditions were 30 s at 94°C, 28 cycles of 30 s at 94°C, 30 s at 54°C, 1 minute at 68°C with a final extension of 10 minutes at 68°C. PCR products were visualised with gel electrophoresis.

4.3.4   Gel electrophoresis

Products to be visualised by gel electrophoresis that did not have any loading dye in PCR mastermix were combined with New England BioLabs® Purple Gel Loading Dye (6X) prior to loading. Agarose gels were made at 1.2% (w/v) with 1X TBE and stained with 3μl Ethidium Bromide per 100ml gel. Gels were run with appropriate ladders in each case to determine band sizes. The ladder used was Thermo Scientific™ GeneRuler™ 1kb Plus DNA Ladder and New England Biolabs® 1KB Plus DNA Ladder. Gels were run at 110V for 60 minutes.

4.3.5   Construct design and building

4.3.5.1   Construct Design

All constructs were designed in Benchling using the Gibson assembly tool. Primers were designed manually or generated by Benchling (Benchling 2023).

4.3.5.2   DNA Synthesis

All synthesised genetic parts were designed in Benchling (Benchling 2023)and ordered from Azenta Life Sciences GENEWIZ® service in a PuC plasmid backbone.

4.3.5.3 Gibson Assembly

4.3.5.3.1 Primer Design

All primers were designed to create a 20bp overlap region between parts to allow for assembly. All primers were obtained from Integrated DNA Technologies and all primer sequences used in the final design are available in the *Supplementary information 4.6.1*.

4.3.5.3.2 Generating Gibson Assembly parts

Individual parts for the Gibson assemblies were generated via PCR with overhanging primers, using New England BioLabs® Q5® High fidelity 2X Master Mix to ensure sequence accuracy. Primer annealing temperatures were calculated using New England BioLabs® Tm Calculator and cycling conditions as outlined in manufacturer's instructions were used. Cycling conditions for all reactions as per the polymerase manufacturer's instructions. Medfly gDNA or plasmid DNA samples were used as templates for all PCR reactions. PCR products were purified and their concentration measured by using the Nanodrop™.

4.3.5.3.3 Gibson Assembly Reaction

An in house solution of Gibson Mastermix was used for the assemblies of plasmid UEA_009 following the method outlined in (Behle and Sakaguchi 2020). After incubation, Gibson reactions were diluted in a ratio of 1:5 and used for transformation into *E. coli* competent cells.

4.3.5.4 Generation of parts with restriction enzymes

New England Biolabs® BtgZI used as per the manufacturer's instructions for the digestion of backbone and insert plasmids or PCR products to generate complementary ends. Reaction mix incubated at 60°C for one hour and heat inactivated at 80°C for 20 minutes. If using PCR products as template DpNI digestion was performed with the direct addition of DpNI to the PCR mix. Incubation for 1h at 37°C followed by heat inactivation at 80°C for 20 minutes. New England Biolabs® Shrimp Alkaline Phosphatase was used as per the manufacturer's instructions to treat digested products to prevent plasmid relinearisation. Incubation at 37°C for 30 minutes and heat inactivated at 65°C for 15 minutes. Products were visualised with gel electrophoresis (4.3.4) and purified through gel extraction.

4.3.5.5   DNA Ligation

New England Biolabs® T4 DNA Ligase was used to anneal complementary DNA digest products using a 3:1 insert to vector DNA ratio. Reactions were incubated overnight at 16°C with 5µl of the completed reaction used for *E.coli* transformation.

4.3.5.6   PCR Purification

The Thermo Scientific™ GeneJET PCR Purification Kit was used for the purification of PCR products, as per the manufacturer's instructions with final elution in 25µl molecular grade water after a 1 minute incubation at room temperature.

4.3.5.7   Gel Extraction

The Thermo Scientific™ GeneJET Gel Extraction Kit was used for purification of excised bands from agarose gels as per the manufacturer's instructions with a final elution in 25µl molecular grade water after a 1 minute incubation at room temperature.

4.3.5.8   Transformation of *E. coli* competent cells

Agilent SURE® 2 Supercompetent Cells were used and the transformation protocol is summarised below:

Table 4.2 Transformation protocols of Agilent SURE® 2 Supercompetent cells used for construct assembly.

	Agilent SURE® 2 Supercompetent Cells
Volume used	5µl diluted Gibson : 50µl cells
Ice incubation	30 mins
Heat shock	30 s at 42°C
Ice incubation	2 mins
Volume of SOC added	700µl
Incubation	1 hr at 37°C, 100rpm
Spin step	2 mins at 4000rpm, pellet resuspended in 100 µl SOC
Final plated volume onto LB Ampicillin (100µg/ml)	100 µl
Plate incubation	Overnight at 37°C

4.3.5.9 Colony PCR

Individual colonies were screened for inserts after Gibson assembly. Colonies of varying sizes were selected that did not overlap with other colonies. A 10µl pipette tip was used to remove the edge of each screened colony, and the material was then resuspended in 10µl of molecular grade water. 1µl of this colony mix was then used as the template in a Thermo Scientific™ DreamTaq Green PCR Master Mix (2X) reaction with a total volume of 20µl. Primers were designed to cover insert boundaries where possible to confirm presence of both the insert and the backbone. The PCR cycling conditions were 2 minutes at 95°C, 28 cycles of 30 s at 95°C, 30 s at 54°C, 3 minutes at 72°C with a final extension of 10 minutes at 72°C All primers used are listed in *Supplementary Information 4.6.1*.

4.3.5.10 Plasmid Miniprep

Positive clones were miniprepped to prepare them for sequencing. Overnight cultures were inoculated with a single positive colony into 4ml LB Ampicillin (100µg/ml) contained within a 15ml falcon tube to allow sufficient aeration. Cultures were incubated overnight at 37°C with constant agitation at 180rpm. After confirming growth, 2ml of the culture was prepared by using the Thermo Scientific™ GeneJET Plasmid Miniprep Kit, as per the manufacturer's instructions. The final elution step was performed with 30µl molecular grade water. The purity and concentration of all samples were checked with Nanodrop™ before being diluted to the correct concentration for sequencing.

4.3.5.11 Sequencing confirmation of plasmids

Whole plasmid sequencing was completed by diluting miniprepped plasmids to 30ng/µl and sending to SNPsaurus© Plasmidsaurus Whole Plasmid Sequencing Service (Nanopore based long-read). Sequence alignment of returned reads was performed using the Benchling sequence alignment tool (Benchling 2023).

4.3.5.12 Plasmid Midiprep

After sequence confirmation, plasmids were prepared for microinjection using the Macherey-Nagel NucleoBond® Xtra Midi EF, Midi kit for endotoxin-free plasmid DNA kit. Cultures were revived from glycerol stocks onto a streak plate which was incubated overnight at 37°C. A single culture from this was used for inoculation of a 4ml LB Ampicillin (100µg/ml) starter culture and incubated for 6-8h at 37°C with constant agitation at 180rpm. After confirming growth in starter cultures, 400µl was used to inoculate 100ml LB Ampicillin (100µg/ml) to be incubated overnight at 37°C with constant agitation at 200rpm. This culture was then used as per the manufacturer's instructions for midiprep. The final pellet was resuspended in 50µl endotoxin-free molecular grade water.

4.3.5.13 Midiprep Purification

A further purification step was performed on midiprep samples for microinjection using Millipore Ultrafree®-MC GV Centrifugal Filters. The full midiprep sample was loaded to the column and centrifuged at 14,000 RPM for 4 minutes.

4.3.5.14 Preparing glycerol stocks

For the preparation of glycerol stocks for long term plasmid storage, single colonies were selected to inoculate 4ml LB Ampicillin (100µg/ml). This was then incubated overnight at 37°C with constant agitation at 180rpm (typically, this was the culture that was also used for plasmid miniprep). From this, 500µl of culture was added to 500µl 50% glycerol and mixed well by inversion. Glycerol stocks were then stored at -80°C, and only the top layer was thawed when reviving them.

4.4   Results

4.4.1   The expression pattern of *ß2-tubulin* determined through RT-PCR

To confirm the testis-specific expression of *ß2-tubulin*, RT-PCR was performed. Three primer sets were designed to amplify *ß2-tubulin,* with each designed to cover an intronic region to allow for the differentiation between DNA and RNA amplicons. All primer sets were tested on medfly TOLIMAN gDNA to determine the most efficient primer pair for use.
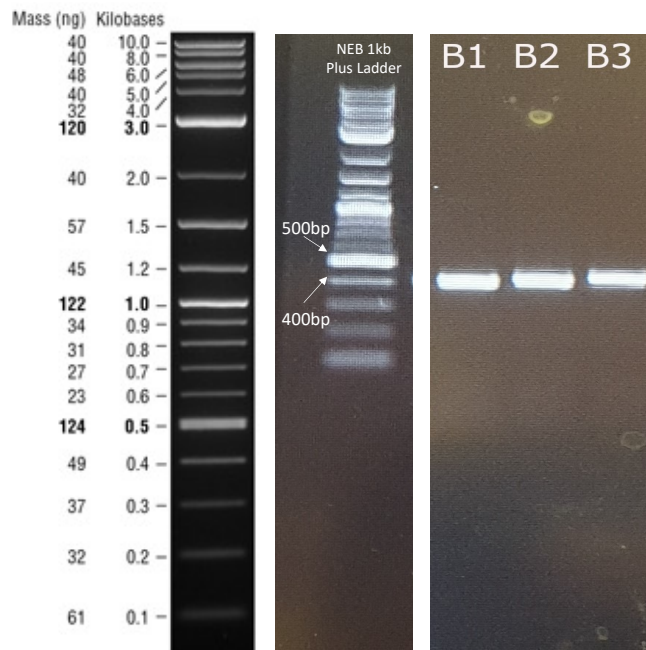
Figure 4.1 Gel electrophoresis of PCR products for multiple primer sets amplifying b2-tubulin from genomic DNA. The three lanes represent each primer set. The gel was run with NEB 1kb plus a DNA Ladder.

As shown in *Figure 4.1*, all three primer sets amplified an appropriately sized fragment (expected sizes were between 400 and 500bp). Primer set 2 was selected as this has the greatest size difference between DNA and RNA products, which allowed for easier identification of DNA contamination.

To determine the expression of *ß2-tubulin* in different tissue types, RNA was extracted from: male somatic tissue, early germline male testes (upper 1/3rd of testes), late germline male testes (lower 2/3rd of testes), female somatic tissue, early germline ovaries (upper 1/3rd of ovary), late germline ovaries (lower 2/3rd of ovary) and eggs at several development times (0h, 1h, 3h, 6h, 24h and 48h). First strand synthesis was performed on the RNA extracted

from all samples to generate cDNA. *ß2-tubulin* was then amplified through PCR. The products were visualised with gel electrophoresis and shown in *Figure 4.2*.
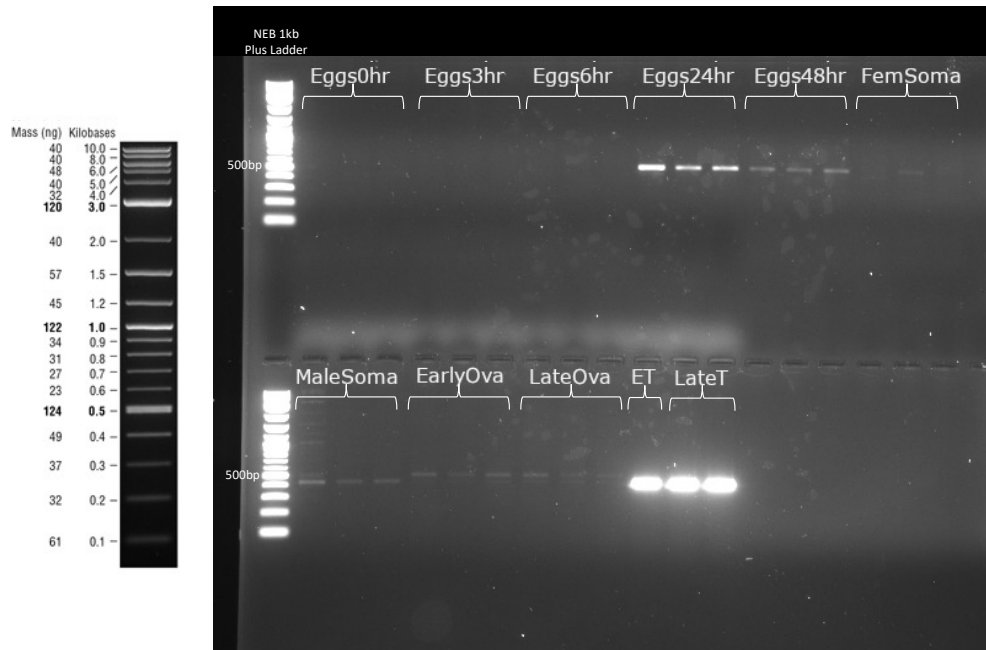
Figure 4.2 RNA expression pattern of *ß2-tubulin.* End point RT-PCR results indicate in which tissue types *ß2-tubulin* mRNA is expressed. The smaller sized band indicates the presence of spliced mRNA, whereas the larger band is unspliced mRNA. The gel was run with a NEB 1kb Plus Ladder with band sizes shown. Order of loading: Ladder, Eggs 0h post collection, Eggs 3h post collection, Eggs 6h post collection, 24h post egg collection, 48h post egg collection,, Female Soma, Male Soma, Early Ovary, Late Ovary, Early Testes, Late Testes. Three biological replicates of every sample type were run except early testes for which only one pooled sample could be produced, and late testes with two pooled samples only.

High expression was seen in the testes samples, as expected, which confirms that the promotor has the potential to be used to drive male germline expression of *MoY*. There was also evidence of expression in male somatic tissue, albeit at a lower level. The bands present in ovary, female somatic tissue and embryos were unexpected. However, these were primarily of the size expected for DNA samples, so may have represented DNA contamination. To test for such gDNA contamination, PCR was performed on an intronic region in a 24h post egg collection sample and early ovary sample with a gDNA sample as control. No bands were present for either cDNA sample. Therefore, DNA contamination appears unlikely. Another possible cause of the larger bands is unspliced or differentially spliced mRNA. High expression in male germline tissues is generally found to be the main requirement for use of a candidate sequence as a promotor, with potential low-level expression in other tissues not excluding use.

4.4.2   Plasmid design

The sequence of *MoY* needed for expression had previously been determined (Meccariello et al. 2019) and the aim was to put *MoY* under the expression of the 868bp *ß2-tubulin* promotor and 5'UTR (Scolari et al. 2008). A *gfp* marker was paired with a cleaved ubiquitin linker to allow for its attachment to the *MoY* component. This would allow visualisation of any successful *gfp* packaging into sperm cells. The fluorophore DsRed2 (Nishizawa et al. 2006) was also used as a marker element, driven by the commonly used insect expression promotor Hr5-IE1 (Jarvis et al. 1996) with the *Drosophila* native K10 acting as the 3'UTR and terminator. A nuclear localisation signal (NLS) flanked the DsRed2, to allow for a clearer phenotype by restricting DsRed2 expression to the nucleus. Two *piggyBac* (Li et al. 2013) arms flanked this region for genome insertion. On the outside of these arms were the elements needed for plasmid propagation in *E. coli.* This included the standard pUC origin of replication with an AmpR sequence coding for Ampicillin resistance. The original design for the *ß2-MoY* also included *gypsy* elements to protect against positional effects of the insert (Geyer and Corces 1992; Sarkar et al. 2006). However, due to difficulties with assembly these were excluded from the final plasmid design. The final plasmid design is shown in *Figure 4.3* and was assigned the designation UEA_009.
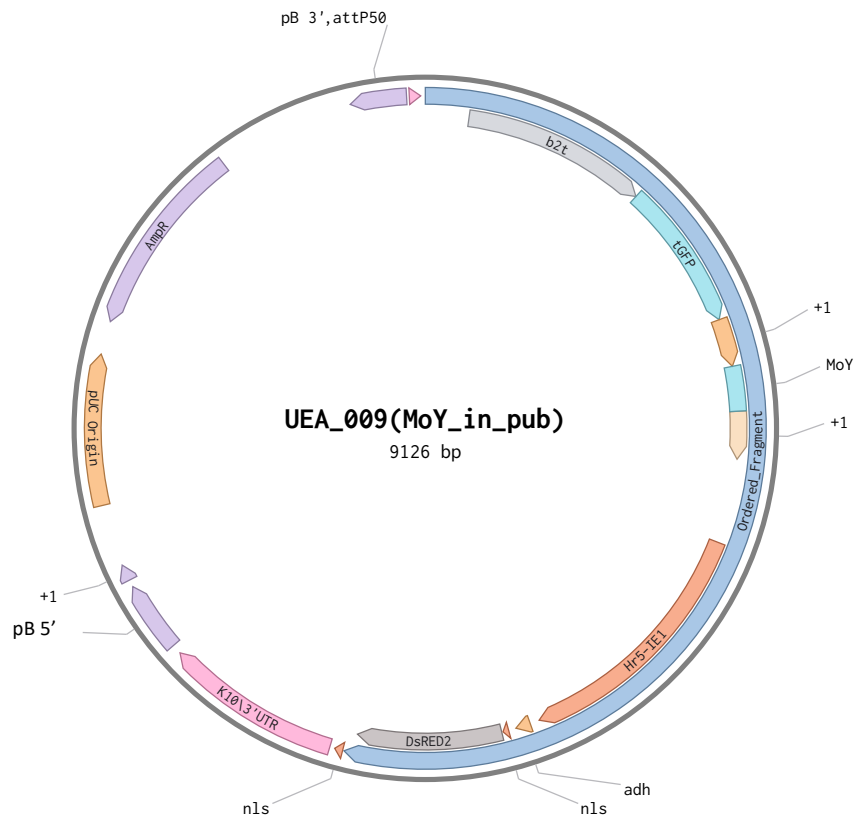
Figure 4.3 The plasmid map of UEA_009 generated in Benchling (Benchling 2023). The full sequence can be found in *Supplementary information 4.6.3*. The section of the plasmid annotated Ordered_Fragment was synthesised externally, the rest of the plasmid was generated as a single part from previously assembled Cas9 expression plasmid as discussed in 4.4.3 Plasmid construction.

4.4.3   Plasmid construction

After design of the *ß2-MoY* plasmid, the construction was attempted with part assembly however this was unsuccessful due to recombination during Gibson Assembly. A 5000bp *ß2-MoY* part was ordered which was designed to allow for the insertion into an already assembled plasmid backbone after digestion with BtgZI. The synthesised *ß2-MoY* part contained the *ß2-MoY-ub-gfp* and HR5-DsRED2 and the aim was to insert this within the *gypsy* elements of the backbone plasmid. Digests were successful, but when ligating the plasmid, only recombined parts of the backbone fragment (containing the *E.coli* survival elements) were present in sequenced colonies. This issue persisted even when shrimp alkaline phosphatase was used to prevent self-annealing. Therefore, the decision was made to assemble the construct through an alternative method of Gibson Assembly. Primers were originally designed to create a final plasmid the same as that which used restriction enzyme

digestion and ligation. However, the PCR products produced multiple banding resulting in products having to be gel cut reducing the concentration and increasing impurities in the final part. Unfortunately, no successful Gibson assemblies were obtained from this procedure.

The plasmid was then slightly redesigned to allow for primers to be placed at different points in the sequence. The plasmid being used as a backbone was changed to one which did not have *gypsy* elements, as these has caused issues with multiple banding and recombination when primers were placed within them. The new backbone was generated with primers 325 and 327 (*Supplementary information 4.6.1*) using UEA_005 (*Supplementary information 4.6.2*) as template. The insert was generated using primers 326 and 328 on the plasmid received from Azenta containing *ß2-MoY-gfp* and HR5-DsRED2 constructs. Products were run on a 1.2% gel at 120V for 1h with NEB 1KB Plus DNA Ladder shown in *Figure 4.4*.
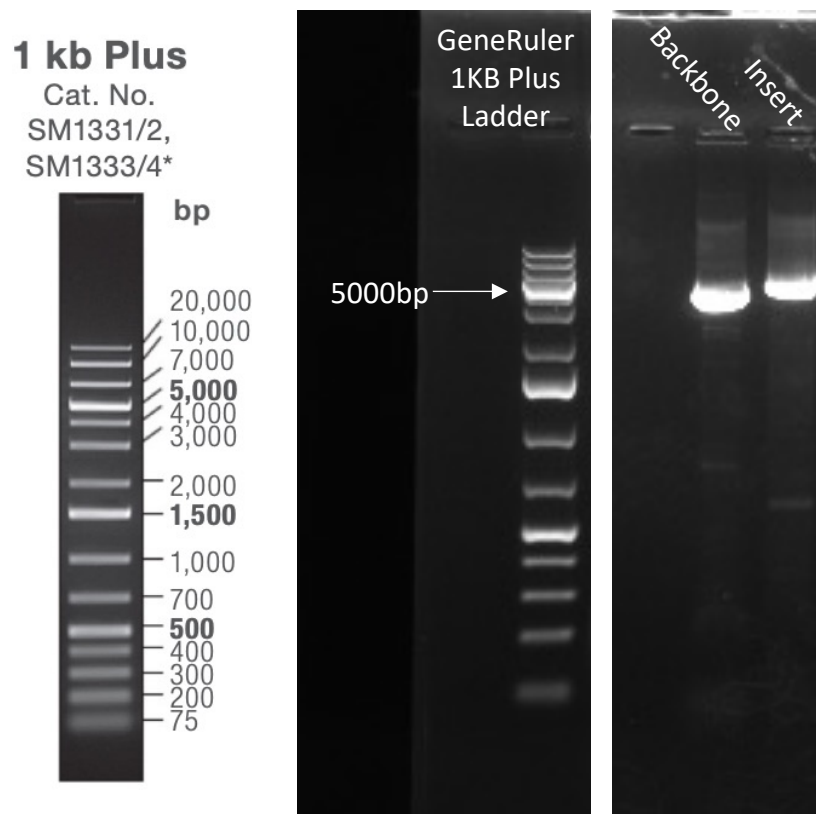
Figure 4.4 Gel electrophoresis of products from the PCR to generate the parts for Gibson assembly of *ß2-MoY*. Products were run on a 1.2% gel at 120V for 1h alongside the GeneRuler 1KB Plus DNA Ladder. Order of loading: Ladder, backbone product (expected size 4204bp), insert product (expected size 4922bp). Both are of expected size and can be used for PCR purification.

The resulting PCR products were of the expected size and showed single banding, hence PCR purification could be used without the need for gel cutting, which optimised the

concentrations achieved. Concentrations of purified products measured with Nanodrop and are shown in *Table 4.3*.

Table 4.3 Nanodrop results of purified PCR products to be used for Gibson assembly.

Product	Concentration (ng/μl)	260/280	260/230
Backbone	149.3	1.81	1.75
Insert	131.1	1.77	0.91

Products were of a sufficiently high concentration for use in Gibson assembly. Plate showed high levels of growth with over 50 colonies present and 20 colonies were selected for colony PCR using primers 158 and 163 (*Supplementary information 4.6.1*). Products were visualised through gel electrophoresis as shown in *Figure 4.5*.
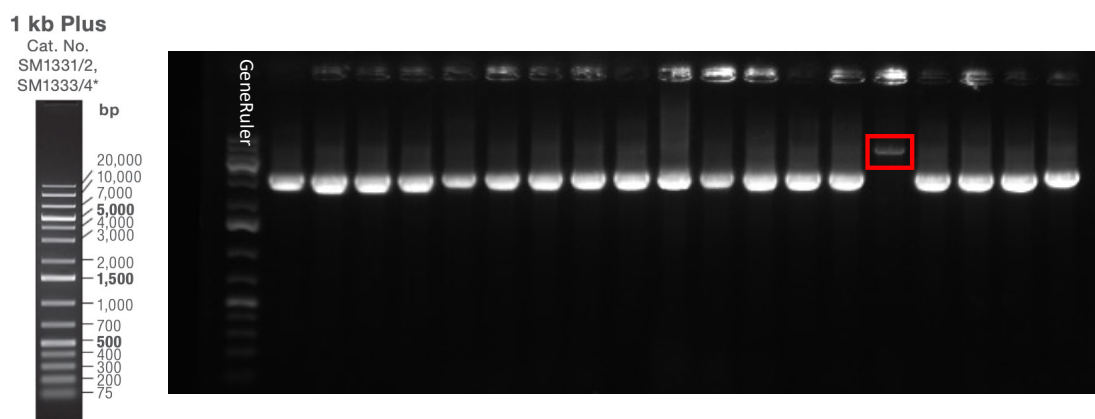
Figure 4.5 Gel electrophoresis of colony PCR results for *MoY* colonies. Samples were run on a 1.2% gel for 1h at 120V alongside the GeneRuler 1KB Plus DNA ladder. Order of loading: DNA Ladder, Colonies 2-20. All colonies showed a positive result except as indicated with the red square, which was a band of the size expected from original plasmid template contamination.

Nearly all colony PCRs returned positive results, 3 plasmids with successful results were randomly selected to be sent for sequence confirmation. Overnight cultures of colonies 2, 5 and 9 were purified through plasmid miniprep and sent for whole plasmid sequencing, the remaining overnight culture was used to create glycerol stocks for plasmid revival in the future. Sequencing results for all three plasmids confirmed correct assembly and can be found in *Supplementary information 4.6.3*. *MoY* colony 2 was selected for further use and is henceforth referred to as UEA_009. A midiprep of UEA_009 was completed with the nanodrop results shown in *Table 4.4*.

Table 4.4 Concentration of UEA_009 midiprep measured by Nanodrop.

Product	Concentration (ng/µl)	260/280	260/230
UEA_009	4360	1.37	1.61

This sample was sent for microinjection at Imperial College. However, there were issues during the injection process due to low concentration of the injection mix. A new midiprep was then generated with a further purification step, the concentration was measured by nanodrop and shown in *Table 4.5*.

Table 4.5 Concentration of purified UEA_009 midiprep measured by Nanodrop.

Product	Concentration (ng/µl)	260/280	260/230
UEA_009	1227	1.86	2.35

This sample was sent for microinjection at Imperial College and was successfully injected.

4.4.4   Microinjection of UEA_009

Two batches of microinjections were completed for UEA_009. The second batch of injections was completed, as a relatively high death rate of injected embryos with very low hatch rate was observed post injection in the first batch compared to usual injection survival numbers. From the injections 28 individuals survived post-eclosion (20 females and 11 males). Some individuals had wing damage incurred during eclosion. Males with damaged wings were not used for crosses and females with damaged wings were all placed in a single cage as they were deemed to have a low chance of producing offspring. The surviving individuals were crossed to *Benavi* wildtype individuals in batch crosses outlined below.

Table 4.6 Crosses set up with microinjection survivors to wildtype individuals. Only healthy male survivors were used and all damaged female survivors were isolated to cage 009.3.

Cage ID	Number of Males	Number of Females
009.1	6 Wildtype	7 Injection survivors
009.2	2 Injection survivors	12 Wildtype
009.3	2 Wildtype	7 Injection survivors with damaged wings
009.4	1 Injection survivor	3 Wildtype
009.5	2 Wildtypes	3 Injection survivors
009.6	2 Injection survivors	8 Wildtype
009.7	2 Injection survivors	10 Wildtype

Individuals placed within the cages were given time to mate and 3 egg collections were then taken from each cage. No eggs were produced by cage 009.3 possibly due to the damaged wing individuals being unable to lay eggs therefore no egg collections were taken from this cage. Egg collections were obtained from cage 009.4 but no eggs hatched from these samples.

4.4.5   Fluorescent screening of G1

Offspring from the crosses set up as described above were screened as both pupae and adult flies. A total of 858 adult flies were checked for fluorescence with no fluorescence observed. There was also no evidence of sex biasing, with 438 males screened and 420 females screened. All pupae were screened pre-eclosion and there was no evidence of fluorescence in pupal stages. With the lack of transgenics, coupled with the low egg hatching rate post microinjection, it was assumed that the construct had a toxic effect and any embryos that had successfully integrated the construct died before hatching.

4.5   Discussion

The use of *MoY* as a potential tool to achieve sex conversion has opened up an interesting avenue of research in medfly since its discovery. This type of isolated male determining factor, which achieves masculinisation of XX individuals to fertile pseudomales when expressed transiently, and feminisation of XY individuals to fertile females when knocked out (Meccariello et al. 2019), could be harnessed for controlling sex determination in pest control settings, provided a more thorough understanding of expression timing can be obtained.

The masculinisation of females (by using *MoY* to create XX fertile pseudomales) is the pathway that would be more beneficial in pest control and rearing facilities. Transient expression of *MoY* recombinant protein is reported to masculinise XX females. However, full conversion to fertile pseudomales occurred only when individuals were injected with linearised or plasmid *MoY* transcripts (Meccariello et al. 2019). Accurate timing and the level of *MoY* expression is difficult to achieve in comparison to a knockout and this is the most likely reason only partial masculinisation could be achieved when injecting with recombinant protein. The injection of recombinant *MoY* protein would only result in transient presence of *MoY* and therefore cannot be present throughout development. To test whether complete conversion can be achieved through genetic modification, *MoY* must be inserted into the genome to allow for expression throughout development in transgenic individuals. This was previously attempted with MoY under the control of its native promotor, in order to see how stable expression would affect XX individuals. However, this was shown to be embryonic lethal and no transgenic survivors were recovered (Meccariello, personal communication).

The embryonic death previously observed by Meccariello could have two explanations. First, overexpression of *MoY* at any stage of development could be lethal, with too much *MoY* being produced causing organism death. Second, the timing of *MoY* expression might need to be tightly controlled temporally, with expression past a certain point in development resulting in death. The rate of death observed indicates off-plasmid production of *MoY* is able to cause death before integration, affecting both males and females where expression should be minimal. The lethality of *MoY* overexpression that was observed in both males *and* females is interesting as this had not been seen in other experiments done with comparable male determining factors in other species, such as Yob in *An.gambiae*, in which overexpression had no observed effect on males (Krzywinska and Krzywinski 2018). This means that male determining factors are not necessarily always lethal when mis-expressed, although this does seem to be the case in medfly. Placing *MoY* expression under the control

of a germline specific promotor might be more successful if timing of expression is the critical issue. *ß2-tubulin* is reported to be a testis-specific gene, with expression restricted to during spermatogenesis (Santel 2000). This promotor has been used extensively for transgene expression during spermatogenesis in a variety of insects (Smith et al. 2007; Zimowska et al. 2009; Yan et al. 2021) and has been used for fluorescent sperm marking in medfly (Scolari et al. 2008).

When investigating the expression pattern of *ß2-tubulin* here*,* expression was expected to be restricted to the male testes and possibly late egg samples. I found that these were indeed the samples in which expression was strongest. However, I also observed low levels of expression in male somatic tissue. This could have been caused by expression of *MoY* in parts of the male reproductive tract outside of the testes, which were not removed when the testes were dissected. Faint bands of a larger size were also present in the ovary and female somatic tissue samples. I identified these bands as potential unspliced full length mRNA transcripts. This suggests that *ß2-tubulin* is not likely to be expressed in these tissues, as the mRNA did not appear to be translation-ready. Alternative splicing in different tissue types is well documented, with the prevention of expression likely due to the splicing complex not being recruited to *ß2-tubulin* mRNA in tissues where it is not needed (Alberts et al. 2002). This does pose a potential problem when using the promotor for expression of a transgene. Transgenes are typically designed to need minimal post translational modification (to reduce the size of the inserted genetic construct, reducing fitness load). Therefore, if translation occurs in these tissue types and expression is prevented by a lack of post translational modifications, transgene expression could still occur in such tissues which is undesirable. The level of *ß2-tubulin* RNA appeared to be much lower outside of the testes (when looking at both spliced and unspliced isoforms) in my study. This suggests that although the *ß2-tubulin* promotor may allow for some expression outside of the testes, it will be at a much lower level than in the testes. As there was a low level of expression in the somatic tissue (spliced or unspliced), the decision was made to use the promotor for *MoY* expression.

If the promotor had worked as envisaged, *MoY* protein would be packaged and transmitted via sperm cells. The activity of a protein packaged in sperm cells is an interesting study in itself. *GFP* can successfully be packaged and show activity in sperm, creating the ability to visualise sperm movement (Scolari et al. 2008). However, the stability of proteins when carried in sperm is relatively unknown. There are several identified proteins carried within sperm cells which are important for aspects of fertilisation, but the persistence of these

proteins beyond the point of fertilisation has not yet been characterised (Barroso et al. 2009; Castillo et al. 2018). *GFP* is a very stable protein with protein degradation shown to take several days (Nash and Lever 2004). The presence of fluorescent sperm cells does not show the protein is still active within the sperm cells, just that it was active and successfully packaged at one point in time. *MoY* is likely to show less stability as a protein, as overexpression or expression beyond a certain point in development is lethal (Meccariello, personal communication). Hence, the likelihood of expression persisting for days, given the development time of medfly, is relatively low.

The incomplete masculinisation of medfly when injected with recombinant *MoY* protein (Meccariello et al. 2019) could have been caused by several factors. The first consideration would be whether *MoY* was introduced early enough in development to influence the entire sex determination pathway. When injected with plasmid or linear *MoY* transcribing DNA, full masculinisation of XX individuals was seen, meaning the timing of injection was sufficiently early in development. The possibilities which cannot yet be ruled out are that the level of injected *MoY* was not sufficiently high for full sex conversion or that *MoY* does not persist for a long enough time to have full effect and thus would need to be translated continuously through sexual development. Both of these problems could be rectified by injecting plasmid DNA as it allows native expression levels and plasmid DNA can persist for much longer than protein.

Assuming the *MoY* protein is successfully transferred through the sperm and persists past fertilisation, we should consider the importance of the timing of *MoY* expression. If *MoY* protein does have a toxic effect, it would be assumed that it has a relatively narrow window of expression. The *Sry* gene is a Y-linked sex determining element in mice tested in the context of timing of expression and its effects on development. The timing of *Sry* expression was limited to a 6 hour window to successfully induce testis development (Hiramatsu et al. 2009). If *MoY* was successfully packaged in sperm cells, the presence of transgenic *MoY* should be as early as possible as it would be present at the moment of fertilisation and would persist through early development. If it persisted for an equal amount of time as native *MoY*, it should not persist for longer than necessary as *MoY* must degrade naturally to prevent the lethality we have seen in overexpression. Several issues could arise with this early expression. Lethality could be caused when present too early in development similar to the lethality observed when expression persists for too long. Lethality if there is overexpression of *MoY* due to the use of a non-native promotor, which could cause higher than native levels, could

be an issue. Finally, if the protein does not persist for long enough to fully act upon the sex determination cascade as it degrades prior to the sex determination window.

MoY being placed under its native promotor but still showing high levels of lethality means that genomic context is likely to be important in expression. Genomic context encompasses how the location of the gene effects its expression both through its chromosome location and the surrounding regulatory regions expanding beyond the 5' and 3' untranslated regions. The native location of *MoY* on the Y chromosome might be particularly important given its location on the Y chromosome. Expression of transgenes off the Y is often more difficult to predict due to chromosome degradation and can be much more varied than is seen for autosomes (Hall et al. 2016). Gene expression from sex chromosomes is also affected by dosage compensation, with X shutdown being a well-documented phenomenon (Morey and Avner 2011) to account for the difference in X chromosome number between XX and XY individuals. As the Y is much smaller and contains significantly fewer genes, its inactivation is not considered as vital. However, it has been shown that Y inactivation does occur at specific developmental stages (Khalil et al. 2004). Although the Y chromosome in medfly seems relatively degenerated (as XX pseudomales are fertile), fitness costs have been observed in other species where XO pseudomales are fertile in the absence of a Y chromosome (Voelker and Kojima 1971). If Y inactivation occurs in medfly, this could be the control mechanism to prevent the expression of *MoY* past a certain developmental point, preventing lethality due to overexpression. This informed the decision to place *MoY* under a germline specific promotor to prevent overexpression. When injecting my construct into medfly, we experienced a similar lethality effect to that of the insertion of native *MoY* off the Y chromosome. It is difficult to judge whether leaky expression with non-integrated, multiple copies of plasmid produces a more toxic effect than might be observed if expression could be silenced until after genomic integration.

To more tightly control the expression of *MoY* and avoid lethality during microinjection, a further construct was designed which utilised the Cre-*loxP* system (*Supplementary information 4.6.4*) (Sternberg and Hamilton 1981). This should allow the *MoY* construct to be integrated without the concern of expression effects during the microinjection process. By inserting a *loxP* stuffer within the coding sequence of *MoY,* we can ensure expression is not occurring during line establishment. Once the transgenic line is established the stuffer would be removed by crossing to a line expressing Cre recombinase. As the stuffer would not be removed until crossed with the Cre line, this should allow a much tighter level of control and

inform whether the *MoY* related mortality goes beyond preventing microinjection survivors. Another potential avenue of research could be the trialling of degradation tags (degrons) flanking the *MoY* (Natsume and Kanemaki 2017)*. The use of degrons coupled with a late-acting sperm specific promotor (such as protamines) would cause *MoY* to degrade quicker and persist for less time. Conditional degrons could also be used to allow for the degradation to be triggered by an external factor that can be controlled, such as temperature. This would allow degradation at specific time points to be tested. If persistence beyond a certain developmental point is the cause of lethality this could allow a transgenic *MoY* line to be established.

Overall, *MoY* provides research opportunities for pest control. However, further research is clearly needed into expression timing and the apparent lethality of overexpression. *MoY* remains a useful tool for behavioural biology as it has a strong ability to generate XX pseudomales which can be studied and the evolution of this protein is also of great general interest. Additional avenues of research include the use of the Cre-*LoxP* construct and more closely examining the genetic context of *MoY* and how this affects its expression. Male determining elements hold great potential in pest control in several species in which they are present. However, as shown here, the usage of these tools may be more complex than originally envisioned.

4.6 Supplementary Materials

4.6.1 Primers sequences

Table 4.7 Sequences of primers used throughout this experiment.

Primer Number	Sequence (5'-3')
17	agctggacaatgtggcaaccaga
18	gttcccatgccggaacccgtac
19	gccacaggcacttactatggcg
20	tcggatacctttggggagggca
21	tgagaagcggtaaaaccttaagt
22	tggcccaattgttacccgctcc
158	tcggtctgtatatcgaggtt
163	cgatcgtgcgttacacgtag
325	ggaaggtggaggacccgtaaagatctcgacccaagaaaaa
326	tttttcttgggtcgagatctttacgggtcctccaccttcc
327	cgacgtcccatggccattcgcagatctcagtctgtcgacc
328	ggtcgacagactgagatctgcgaatggccatgggacgtcg

4.6.2 UEA_005 plasmid map
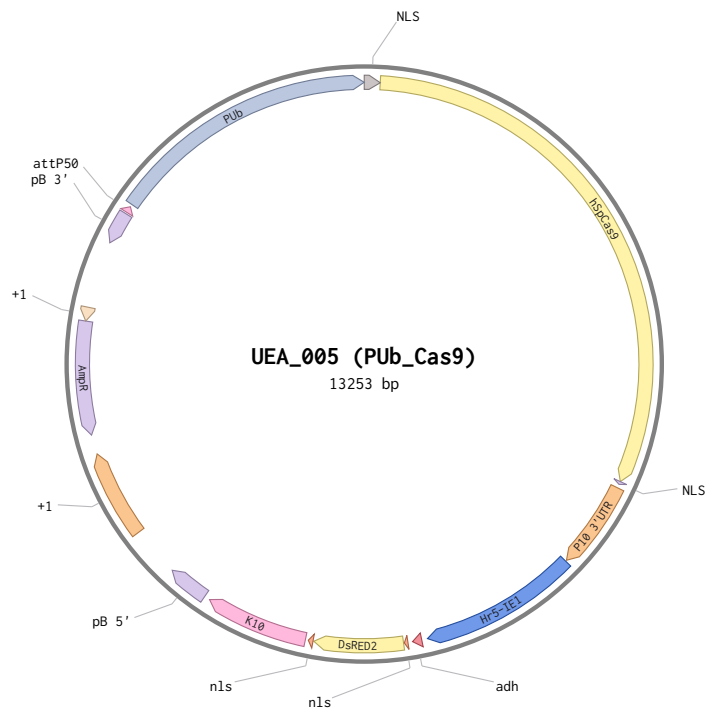
Figure 4.6 Plasmid map of UEA_005 used as backbone for UEA_009. Generated in Benchling (Benchling 2023).

4.6.3 Sequences

All sequences and aligned reads can be obtained at

https://benchling.com/siddall/f_/sqcsnseT-alex-siddall-thesis-supplementary-info/

MoY_TRE_CMV (11158 bp)
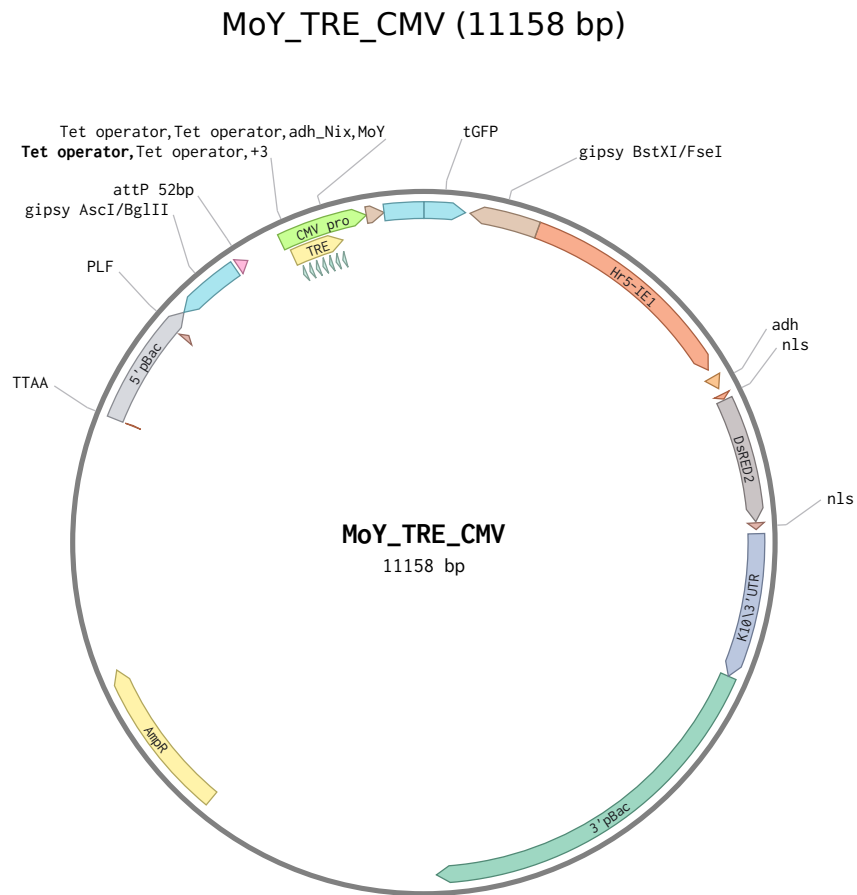
Figure 4.7 The plasmid design for inserting MoY with Lox stuffers which would prevent transcription until crossed with a line carrying Cre recombinase.

4.7   References

Adolfi, Mateus Contar, Amaury Herpin, and Manfred Schartl. 2021. 'The Replaceable Master of Sex Determination: Bottom-up Hypothesis Revisited'. *Philosophical Transactions of the Royal Society B: Biological Sciences* 376 (1832): 20200090. https://doi.org/10.1098/rstb.2020.0090.

Alberts, B, A Johnson, J Lewis, and et al. 2002. 'From DNA to RNA'. In *Molecular Biology of the Cell*, 4th ed. New York: Garland Science. https://www.ncbi.nlm.nih.gov/books/NBK26887/.

Aryan, Azadeh, Michelle A. E. Anderson, James K. Biedler, Yumin Qi, Justin M. Overcash, Anastasia N. Naumenko, Maria V. Sharakhova, Chunhong Mao, Zach N. Adelman, and Zhijian Tu. 2020. '*Nix* Alone Is Sufficient to Convert Female *Aedes aegypti* into Fertile Males and *Myo-Sex* Is Needed for Male Flight'. *Proceedings of the National Academy of Sciences* 117 (30): 17702–9. https://doi.org/10.1073/pnas.2001132117.

Barroso, Gerardo, Carlos Valdespin, Eva Vega, Ruben Kershenovich, Rosaura Avila, Conrado Avendaño, and Sergio Oehninger. 2009. 'Developmental Sperm Contributions: Fertilization and Beyond'. *Fertility and Sterility* 92 (3): 835–48. https://doi.org/10.1016/j.fertnstert.2009.06.030.

Beaghton, Andrea, Pantelis John Beaghton, and Austin Burt. 2017. 'Vector Control with Driving Y Chromosomes: Modelling the Evolution of Resistance'. *Malaria Journal* 16 (1): 286. https://doi.org/10.1186/s12936-017-1932-7.

Behle, Anna, and Tatsuya Sakaguchi. 2020. 'Homemade Gibson Mastermix V1'. Preprint. https://doi.org/10.17504/protocols.io.ba2yigfw.

Benchling. 2023. 'Biological Software'. Benchling.com.

Bernardini, Federica, Roberto Galizi, Miriam Menichelli, Philippos-Aris Papathanos, Vicky Dritsou, Eric Marois, Andrea Crisanti, and Nikolai Windbichler. 2014. 'Site-Specific Genetic Engineering of the *Anopheles gambiae* Y Chromosome'. *Proceedings of the National Academy of Sciences* 111 (21): 7600–7605. https://doi.org/10.1073/pnas.1404996111.

Castillo, Judit, Meritxell Jodar, and Rafael Oliva. 2018. 'The Contribution of Human Sperm Proteins to the Development and Epigenome of the Preimplantation Embryo'. *Human Reproduction Update* 24 (5): 535–55. https://doi.org/10.1093/humupd/dmy017.

Eckermann, Kolja N., Hassan M.M. Ahmed, Mohammad KaramiNejadRanjbar, Stefan Dippel, Christian E. Ogaugwu, Peter Kitzmann, Musa D. Isah, and Ernst A. Wimmer. 2018. 'Hyperactive piggyBac Transposase Improves Transformation Efficiency in Diverse Insect Species'. *Insect Biochemistry and Molecular Biology* 98 (July): 16–24. https://doi.org/10.1016/j.ibmb.2018.04.001.

Galizi, Roberto, Andrew Hammond, Kyros Kyrou, Chrysanthi Taxiarchi, Federica Bernardini, Samantha M O'Loughlin, Philippos-Aris Papathanos, Tony Nolan, Nikolai Windbichler, and Andrea Crisanti. 2016. 'A CRISPR-Cas9 Sex-Ratio Distortion System for Genetic Control'. *Sci. Rep.* 6 (August): 31139. https://doi.org/10.1038/srep31139.

Gamez, Stephanie, Duverney Chaverra-Rodriguez, Anna Buchman, Nikolay P. Kandul, Stelia C. Mendez-Sanchez, Jared B. Bennett, Héctor M. Sánchez C., et al. 2021. 'Exploiting a Y Chromosome-Linked Cas9 for Sex Selection and Gene Drive'. *Nature Communications* 12 (1): 7202. https://doi.org/10.1038/s41467-021-27333-1.

Gempe, Tanja, and Martin Beye. 2011. 'Function and Evolution of Sex Determination Mechanisms, Genes and Pathways in Insects'. *Bioessays* 33 (1): 52–60. https://doi.org/10.1002/bies.201000043.

Geyer, P K, and V G Corces. 1992. 'DNA Position-Specific Repression of Transcription by a *Drosophila* Zinc Finger Protein.' *Genes & Development* 6 (10): 1865–73. https://doi.org/10.1101/gad.6.10.1865.

Hall, Andrew Brantley, Philippos-Aris Papathanos, Atashi Sharma, Changde Cheng, Omar S. Akbari, Lauren Assour, Nicholas H. Bergman, et al. 2016. 'Radical Remodeling of the Y Chromosome in a Recent Radiation of Malaria Mosquitoes'. *Proceedings of the National Academy of Sciences* 113 (15). https://doi.org/10.1073/pnas.1525164113.

Hamilton, W. D. 1967. 'Extraordinary Sex Ratios: A Sex-Ratio Theory for Sex Linkage and Inbreeding Has New Implications in Cytogenetics and Entomology.' *Science* 156 (3774): 477–88. https://doi.org/10.1126/science.156.3774.477.

Hiramatsu, Ryuji, Shogo Matoba, Masami Kanai-Azuma, Naoki Tsunekawa, Yuko Katoh-Fukui, Masamichi Kurohmaru, Ken-ichirou Morohashi, Dagmar Wilhelm, Peter Koopman, and Yoshiakira Kanai. 2009. 'A Critical Time Window of *Sry* Action in Gonadal Sex Determination in Mice'. *Development* 136 (1): 129–38. https://doi.org/10.1242/dev.029587.

Hopkins, Ben R, and Artyom Kopp. 2021. 'Evolution of Sexual Development and Sexual Dimorphism in Insects'. *Current Opinion in Genetics & Development* 69 (August): 129–39. https://doi.org/10.1016/j.gde.2021.02.011.

Jarvis, Donald L., Carla Weinkauf, and Linda A. Guarino. 1996. 'Immediate-Early Baculovirus Vectors for Foreign Gene Expression in Transformed or Infected Insect Cells'. *Protein Expression and Purification* 8 (2): 191–203. https://doi.org/10.1006/prep.1996.0092.

Khalil, Ahmad M., Fatih Z. Boyar, and Daniel J. Driscoll. 2004. 'Dynamic Histone Modifications Mark Sex Chromosome Inactivation and Reactivation during Mammalian Spermatogenesis'. *Proceedings of the National Academy of Sciences* 101 (47): 16583–87. https://doi.org/10.1073/pnas.0406325101.

Krzywinska, Elzbieta, and Jaroslaw Krzywinski. 2018. 'Effects of Stable Ectopic Expression of the Primary Sex Determination Gene Yob in the Mosquito *Anopheles gambiae*'. *Parasites & Vectors* 11 (S2): 648. https://doi.org/10.1186/s13071-018-3211-z.

Labbé, Geneviève M C, Sarah Scaife, Siân A Morgan, Zoë H Curtis, and Luke Alphey. 2012. 'Female-Specific Flightless (fsRIDL) Phenotype for Control of *Aedes albopictus*'. *PLoS Negl. Trop. Dis.* 6 (7): e1724. https://doi.org/10.1371/journal.pntd.0001724.

Li, Xianghong, Erin R. Burnight, Ashley L. Cooney, Nirav Malani, Troy Brady, Jeffry D. Sander, Janice Staber, et al. 2013. '*piggyBac* Transposase Tools for Genome Engineering'. *Proceedings of the National Academy of Sciences* 110 (25). https://doi.org/10.1073/pnas.1305987110.

Lutrat, Célia, Roenick P. Olmo, Thierry Baldet, Jérémy Bouyer, and Eric Marois. 2022. 'Transgenic Expression of Nix Converts Genetic Females into Males and Allows Automated Sex Sorting in *Aedes albopictus*'. *Communications Biology* 5 (1): 210. https://doi.org/10.1038/s42003-022-03165-7.

Meccariello, Angela, Marco Salvemini, Pasquale Primo, Brantley Hall, Panagiota Koskinioti, Martina Dalíková, Andrea Gravina, et al. 2019. '*Maleness-on-the-Y* (*MoY*) Orchestrates Male Sex Determination in Major Agricultural Fruit Fly Pests'. *Science* 365 (6460): 1457–60. https://doi.org/10.1126/science.aax1318.

Morey, Céline, and Philip Avner. 2011. 'The Demoiselle of X-Inactivation: 50 Years Old and As Trendy and Mesmerising As Ever'. Edited by Wendy A. Bickmore. *PLoS Genetics* 7 (7): e1002212. https://doi.org/10.1371/journal.pgen.1002212.

Nash, K L, and A M L Lever. 2004. 'Green Fluorescent Protein: Green Cells Do Not Always Indicate Gene Expression'. *Gene Therapy* 11 (11): 882–83. https://doi.org/10.1038/sj.gt.3302246.

Natsume, Toyoaki, and Masato T. Kanemaki. 2017. 'Conditional Degrons for Controlling Protein Expression at the Protein Level'. *Annual Review of Genetics* 51 (1): 83–102. https://doi.org/10.1146/annurev-genet-120116-024656.

Nishizawa, Keito, Yoichi Kita, Masahiko Kitayama, and Masao Ishimoto. 2006. 'A Red Fluorescent Protein, DsRed2, as a Visual Reporter for Transient Expression and Stable Transformation in Soybean'. *Plant Cell Reports* 25 (12): 1355–61. https://doi.org/10.1007/s00299-006-0210-x.

Pane, Attilio, Marco Salvemini, Pasquale Delli Bovi, Catello Polito, and Giuseppe Saccone. 2002. 'The *Transformer* Gene in *Ceratitis capitata* Provides a Genetic Basis for Selecting and Remembering the Sexual Fate'. *Development* 129 (15): 3715–25. https://doi.org/10.1242/dev.129.15.3715.

Perrotta, Maryanna Martina, Francesca Lucibelli, Sarah Maria Mazzucchiello, Nicole Fucci, Bruno Hay Mele, Ennio Giordano, Marco Salvemini, et al. 2023. 'Female Sex Determination Factors in *Ceratitis capitata*: Molecular and Structural Basis of *TRA* and *TRA2* Recognition'. *Insects* 14 (7): 605. https://doi.org/10.3390/insects14070605.

Rendon Arana, Pedro Alfonso., The University of Manchester (United Kingdom), and Department not provided. 1996. 'Development and Evaluation of a Temperature Sensitive Lethal (TSL) Genetic Sexing Strain of the Mediterranean Fruit Fly, *Ceratitis capitata* (Wied.)'.

Rubin, Gerald M., and Allan C. Spradling. 1982. 'Genetic Transformation of *Drosophila* with Transposable Element Vectors'. *Science* 218 (4570): 348–53. https://doi.org/10.1126/science.6289436.

Salvemini, Marco, Mark Robertson, Benjamin Aronson, Peter Atkinson, Lino C Polito, and Giuseppe Saccone. 2009. '*Ceratitis capitata Transformer-2* Gene Is Required to Establish and Maintain the Autoregulation of Cctra, the Master Gene for Female Sex Determination'. *Int. J. Dev. Biol.* 53 (1): 109–20. https://doi.org/10.1387/ijdb.082681ms.

Santel, A. 2000. 'The Initiator Element of the Drosophila Beta2 Tubulin Gene Core Promoter Contributes to Gene Expression in Vivo but Is Not Required for Male Germ-Cell Specific Expression'. *Nucleic Acids Research* 28 (6): 1439–46. https://doi.org/10.1093/nar/28.6.1439.

Sarkar, Abhimanyu, Asela Atapattu, Esther J Belikoff, Jörg C Heinrich, Xuelei Li, Carsten Horn, Ernst A Wimmer, and Maxwell J Scott. 2006. 'Insulated piggyBac Vectors for Insect Transgenesis'. *BMC Biotechnology* 6 (1): 27. https://doi.org/10.1186/1472-6750-6-27.

Scolari, Francesca, Marc F. Schetelig, Sabrina Bertin, Anna R. Malacrida, Giuliano Gasperi, and Ernst A. Wimmer. 2008. 'Fluorescent Sperm Marking to Improve the Fight against the Pest Insect *Ceratitis capitata* (Wiedemann; Diptera: Tephritidae)'. *New Biotechnology* 25 (1): 76–84. https://doi.org/10.1016/j.nbt.2008.02.001.

Siddall, Alex, Tim Harvey-Samuel, Tracey Chapman, and Philip T. Leftwich. 2022. 'Manipulating Insect Sex Determination Pathways for Genetic Pest Management: Opportunities and Challenges'. *Frontiers in Bioengineering and Biotechnology* 10 (June): 867851. https://doi.org/10.3389/fbioe.2022.867851.

Smith, R. C., M. F. Walter, R. H. Hice, D. A. O'Brochta, and P. W. Atkinson. 2007. 'Testis-Specific Expression of the ?2 Tubulin Promoter of *Aedes aegypti* and Its Application as a Genetic Sex-Separation Marker'. *Insect Molecular Biology* 16 (1): 61–71. https://doi.org/10.1111/j.1365-2583.2006.00701.x.

Sternberg, Nat, and Daniel Hamilton. 1981. 'Bacteriophage P1 Site-Specific Recombination'. *Journal of Molecular Biology* 150 (4): 467–86. https://doi.org/10.1016/0022-2836(81)90375-2.

Voelker, Robert A., and Ken-Ichi Kojima. 1971. 'Fertility and Fitness of XO Males in *Drosophila* I. Qualitative Study'. *Evolution* 25 (1): 119. https://doi.org/10.2307/2406505.

White, Ian M., and Marlene M. Elson-Harris. 1992. *Fruit Flies of Economic Significance: Their Identification and Bionomics*. Wallingford, Oxon, UK: CAB International in association with ACIAR.

Windbichler, Nikolai, Philippos Aris Papathanos, and Andrea Crisanti. 2008. 'Targeting the X Chromosome during Spermatogenesis Induces Y Chromosome Transmission Ratio Distortion and Early Dominant Embryo Lethality in *Anopheles gambiae*'. Edited by David L. Stern. *PLoS Genetics* 4 (12): e1000291. https://doi.org/10.1371/journal.pgen.1000291.

Yan, Ying, Jonas Schwirz, and Marc F. Schetelig. 2021. 'Characterization of the Drosophila Suzukii B2-Tubulin Gene and the Utilization of Its Promoter to Monitor Sex Separation and Insemination'. *Gene* 771 (March): 145366. https://doi.org/10.1016/j.gene.2020.145366.

Zimowska, Grazyna J., Xavier Nirmala, and Alfred M. Handler. 2009. 'The *B2-Tubulin* Gene from Three Tephritid Fruit Fly Species and Use of Its Promoter for Sperm Marking'. *Insect Biochemistry and Molecular Biology* 39 (8): 508–15. https://doi.org/10.1016/j.ibmb.2009.05.004.

5 Chapter Five - Designing guides for targeting sex determination genes

5.1 Abstract

Sex determination mechanisms are extremely varied across animal and plant taxa. The sex determination pathway is a well-studied cascade in insects where the presence or ratio of sex chromosomes is the sex determining factor. Despite variation is some aspects of these pathways, there are many commonalities across insect species in both the genes involved and the use of alternative splicing to dictate male or female development. The genes involved in sex determination pathways make appealing targets for genetic pest control strategies which aim to increase the production of males or induce female sterility. Two key genes in the sex determination pathway of the medfly *Ceratitis capitata* are *transformer* (*tra*) and *doublesex* (*dsx*). *Tra* is a midstream element in the sex determination pathway that is present across several, but not all, insect orders. It has potential uses in medfly in particular, and even potential portability once its full mechanisms are understood, as *tra* knockout in medfly generates fertile pseudomales. *Dsx* is the final element in the sex determination pathway. It is highly conserved across insects, offering additional potential for use in control strategies based on sex determination pathway manipulations. In this chapter, experimental designs are developed which target these key genes to generate sex conversion in the medfly. Two *tra* guide carrying lines were developed and established, and tested for efficacy by crossing them with previously developed Cas9 lines. However, none of the offspring cohorts from these crosses showed strong deviation from a 50:50 sex ratio in either the F1 or F2 generations. Thus, there was no evidence for any sex conversion, suggesting that the guide was not effective at inducing *tra* knockout. Due to the lack of sex conversion seen from this guide, lines containing a different guide were established and are currently being tested. An issue with the guide design was subsequently discovered which explained the lack of cutting and the correction for this has been discussed, along with future research directions for this type of pest control strategy.

5.2 Introduction

Genes within the sex determination pathway present promising targets for genetic control strategies. Sex determination pathways are well-suited to this aim, as there are both downstream commonalities between species, which offers portability, but also upstream differences offering specificity (Siddall et al. 2022; Geuverink and Beukeboom 2014). There are many different mechanisms of sex determination across taxa. Here I focus on those employed in several insects including the medfly *Ceratitis capitata*, which are based upon the presence of sex chromosomes in an XX/XY sex determination system. These sex determination pathways are tiggered by a 'master regulator' such as a masculinizing factor which is often carried on the Y and which then triggers a genetic cascade of sexually dimorphic effects. Interestingly, master regulators are one of the most variable features of sex determination pathways. The cascade they induce utilises alternative splicing for male and female development to either express or repress genes determining sex-specific phenotypes, to result in the *doublesex* (*dsx*) mRNA being produced in either the male or female isoform. The genes involved in the sex determination cascade can vary; however, the downstream elements are highly conserved between species. Two of the most highly conserved downstream genes across insect species from Diptera, Coleoptera and Hymenoptra are *dsx* (Price et al. 2015) and *transformer* (*tra*) (Geuverink and Beukeboom 2014; Perrotta et al. 2023). *Tra* is a midstream element, and is often the highest-level gene identified in these pathways, just downstream of the master regulator. The presence of *tra* triggers female development and its absence results in male development (Verhulst et al. 2010; Perrotta et al. 2023). *Dsx* is active lower in the sex determination cascade and determines sex through alternative splicing in males versus females. Depending upon the upstream signals acting upon it, *dsx* is alternatively spliced into either male-specific *dsx* (*dsxm*) isoform or female-specifc *dsx* (*dsxf*) isoform (Kyriacou 1992). Alternative splicing of *dsx* is the part of the somatic pathway in a wide variety of species and is a well conserved phenomenon even though the exact pattern of sex-specific splicing may vary between species.

In the medfly specifically, several upstream elements have also been identified – for example, both the master regulator is known as well as details of the mechanism of action of both *tra* and *dsx* (Salvemini et al. 2009; Primo et al. 2020; Perrotta et al. 2023; Saccone et al. 2008). *Tra* expression is repressed by the male determining factor (*Figure 5.1*; *see also Chapter 4*).
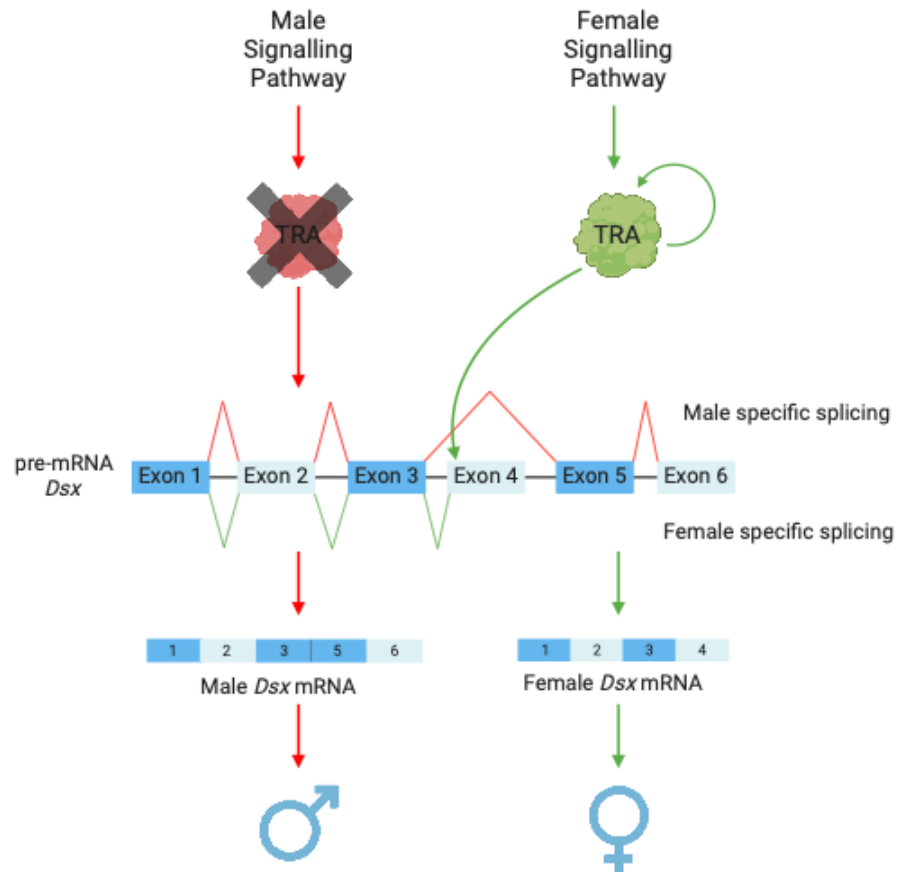
Figure 5.1 Mechanisms of *tra* and *dsx* in the sex determination pathway of *Ceratitis capitata*. Figure reproduced from Siddall et al. (2022) with permission and generated in Biorender.

The lack of *tra* results in splicing regulators not being recruited to the weak splice point of exon 4 of *dsx.* Therefore, exon 4 is not retained in the final *dsx* mRNA to produce *dsxm*. This then drives male development. When the male-determining factor is not present, *tra* is expressed (upregulating itself and other sex determination genes such as *tra-2* (Salvemini et al. 2009)) which recruits the necessary splicing machinery to retain exon 4. This results in *dsxf* which drives female development. Therefore, both *tra* and *dsx* represent promising genes for targeting to manipulate sexual development for pest control strategies.

Given their difference in mechanism of action, *dsx* and *tra* must be targeted in different ways to achieve the same goal (producing only males which are fertile). A knockout of *tra* would be necessary to force male development as this is repressed in male development and serves no known function in male sexual development. *Dsx* cannot be knocked out completely, as

both males and females utilise *dsx* for sex determination. The knocking out of *dsxf* is one option. However, this is not predicted to result in sex conversion to male. The complete absence of *dsxf* would be lethal or produce only sterile intersex individuals (Kyrou et al. 2018). If this was deployed for the development of insect control, only male offspring would be fertile. However, fitness costs are incurred if 50% of offspring are non-viable or infertile. This would impose a significant fitness cost to carrying the transgene, which is not desirable in a control context as it rapidly selects for resistance. The alternative option is to prevent female specific splicing. This would allow *dsxm* to be formed as this is the default splicing pattern in the absence of signals from additional regulatory elements. In principle, targeting the splice site and splicing regulatory regions of exon 4 could be used to prevent female specific splicing by preventing the retention of exon 4, forcing a male specific splicing pattern.

Though manipulations of *dsx* could be used for sex conversion, it is generally thought that the targeting of *tra* is a safer and more promising target option. *Tra* interference has already been shown to induce masculinisation and produce fertile pseudomales (Salvemini et al. 2009; Meccariello et al. 2023; Primo et al. 2020). In contrast, providing *dsxm* has been shown to result in only partial masculinisation (Saccone et al. 2008). Targeting upstream elements of the sex determination pathway is more likely to result in full conversion if dosage compensation or early sex specific gene expression has an effect on sexual development. It is difficult to determine if the lack of full conversion seen when providing *dsxm* was due to competition with the native *dsxf* still being produced or because the upstream elements have more of a role in sex determination than inducing *dsx* splicing. Although the targeting of *tra* is more likely to be effective at sex conversion in *Ceratitis,* the targeting of *dsx* has more cross species applicability if effective. *Dsx* alternate splicing is utilised by a wide variety of pest species of interest and although the splicing patterns can differ, if the targeting of splice sites is successful, this will have a wider portability than for *tra*, for which homologues are missing in some species. If only partial sex conversion occurred, this would still have utility, as the ability to ensure no fertile females are produced would still be useful in applied genetic pest control methods (Labbé et al. 2012).

The work outlined in this chapter aimed to further understand how *dsx* and *tra* could be effectively targeted for the generation of pseudomales in medfly. Multiplexed guides were designed as if these would both allow the testing of multiple guides efficacies and would be more applicable to usage in gene drive constructs. A *tra* guide line was established and subsequently crossed with Cas9 lines to determine if this guide was effective. Two established

tra lines were crossed to four Cas9 lines with the sex of offspring of the first generation measured. When no bias was seen these offspring were then screened for individuals with both transgenes and these were inbred. Sex conversion was not seen in the subsequent offspring. This work opened up several potential avenues which are currently being explored.

5.3 Methods

5.3.1 Fly Husbandry

5.3.1.1 Medfly stock lines

Three wildtype medfly stock lines were used throughout these experiments. 'Cepa Petapa' is a mass-reared wildtype strain originally collected from the wild in Guatemala, Central America (Rendon, 1996). TOLIMAN is a wildtype strain from Guatemala, Central America originally collected in 1990. Both strains have been reared at Oxitec LTD (Milton Park, Abingdon) from 2004 and a sub-culture of each was established at the University of East Anglia (UEA) in 2010. The wild-type Benakeion strain (originally isolated in Athens, Greece) was provided by A. Meccariello (Imperial College, London) and maintained at UEA from 2022.

5.3.1.2 Medfly rearing

All wild type medfly cultures were reared in a controlled environment with temperature set at 25°C ±0.5℃, humidity at 60% ± 10% RH and on a 12h light/dark schedule. When development needed to be slowed, temperature was lowered to 22°C ±0.5℃. Depending upon the demand for eggs from cultures, adult populations were kept in 1 of 3 different cage sizes: small cages (10cm x 10cm x 10cm) initiated with 50 pupae, medium cages (13cm x 13cm x 14cm) with 100 pupae and large cages (33cm x 33cm x 16cm) with 500 pupae. The emerging adult flies held in these cages were fed on an *ad libitum* diet of 3:1 sucrose:hydrolysed yeast mixture, changed every seven days. Water was supplied through the side of each cage via dental rolls soaked in RO water, which were renewed every 4 days.

All cages were designed with a mesh layer covering the majority of the surface area of one vertical wall of the cage, to allow egg deposition. Water pots containing RO water were placed under the mesh side of each cage to collect eggs. Eggs laid through the cage mesh were collected after a period of no more than 24h of egg laying, by removing eggs with a Pasteur pipette and transferring up to two drops of the eggs contained in the water solution to filter paper. The filter papers with eggs were then placed into 1/3 pint milk bottles containing 100ml of ASG larval food (1L ASG food: 850ml RO water, 12.5g agar, 73.5g sucrose, 67g maize, 47.5g Brewer's Yeast, 2ml Propionic acid, 25ml Nipagin). 7 days post egg collection, bottles were prepared to allow the third instar larvae to exit and start pupation. To do this, 2 filter paper folded strips were placed in each bottle before laying the bottles down on a thin layer

of sand within a pupation box (170mm x 130mm x 50mm). The box was then sealed with a lid containing a mesh for a further 7-9 days. During this time third instar 'jumping' larvae exited the bottles and pupated within the sand. After this period, pupae were sieved from the sand through a standard metal sieve and transferred into petri dishes. From this the appropriate number of pupae required for each procedure were randomly selected and placed into a fresh cage or used in experiments.

5.3.2   DNA Extraction

DNA extraction was performed on whole flies of both TOLIMAN and Cepa Petapa wildtype strains. Whole flies were frozen at -20°C for a minimum of 48 hours with 3 individuals per sample. Samples were removed from freezer to be placed onto dry ice. Samples were homogenised with a sterilised, plastic pestle over dry ice. Qiagen DNeasy® Blood & Tissue Kit was used as per the manufacturer's instructions with the addition of an extra elution step with 25µl molecular grade water after the original 50µl elution. Concentration of the sample was measured using a Thermo Fisher Scientific Nanodrop™ 8000 Spectrophotometer.

5.3.3   PCR amplification for sequencing

Genomic regions for targeting were checked for sequence divergence through sequence confirmation, as follows. New England BioLabs® Q5® High fidelity 2X Master Mix was used to amplify target regions with genomic DNA used as template. Primers 128 and 129 were used for the amplification of *tra*. Primers 130 and 131 were used for the amplification of *dsx*. Primer sequences can be found in *Supplementary information 5.6.1*. Both reactions used the following cycling conditions 30s at 98°C, 28 cycles of 10s at 98°C, 20s at 68°C, 30s at 72°C with final extension of 2 minutes at 72°C. Products were visualised with gel electrophoresis and purified through PCR purification. Final concentration was measured using a Thermo fisher Scientific Nanodrop™ 8000 Spectrophotometer and sent for Sanger sequencing with Eurofins Genomics TubeSeq Sequencing after dilution to 15ng/µl with sequencing primers (same as those used for amplification).

5.3.4   General Molecular Techniques

5.3.4.1   Gel electrophoresis

Products to be visualised by gel electrophoresis that did not have any loading dye in PCR mastermix were combined with New England BioLabs® Purple Gel Loading Dye (6X) prior to

loading. Agarose gels were made at 1.2% (w/v) with 1X TBE buffer and stained with 3μl Ethidium Bromide per 100ml gel. Gels were run with appropriate ladders in each case to determine band sizes. The ladders used were Thermo Scientific™ GeneRuler™ 1kb Plus DNA Ladder and Thermo Scientific™ GeneRuler™ 1kb DNA Ladder. Gels were run at 90-120V for 60-90 minutes depending on product size.

5.3.4.2 PCR Purification

The Thermo Scientific™ GeneJET PCR Purification Kit was used for the purification of PCR products, as per the manufacturer's instructions, with final elution in 25μl molecular grade water after a 1 minute incubation at room temperature.

5.3.5 Construct Building

5.3.5.1 Selection of guides

CHOPCHOP software (Labun et al. 2019) using the medfly genome assembly Ccap_2.1 (Poelchau et al. 2015; Spanos et al. 2000) was used to design *tra* guides. A further *tra* guide was provided by Meccariello (Meccariello et al. 2023). *Doublesex* guides were designed manually to target the splice site, regulatory elements and purine rich element previously identified (Saccone et al. 2008). All guides were checked for off-target effects with BLAST (Madden 2002).

5.3.5.2 Construct Design

All constructs were designed in Benchling using the Gibson assembly tool. Primers were designed manually or generated by Benchling (Benchling 2023). *Tra* guide cassette and flanking elements were ordered in a pUC backbone from GENEWIZ® from Azenta Life Sciences. Ordered parts can be found in *Supplementary information 5.6.2*.

5.3.5.3 Restriction Digestion

Restriction digests were carried out on both insert and backbone plasmids with New England Biolabs® *BstEII* with a 60 minute incubation at 37°C. To prevent self-ligation restriction digests were then incubated with New England Biolabs® Shrimp Alkaline Phosphatase at 37°C for 30 minutes and heat inactivated at 65°C for 15 minutes. New England Biolabs® T4 DNA Ligase was used for annealing of digested products with a 2:1 insert:backbone ratio. Incubation was

completed at room temperature for a minimum of 60 minutes with heat inactivation at 65°C for 15 minutes.

5.3.5.4 Gibson Assembly

5.3.5.4.1 Primer Design

All primers were designed to create a 20bp overlap region between parts to allow for assembly. All primers were ordered from Integrated DNA Technologies and are shown in *Supplementary information 5.6.1.*

5.3.5.4.2 Generating Gibson Assembly parts

Individual parts for the Gibson assemblies were generated via PCR with overhanging primers, using New England BioLabs® Q5® High fidelity 2X Master Mix to ensure sequence accuracy. Primer annealing temperatures were calculated using New England BioLabs® Tm Calculator and cycling conditions as outlined in manufacturer's instructions were used. Primers 319 and 324 were used for the amplification of the *tra* insert (cycling conditions; initial denaturation of 30s at 98°C, 28 cycles of 10s at 98°C, 20s at 66°C, 1 minute and 40s at 72°C with a final extension of 2 minutes at 72°C). Primers 320 and 323 were used for the amplification of backbone from line UEA_005 (cycling conditions; initial denaturation of 30s at 98°C, 28 cycles of 10s at 98°C, 20s at 65°C, 1 minute and 40s at 72°C with a final extension of 2 minutes at 72°C). Plasmid DNA was used as templates for all PCR reactions. All plasmids used can be found in *Supplementary information 5.6.2*. PCR products were purified and their concentration measured using the Nanodrop™ 8000 Spectrophotometer.

5.3.5.4.3 Touchdown PCR

Touchdown PCR was used for Gibson assembly part generation. This uses the standard Q5 cycling conditions outlined above with the modification of the annealing temperature beginning at 70°C and decreasing by 1°C each subsequent cycle until 60°C is reached.

5.3.5.4.4 Gibson Assembly Reaction

An in house Gibson Mastermix was used for the assemblies of *tra* guide plasmid following the method outlined in (Behle and Sakaguchi 2020). After incubation, Gibson reactions were diluted in a ratio of 1:5 and used for transformation into *E. coli* competent cells.

5.3.5.5 Transformation of *E. coli* competent cells

Three competent cell types were used and the transformation protocols are summarised below.

Table 5.1 Transformation protocols of Agilent SURE® 2 Supercompetent cells used for construct assembly.

Input cells	Agilent SURE® 2 Supercompetent Cells
Volume used	5µl diluted Gibson : 50µl cells
Ice incubation	30 mins
Heat shock	30 s at 42°C
Ice incubation	2 mins
Volume of SOC added	700µl
Incubation	1 hr at 37°C, 100rpm
Spin step	2 mins at 4000rpm, pellet resuspended in 100 µl SOC
Final plated volume onto LB Ampicillin (100µg/ml)	100 µl
Plate incubation	Overnight at 37°C

5.3.5.6 Colony PCR

Individual colonies were screened for inserts after Gibson assembly. Colonies of varying sizes were selected that did not overlap others. A 10µl pipette tip was used to remove the edge of a colony and then resuspended in 10µl of molecular grade water. 1µl of this colony mix was then used as the template in a Thermo Scientific™ DreamTaq Green PCR Master Mix (2X) reaction with a total volume of 20µl. Primers used were designed to cover insert boundaries where possible to confirm presence of both the insert and the backbone. Primers 158 and 163 (*Supplementary information 5.6.1*) used for all reactions with the following cycling

conditions 2 minutes at 95°C, 28 cycles of 30s at 95°C, 30s at 54°C, 3 minutes at 72°C with final extension of 10 minutes at 72°C.

5.3.5.7 Plasmid Miniprep

Positive clones were miniprepped to prepare them to be sent for sequencing. Overnight cultures were inoculated with a single positive colony into 4ml LB Ampicillin (100μg/ml) contained within a 15ml falcon tube to allow sufficient aeration. Cultures were incubated overnight at 37°C with constant agitation at 180rpm. After confirming growth, 2ml of the culture was prepared by using the Thermo Scientific™ GeneJET Plasmid Miniprep Kit, as per the manufacturer's instructions. The final elution step was performed with 30μl molecular grade water. The purity and concentration of all samples were checked with Nanodrop™ before being diluted to the correct concentration for sequencing.

5.3.5.8 Plasmid Sequencing

Nanopore based long-read sequencing was completed by diluting miniprepped plasmids to 30ng/μl and sending to SNPsaurus© Plasmidsaurus Whole Plasmid Sequencing Service. Sequence alignment of returned reads was performed using the Benchling sequence alignment tool (Benchling 2023).

5.3.5.9 Plasmid Midiprep

After sequence confirmation, plasmids were prepared for microinjection using the Macherey-Nagel NucleoBond® Xtra Midi EF, Midi kit for endotoxin-free plasmid DNA kit. Cultures were revived from glycerol stocks onto a streak plate which was incubated overnight at 37°C. A single culture from this was used for inoculation of a 4ml LB Ampicillin (100μg/ml) starter culture and incubated for 6-8 hours at 37°C with constant agitation at 180rpm. After confirming growth in starter cultures, 400μl was used to inoculate 100ml LB Ampicillin (100μg/ml) to be incubated overnight at 37°C with constant agitation at 200rpm. This culture was then used as per the manufacturer's instructions for midiprep. The final pellet was resuspended in 50μl endotoxin-free molecular grade water.

5.3.5.10 Preparing glycerol stocks

For the preparation of glycerol stocks for long term plasmid storage, single colonies were selected to inoculate 4ml LB Ampicillin (100μg/ml). This was then incubated overnight at

37°C with constant agitation at 180rpm (this was the culture that was also used for plasmid miniprep). From this, 500µl of culture was added to 500µl 50% glycerol and mixed well by inversion. Glycerol stocks were then stored at -80°C, and only the top layer was thawed when reviving them for use in subsequent experiments.

5.3.6 Transgenic Line Establishment

5.3.6.1 Microinjection of medfly eggs to generate transgenic strains

Medfly microinjection was performed at the Imperial College London insectary by A. Meccariello. A further purification step was performed on midiprep samples for microinjection using Millipore Ultrafree®-MC GV Centrifugal Filters. The full midiprep sample was loaded to the column and centrifuged at 14,000 RPM for 4 minutes. 40µl of injection mix was used per microinjection session with a final concentration of 500ng/µl plasmid of interest, 200ng/µl *hyPBase* transposase helper plasmid (Eckermann et al. 2018). This was made up to a final volume with injection buffer (Rubin and Spradling 1982). Embryos were collected after an egg laying period of 45 mins, transferred to double sided sticky tape and allowed to dry for 6 mins. Eggs were manually dechorionated using the tip of a blunted needle. Once the chorion was removed, eggs were transferred to slides coated in a thin layer of glue (double sided sticky tape dissolved in heptane) with the posterior of the egg facing outwards. These egg slides were desiccated by placing them in a petri dish containing calcium chloride for 6 minutes, and then covered with a thin layer of Halocarbon™ 700 Oil. Eggs on these slides were then injected using WPI Sutter Quartz needles (outer diameter (OD) = 1.00mm, inner diameter (ID) = 0.70mm) that had been drawn on a Sutter P-200/G laser needle puller. Needles were loaded with injection mix using an Eppendorf Microloader. Microinjections were done using an inverted stereo microscope with an Eppendorf microinjection set-up and compressed air-driven pump.

Post injection, eggs were left on slides until the larvae hatched (typically after 48 hours) at which point the larvae were placed onto larval food (30g paper, 30g sucrose, 30g yeast extract, 10ml cholesterol stock (5g cholesterol, 140 ml distilled water, 30ml 95% EtOH), 2ml HCl Stock (384ml distilled water, 66ml 37% HCl), 8.5ml Benzoic Stock (50g Benzoic acid, 300ml 95% EtOH, 150ml distilled water) made up to 400ml with distilled water). After 7 days larval culture bottles were laid down in pupation boxes with a thin layer of fine sand and allowed to pupate. Surviving pupae were then sent to UEA within three sealed layers.

5.3.6.2   Fluorescence screening for medfly transgenics

Transgenic lines were established and maintained by screening flies for fluorophore marker expression at each generation. When establishing transgenics, live adults were screened on the day of eclosion from the pupae. Flies were anaesthetised over ice for screening. A Leica MZ7.5 dissection microscope was used for screening and was fitted with fluorescence filters to visualise DsRed2 (excitation wavelength 558nm, emission wavelength 583nm) or AmCyan (excitation wavelength 458nm, emission wavelength 485nm) depending on the marker being screened. Established transgenic lines were screened as pupae as the fluorophore expression was clearer at this stage.

5.3.7   Transgenic crosses

Guide carrying lines were crossed to four of the previously established Cas9 lines (M68.3, M68.4, bgcn.2A and bgcn.2C Chapter 2) to two of the newly established guide lines (010.1 and 010.2). Crosses were set up using 2 males and 5 females per cage screened and positive for their respective transgenic marker. Cas9 lines were enriched over several generations and therefore mostly homozygous for the insertion. All crosses were set up in small cages as outlined in *Table 5.2*.

Table 5.2 Crosses set up between Cas9 lines and guide lines. In case of maternal deposition all crosses were set up reciprocally so the Cas9 provider line could be both male and female. Those highlighted in grey were the cages for line continuation, not crosses between Cas9 and guide.

Cage Number	Line of male flies	Line of female flies
1 (010.1♂ x M68.3♀)	010.1	M68.3
2 (M68.3♂ x 010.1♀)	M68.3	010.1
10 (010.2♂ x M68.3♀)	010.2	M68.3
11 (M68.3♂ x 010.2♀)	M68.3	010.2
3 (010.1♂ x M68.4♀)	010.1	M68.4
4 (M68.4♂ x 010.1♀)	M68.4	010.1
12 (010.2♂ x M68.4♀)	010.2	M68.4
13 (M68.4♂ x 010.2♀)	M68.4	010.2
5 (010.1♂ x bgcn.2A♀)	010.1	bgcn.2A
6 (bgcn.2A♂ x 010.1♀)	bgcn.2A	010.1
14 (010.2♂ x bgcn.2A♀)	010.2	bgcn.2A
15 (bgcn.2A♂ x 010.2♀)	bgcn.2A	010.2
7 (010.1♂ x bgcn.2C♀)	010.1	bgcn.2C
8 (bgcn.2C♂ x 010.1♀)	bgcn.2C	010.1
16 (010.2♂ x bgcn.2C♀)	010.2	bgcn.2C
17 (bgcn.2C♂ x 010.2♀)	bgcn.2C	010.2
9 (010.1♂ x 010.1♀)	010.1	010.1
18 (010.2♂ x 010.2♀)	010.2	010.2

Three egg collections were taken from each cross, except cage number one which was lost due to a technical issue. Offspring pupae were screened for dsRED and mCyan fluorescence to select individuals for setting up inbreeding cages. Pupae were screened within a day of pupation to enable the detection of mCyan. Due to the high amount of crosses and low pupal numbers, some lines did not have sufficient positive screened pupae to set up inbreeding cages. A minimum of 2 males and 3 females was needed to set up a small inbreeding cage with the cage numbers shown in Table 5.3.

Table 5.3 Inbreeding crosses set up. Cage number reflects the original F1 cross the offspring are from. The number of males was kept equal to or less than the number of females to avoid female harassment. Cages required a minimum of 2 males and 3 females to be set up and produce viable egg numbers.

Cage Number	Number of Males	Number of Females
1 (010.1♂ x M68.3♀)	3	6
2 (M68.3♂ x 010.1♀)	3	9
10 (010.2♂ x M68.3♀)	12	12
11 (M68.3♂ x 010.2♀)	10	10
3 (010.1♂ x M68.4♀)	2	11
5 (010.1♂ x bgcn.2A♀)	5	9
14 (010.2♂ x bgcn.2A♀)	8	8
15 (bgcn.2A♂ x 010.2♀)	3	4
7 (010.1♂ x bgcn.2C♀)	2	3
16 (010.2♂ x bgcn.2C♀)	5	4
17 (bgcn.2C♂ x 010.2♀)	2	4

The sex of all eclosed offspring was monitored to check for male bias. When none was observed, the generation F2 was similarly checked. A minimum of one egg collection was taken from each cage with three egg collections taken where possible. Sex of all the eclosed offspring was recorded with pupal dishes briefly placed on ice to anesthetise live flies for their removal and sex scoring.

5.4   Results

5.4.1   Design of guides targeting *doublesex* and *transformer*

Doublesex (*dsx*) and *transformer* (*tra*) have different mechanisms of action in sex determination and therefore the guides targeting these genes had specific designs. The presence of *tra* induces the cascade for female development whereas in males *tra* is not active. This means a knockout of *tra* is desired here to drive male development. Guides could therefore be designed with CHOPCHOP to induce a knockout. The top three ranked guides were chosen and after screening of the guides with a 75% match in BLAST to check for off target effects and finding none, these were deemed usable. As *dsx* is used in both the male and female developmental pathway a knockout would not be suitable here. Differential splicing of *dsx* determines male or female development with exon 4 being retained in female *dsx* and not male *dsx*. The aim of these guides was to disrupt the splice site which results in the retention of exon 4. These were designed manually based upon the predicted splice site, regulatory regions and purine rich elements outlined by Saccone (Saccone et al. 2008). Disruption of this splice site should force into the male splicing pattern which is the default without the recruitment of additional splicing elements. Five *dsx* guides were designed in total, one of which targeted the splice site and four of which targeted regulatory elements including the purine rich elements. All designed guides can be found in *Supplementary information 5.6.2*.

The region in which these potential guides fell were sequenced to check for any single nucleotide polymorphisms that could affect guide efficacy. Both genes were expected to be relatively well conserved due to their essential function. Six individuals were used for gDNA extraction, the samples and concentrations obtained are outlined in *Table 5.4*.

Table 5.4 Genomic DNA extractions and the concentration obtained from each sample.

Sample	Concentration (ng/µl)
TOLIMAN male sample 1 (TM1)	352.0
TOLIMAN male sample 2 (TM2)	240.3
TOLIMAN female sample (TF)	310.6
Cepa petapa female sample 1 (CF1)	640.2
Cepa petapa female sample 2 (CF2)	748.1
Cepa petapa male sample (CM)	293.8

PCR to amplify regions in which guides were designed to bind was completed on all samples. Products were visualised through gel electrophoresis for size confirmation shown in *Figure 5.2*.
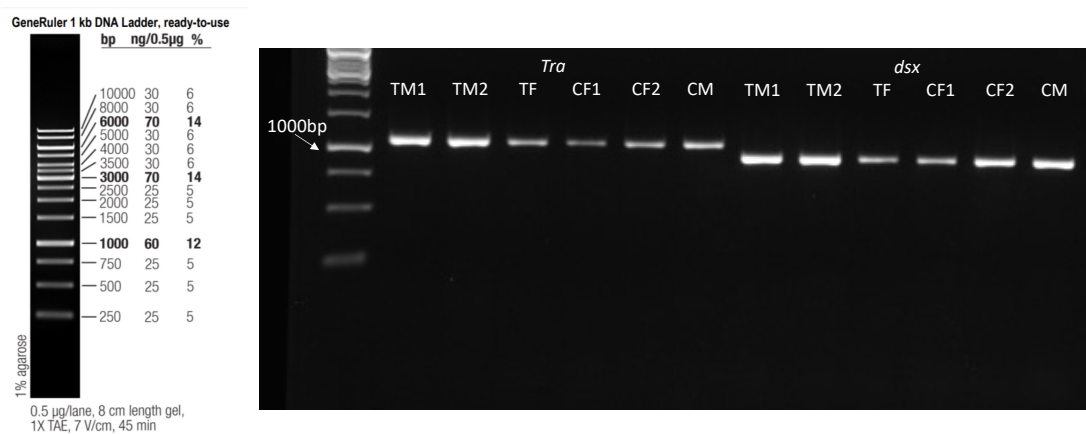
Figure 5.2 PCR products for the sequencing of *dsx* and *tra*. Samples were run on a 1.2% agarose gel at 110V for 60 minutes alongside the Thermo Scientific GeneRuler 1kb. Order of loading : *Tra* (TOLIMAN male 1, TOLIMAN male 2, TOLIMAN female, Cepa petapa female 1, Cepa petapa female 2, Cepa petapa male), *dsx* (TOLIMAN male 1, TOLIMAN male 2, TOLIMAN female, Cepa petapa female 1, Cepa petapa female 2, Cepa petapa male). All bands were of expected size and all samples carried through to PCR purification.

All products were purified through PCR purification with the concentrations obtained shown in *Table 5.5*.

Table 5.5 Concentrations of purified PCR products amplifying *tra* and *dsx* to be used for sequencing.

Sample	Concentration (ng/µl)
Tra TOLIMAN male 1 (T1)	79.37
Tra TOLIMAN male 2 (T2)	64.30
Tra TOLIMAN female (T3)	31.19
Tra Cepa petapa female 1 (T4)	17.91
Tra Cepa petapa female 2 (T5)	31.65
Tra Cepa petapa male (T6)	27.66
dsx TOLIMAN male 1 (D1)	99.76
dsx TOLIMAN male 2 (D2)	68.36
dsx TOLIMAN female (D3)	29.32
dsx Cepa petapa female 1 (D4)	41.10
dsx Cepa petapa female 2 (D5)	117.4
dsx Cepa petapa male (D6)	39.05

All samples were sent for short read Sanger sequencing and the results were aligned to the respective genomic region on Benchling (Benchling 2023). Aligned sequences are available in *Supplementary information 5.6.2*. Several mutations were present in the two sequenced regions of approximately 500 base pairs. Mutations were only of interest when they fell within a region which a guide had been designed to target. Two *dsx* guides (*dsx*_RE3 and *dsx*_RE4) had a single nucleotide polymorphism present in all sequenced individuals (*dsx*_RE3 C>T and *dsx*_RE4 A>G). As five guides in total had been designed for *dsx,* these two guides could be excluded from the final guide cassette as there were three other suitable guides. *Tra* had a single SNP in the third ranked guide created by CHOPCHOP (Labun et al. 2019) (Tra_Rank_3). This was a T>A that was present in all sequenced samples. This would have presented an issue as only three guides had been designed for *tra* however our colleague Dr A Meccariello (Imperial) had been working on *tra* knockouts and kindly provided a guide sequence that had been tested and shown to result in sex conversion (Meccariello et al. 2023). This guide was used in addition to the two designed (TRA_Rank_1 and TRA_RANK_3) for a multiplex cassette.

5.4.2   Guide cassette design

Two guide cassettes were designed with a multiplexed guide system of three guides to reduce the risk of resistance allele formation. Each consists of three guides preceded by a native U6 promotor (Meccariello et al. 2021) followed by a gRNA scaffold amplified from AGG1728 (*Supplementary information 5.6.2*) (plasmid designed at Pirbright Institute Arthropods Genetic Group) and terminated by a repetitive Pol3 sequence. These guide cassette designs

can be found in supp info. For assembly of these guide cassettes, a single element which contained the scaffold-pol3-U6 was to be ordered with overhang PCR used to insert the different guides to reduce costs of ordering two large, repetitive parts. This part can be found in *Supplementary information 5.6.2*. Due to time constraints, the decision was made to focus on the assembly of the *tra* construct as there was not enough time to complete both. To further save on time the entire *tra* guide cassette was ordered as a single part with the mCyan marker elements from AGG1041 (*Supplementary information 5.6.2*) (plasmid designed at Pirbright Institute Arthropods Genetic Group) was ordered with flanking regions complementary to previously constructed plasmids. This was ordered in a pUC backbone from Genewiz® by Azenta Life Sciences which could be digested with the restriction enzyme *BstEII* (UEA_006 *Supplementary information 5.6.2*).

5.4.3   Synthesising the *tra* guide plasmid

The *tra* part ordered was designed to be inserted into a modified Scolari backbone (Scolari et al. 2008) with digestion of both backbone and part with *BstEII*. Restriction digests were successful: however, no positive colonies were recovered. Several colonies were present on transformant plates, but screening of a random selection of these showed them to be recombined plasmids with only the backbone plasmid part present, often repeated. Due to this the rSAP digestion was repeated for a longer time period and a high volume of insert to backbone was used however this did not resolve the issue. Due to this, primers were designed to assemble the plasmid with Gibson Assembly.

The original Gibson Assembly design was to insert the *tra* part into the same plasmid which was being used for restriction digest. However, this was unsuccessful due to difficulty in obtaining a high concentration of the backbone part as both parts had to be gel cut due to issues with multiple banding. *Gypsy* elements had caused issues with multiple banding and recombination in other plasmid constructions, so the plasmid was once again redesigned for insertion into a previously constructed Cas9 plasmid (UEA_005 *Supplementary information 5.6.2*). The final plasmid design shown in *Tra* cassette in PuB backbone can be found in *Supplementary information 5.6.2*. Primers 319 and 324 (*Supplementary information 5.6.1)* were used for the amplification of the *tra* insert with primers 320 and 323 (*Supplementary information 5.6.1)* used for the amplification of backbone from UEA_005. PCR products visualised through gel electrophoresis and shown in *Figure 5.3*.
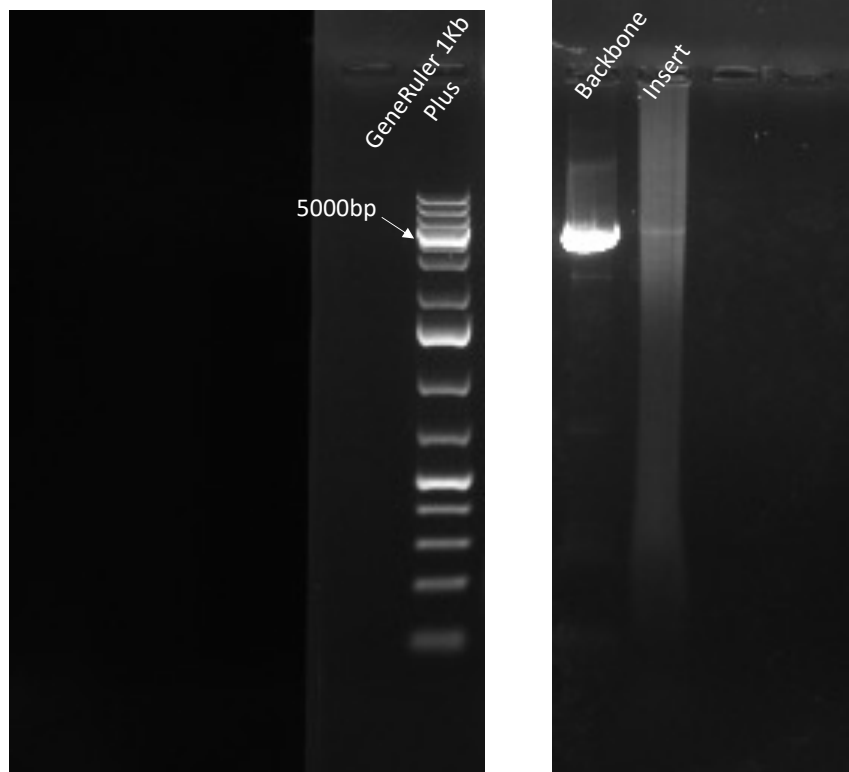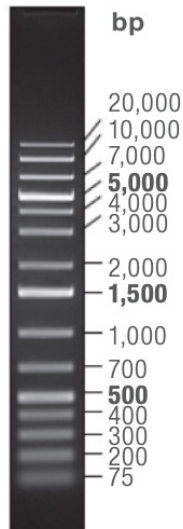
Figure 5.3 Gel electrophoresis of PCR products to amplify backbone and insert parts for *tra* plasmid assembly. Samples were run on a 1.2% agarose gel at 110V for 60 minutes alongside the ThermoScientific GeneRuler 1kb Plus. Gel photographed at slight angle so gel wells have been manually aligned upon editing. Order of loading: Ladder, *tra* backbone, *tra* insert. Both bands of expected size with *tra* backbone expected size 4916bp and *tra* insert expected size 4047bp. Insert product band is very weak and may not produce high enough concentration post purification.

PCR products were purified with their concentration measured with Nanodrop and shown in *Table 5.6*.

Table 5.6 Concentration of PCR products to amplify parts for the assembly of *tra* plasmid.

PCR Product	Concentration (ng/µl)	260/280	260/230
Tra backbone	172.7	1.75	1.24
Tra insert	116.2	1.73	0.97

The insert PCR product was of higher concentration than anticipated so the decision was made to proceed with Gibson assembly and transformation. There were over 50 colonies present on the subsequent transformation plate however colony PCR indicated none of these were successfully assembled as no bands were seen. It was assumed that the insert part was the issue due to the weakness of this band. Amplification of this part was repeated through

two modified PCR methods. The first modification was the use of nested PCR using the purified PCR product as the PCR template with a lower annealing temperature of 60°C. Concurrently, touchdown PCR was also used on both the original template and the PCR product template. All four PCRs were also run for 35 cycles as opposed to the original 28. The resulting PCR products were visualised through gel electrophoresis and shown in *Figure 5.4*.
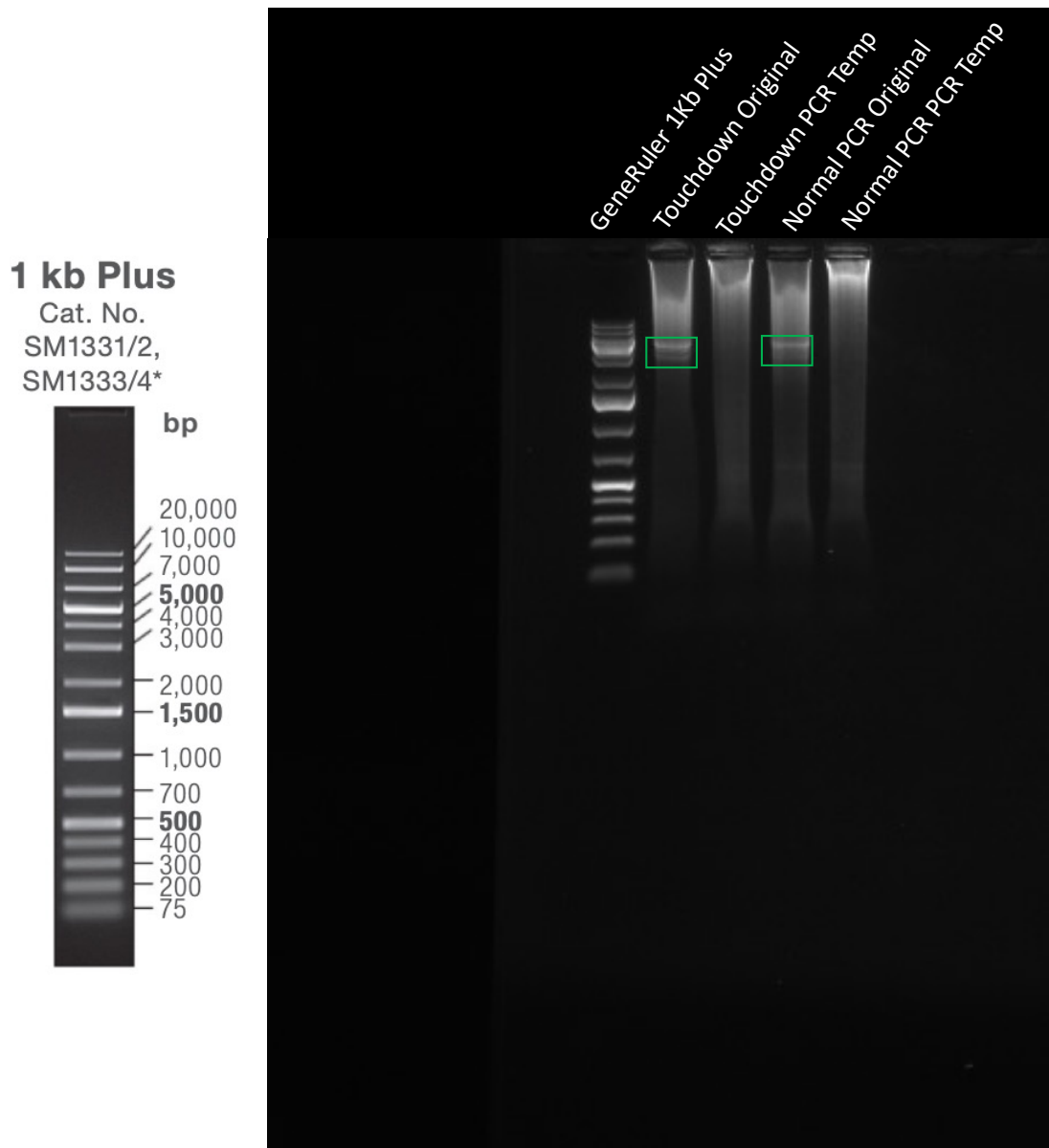
Figure 5.4 Gel electrophoresis of PCR products to amplify insert parts for *tra* plasmid assembly. Samples were run on a 1.2% agarose gel at 110V for 60 minutes alongside the Thermo Scientific GeneRuler 1kb Plus. Order of loading: Ladder, touchdown PCR on plasmid template, touchdown PCR on PCR product template, normal PCR with extended cycles on plasmid template, normal PCR with extended cycles on PCR template. Correct sized band of 4074bp is indicated in green, a slightly smaller band is very faintly present in both samples however obtaining a high enough concentration from gel cutting would be very difficult with this product and the bands are very close together so this was not attempted.

The touchdown PCR on plasmid template showed the strongest band which was used for subsequent PCR purification. There was a slightly smaller band present in both samples where the PCR had been successful. However, due to difficulties in obtaining high concentrations from gel extractions and the difficulty in amplifying this product, I decided to proceed with the assembly. PCR purification yielded a concentration of 180.7ng/μl with a 260/280 of 1.78 and a 260/230 of 1.54 which is a much cleaner sample than originally obtained.

This newly obtained insert was used for Gibson assembly and subsequent transformation. Approximately 30 colonies were present on the transformant plate of which 20 were randomly selected for colony PCR, the results of which can be seen in *Figure 5.5*.
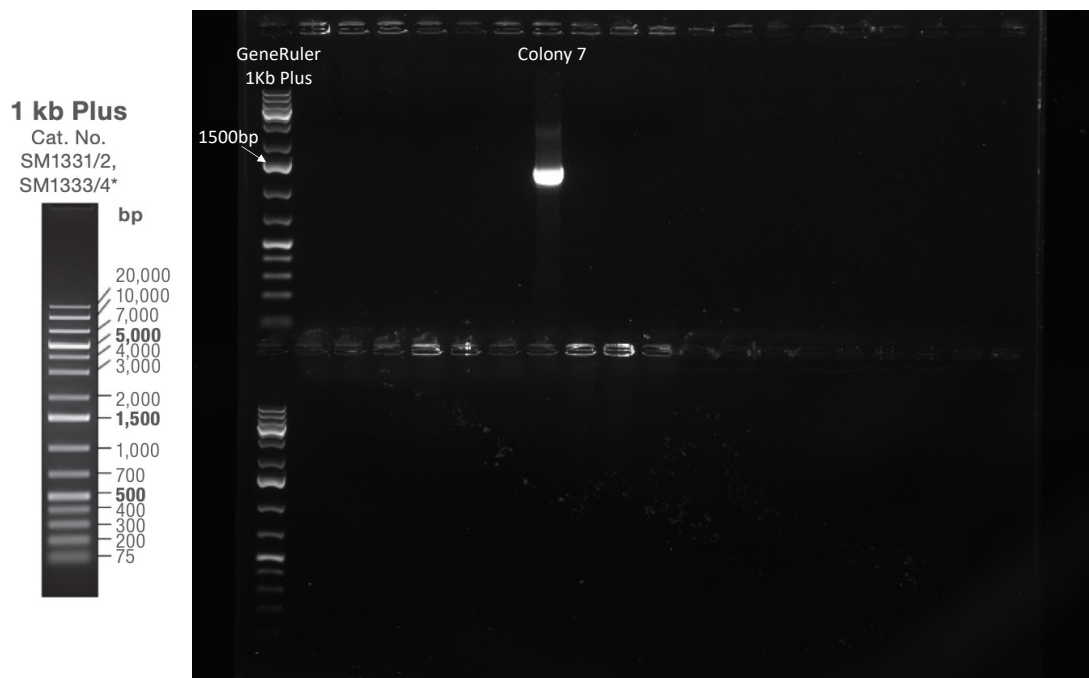
Figure 5.5 Gel electrophoresis of colony PCR products of 20 potential *tra* plasmids. Samples were run on a 1.2% agarose gel at 110V for 60 minutes alongside the ThermoScientific GeneRuler 1kb Plus. Only one colony PCR produced a result, that seen in lane 8 representing colony 7. This was smaller than the expected colony PCR product size of 2364bp with the band being between 1000bp and 1500bp. As this was the only colony to produce a result it was selected for sequencing to check assembly.

Only one colony produced a result in the colony PCR, namely colony 7. This band was smaller than the expected positive colony PCR result of 2364bp with its size being approximated between 1000bp and 1500bp. As this was the only colony to produce a band of the 30 tested, it proceeded to the sequencing step. Colony 7 was miniprepped and sent for sequencing with two other colonies which had no positive bands. Sequencing results are available in *Supplementary information 5.6.2*.

The sequencing results showed that the two colonies with no bands did not contain the insert. Colony 7 sequencing results showed that the insert part was present. However, recombination had occurred between the first and last U6 resulting in the loss of two of the guides in the cassette. Only the third guide had been retained (TRA_RANK_2). Due to time constraints the decision was made to proceed with this construct as recombination between these repetitive elements would be difficult to overcome. This construct was assigned the designation UEA_010 and can be viewed in *Figure 5.6.*
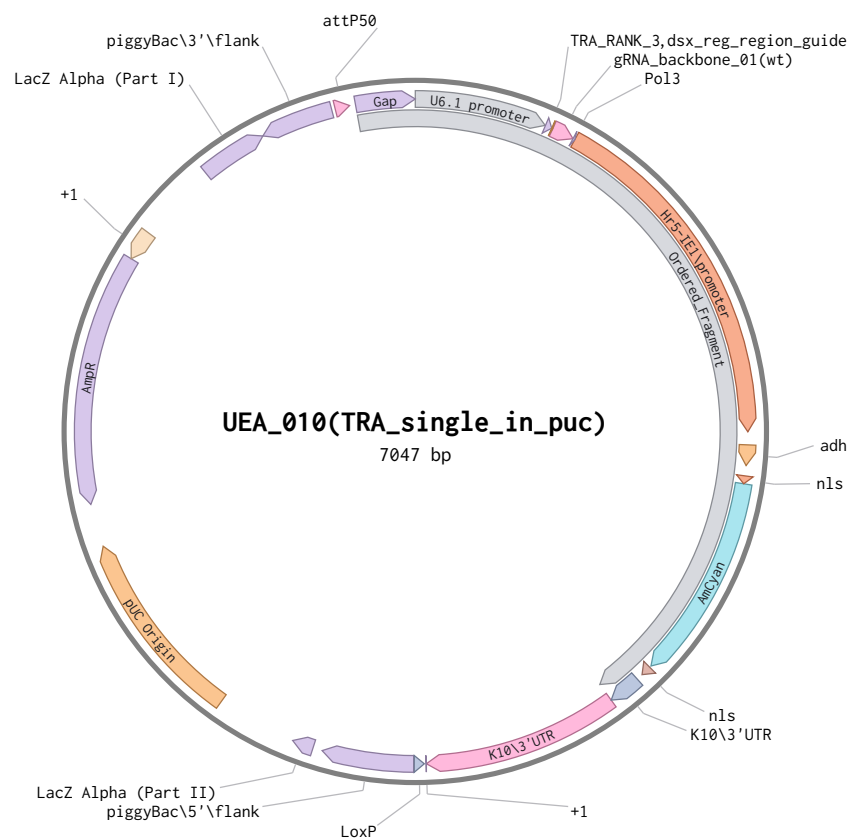
Figure 5.6 Final design of UEA_010 designed and visualised in Benchling.

5.4.4   Establishment of *tra* guide transgenic line

The plasmid was midiprepped achieving a concentration of 2744ng/µl with a 260/280 of 0.67 and a 260/230 of 0.74. This sample was sent for microinjection (by A Meccariello, Imperial College London). 63 pupae were obtained from microinjections, of these 48 eclosed and the following cages shown in *Table 5.7* were set up from microinjection survivors.

Table 5.7 Cage set up of microinjections survivors. Cages with female survivors were provided with 13 male wildtypes to avoid female harassment. Male survivors were provided with a surplus of female wildtypes.

Cage Designation	Number of Males	Number of Females
010.1	13 Wildtype	14 Injection survivors
010.2	13 Wildtype	14 Injection survivors
010.3	5 Injection survivors	29 Wildtype
010.4	5 Injection survivors	28 Wildtype
010.5	5 Injection survivors	36 Wildtype
010.6	5 Injection survivors	35 Wildtype

Three egg collections were taken from each cage and allowed to develop normally. Eclosed flies were sex separated and screened for fluorescence. Eight transgenics were found, six originated from cage 010.4 (3 male and 3 females) and 2 originated from cage 010.3 (2 females). These were set up as three new cages crossed to wildtypes outlined in *Table 5.8*.

Table 5.8 Crosses set up with the transgenic individuals found.

Cage Designation	Number of Males	Number of Females
010.4A	3 Transgenic	14 Wildtype
010.4B	3 Wildtype	3 Transgenic
010.3	2 Wildtype	2 Transgenic

Egg collections were taken daily over a week period to obtain as many offspring as possible. Fluorescence was visible as early pupae as shown in *Figure 5.7* so all pupae were screened on day of pupation or the following day.

Figure 5.7 GFP fluorescent pupa from offspring of microinjection survivors crossed to wildtype. Spotting patten caused by GFP being under nuclear localisation signals to allow for easy identification.

The number of screened individuals and transgenics recovered from each line is shown in *Table 5.9*.

Table 5.9 Transgenic offspring obtained from crosses. There did appear to be a larger number of non-transgenics than transgenics. However, this was more likely to be due to screening not detecting fluorescence due to pupal case closure and thickening.

Cage Designation	Number of Transgenics	Number of Non-transgenics
010.4A	84	114
010.4B	116	151
010.3	22	32

The ratio of transgenic to non-transgenic was slightly skewed towards non-transgenic individuals. However, this is more likely due to missed fluorescence during screening as the fluorescence was difficult to detect once the pupal casing was fully formed. There were insufficient numbers of 010.3 transgenic individuals to be used for crosses so these were crossed to wildtype to establish a stable line. Inbreeding cages were set up for lines 010.4A and 010.4B by crossing 5 male transgenic flies to 5 female transgenic flies.

5.4.5 Crossing *tra* guide lines with Cas9 lines

Crosses between the 010.4 lines to 4 different previously constructed Cas9 lines were completed as per the method outlined in *5.3.7.* Sex was recorded of adults emerging from all eclosed pupae including those screened to be used for inbreeding crosses, and all unscreened individuals. The sex of adults from eclosed pupae is shown in *Table 5.10*.

Table 5.10 Sex of all eclosed pupae from the crosses between *tra* guide lines and Cas9 lines. Cage 18 indicated in grey is an inbreeding cross to show the sex distribution of a cage in which there was predicted to be no sex bias.

Cage Number	Number of males	Number of females	Male Percentage
1 (010.1♂ x M68.3♀)	41	29	58.6%
2 (M68.3♂ x 010.1♀)	45	51	46.9%
10 (010.2♂ x M68.3♀)	135	135	50.0%
11 (M68.3♂ x 010.2♀)	24	21	53.3%
3 (010.1♂ xM68.4♀)	61	83	43.0%
4 (M68.4♂ x 010.1♀)	1	0	N/A
12 (010.2♂ x M68.4♀)	0	2	N/A
13 (M68.4♂ x 010.2♀)	32	22	59.3%
5 (010.1♂ x bgcn.2A♀)	90	86	51.1%
6 (bgcn.2A♂ x 010.1♀)	0	0	N/A
14 (010.2♂ x bgcn.2A♀)	91	76	54.5%
15 (bgcn.2A♂ x 010.2♀)	47	45	51.1%
7 (010.1♂ x bgcn.2C♀)	2	3	N/A
8 (bgcn.2C♂ x 010.1♀)	41	33	55.4%
16 (010.2♂ x bgcn.2C♀)	116	107	52.0%
17 (bgcn.2C♂ x 010.2♀)	40	34	54.1%
18 (010.2♂ x 010.2♀)	138	128	51.9%

No offspring appeared intersex and none of the lines presented a sufficiently strong male bias to suggest that sex conversion was occurring (the male percentage is expected to be 75% in the case of full conversion). A Paired Welch Two Sample t-test showed a non-significant and small (difference = -2.44, t(15) = -1.13, p = 0.275) between male and female offspring. The lines which did show male bias were those with low offspring numbers, and a similar level of variation towards female bias was seen in other lines suggesting this was indicative of natural

variation. Pupal eclosion rates of the confirmed transgenics to be used for inbreeding cages was also monitored to obtain an estimation of pupal eclosion rates. Of the 196 confirmed transgenic pupae, 138 eclosed (70.4%). This is lower than the eclosion numbers seen in other lines with fitness tests on other lines including transgenics typically yielding >96% eclosion rates. This could be due to temperature fluctuations during pupal developments as the room temperature had to be lowered to 22°C during this experiment or pupae being kept in smaller groups within their petri dishes. It could also be indicative of fitness cost in expressing two transgenic constructs.

As there was no clear indication of sex conversion in the F1 generation, inbreeding cages were set up where possible as per the method outlined in *5.3.7.* Eclosions from these crosses were monitored for sex bias until no further pupae had eclosed for two days. Due to low numbers some cages were not productive and produced minimal egg collections. The number of eclosions from each inbreeding cross and the sex of the offspring is shown in *Table 5.11*.

Table 5.11 Sex of eclosed individuals from inbreeding crosses with the percentage of males shown to determine level of male bias.

Cage Number	Number of Males	Number of Females	Percentage of males
1 (010.1♂ x M68.3♀)	2	8	20.0%
2 (M68.3♂ x 010.1♀)	51	46	47.4%
10 (010.2♂ x M68.3♀)	59	80	42.4%
11 (M68.3♂ x 010.2♀)	149	138	51.9%
3 (010.1♂ x M68.4♀)	106	88	54.6%
5 (010.1♂ x bgcn.2A♀)	38	51	42.7%
14 (010.2♂ x bgcn.2A♀)	95	96	49.7%
15 (bgcn.2A♂ x 010.2♀)	9	7	56.3%
7 (010.1♂ x bgcn.2C♀)	15	15	50.0%
16 (010.2♂ x bgcn.2C♀)	5	9	35.7%
17 (bgcn.2C♂ x 010.2♀)	40	35	53.3%

A male bias was not observed in any of the crosses. A Paired Welch Two Sample t-test showed a non-significant and small (difference = 0.36, t(10) = 0.11, p = 0.913) between male and female offspring. As the Cas9 lines have been shown to be expressing through qPCR, there is

most likely an issue with the guide line. As both guide lines come from the same founder cage the possibility of an insertion effect cannot be ruled out. However, the marker is behaving as expected and there was no indication for inactivation at the insertion site. The other possibility is that the guide was not binding to *tra* or that the cut it is inducing was not sufficient to result in *tra* knockout.

5.4.6   Generating a further *tra* construct

To confirm the Cas9 guides are functional we had the potential to cross them with an additional, existing guide that is known to be functional. The guide provided by Meccariello (Meccariello et al. 2023) has been shown to be functional and was part of the original guide construct. Using overhang PCR the single guide present in the construct can be replaced with the new guide. Primers 329 and 330 (*Supplementary information 5.6.1*) were used on the designed *tra* plasmid (UEA_010) to generate a fragment which removed the current guide and replaced it with the functional guide. Gel electrophoresis used for size confirmation. Gibson assembly was used to circularise the plasmid. PCR product purified through gel purification and concentration measured with nanodrop. Final concentration of 19.32ng/µl with a 260/280 of 1.99 and a 260/230 of 0.78 was quite weak however as it was a simple assembly (only one part), Gibson assembly was used for recircularization. Over 30 colonies were present on the transformation plate and four were randomly chosen for miniprep. The concentrations obtained are shown in *Table 5.12*. All samples were diluted and sent for full plasmid sequencing.

Table 5.12 Concentrations of miniprepped randomly selected colonies from Gibson transformation plate containing potential replaced guide plasmids.

Colony	Concentration (ng/µl)	260/280	260/230
1	340.1	1.86	2.30
2	391.0	1.86	2.26
3	510.9	1.86	2.31
4	150.7	1.87	2.28

Sequencing results were received and aligned to the new plasmid design on Benchling (*Supplementary information 5.6.2*). Colonies 1 and 4 showed successful assembly. Colony 1 was selected and midiprepped to be used for microinjection and given the designation

UEA_011 (*Supplementary information 5.6.2*). After the extra purification step the following concentration of sample was obtained.

Table 5.13 Concentration obtained after midiprep of UEA_011 to be sent for microinjection.

Colony	Concentration (ng/µl)	260/280	260/230
1	3769	1.05	1.22

Concentration was high enough to be used for microinjection and was sent to Imperial.

5.4.7   Establishment of new *tra* guide transgenic line

37 microinjection survivors eclosed, 25 males and 12 females which were subsequently set up in the following crosses *Table 5.14*.

Table 5.14 Microinjection survivors crossed to wildtype with the cage designation of each shown.

Cage Designation	Number of Males	Number of Females
011.1	6 Injection survivors	15 Wildtype
011.2	6 Injection survivors	15 Wildtype
011.3	6 Injection survivors	15 Wildtype
011.4	7 Injection survivors	15 Wildtype
011.5	3 Wildtype	6 Injection survivors
011.6	3 Wildtype	6 Injection survivors

Three egg collections were taken from each cage with the pupae allowed to develop normally. Pupae were screened for fluorescence with the number of transgenic pupae found shown in *Table 5.15*.

Table 5.15 Number of transgenic pupae produced by each cage.

Cage Designation	Number of Transgenic pupae
011.1	0
011.2	6
011.3	4
011.4	26
011.5	2
011.6	8

A high number of transgenic pupae were obtained. However, eclosion numbers were once again very low. This could once again be due to temperature fluctuations as these pupae were

eclosing during a similar time to the inbreeding crosses or fitness cost of expressing the transgenic. The following cages were set up with the pupae which eclosed *Table 5.17*.

Table 5.16 Eclosed transgenic pupae and the number of wildtypes to which they were crossed.

Cage Designation	Number of Males	Number of Females
011.2	2 Injection survivors	5 Wildtype
011.4A	3 Wildtype	2 Injection survivors
011.4B	3 Injection survivors	15 Wildtype

Despite the low eclosion numbers, sufficient transgenics were obtained to set up crosses to wildtypes for line establishment. At this point I finished my experimental work so these lines were taken over by other researchers for continuation.

5.4.8   Editing the guide designs

Post my finishing of experimental work, after the second tested guide did not work an issue was noticed with the guide designs. PAM sequences were retained in the guide constructs when these should be removed from the finalised guides. In addition to this when recombining UEA_010 had retained the PAM element of another guide in addition to its own. This explains why neither guide construct generated cutting. These construct designs have now been fixed and can be found in *Supplementary information 5.6.2*. Overhang PCR will be used to remove the PAM sequences from UEA_010 and UEA_011 and these shall be retested in future work.

5.5   Discussion

Manipulation of the sex determination pathway in medfly has several potential uses in the both the field of pest control and in furthering our understanding of the sex determination pathway (Siddall et al. 2022). Medfly is a particularly suitable species for manipulation within the sex determination pathway due to the apparent lack of necessity of the Y chromosome for male fertility (Meccariello et al. 2019). There are several points in the sex determination pathway which can be targeted. Targeting of upstream elements is more likely to result in full sex conversion which is of particular interest in medfly as pseudomales are fertile (Meccariello et al. 2019). These elements do however have more variability between species so if the aim is to develop systems more easily transferred to other species, downstream elements such as *dsx* could represent better targets although interference is more likely to result in sterilisation than full sex conversion. *Tra* and *dsx* were both potential targets in this work as they both have promise as tools to effect masculinisation of XX individuals (Saccone et al. 2008; Meccariello et al. 2023; Salvemini et al. 2009; Primo et al. 2020).

Once the targets had been decided, the next consideration was how to target these genes. This required two different strategies as their mechanisms of use differs in the sex determination pathway. *Tra* is relatively simple as its presence drives female development whereas it is repressed in male development. A knockout of *tra* could therefore be used to force male development. Designing guides for knockout can be automated using CHOPCHOP (Labun et al. 2019; Montague et al. 2014) making this a relatively simple process. Three guides were chosen for use as multiplexing can both reduce the formation of resistance alleles (Champer et al. 2020) and would allow for knockout if one of the guides designed were non-functional.

The targeting of *dsx* was slightly more complex. As both males and females utilise *dsx*, a knockout would not be appropriate in this case. One option would be the targeted knockout of only the female specific *dsx* (*fs-dsx*), this however would not result in masculinisation unless the male isoform was produced as complete *dsx* knockout would result in lethality or sterile intersex (Kyrou et al. 2018). Although this would have the desired effect of only fertile male offspring it would mean that 50% of offspring would be non-viable, this would not be an issue if the lethality was early enough that non-viable embryos were replaced, but this design would introduce the need to study where lack of *dsx* causes lethality. As the male splicing pattern of *dsx* is the default this does present the opportunity to avoid this scenario. By preventing female specific splicing from occurring, this will allow for the formation of

male-specific *dsx*. The female specific splicing pattern requires the recruitment of splicing enhancers to the splice site which allows for the retention of exon 4 (Saccone et al. 2008). By removing the sites which these splicing enhancers bind to and the splice site, this was predicted to prevent the female-specific splicing even in the presence of the splicing enhancers (such as *tra* and *tra-2* (Salvemini et al. 2009)). It is unclear what level of masculinisation would be seen in this case. The likelihood of fertile pseudomales being formed is low as this would mean all intermediary genes between *tra* and dsx, have no function ins sexual development other than driving the splicing of *dsx*. Producing masculinised sterile individuals would however be preferable to sterile intersex individuals as individuals with the ability to oviposit are not desirable in pest control strategies. These guides were once again multiplexed as although the removal of the splice site should be sufficient to prevent splicing on its own, removal of further regulatory regions decreases the likelihood of splicing. Resistance allele formation was predicted to be relatively low for this region as it is highly conserved.

The sequencing of the *tra* and *dsx* regions showed more mutations than was expected for such a highly conserved sequence. The SNP in the 3rd ranked *tra* guide was not an issue as this guide was not used in the end and mutations were expected in these guides as they are not in key regulatory elements. The presence of mutations in one of the regulatory regions in *dsx* was surprising as this was predicted to be a highly conserved region as it is responsible for enhancer recruitment and binding. This could indicate the region is tolerant to more mutations than we would expect, making the multiplexing strategy more necessary than relying on the high level of gene conservation maintaining the sequence.

When it became clear that time constraints would not allow for the development of both guide cassettes, a decision was made to rationalise the subsequent tests. The design of *dsx* was more novel and could have been an interesting study in how splicing patterns can be manipulated. However, the manipulation of *tra* had been shown to be successful in generating fertile pseudomales (Salvemini et al. 2009; Meccariello et al. 2023; Primo et al. 2020). Therefore, this was the more promising construct for future use in genetic pest control strategies. The development of a multiplexed guide cassette for *tra* was nevertheless a novel element that had not yet been developed in medfly. As gene drive development in medfly is in the relatively early stages, the elements needed for a guide cassette were not always available.

The promotor of the guide elements is very important for the successful guide production. However, at the time of starting this work there was no published *Ceratitis* U6 or 7sk promotors. These promotors are RNA Polymerase III promotors which are derived from U6 or 7sk genes and are ideal for expression of short non-coding RNA (Anderson et al. 2020). Meccariello kindly provided the U6 (the only native U6 tested in medfly) which is now published (Meccariello et al. 2023). There are more identified U6 promotors in other species that could potentially be used, but cross-species conservation of function is variable and difficult to predict (Anderson et al. 2020). This presents an issue when multiplexing guides. Although there is no concern with using the same U6 promotor to drive multiple guides, it does present a technical issue in that there are 3 repetitive elements separated by a short sequence length. This means when constructing and inserting these cassettes, recombination risk is high. It also drives up synthesis costs as flanking regions must be ordered for synthesis as the cassette on its own is too repetitive a sequence. Recombination occurred during construction for this guide cassette which resulting in two guides being removed from the cassette, negating the multiplexed construct. Without further research into these elements, producing a multiplexed guide cassette will remain technically difficult.

The final guide construct used only had one guide present but the decision was made to proceed - as a single guide (if functional) should still be sufficient for a knockout of *tra*. This would allow the testing of the previously developed Cas9 lines to test which are most effective at cutting. Ideally the guide provided by Meccariello would have been preferable for the final construct, as this would give us a functionally validated guide and higher confidence regarding crosses to the Cas9 lines. Although there was insufficient time to test both guides in this project, I did create and establish a transgenic line carrying the Meccariello guide, with the aim that this could be used if there were no observed sex conversion effects from the crosses of my original guide and Cas9 lines.

There was no male bias or intersex individuals in the first generation of crosses. However, this was not immediately discouraging, as the effect of the transgene is often not present at this stage, with inbreeding being necessary to produce homozygous individuals in which the phenotype is observable. If sex conversion had been seen in the first generation, this would have indicated a high level of Cas9 deposition into the embryo, given that *tra* is haplosufficient. The crosses were set up reciprocally to test for Cas9 deposition effects. By setting up crosses where both the males and females are the Cas9 providers, it should identify if Cas9 deposition was increasing the level of cutting seen in the offspring either maternally

or paternally. Maternal Cas9 deposition is well documented (Champer et al. 2018; Hammond et al. 2017) with paternal deposition being observed less but still possible (Xu et al. 2022). Inbreeding crosses were maintained, with the original cross information being retained for each cross. This is due to the sex of the original Cas9 carrier being shown to have an effect on transgene inheritance in constructed gene drives in subsequent generations (Verkuijl et al. 2022).

However, after examining the phenotype of the offspring of the inbreeding crosses there was still no evidence of masculinisation (partial or complete). Originally it was thought that as the guide line was crossed with four separate Cas9 lines (with two different promotors) which have been shown to express Cas9 (Chapter 2), the issue was more likely to be with the guide construct than with a lack of Cas9 expression. There are several avenues of research which could be pursued to determine if this is the case. The validated *tra* guide with demonstrated function in medfly had already been established so this was a good next step to pursue and such tests were completed by another lab member. This would both show if the original guide was defective, if cutting was observed in the new crosses – and this would allow for continuation of the work studying *tra* manipulation for pest control. If unsuccessful, this would identify an issue with the Cas9 lines or the construct design of the elements surrounding the guide, as these were shared across both types of guides. If it is discovered the guide was the cause of cutting failure there are several potential reasons. The first is that the guide does not bind where predicted and is therefore not recruiting Cas9 for cutting. Alternatively binding could be occurring but the cutting which is occurring is not inducing the knockout of *tra* that was predicted by CHOPCHOP. The first of these issues could be tested for with an *in vitro* incubation of *tra* sequence with Cas9 protein and the guide. If cutting is seen this means the guide is behaving as expected. If this was the case a T7 assay could be completed on inbred individuals expressing both fluorescence markers. If this showed cutting this would mean that the issue is that the mutation does not result in loss of function. If cutting does not occur, this means the guide is not working *in vivo*. This could be due to the insertion site of the transgene. As *piggybac* insertion was used, the genomic location of the transgene is random and therefore the insertion site could be on a transcriptional inactive section of a chromosome. The strong expression of the marker element which is carried with the rest of the construct makes this relatively unlikely but it cannot be ruled out.

The design problem arising from the retention of a PAM sequence was then discovered which is the most likely explanation for the lack of cutting observed. Unfortunately, this error was

not found until after the completion of my PhD work. The error can easily be rectified in the future with the use of an overhang PCR to remove this PAM sequence. The resulting guides can then be tested with the Cas9 promotors. The discovery of the error and route to fix it suggest that both the Cas9 and guide lines will yield positive future results to facilitate the development of genetic control strategies in medfly.

Overall, this work highlights the need for expanded toolbox elements in medfly to fully realise the potential of genetic control methods in this species. Further research into U6 and 7sk promotors is necessary to construct multiplexed guide cassettes without risk of recombination during plasmid construction, transgene insertion and transgene propagation. The use of computational software such as CHOPCHOP to design guides is common. However, the need for guide validation is clearly demonstrated by the results described here. As the original cassette had a validated guide, it was not deemed necessary to validate in this way as single guide failure would have been tolerated. *Tra* remains a promising target for pest control strategies and is indeed the target used in the first developed gene drive in medfly (Meccariello et al. 2023) and a medfly transgenic sorting strain (Davydova et al. 2023). The constructs developed here are still being tested and may yet yield positive results which will continue to expand the toolbox for genetic modification in medfly.

5.6 Supplementary Materials

5.6.1 Primers sequences

Table 5.17 Sequences of primers used throughout this experiment.

Primer Number	Sequence (5'-3')
128	tcaacgcgacgacatcgttg
129	tgtttgagttggccaagtggac
130	gctggagctgacattgaagaggca
131	tgcaaatcaactccgctttgct
158	tcggtctgtatatcgaggtt
163	cgatcgtgcgttacacgtag
319	ggtcgacagactgagatctgcgaatggccatgggacgtcg
320	cgacgtcccatggccattcgcagatctcagtctgtcgacc
323	tagcgccataaacatagcacc
324	ggtgctatgtttatggcgcta
329	gttgttattaaacgtagattcgggttttagagctagaaat
330	aatctacgtttaataacaaccaaatcacattaaaaacaca

5.6.2 Plasmid maps, parts and sequencing results

All sequences and aligned reads can be obtained at
https://benchling.com/siddall/f_/sqcsnseT-alex-siddall-thesis-supplementary-info/

5.7   References

Anderson, Michelle A. E., Jessica Purcell, Sebald A. N. Verkuijl, Victoria C. Norman, Philip T. Leftwich, Tim Harvey-Samuel, and Luke S. Alphey. 2020. 'Expanding the CRISPR Toolbox in *Culicine* Mosquitoes: *In Vitro* Validation of Pol III Promoters'. *ACS Synthetic Biology* 9 (3): 678–81. https://doi.org/10.1021/acssynbio.9b00436.

Behle, Anna, and Tatsuya Sakaguchi. 2020. 'Homemade Gibson Mastermix V1'. Preprint. https://doi.org/10.17504/protocols.io.ba2yigfw.

Benchling. 2023. 'Biological Software'. Benchling.com.

Champer, Jackson, Jingxian Liu, Suh Yeon Oh, Riona Reeves, Anisha Luthra, Nathan Oakes, Andrew G Clark, and Philipp W Messer. 2018. 'Reducing Resistance Allele Formation in CRISPR Gene Drive'. *Proc. Natl. Acad. Sci. U. S. A.* 115 (21): 5522–27. https://doi.org/10.1073/pnas.1720354115.

Champer, Samuel E., Suh Yeon Oh, Chen Liu, Zhaoxin Wen, Andrew G. Clark, Philipp W. Messer, and Jackson Champer. 2020. 'Computational and Experimental Performance of CRISPR Homing Gene Drive Strategies with Multiplexed gRNAs'. *Science Advances* 6 (10): eaaz0525. https://doi.org/10.1126/sciadv.aaz0525.

Davydova, S., J. Liu, N.P. Kandul, W. Evan Braswell, Omar S. Akbari, and Angela Meccariello. 2023. 'Next-Generation Genetic Sexing Strain Establishment in the Agricultural Pest *Ceratitis capitata*'. *Sci. Rep.* 13 (19866). https://doi.org/10.1038/s41598-023-47276-5.

Eckermann, Kolja N., Hassan M.M. Ahmed, Mohammad KaramiNejadRanjbar, Stefan Dippel, Christian E. Ogaugwu, Peter Kitzmann, Musa D. Isah, and Ernst A. Wimmer. 2018. 'Hyperactive *piggyBac* Transposase Improves Transformation Efficiency in Diverse Insect Species'. *Insect Biochemistry and Molecular Biology* 98 (July): 16–24. https://doi.org/10.1016/j.ibmb.2018.04.001.

Geuverink, E., and L.W. Beukeboom. 2014. 'Phylogenetic Distribution and Evolutionary Dynamics of the Sex Determination Genes *Doublesex* and *Transformer* in Insects'. *Sexual Development* 8 (1–3): 38–49. https://doi.org/10.1159/000357056.

Hammond, Andrew M, Kyros Kyrou, Marco Bruttini, Ace North, Roberto Galizi, Xenia Karlsson, Nace Kranjc, et al. 2017. 'The Creation and Selection of Mutations Resistant to a Gene Drive over Multiple Generations in the Malaria Mosquito'. *PLoS Genet.* 13 (10): e1007039. https://doi.org/10.1371/journal.pgen.1007039.

Kyriacou, C. 1992. 'Sex Variations'. *Trends in Genetics* 8 (8): 261–63. https://doi.org/10.1016/0168-9525(92)90241-U.

Kyrou, Kyros, Andrew M Hammond, Roberto Galizi, Nace Kranjc, Austin Burt, Andrea K Beaghton, Tony Nolan, and Andrea Crisanti. 2018. 'A CRISPR–Cas9 Gene Drive Targeting Doublesex Causes Complete Population Suppression in Caged *Anopheles gambiae* Mosquitoes'. *Nature Biotechnology* 36 (11): 1062–66. https://doi.org/10.1038/nbt.4245.

Labbé, Geneviève M C, Sarah Scaife, Siân A Morgan, Zoë H Curtis, and Luke Alphey. 2012. 'Female-Specific Flightless (fsRIDL) Phenotype for Control of *Aedes albopictus*'. *PLoS Negl. Trop. Dis.* 6 (7): e1724. https://doi.org/10.1371/journal.pntd.0001724.

Labun, Kornel, Tessa G Montague, Maximilian Krause, Yamila N Torres Cleuren, Håkon Tjeldnes, and Eivind Valen. 2019. 'CHOPCHOP v3: Expanding the CRISPR Web Toolbox beyond Genome Editing'. *Nucleic Acids Research* 47 (W1): W171–74. https://doi.org/10.1093/nar/gkz365.

Madden, T. 2002. 'The BLAST Sequence Analysis Tool'. In *The NCBI Handbook*. Chapter 16. http://www.ncbi.nlm.nih.gov/books/NBK21097/.

Meccariello, Angela, Shibo Hou, Serafima Davydova, James Fawcett, Alexandra Siddall, Philip T. Leftwich, Flavia Krsticevic, Philippos Aris Papathanos, and Nikolai Windbichler. 2023. 'Gene Drive and Genetic Sex Conversion in the Global Agricultural Pest *Ceratitis capitata*'. Preprint. Synthetic Biology. https://doi.org/10.1101/2023.08.16.553191.

Meccariello, Angela, Flavia Krsticevic, Rita Colonna, Giuseppe Del Corsano, Barbara Fasulo, Philippos Aris Papathanos, and Nikolai Windbichler. 2021. 'Engineered Sex Ratio Distortion by X-Shredding in the Global Agricultural Pest *Ceratitis capitata*'. *BMC Biology* 19 (1): 78. https://doi.org/10.1186/s12915-021-01010-7.

Meccariello, Angela, Marco Salvemini, Pasquale Primo, Brantley Hall, Panagiota Koskinioti, Martina Dalíková, Andrea Gravina, et al. 2019. '*Maleness-on-the-Y* (*MoY*) Orchestrates Male Sex Determination in Major Agricultural Fruit Fly Pests'. *Science* 365 (6460): 1457–60. https://doi.org/10.1126/science.aax1318.

Montague, Tessa G., José M. Cruz, James A. Gagnon, George M. Church, and Eivind Valen. 2014. 'CHOPCHOP: A CRISPR/Cas9 and TALEN Web Tool for Genome Editing'. *Nucleic Acids Research* 42 (W1): W401–7. https://doi.org/10.1093/nar/gku410.

Perrotta, Maryanna Martina, Francesca Lucibelli, Sarah Maria Mazzucchiello, Nicole Fucci, Bruno Hay Mele, Ennio Giordano, Marco Salvemini, et al. 2023. 'Female Sex Determination Factors in *Ceratitis capitata*: Molecular and Structural Basis of *TRA* and *TRA2* Recognition'. *Insects* 14 (7): 605. https://doi.org/10.3390/insects14070605.

Poelchau, Monica, Christopher Childers, Gary Moore, Vijaya Tsavatapalli, Jay Evans, Chien-Yueh Lee, Han Lin, Jun-Wei Lin, and Kevin Hackett. 2015. 'The I5k Workspace@NAL— Enabling Genomic Data Access, Visualization and Curation of Arthropod Genomes'. *Nucleic Acids Research* 43 (D1): D714–19. https://doi.org/10.1093/nar/gku983.

Price, Dana C., Andrea Egizi, and Dina M. Fonseca. 2015. 'The Ubiquity and Ancestry of Insect *Doublesex*'. *Scientific Reports* 5 (1): 13068. https://doi.org/10.1038/srep13068.

Primo, Pasquale, Angela Meccariello, Maria Grazia Inghilterra, Andrea Gravina, Giuseppe Del Corsano, Gennaro Volpe, Germano Sollazzo, et al. 2020. 'Targeting the Autosomal *Ceratitis capitata Transformer* Gene Using Cas9 or dCas9 to Masculinize XX Individuals without Inducing Mutations'. *BMC Genetics* 21 (S2): 150. https://doi.org/10.1186/s12863-020-00941-4.

Rendon Arana, Pedro Alfonso., The University of Manchester (United Kingdom), and Department not provided. 1996. 'Development and Evaluation of a Temperature

Sensitive Lethal (TSL) Genetic Sexing Strain of the Mediterranean Fruit Fly, *Ceratitis capitata* (Wied.)'.

Rubin, Gerald M., and Allan C. Spradling. 1982. 'Genetic Transformation of *Drosophila* with Transposable Element Vectors'. *Science* 218 (4570): 348–53. https://doi.org/10.1126/science.6289436.

Saccone, Giuseppe, Marco Salvemini, Attilio Pane, and Lino C. Polito. 2008. 'Masculinization of XX Drosophila Transgenic Flies Expressing the *Ceratitis capitata DoublesexM* Isoform'. *The International Journal of Developmental Biology* 52 (8): 1051–57. https://doi.org/10.1387/ijdb.082657gs.

Salvemini, Marco, Mark Robertson, Benjamin Aronson, Peter Atkinson, Lino C Polito, and Giuseppe Saccone. 2009. '*Ceratitis capitata Transformer-2* Gene Is Required to Establish and Maintain the Autoregulation of *Cctra*, the Master Gene for Female Sex Determination'. *Int. J. Dev. Biol.* 53 (1): 109–20. https://doi.org/10.1387/ijdb.082681ms.

Scolari, Francesca, Marc F. Schetelig, Sabrina Bertin, Anna R. Malacrida, Giuliano Gasperi, and Ernst A. Wimmer. 2008. 'Fluorescent Sperm Marking to Improve the Fight against the Pest Insect *Ceratitis capitata* (Wiedemann; Diptera: Tephritidae)'. *New Biotechnology* 25 (1): 76–84. https://doi.org/10.1016/j.nbt.2008.02.001.

Siddall, Alex, Tim Harvey-Samuel, Tracey Chapman, and Philip T. Leftwich. 2022. 'Manipulating Insect Sex Determination Pathways for Genetic Pest Management: Opportunities and Challenges'. *Frontiers in Bioengineering and Biotechnology* 10 (June): 867851. https://doi.org/10.3389/fbioe.2022.867851.

Spanos, L., G. Koutroumbas, M. Kotsyfakis, and C. Louis. 2000. 'The Mitochondrial Genome of the Mediterranean Fruit Fly, *Ceratitis capitata*'. *Insect Molecular Biology* 9 (2): 139–44. https://doi.org/10.1046/j.1365-2583.2000.00165.x.

Verhulst, Eveline C, Louis Van De Zande, and Leo W Beukeboom. 2010. 'Insect Sex Determination: It All Evolves around *Transformer*'. *Current Opinion in Genetics & Development* 20 (4): 376–83. https://doi.org/10.1016/j.gde.2010.05.001.

Verkuijl, Sebald A. N., Estela Gonzalez, Ming Li, Joshua X. D. Ang, Nikolay P. Kandul, Michelle A. E. Anderson, Omar S. Akbari, Michael B. Bonsall, and Luke Alphey. 2022. 'A CRISPR Endonuclease Gene Drive Reveals Distinct Mechanisms of Inheritance Bias'. *Nature Communications* 13 (1): 7145. https://doi.org/10.1038/s41467-022-34739-y.

Xu, Xuejiao, Tim Harvey-Samuel, Hamid Anees Siddiqui, Joshua Xin De Ang, Michelle Ellis Anderson, Christine M. Reitmayer, Erica Lovett, Philip T. Leftwich, Minsheng You, and Luke Alphey. 2022. 'Toward a CRISPR-Cas9-Based Gene Drive in the Diamondback Moth *Plutella xylostella*'. *The CRISPR Journal* 5 (2): 224–36. https://doi.org/10.1089/crispr.2021.0129.

6 Chapter Six – The identification of essential genes for targeting

6.1 Abstract

Resistance alleles are a major hurdle in the development of gene drive in *Ceratitis capitata*. These mutations within the targeted region result in guide RNAs no longer being able to bind, disrupting the action of the gene drive. There are several proposals on how to reduce resistance allele formation, one of which is the targeting of an essential gene. This will inhibit resistance allele formation because mutations within an essential gene cannot be tolerated by the organism if they cause loss of function. Toxin/antidote (TA) systems are the main types of gene drive systems that primarily utilise this strategy. The underlying element of TA gene drive constructs is based upon the targeted disruption and subsequent rescue of an essential gene. Therefore, the identification of additional essential genes for targeting in TA systems is important for further developments of this promising control strategy. The aim of the work in this chapter was to identify and investigate new essential genes for TA targeting in medfly. Guide (g) gRNAs for targeted disruption of the essential genes identified as promising targets were designed alongside recoded rescue versions of the essential genes identified as targets. I then discuss how these tools could be used in gene drive designs and also how they could provide key new information on how genetic context (i.e. the surrounding genetic location) affects gene expression.

6.2   Introduction

Upon the discovery of the CRIPSR system (Jinek et al. 2012), homing gene drives have become one of the most promising prospects for genetic pest control (Esvelt et al. 2014; Raban et al. 2020). These drives have been constructed in several important pest species including *Ceratitis capitata* (Meccariello et al. 2023), *Drosophila suzukii (Yadav et al. 2023), Anopheles gambiae* (Hammond et al. 2016; Kyrou et al. 2018), *Aedes aegypti* (M. Li et al. 2020) and *Anopheles stephensi (Adolfi et al. 2020).* The major challenge to the effective implementation of homing gene drives is non-homologous end joining (NHEJ), which causes the formation of resistance alleles (Champer et al. 2017; Carrami et al. 2018) and reduces the amount of homing occurring, disrupting drive spread. When cutting with Cas9 in homing gene drives, the intention is for repair through homology directed repair (HDR). This HDR allows for the copying of the drive construct from one chromosome to the other during meiosis, resulting in the generation of offspring of which 100% are homozygous for the intended mutation. HDR is favoured over NHEJ at certain time points in the cell cycle and within specific tissues (such as the germline) (Raban et al. 2020). This means if we can limit Cas9 action to the periods where HDR is favoured, repairs by NHEJ can be reduced. Resistance alleles are formed when the NHEJ repair pathway is deployed as opposed to the homology directed repair, this can result in mutations at the target region as the two ends are joined eliminating the target region which was originally between them. These resistance alleles are detrimental to gene drive spread as Cas9 can no longer cut the targeted regions once the guide RNAs are no longer fully complementary to these target sequences. Accumulation of these resistance alleles within a population severely challenges the ability of gene drives to spread.

There are two classifications of resistance alleles within CRISPR gene editing (Champer et al. 2018), r1 resistance alleles preserve the function of the gene, whereas r2 resistance alleles result in disrupted gene function. In homing-based control strategies, there can be an accumulation of both r1 and r2 resistance alleles. R1 resistance alleles are predicted to be much less common as typically they can only be caused by a synonymous mutation – hence must retain the original amino acid. There have been several proposed mechanisms to prevent the accumulation of resistance alleles. These include restriction of Cas9 expression to the germline thereby reducing cutting in stages of cell development where NHEJ is favoured (Kondo and Ueda 2013; Champer et al. 2018; Meccariello et al. 2023) (also covered in Chapter 3), multiplexing of guide RNAs (Champer et al. 2018; Yang et al. 2022; Marshall et al. 2017) (which modelling shows can still result in resistance (Champer et al. 2020c)) and the

targeting of essential genes within the guide design. R2 mutations within essential genes are not tolerated as they lead to organism death and therefore these more common resistance alleles cannot accumulate in a strategy that involves targeting an essential gene.

Essential genes have uses in both homing strategies and toxin/antidote systems. Direct targeting of an essential gene with a homing drive was achieved by Kyrou et al. (2018) who lowered resistance allele formation by targeting *dsx* which is a highly conserved essential gene. This was achievable due to the action of *dsx* being influenced by different splicing patterns in males and females which allowed the targeting of exon/intron boundaries making the technique highly specific to *dsx*. Recoded (Rec) (Adolfi et al. 2020) and Home-and-Rescue (HomeR) (Kandul et al. 2021) both utilise essential genes and achieved high levels of drive transmission with reduced levels of resistance allele formation compared to standard homing drives. These drives insert the homing drive into the selected essential gene, disrupting its native expression. The rescue version is then provided within the drive construct. The homing drive then copies itself to the non-drive allele in germline through homology directed repair, creating efficient driving. The difference here is that leaky expression of Cas9 causing cuts when homology directed repair is not possible and forcing the DNA to repair through non-homologous end joining does not cause the usual accumulation of resistance alleles in a population. This is because the drive is located on an essential gene where mutations resulting in loss of function are not tolerated.

Toxin-antidote systems also use the targeting of an essential gene to ensure drive spread (Oberhofer et al. 2018; Champer et al. 2020b; Champer et al. 2020a; Beaghton et al. 2017). These systems are fitness based (they cannot drive if they impose an immediate high fitness cost) and are therefore primarily limited to population modification (the exception being TADE drives (Champer et al. 2020a)), whereas homing drives are more easily suited to population suppression. Population modification has several potential uses in medfly. For example, insecticide resistance has often been detected in medfly (Castells-Sierra et al. 2023) and this is a major issue as insecticides deployed through cover spraying or bait spraying are currently the most commonly used control method (Vontas et al. 2011). Population modification strategies can also be used to remove insecticide resistance in populations (Kaduskar et al. 2022). The sterile insect technique (SIT) is still a widely used method of control in medfly. However, improvements to SIT are urgently needed to improve the production rate of sterile males while minimising costly techniques such as the need for fluorescent sorting and the rearing of females which are superfluous in such control methods.

A temperature-dependent point mutation in *transformer-2* has been used in *Drosophila* to create only sterile, predominantly male offspring (Li and Handler 2017). This would allow for efficient production of sterile males for SIT by removing the need for sorting and screening while allowing for line maintenance at lower temperatures. This has been repeated in medfly however with limited success due to an inability to rear the flies at a low enough temperature to not trigger the temperature-dependent point mutation (Aumann et al. 2018). Population modification strategies could also theoretically allow for other methods of control through sex conversion. By targeting sex determination genes, offspring could be directed down the male developmental pathway, decreasing the number of females within a population and thus the damage they cause (Siddall et al. 2022).

There are several types of drives which utilise essential genes for a toxin/antidote system. *ClvR* (Cleave and Rescue) (Oberhofer et al. 2019) and TARE (toxin-antidote recessive embryo) (Champer et al. 2020b) constructs operate by having a 'toxic' element which relies upon guides targeting an essential gene, Cas9 to cut and disrupt that gene, an 'antidote' element consisting of a recoded version of the essential gene resistant to Cas9 cutting and the cargo. The cargo element carries the gene or guides which allow for population modification. With the exception of TADE designs (Champer et al. 2020a) which have a dominant lethality effect, the essential genes targeted must be haplosufficient. The targeting of a haploinsufficient gene allows for the system to be used for population suppression (Zhu and Champer 2023) whereas those targeting haplosufficient genes are more limited to population modification. Haplosufficiency is when a single copy of the gene is adequate for function. Targeting of a haplosufficient gene creates the threshold dependency demonstrated by these drives and negates the need for tightly controlled germline expression as resistance alleles are much more likely to accumulate due to loss of function mutations not being tolerated in essential genes. The threshold dependency is the result of a high number of disrupted essential gene copies needing to be present in the population before the drive begins to be selected for. The recessive essential gene also allows for Cas9 expression outside of the germline to be tolerated as disrupted alleles are present in non-drive carrying offspring if a functional copy is still present. As these designs do not rely on homology directed repair, the disruption of more essential gene alleles aids in the reliance on the drive rescue copy.

The meaning of essential gene which is used in this work is a gene which is necessary for the survival of the organism or cell. Due to their importance for cell survival, essential genes are often highly conserved across species, which aids in their identification. A common

characteristic of genes with ubiquitous expression is their essential nature. This ubiquitous expression also leads them to be employed as reference genes in gene expression studies. These well-characterised genes can be harnessed for other research with the primary use envisioned here is as a functional component of a gene drive. Targeted knockout out of essential genes can be used to ensure retention of the drive construct by providing a necessary rescue version of the essential gene which is linked to the drive payload.

The use of haplosufficient essential genes allows more flexibility in essential gene choice as haploinsufficiency is difficult to identify and rare (Morrill and Amon 2019; Deutschbauer et al. 2005; Lindsley et al. 1972). However, the frequency of haploinsufficiency in essential genes has been suggested to be higher than that of other genes, with over half of the profiled essential genes in yeast showing haplosufficient phenotype (Ohnuki and Ohya 2018). There is also difficulty in disentangling haploinsufficiency from dosage compensation. However, if there is any effect on fitness to the organism if there is only a single functional copy of an essential gene, this reduces the likelihood of the drive reaching fixation (Marshall and Hay 2012). As essential genes are often more highly studied than other genes, expression profiles and their usage in other work can help inform on the extent of haplosufficiency. There are also predictive approaches for determining haploinsufficiency (Steinberg et al. 2015; Norris et al. 2013; Huang et al. 2010), although these studies are primarily focused on human medical genetics.

To construct drives that target essential genes to prevent the formation of resistance alleles, identification of suitable candidates must be completed. The work described in this chapter investigates potential essential genes for use in *Ceratitis capitata* gene drive systems. The identification of suitable candidates was completed and I narrowed down these candidates based on their suitability for use in a genetic construct. Guides for targeted knockout of these essential genes were designed and the sequence of these genes in multiple *Ceratitis capitata* individuals was determined to check for conservation. A recoded rescue version was then designed for all genes to be used, to which the guides are non-complementary and should be cleavage-resistant.

6.3   Methods

6.3.1   Selecting essential genes

Essential genes were selected from a publication on housekeeping genes in medfly (Sagri et al. 2017). To reduce the list to suitable candidates, those with other genes within 2kb and those with a sequence size below 500bp were excluded. This ensured that non-target genes regulatory regions were not at risk of disruption and allowed a sufficiently long sequence for multiple guide targets. After filtering, three candidate genes (*GAPDH, alpha-tubulin* and *RPL19*) were selected from those outlined in Sagri et al. (2017).

6.3.2   Guide RNA design

CHOPCHOP software (Labun et al. 2019) using the medfly genome assembly Ccap_2.1 (Poelchau et al. 2015; Spanos et al. 2000) was used to design all PCR primers, and to identify guide targets in the three selected candidate essential genes. The top five ranked gRNAs generated for each gene were then checked for off target similarity using BLAST (Madden 2002). From this, the top three candidate guide sequences with the highest rank and no identified risk of off target effects were selected for each essential gene.

6.3.3   Fly Husbandry

6.3.3.1   Medfly stock lines

Two wildtype medfly stock lines were used throughout these experiments. 'Cepa Petapa' is a mass-reared wildtype strain originally collected from the wild in Guatemala, Central America (Rendon, 1996). TOLIMAN is a wildtype strain from Guatemala, Central America originally collected in 1990. Both strains were reared at Oxford Insect Technologies (Oxitec LTD, Milton Park, Abingdon) from 2004 and a sub-culture of each was established at the University of East Anglia (UEA) in 2010.

6.3.3.2   Medfly rearing

All wild type medfly cultures were reared in a controlled environment with temperature set at 25°C ±0.5°C, humidity at 60% ± 10% RH and on a 12h light/dark schedule. Depending upon the demand for eggs from cultures, adult populations were kept in 1 of 3 different sizes: small cages (10cm x 10cm x 10cm) initiated with 50 pupae, medium cages (13cm x 13cm x 14cm)

with 100 pupae and large cages (33cm x 33cm x 16cm) with 500 pupae. The emerging adult flies held in these cages were fed on an *ad libitum* diet of 3:1 sucrose:hydrolysed yeast mixture, changed every seven days. Water was supplied through the side of each cage via dental rolls soaked in RO water, which were renewed every 4 days.

All cages were designed with a mesh layer covering the majority of the surface area of one vertical wall of the cage to allow egg deposition. Water pots containing RO water were placed under the mesh side of each cage to collect eggs. Eggs laid through the cage mesh were collected after a period of no more than 24h of egg laying by removing eggs with a Pasteur pipette and transferring up to two drops of the eggs in water solution to filter paper. The filter papers containing eggs were then placed into 1/3 pint milk bottles containing 100ml of ASG larval food (1L ASG food: 850ml RO water, 12.5g agar, 73.5g sucrose, 67g maize, 47.5g Brewer's Yeast, 2ml Propionic acid, 25ml Nipagin). 7 days post egg collection, 2 filter paper folded strips were placed in each bottle before laying the bottles down on a thin layer of sand within a pupation box (170mm x 130mm x 50mm). The box was then sealed with a lid containing a mesh for a further 7-9 days. During this time third instar 'jumping' larvae exited the bottles and pupated in the sand. After this period, pupae were removed by sieving the sand through a standard metal sieve and transferring the pupae into petri dishes. From this the appropriate number of pupae were randomly selected and placed into a fresh cage or used in experiments.

6.3.4 DNA Extraction

DNA extraction was performed on whole flies of both TOLIMAN and Cepa Petapa wildtype strains. Whole flies were frozen at -20°C for a minimum of 48 hours with 3 individuals per sample. For the dissections, the frozen flies were removed from the freezer and placed onto dry ice, then homogenised with a sterilised, plastic pestle over dry ice. The Qiagen DNeasy® Blood & Tissue Kit was used for DNA extractions, as per the manufacturer's instructions with the addition of an extra elution step with 25µl molecular grade water after the original 50µl elution. The concentration of each sample was measured using a Thermo Fisher Scientific Nanodrop™ 8000 Spectrophotometer.

6.3.5 Amplification of target regions

DNA guide target regions were amplified and extracted from TOLIMAN and Cepa petapa strain gDNA using a high-fidelity polymerase (NEB Q5®) with the following primers: *GAPDH*

amplified with primers 57 and 60; *RPL19* amplified with primers 61 and 62; *alpha-tubulin* amplified with primers 63 and 66. Primer sequences can be found in *Supplementary information 6.6.1*. PCR products were confirmed as being the correct size by gel electrophoresis. 5µl PCR product stained with Thermo Scientific™ TriTrack DNA Loading Dye (6X) was run alongside the Thermo Scientific™ GeneRuler™ 1kb DNA Ladder on a 1.2% agarose gel at 110V for 60 minutes.

6.3.6   PCR Purification

The Thermo Scientific™ GeneJET PCR Purification Kit was used for the purification of PCR products, as per the manufacturer's instructions with final elution in 25µl molecular grade water after a 1 minute incubation at room temperature. The concentration of each sample was measured using a Thermo fisher Scientific Nanodrop™ 8000 Spectrophotometer.

6.3.7   Sequence Confirmation of PCR products

To confirm that the amplified PCR products were of the correct sequences, purified PCR samples were diluted to 10ng/µl and sent for sequencing using the Eurofins Genomics TubeSeq Sequencing service. Sequencing primers were sent with samples; primer 60 with GAPDH samples; primer 61 with *RPL19*; primer 66 with *alpha-tubulin*. Primer sequences can be found in *Supplementary information 6.6.1*. Sequence alignment of the resulting reads was performed using the Benchling sequence alignment tool (Benchling 2023).

6.3.8   Designing recoded genes

Recoded versions of essential genes were designed by changing every 3rd base pair in a reading frame for all residues for which a synonymous nucleotide change could be made. A codon usage table for *Ceratitis capitata* (Nakamura 2000) was used to ensure that codons with a similar frequency were selected in the design of the recoded gene sequences.

6.4   Results

6.4.1   Selecting essential gene targets

The Sagri et al. (2017) paper was used to identify suitable essential genes for targeted knockout and rescue. Genes which resided within 2kb from another gene (to prevent the disruption of regulatory elements of a non-target gene) or were less than 500bp in sequence length (to ensure multiple guide target sites) were excluded. This left three essential gene targets: *GAPDH (glyceraldehyde 3-phosphate dehydrogenase*, LOC:105664420), *RPL19 (ribosome protein L19,* LOC:101460909) and *alpha-tubulin* (LOC:101459628). *RPL19* has been shown to be haploinsufficient in *Drosopholia* however this has yet to be demonstrated in medfly.

6.4.2   Confirming guide RNA design

The use of the CHOPCHOP software (Labun et al. 2019) to design guides that would produce knockouts was effective and produced several candidates for all three essential genes. The CHOPCHOP algorithm checks for possible off target sites – however, to provide additional confirmation the top ranked guides were also run through BLAST against the *Ceratitis capitata* genome. When guide target sites overlapped, the highest ranking of the overlapping guides was chosen, and other overlaps then excluded. This was done to avoid targeting the same region multiple times, and thus reduce likelihood of resistance mutations arising. When analysing BLAST results, off target matches of 13 base pairs or more were noted. Three guides were checked for all target essential genes. None of the guide base pairs showed a 100% match with any off-target genome site. The BLAST check results are shown in *Table Table* 6.1.

Table 6.1 BLAST results of potential guides with off target matches of 13bp or more.

Essential gene target	Guide name	Results of BLAST check
GAPDH	GAPDH_RANK_1	13bp match in mRNA-fibrillin NW_019378519.1:869,301 Scaffold 111
GAPDH	GAPDH_RANK_2	No identified off target
GAPDH	GAPDH_RANK_4	No identified off target
RPL19	RPL19_RANK_1	14bp match in uncharacterised locus LOC101459813 scaffold 31
RPL19	RPL19_RANK_2	No identified off target
RPL19	RPL19_RANK_3	• 13bp match in mRNA-uncharacterized LOC101455552 scaffold 98 • 13bp match in mRNA-tetratricopeptide repeat protein 28 scaffold 219 • 13bp match in mRNA-uncharacterized LOC101463093 in scaffold 140 overlap
alpha-tubulin	ALPHA_RANK_1	No identified off target
alpha-tubulin	ALPHA_RANK_2	No identified off target
alpha-tubulin	ALPHA_RANK_3	13 bp match in 3' end mRNA-fibroblast growth factor receptor 3 scaffold 80

For all genes there was one or more guide which did not have any identified off target potential binding sites. The maximum number of base pair matches was 14 base pairs in an off-target region for RPL19_RANK_1. RPL19_RANK_3 returned the greatest number of potential off target binding sites with 13 base pair matches in 3 different genes. The likelihood

of these off-target locations causing off target cuts is low. Wherever possible, alternative guides without any off targets sequence matches were used.

6.4.3 Amplifying the target regions

The region in which the guides would bind were sequenced to check for the presence of single nucleotide polymorphisms (SNPs) which could indicate low conservation of region, increasing the risk of resistance alleles both already present in other populations and occurring over time. To assess this, I sampled individuals from two independent wildtype populations (TOLIMAN and Cepa petapa) which have been maintained in the lab for several years. DNA extractions were performed on three individuals from each population, ensuring a minimum of one of each sex from each population. The concentration of the DNA extractions was measured with Nanodrop and the results are outlined in *Table 6.2*.

Table 6.2 Concentration of DNA extractions of TOLIMAN and Cepa petapa populations. 3 individuals were used for each sample with a minimum of one male and one female sample for both populations.

Sample Shorthand	Sample	Concentration (ng/µl)
TM1	TOLIMAN Male Sample 1	352.0
TM2	TOLIMAN Male Sample 2	240.3
TF	TOLIMAN Female	310.6
CF1	Cepa petapa Female Sample 1	640.2
CF2	Cepa petapa Female Sample 2	748.1
CF3	Cepa petapa Male	293.8

Samples were diluted and to amplify genetic regions of interest by using PCR. Gel electrophoresis was used for size confirmation of PCR products, the resulting image is shown in *Figure 6.1*.
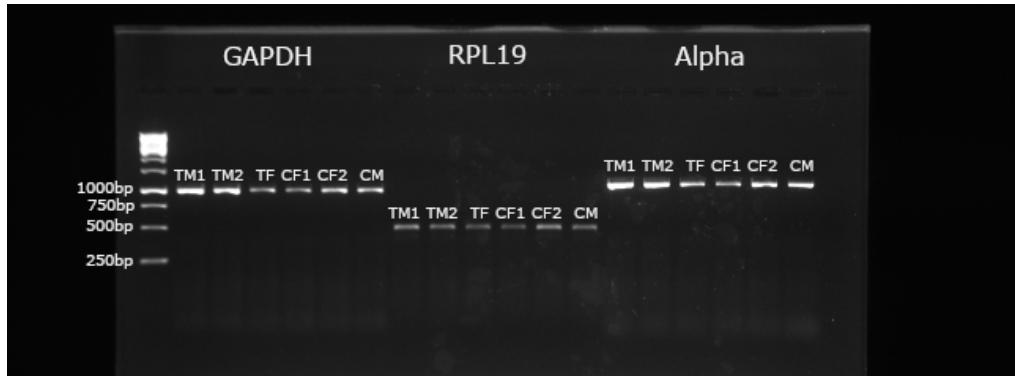
Figure 6.1 Gel electrophoresis for size confirmation of PCR products of amplified candidate gene fragments. Samples were run on a 1.3% agarose gel at 110V for 60 minutes alongside the ThermoScientific GeneRuler 1kb. Loading order: **GAPDH** (TOLIMAN Male Sample 1 (TM1), TOLIMAN Male Sample 2 (TM2), TOLIMAN Female (TF), Cepa Petapa Female Sample 1 (CF1), Cepa Petapa Female Sample 2 (CF2), Cepa Petapa Male (CM)); **RPL19** (TOLIMAN Male Sample 1, TOLIMAN Male Sample 2, TOLIMAN Female, Cepa Petapa Female Sample 1, Cepa Petapa Female Sample 2, Cepa Petapa Male); Alpha:**alpha-tubulin** (TOLIMAN Male Sample 1, TOLIMAN Male Sample 2, TOLIMAN Female, Cepa Petapa Female Sample 1, Cepa Petapa Female Sample 2, Cepa Petapa Male).

All products were of expected size, and the PCR products were then purified using a GeneJet PCR purification kit. The concentration of each sample was then measured with a Nanodrop (*Table 6.3*).

Table 6.3 Concentration of PCR products of amplified genes of interest in the different population samples.

Sample	Concentration (ng/μl)
GAPDH TM1	52.62
GAPDH TM2	54.28
GAPDH TF	19.28
GAPDH CF1	20.47
GAPDH CF2	51.77
GAPDH CM	35.58
RPL19 TM1	64.04
RPL19 TM2	32.94
RPL19 TF	27.99
RPL19 CF1	30.76
RPL19 CF2	31.07
RPL19 CM	26.63
Alpha-tubulin TM1	86.47
Alpha-tubulin TM2	48.24
Alpha-tubulin TF	35.09
Alpha-tubulin CF1	22.25
Alpha-tubulin CF2	36.33
Alpha-tubulin CM	25.23

6.4.4   Sequencing results

All samples were diluted to the appropriate level and sent for sequencing. The sequencing results were then aligned to their respective genomic regions on Benchling to check the accuracy of the genome annotation. Sequencing results can be found in *Supplementary information 6.6.2*. *GAPDH* and *RPL19* had sequence matches across all guide choices, implying they are highly conserved and thus good guide target choices. *Alpha-tubulin* had no sequence variation within the regions to which the guides would bind. However, there was some unexpected sequence divergence from the annotated genome, with three TOLIMAN-specific SNPs present throughout the genome sequence of *alpha-tubulin* (approximately 1kb).
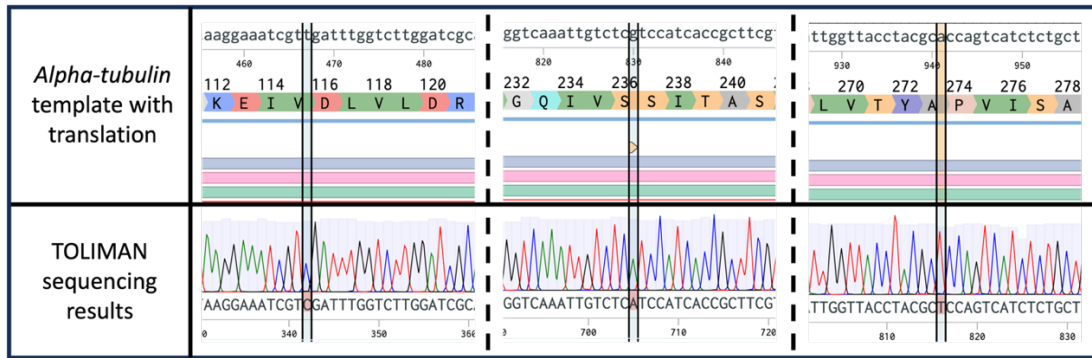
Figure 6.2 Sequencing results of TOLIMAN samples for *alpha-tubulin*. The three SNPs are highlighted with their position relative to the translated protein sequence shown. All mutations were synonymous and thus have no effect on the amino acid sequence.

All of the TOLIMAN-specific SNPs were substitutions of the third base pair in a codon, and all were silent, synonymous mutations that had no effect on the amino acid sequence. This means that although protein structure is highly conserved, there are some population-specific point mutations. Therefore, it is possible that guides might not be fully effective in all populations for targeting the *alpha-tubulin* gene, and if sufficient point mutations accumulated, these could form resistance alleles. The guides I designed do not fall in the region of the SNPs that were identified. However, this does indicate that the likelihood of resistance alleles already being present in populations is higher for *alpha-tubulin* than for the other two essential gene targets as it appears less highly conserved.

6.4.5 Designing the rescue versions of essential genes

GAPDH and *RPL19* had no identified SNPs within the targeted region, and as there were no potential drawbacks with these two genes, I decided to continue with only these and not with *alpha-tubulin,* due to the risk of resistance allele formation. Recoded gene sequences to act as rescues were designed for both *GAPDH* and *RPL19.* The annotated coding region where the chosen guides were expected to bind were recoded, and modifying only these regions reduced the risk of creating non-functional proteins. To create the recoded gene, the coding region was translated into the amino acid sequence and every 3rd base pair in the codon was altered in the regions the guides have homology to. This was done for all codons where possible, and if there was no alternative synonymous codon, the sequence was left unaltered. As the residues where this occurred were not tightly clustered, this is not expected to represent a problem, as sufficient of nearby base pairs will have been changed to the extent that prevents binding by any of the guides.

To ensure codon usage bias did not affect translation of the recoded gene, a codon usage wheel for *Ceratitis capitata* was used in the recoding exercise, so that codons were used at a similar frequency to the original sequence (Nakamura 2000). Once the rescue gene had been generated, guides were checked against the new sequence to confirm there were sufficient changes to prevent binding of the guide. An example of a recoded region of *GAPDH* is shown in *Figure 6.3*.
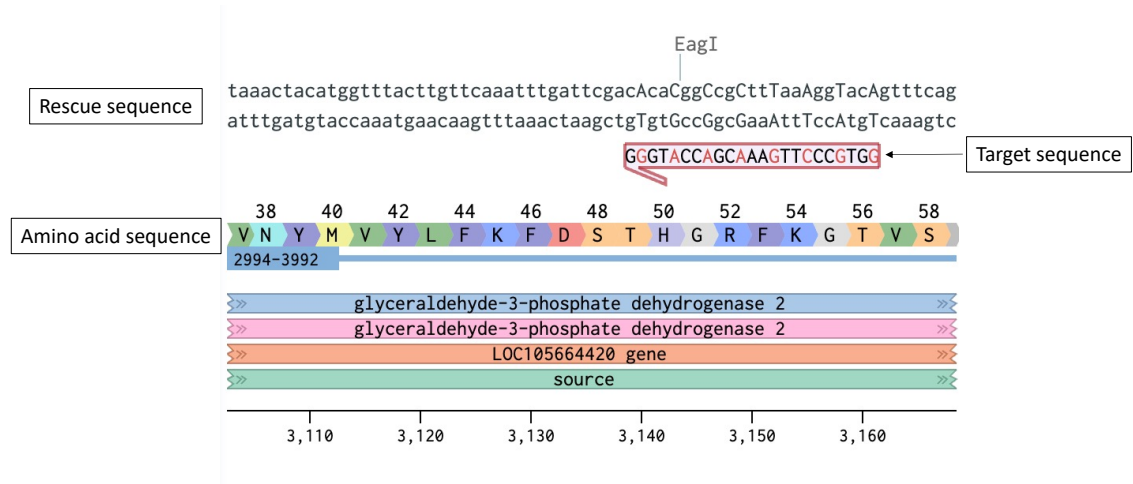
Figure 6.3 A recoded section of *GAPDH* with the modifications introduced shown in upper case letters. The amino acid sequence is shown below the genetic sequence of the rescue gene. The red lettering on the guide sequence highlights where the sequence is no longer complementary.

The sequences of the rescue genes of *GAPDH* and *RPL19* can be found in *Supplementary information 6.6.2*.

6.4.6   The design of the guide cassette

The guides designed in this part of my thesis work were aimed to be part of a TARE drive strategy in which there is multiplexing of guides to reduce the chance of resistance allele formation (Champer et al. 2018). Three guides were designed for each essential gene, all of which were planned to be integrated into a single guide cassette. This could be achieved by ordering the entire cassette, but also by ordering a single guide cassette part designed with PCR overhangs so that Gibson assembly could be used upon to insert the appropriate guides. The sequence of this guide cassette with the envisioned plasmid backbone can found in *Supplementary information 6.6.2*. The design of these guide elements is outlined in *Figure 6.4*.
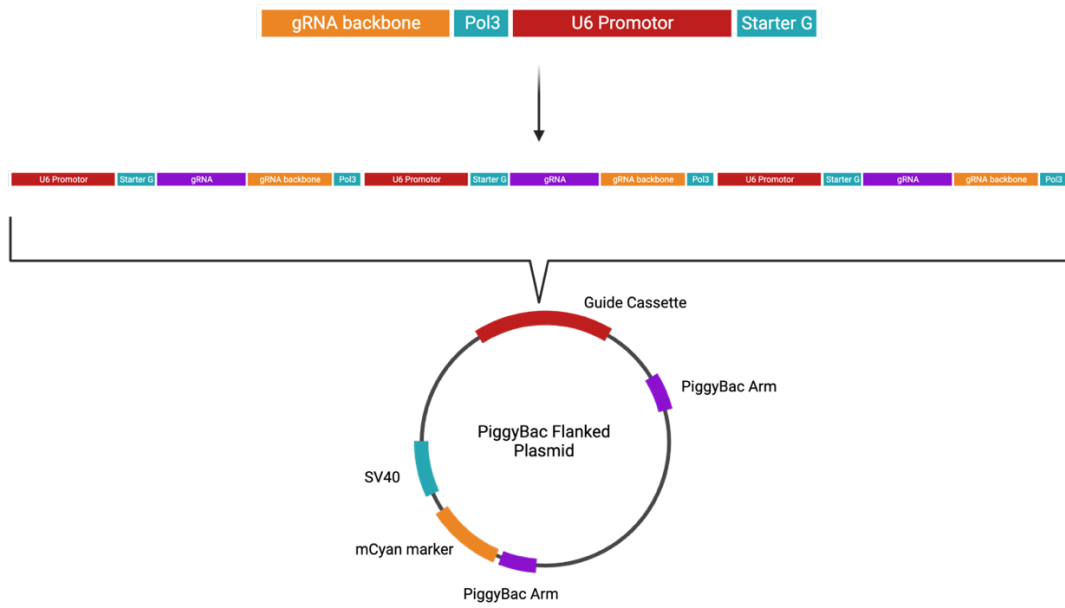
Figure 6.4 The design of the guide part for synthesis and the final construct of the multiplexed guide cassette inserted into a mCyan marker backbone. The synthesised part consists of the elements present in each guide in an order which would allow rearrangement via Gibson Assembly. The specific order of parts needed for each guide would be generated with overhang PCR to slot in the guides to generate the multiplex guide cassette seen below. The entire guide cassette would then be inserted into a PiggyBac flanked region of a plasmid with a mCyan marker (including a promotor and terminator) for transgenic identification. Generated in Biorender.

6.5 Discussion

The choice of essential gene to target for a rescue construct is an important one and can have a large impact on the dynamics of the drive. When targeting essential genes as part of a toxin-antidote system (Champer et al. 2020b; Champer et al. 2020a; Beaghton et al. 2017; Oberhofer et al. 2018), the essential gene targeted must be haplosufficient to ensure population invasion of the drive is threshold-dependent (Champer et al. 2020b). Targeting haplosufficient genes increases the number available for targeting, as approximately 97% of genes tested in *Saccharomyces cerevisiae* were estimated to be haplosufficient (Deutschbauer et al. 2005). Haploinsufficiency is consistently the minority phenotype across diploid organisms (Morrill and Amon 2019). There has been a significant amount of research into algorithms which can predict whether a gene is haploinsufficient (Shihab et al. 2017) (Huang et al. 2010), though most of this work is based primarily on human genome sequences, for applications in treating genetic diseases.

The use of haplosufficient genes allows a wide choice of toxin/antidote targets. The gene chosen must be essential for survival and should be necessary early in development. The earlier in development that a lack of the essential gene results in lethal, the more promising a target it is for use in a toxin/antidote drive. This is due to the offspring which will not survive not taking resources from those who will, imposing less of a fitness cost on drive carrying parents (Champer et al. 2020a). If the gene is not essential (imposing a fitness cost instead of lethality) this would not be suitable as imposing a high fitness cost on offspring results in the drive being unable to spread.

Choosing candidate genes is easier in species that have a well-annotated genome and a wealth of studies of gene functions. Expression in a wide variety of tissue types at all stages of development is a good indication of a gene being essential for organismal survival. Though medfly is well-studied, there is a no definitive list of essential genes used previously in gene drive studies. One strategy is to select orthologues of essential genes used successfully for gene drives in other species. This approach can be successful across closely related species and for genes with well conserved functions. However, gene behaviour can be very variable between species and replicating results in another species is often difficult. A gene which has a well-defined function in one species can have no or limited function in another. An alternative option is to select genes of unknown functions but with well-studied gene expression in the focal model. Another strategy is to use qPCR reference genes to compile a candidate list of genes. The characteristics of a good reference gene in qPCR are similar to

those required for a good essential gene candidate, especially when the qPCR is being used for multiple tissue types as this requires ubiquitous expression. They have stable expression levels across several tissue types and typically have quite high levels of expression to ensure their detection.

For the work described in this chapter, I derived information from a previous study that had quantified the levels of expression of many commonly used housekeeping genes in medfly (Sagri et al. 2017). I selected those genes which showed the most ubiquitous expression across tissues. I also considered technical aspects such as gene length. In order to provide a recoded version of any chosen essential gene, a full length insertion of that recoded gene has to be made into the transgenic construct. When inserting genetic elements in this way, the inserted sequence length should be kept as short as possible. This is to reduce costs and to reduce fitness costs in the resulting transgenic individuals, which tend to increase with the size of the constructs inserted. Another consideration is whether knocking out of the essential gene has any adverse effect on the surrounding genes, which may be difficult to predict given that transcriptional elements can be shared. To minimise this, I excluded candidates for which there was another located within 2kbs. I thus selected three candidate genes with the shortest sequence (*GAPDH, RPL19 and Alpha tubulin*).

GAPDH (glyceraldehyde-3-phosphate dehydrogenase) is necessary for glycolysis and has also been shown to have several additional functions outside of the glycolytic pathway (Nicholls et al. 2012) making it an essential gene for cell survival. *RPL19* (ribosomal protein L19) encodes a structural component of the 60S ribosomal subunit (Feo et al. 1992) with interference preventing the production of *RPL19* resulting in death (Chen et al. 2015). *Alpha-tubulin* is considered constitutively expressed as it is necessary to maintain cell function (Shen et al. 2010). All three genes are deemed essential across all cell types, which predicted that a knockout should be lethal early in development.

As these genes were to be targeted for knockouts it was necessary to determine the extent of their conservation within medfly. Essential genes are expected to be highly conserved due to their roles in promoting cell survival. However, silent or missense mutations could still be tolerated. Should there be a high level of genetic variation in an essential gene sequence it would set up difficulties in designing the original guides and increase the likelihood of resistance alleles evolving. *GAPDH* and *RPL19* lacked any SNPs in the coding region and were thus retained as essential gene targets. A lack of population specific SNPs implies a well

conserved gene in which mutations are not tolerated. The *Alpha-tubulin* sequence had 3 population specific SNPs. However, none of these would cause changes to the amino acid sequence and they were also not within the sequence targeted by the guides. This sequence divergence does present some level of concern, as it could indicate more divergent variation in other wildtype populations. Although none of the mutations were non-synonymous, such sequence differences can nevertheless represent a challenge in creating recoded rescue genes. This is because it is the third base pair in a codon which is often recoded to prevent guides binding. As there were two other suitable options the decision was made to discard *alpha-tubulin* as a candidate to take further.

Using laboratory populations to determine whether SNPs are common within a natural population can also present some concerns. For example, laboratory strains may be inbred and subject to relaxed selection pressures in comparison to their wild counterparts. In this experiment I used two different wildtype populations to identify candidate essential genes that show high conservation. Testing only two populations gives only a small sample of the potential genetic variation that most likely exists globally in this species. Sampling a larger number of individuals across many different populations could give a better indication of gene conservation. If these drives were to move beyond proof of principle, investigations could be done into where the target pest population to test for population specific SNPs.

Targeting a gene with high numbers of SNPs could also be a potential method of drive limitation. If the guides are not effective outside of the target population, this would prevent global spread, which could represent an additional safeguard. In this situation a drive would also most likely be self-limiting over time as the likelihood of resistance allele formation would be high.

The parts designed here are envisioned for usage in a TARE drive. The issue of resistance allele formation is reduced for this type of system in comparison to a homing drive. However, guides are still at risk of becoming ineffective if natural sequence variation occurs at the target site. Three guides were designed for all genes targeted in the work described in this chapter, and recoded genes altered the sequences of all three target sites. This gives the potential for the use of multiplexed guides which greatly reduces the chance of resistance allele evolution (Marshall et al. 2017). Another consideration when designing recoded genes is how much sequence needs to be recoded. Here, only the guide target regions were recoded. This has been theorised to present a concern if the rescue element is used for homology directed repair as the sequence similarity is sufficient for it to be used as the template (Champer et al.

2020b). This would uncouple the driving element from the payload of the drive, creating resistance problems. To reduce the risk of this, the construct would not be inserted as a same site rescue as increased distance between cut sites and template reduces the efficiency of homology directed repair (Oberhofer et al. 2018). This would prevent the rescue element being used as a repair template as an insertion of the rescue construct without the drive payload would render the drive obsolete if this allele spread.

The constructs designed here also have potential use outside of gene drive research. Expression of a recoded version of an essential gene could provide information on how important genomic location is for efficient and correctly timed expression. Synthetic biology has been encountering this problem for some time in that using so called 'plug and play' modules which included regulatory features such as promotors and terminators were not behaving as they did in their native form. When the genetic context, which includes the neighbouring transcriptional environment, was investigated it was found these could have direct influence on the levels of gene expression (Brooks et al. 2022). The relocation of essential genes outside of their genetic context could provide interesting information on whether the surrounding transcriptional regions have any effect on the expression of these essential genes. Changes to expression levels or timing of expression is likely to have visible effects on the organisms due to the importance of these genes for cell function. As they are often ubiquitously expressed it would also be interesting to investigate essential genes with tightly controlled timings or location of expression to see how much of the expression control is a result of the activity of the promotor and terminator or if the surrounding genetic context has any control here.

6.6 Supplementary information

6.6.1 Primers

Table 6.4 Sequences of primers used in Chapter 6.

Primer Number	Sequence
57	caccgggtgaattagagttctcagt
60	gatgaatgtgtgtcactgacgaaatc
61	aatcgctaacaccaattctcgt
62	gttttctggcctcacgtacttt
63	gcgtgaatgtatttctatccatgtc
66	cacagttggtggctggtagtta

6.6.2 Sequences

All sequences and designs can be found at https://benchling.com/siddall/f_/sqcsnseT-alex-siddall-thesis-supplementary-info/ .

6.7 References

Adolfi, Adriana, Valentino M. Gantz, Nijole Jasinskiene, Hsu-Feng Lee, Kristy Hwang, Gerard Terradas, Emily A. Bulger, et al. 2020. 'Efficient Population Modification Gene-Drive Rescue System in the Malaria Mosquito *Anopheles stephensi*'. *Nature Communications* 11 (1): 5553. https://doi.org/10.1038/s41467-020-19426-0.

Aumann, Roswitha A., Marc F. Schetelig, and Irina Häcker. 2018. 'Highly Efficient Genome Editing by Homology-Directed Repair Using Cas9 Protein in *Ceratitis capitata*'. *Insect Biochemistry and Molecular Biology* 101 (October): 85–93. https://doi.org/10.1016/j.ibmb.2018.08.004.

Beaghton, Andrea, Andrew Hammond, Tony Nolan, Andrea Crisanti, H Charles J Godfray, and Austin Burt. 2017. 'Requirements for Driving Antipathogen Effector Genes into Populations of Disease Vectors by Homing'. *Genetics* 205 (4): 1587–96. https://doi.org/10.1534/genetics.116.197632.

Benchling. 2023. 'Biological Software'. Benchling.com.

Brooks, Aaron N., Amanda L. Hughes, Sandra Clauder-Münster, Leslie A. Mitchell, Jef D. Boeke, and Lars M. Steinmetz. 2022. 'Transcriptional Neighborhoods Regulate Transcript Isoform Lengths and Expression Levels'. *Science* 375 (6584): 1000–1005. https://doi.org/10.1126/science.abg0162.

Carrami, Eli M., Kolja N. Eckermann, Hassan M. M. Ahmed, Héctor M. Sánchez C., Stefan Dippel, John M. Marshall, and Ernst A. Wimmer. 2018. 'Consequences of Resistance Evolution in a Cas9-Based Sex Conversion-Suppression Gene Drive for Insect Pest Management'. *Proceedings of the National Academy of Sciences* 115 (24): 6189–94. https://doi.org/10.1073/pnas.1713825115.

Castells-Sierra, Javier, Ana Guillem-Amat, Elena López-Errasquín, Lucas Sánchez, and Félix Ortego. 2023. 'First Detection of Resistance to Deltamethrin in Spanish Populations of the Mediterranean Fruit Fly, *Ceratitis capitata*'. *Journal of Pest Science* 96 (3): 1229–42. https://doi.org/10.1007/s10340-022-01578-1.

Champer, Jackson, Isabel K. Kim, Samuel E. Champer, Andrew G. Clark, and Philipp W. Messer. 2020a. 'Performance Analysis of Novel Toxin-Antidote CRISPR Gene Drive Systems'. *BMC Biology* 18 (1): 27. https://doi.org/10.1186/s12915-020-0761-2.

Champer, Jackson, Esther Lee, Emily Yang, Chen Liu, Andrew G Clark, and Philipp W Messer. 2020b. 'A Toxin-Antidote CRISPR Gene Drive System for Regional Population Modification'. *Nat. Commun.* 11 (1): 1082. https://doi.org/10.1038/s41467-020-14960-3.

Champer, Jackson, Jingxian Liu, Suh Yeon Oh, Riona Reeves, Anisha Luthra, Nathan Oakes, Andrew G Clark, and Philipp W Messer. 2018. 'Reducing Resistance Allele Formation in CRISPR Gene Drive'. *Proc. Natl. Acad. Sci. U. S. A.* 115 (21): 5522–27. https://doi.org/10.1073/pnas.1720354115.

Champer, Jackson, Riona Reeves, Suh Yeon Oh, Chen Liu, Jingxian Liu, Andrew G. Clark, and Philipp W. Messer. 2017. 'Novel CRISPR/Cas9 Gene Drive Constructs Reveal Insights

into Mechanisms of Resistance Allele Formation and Drive Efficiency in Genetically Diverse Populations'. Edited by Harmit S. Malik. *PLOS Genetics* 13 (7): e1006796. https://doi.org/10.1371/journal.pgen.1006796.

Champer, Samuel E, Suh Yeon Oh, Chen Liu, Zhaoxin Wen, Andrew G Clark, Philipp W Messer, and Jackson Champer. 2020c. 'Computational and Experimental Performance of CRISPR Homing Gene Drive Strategies with Multiplexed gRNAs'. *Sci Adv* 6 (10): eaaz0525. https://doi.org/10.1126/sciadv.aaz0525.

Chen, Aie, Weiwei Zheng, Wenping Zheng, and Hongyu Zhang. 2015. 'The Effects of RNA Interference Targeting *Bactrocera Dorsalis* Ds-Bdrpl19 on the Gene Expression of Rpl19 in Non-Target Insects'. *Ecotoxicology* 24 (3): 595–603. https://doi.org/10.1007/s10646-014-1407-3.

Deutschbauer, Adam M, Daniel F Jaramillo, Michael Proctor, Jochen Kumm, Maureen E Hillenmeyer, Ronald W Davis, Corey Nislow, and Guri Giaever. 2005. 'Mechanisms of Haploinsufficiency Revealed by Genome-Wide Profiling in Yeast'. *Genetics* 169 (4): 1915–25. https://doi.org/10.1534/genetics.104.036871.

Esvelt, Kevin M, Andrea L Smidler, Flaminia Catteruccia, and George M Church. 2014. 'Concerning RNA-Guided Gene Drives for the Alteration of Wild Populations'. *Elife* 3 (July). https://doi.org/10.7554/eLife.03401.

Feo, Salvatore, Brendan Davies, and Mike Fried. 1992. 'The Mapping of Seven Intron-Containing Ribosomal Protein Genes Shows They Are Unlinked in the Human Genome'. *Genomics* 13 (1): 201–7. https://doi.org/10.1016/0888-7543(92)90221-D.

Hammond, Andrew, Roberto Galizi, Kyros Kyrou, Alekos Simoni, Carla Siniscalchi, Dimitris Katsanos, Matthew Gribble, et al. 2016. 'A CRISPR-Cas9 Gene Drive System Targeting Female Reproduction in the Malaria Mosquito Vector *Anopheles gambiae*'. *Nat. Biotechnol.* 34 (1): 78–83. https://doi.org/10.1038/nbt.3439.

Huang, Ni, Insuk Lee, Edward M. Marcotte, and Matthew E. Hurles. 2010. 'Characterising and Predicting Haploinsufficiency in the Human Genome'. Edited by Mikkel H. Schierup. *PLoS Genetics* 6 (10): e1001154. https://doi.org/10.1371/journal.pgen.1001154.

Jinek, Martin, Krzysztof Chylinski, Ines Fonfara, Michael Hauer, Jennifer A. Doudna, and Emmanuelle Charpentier. 2012. 'A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity'. *Science* 337 (6096): 816–21. https://doi.org/10.1126/science.1225829.

Kaduskar, Bhagyashree, Raja Babu Singh Kushwah, Ankush Auradkar, Annabel Guichard, Menglin Li, Jared B. Bennett, Alison Henrique Ferreira Julio, John M. Marshall, Craig Montell, and Ethan Bier. 2022. 'Reversing Insecticide Resistance with Allelic-Drive in *Drosophila melanogaster*'. *Nature Communications* 13 (1): 291. https://doi.org/10.1038/s41467-021-27654-1.

Kandul, Nikolay P, Junru Liu, Jared B Bennett, John M Marshall, and Omar S Akbari. 2021. 'A Confinable Home-and-Rescue Gene Drive for Population Modification'. *eLife* 10 (March): e65939. https://doi.org/10.7554/eLife.65939.

Kondo, Shu, and Ryu Ueda. 2013. 'Highly Improved Gene Targeting by Germline-Specific Cas9 Expression in *Drosophila*'. *Genetics* 195 (3): 715–21. https://doi.org/10.1534/genetics.113.156737.

Kyrou, Kyros, Andrew M Hammond, Roberto Galizi, Nace Kranjc, Austin Burt, Andrea K Beaghton, Tony Nolan, and Andrea Crisanti. 2018. 'A CRISPR–Cas9 Gene Drive Targeting Doublesex Causes Complete Population Suppression in Caged *Anopheles gambiae* Mosquitoes'. *Nature Biotechnology* 36 (11): 1062–66. https://doi.org/10.1038/nbt.4245.

Labun, Kornel, Tessa G Montague, Maximilian Krause, Yamila N Torres Cleuren, Håkon Tjeldnes, and Eivind Valen. 2019. 'CHOPCHOP v3: Expanding the CRISPR Web Toolbox beyond Genome Editing'. *Nucleic Acids Research* 47 (W1): W171–74. https://doi.org/10.1093/nar/gkz365.

Li, Jianwei, and Alfred M. Handler. 2017. 'Temperature-Dependent Sex-Reversal by a Transformer-2 Gene-Edited Mutation in the Spotted Wing *Drosophila, Drosophila suzukii*'. *Scientific Reports* 7 (1): 12363. https://doi.org/10.1038/s41598-017-12405-4.

Li, Ming, Ting Yang, Nikolay P Kandul, Michelle Bui, Stephanie Gamez, Robyn Raban, Jared Bennett, et al. 2020. 'Development of a Confinable Gene Drive System in the Human Disease Vector *Aedes aegypti*'. *eLife* 9 (January): e51701. https://doi.org/10.7554/eLife.51701.

Lindsley, Dan L, L Sandler, Bruce S Baker, Adelaide T C Carpenter, R E Denell, Jeffrey C Hall, Patricia A Jacobs, et al. 1972. 'SEGMENTAL ANEUPLOIDY AND THE GENETIC GROSS STRUCTURE OF THE *DROSOPHILA* GENOME'. *Genetics* 71 (1): 157–84. https://doi.org/10.1093/genetics/71.1.157.

Madden, T. 2002. 'The BLAST Sequence Analysis Tool'. In *The NCBI Handbook*. Chapter 16. http://www.ncbi.nlm.nih.gov/books/NBK21097/.

Marshall, John M, Anna Buchman, Héctor M Sánchez C, and Omar S Akbari. 2017. 'Overcoming Evolved Resistance to Population-Suppressing Homing-Based Gene Drives'. *Sci. Rep.* 7 (1): 3776. https://doi.org/10.1038/s41598-017-02744-7.

Marshall, John M., and Bruce A. Hay. 2012. 'Confinement of Gene Drive Systems to Local Populations: A Comparative Analysis'. *Journal of Theoretical Biology* 294 (February): 153–71. https://doi.org/10.1016/j.jtbi.2011.10.032.

Meccariello, Angela, Shibo Hou, Serafima Davydova, James Fawcett, Alexandra Siddall, Philip T. Leftwich, Flavia Krsticevic, Philippos Aris Papathanos, and Nikolai Windbichler. 2023. 'Gene Drive and Genetic Sex Conversion in the Global Agricultural Pest *Ceratitis capitata*'. Preprint. Synthetic Biology. https://doi.org/10.1101/2023.08.16.553191.

Morrill, Summer A., and Angelika Amon. 2019. 'Why Haploinsufficiency Persists'. *Proceedings of the National Academy of Sciences* 116 (24): 11866–71. https://doi.org/10.1073/pnas.1900437116.

Nakamura, Y. 2000. 'Codon Usage Tabulated from International DNA Sequence Databases: Status for the Year 2000'. *Nucleic Acids Research* 28 (1): 292–292. https://doi.org/10.1093/nar/28.1.292.

Nicholls, Craig, He Li, and Jun-Ping Liu. 2012. '*GAPDH*: A Common Enzyme with Uncommon Functions: New Functions of *GAPDH*'. *Clinical and Experimental Pharmacology and Physiology* 39 (8): 674–79. https://doi.org/10.1111/j.1440-1681.2011.05599.x.

Norris, Matthew, Simon Lovell, and Daniela Delneri. 2013. 'Characterization and Prediction of Haploinsufficiency Using Systems-Level Gene Properties in Yeast'. *G3 Genes|Genomes|Genetics* 3 (11): 1965–77. https://doi.org/10.1534/g3.113.008144.

Oberhofer, Georg, Tobin Ivy, and Bruce A Hay. 2018. 'Behavior of Homing Endonuclease Gene Drives Targeting Genes Required for Viability or Female Fertility with Multiplexed Guide RNAs'. *Proc. Natl. Acad. Sci. U. S. A.* 115 (40): E9343–52. https://doi.org/10.1073/pnas.1805278115.

Oberhofer, Georg, Tobin Ivy, and Bruce A. Hay. 2019. 'Cleave and Rescue, a Novel Selfish Genetic Element and General Strategy for Gene Drive'. *Proceedings of the National Academy of Sciences* 116 (13): 6250–59. https://doi.org/10.1073/pnas.1816928116.

Ohnuki, Shinsuke, and Yoshikazu Ohya. 2018. 'High-Dimensional Single-Cell Phenotyping Reveals Extensive Haploinsufficiency'. Edited by Csaba Pal. *PLOS Biology* 16 (5): e2005130. https://doi.org/10.1371/journal.pbio.2005130.

Poelchau, Monica, Christopher Childers, Gary Moore, Vijaya Tsavatapalli, Jay Evans, Chien-Yueh Lee, Han Lin, Jun-Wei Lin, and Kevin Hackett. 2015. 'The I5k Workspace@NAL—Enabling Genomic Data Access, Visualization and Curation of Arthropod Genomes'. *Nucleic Acids Research* 43 (D1): D714–19. https://doi.org/10.1093/nar/gku983.

Raban, Robyn R., John M. Marshall, and Omar S. Akbari. 2020. 'Progress towards Engineering Gene Drives for Population Control'. Edited by Michael H. Dickinson, Leslie B. Vosshall, and Julian A. T. Dow. *Journal of Experimental Biology* 223 (Suppl_1): jeb208181. https://doi.org/10.1242/jeb.208181.

Rendon Arana, Pedro Alfonso., The University of Manchester (United Kingdom), and Department not provided. 1996. 'Development and Evaluation of a Temperature Sensitive Lethal (TSL) Genetic Sexing Strain of the Mediterranean Fruit Fly, *Ceratitis capitata* (Wied.)'.

Sagri, Efthimia, Panagiota Koskinioti, Maria-Eleni Gregoriou, Konstantina T. Tsoumani, Yiannis C. Bassiakos, and Kostas D. Mathiopoulos. 2017. 'Housekeeping in Tephritid Insects: The Best Gene Choice for Expression Analyses in the Medfly and the Olive Fly'. *Scientific Reports* 7 (1): 45634. https://doi.org/10.1038/srep45634.

Shen, Guang-Mao, Hong-Bo Jiang, Xiao-Na Wang, and Jin-Jun Wang. 2010. 'Evaluation of Endogenous References for Gene Expression Profiling in Different Tissues of the Oriental Fruit Fly *Bactrocera dorsalis* (Diptera: Tephritidae)'. *BMC Molecular Biology* 11 (1): 76. https://doi.org/10.1186/1471-2199-11-76.

Shihab, Hashem A, Mark F Rogers, Colin Campbell, and Tom R Gaunt. 2017. 'HIPred: An Integrative Approach to Predicting Haploinsufficient Genes'. Edited by Alfonso Valencia. *Bioinformatics* 33 (12): 1751–57. https://doi.org/10.1093/bioinformatics/btx028.

Siddall, Alex, Tim Harvey-Samuel, Tracey Chapman, and Philip T. Leftwich. 2022. 'Manipulating Insect Sex Determination Pathways for Genetic Pest Management:

Opportunities and Challenges'. *Frontiers in Bioengineering and Biotechnology* 10 (June): 867851. https://doi.org/10.3389/fbioe.2022.867851.

Spanos, L., G. Koutroumbas, M. Kotsyfakis, and C. Louis. 2000. 'The Mitochondrial Genome of the Mediterranean Fruit Fly, *Ceratitis capitata*'. *Insect Molecular Biology* 9 (2): 139–44. https://doi.org/10.1046/j.1365-2583.2000.00165.x.

Steinberg, Julia, Frantisek Honti, Stephen Meader, and Caleb Webber. 2015. 'Haploinsufficiency Predictions without Study Bias'. *Nucleic Acids Research* 43 (15): e101–e101. https://doi.org/10.1093/nar/gkv474.

Vontas, John, Pedro Hernández-Crespo, John T. Margaritopoulos, Félix Ortego, Hai-Tung Feng, Kostas D. Mathiopoulos, and Ju-Chun Hsu. 2011. 'Insecticide Resistance in Tephritid Flies'. *Pesticide Biochemistry and Physiology* 100 (3): 199–205. https://doi.org/10.1016/j.pestbp.2011.04.004.

Yadav, Amarish K., Cole Butler, Akihiko Yamamoto, Anandrao A. Patil, Alun L. Lloyd, and Maxwell J. Scott. 2023. 'CRISPR/Cas9-Based Split Homing Gene Drive Targeting *Doublesex* for Population Suppression of the Global Fruit Pest *Drosophila suzukii*'. *Proceedings of the National Academy of Sciences* 120 (25): e2301525120. https://doi.org/10.1073/pnas.2301525120.

Yang, Emily, Matthew Metzloff, Anna M Langmüller, Xuejiao Xu, Andrew G Clark, Philipp W Messer, and Jackson Champer. 2022. 'A Homing Suppression Gene Drive with Multiplexed gRNAs Maintains High Drive Conversion Efficiency and Avoids Functional Resistance Alleles'. Edited by M Ramaswami. *G3 Genes|Genomes|Genetics* 12 (6): jkac081. https://doi.org/10.1093/g3journal/jkac081.

Zhu, Yutong, and Jackson Champer. 2023. 'Simulations Reveal High Efficiency and Confinement of a Population Suppression CRISPR Toxin-Antidote Gene Drive'. *ACS Synthetic Biology* 12 (3): 809–19. https://doi.org/10.1021/acssynbio.2c00611.

7 Chapter Seven – General Discussion

7.1 Discussion

The control of pest species is an increasingly important issue as climate change expands their habitable regions (Deutsch et al. 2018; Gutierrez et al. 2021; Sultana et al. 2017). Agricultural pests are of particular concern as crop production is expected to be reduced due to the effects of climate change (Challinor et al. 2014) even before the impact of agricultural pests is taken into consideration. This is at a time where human population growth continues to demand increases in crop production (Lotze-Campen 2011). Medfly has the ability to damage over 300 crop species (White and Elson-Harris 1992; Liquido, NJ et al. 2017) through larval feeding and by creating oviposition holes in fruits through which bacterial and fungal infections can occur. The medfly has a wide geographical spread (Weldon 2022) with both direct crop damage and stringent control measures resulting in severe economic loss. The global cost of invasive insects has been estimated as high as $77 billion per annum (Bradshaw et al. 2016) with medfly costing the US alone $2 billion annually (Thomas et al. 2022). This highlights the need to expand and improve upon the current methods of genetic pest control.

To this aim, in this thesis I worked on expanding the toolbox for genetic modification in medfly to create the tools necessary to develop improved genetic control strategies. The discovery of CRISPR/Cas9 allowed for the expansion in the field of genetic pest control by (i) allowing genetic modifications to be induced with more ease for the study of insects (Chen et al. 2016), (ii) the development of new transgenics for genetic sexing strains (Nguyen et al. 2021) and (iii) through making designs such as gene drive achievable (McFarlane et al. 2018). To fully utilise CRISPR/Cas9 for pest control strategies such as gene drive, the appropriate tissue- and temporal-specific expression of Cas9 is an important factor. When inducing stable mutations, the injection of Cas9 can be used. However, for gene drives the organism must express Cas9 to induce mutations in future generations. To express the Cas9, unlike many transgenes which use viral promotors, an endogenous promotor is necessary (Carrami et al. 2018) as ubiquitous, high level expression is often undesirable.

Endogenous Cas9 promotors were a clear requirement needed to develop a genetic control toolbox in *Ceratitis capitata* in order to develop new methods of genetic pest control. Germline promotors of Cas9 have highest utility, as early expression reduces mosaicism and resistance allele formation (Champer et al. 2018). Several germline specific genes were identified as potential promotors, and the development of these is outlined in Chapter 2 and

Chapter 3. These Cas9 promotors were designed with slightly different purposes in mind and therefore differed in structure. All promotors were from an original set of candidate genes which have been shown to be germline-specific, and most had previous usage as transgene promotors in other species. In Chapter 2, two Cas9 promotors were developed, one was driven by *bgcn (Benign gonial cell neoplasm)* which had previously been tested in *Drosophila melanogaster* (Chan et al. 2013) and the other by *mei-W68* which had no known previous usage but displayed the germline restricted expression in addition to being well annotated in the medfly genome. The qPCR results I obtained showing Cas9 expression levels across different tissues emphasise the importance of understanding the variation observed as well as investigating additional promotors. The expression of Cas9 produced by the *mei-W68* was consistently higher than that of *bgcn*. This did not render *bgcn* as an unsuitable promotor as the level of Cas9 expression needs to be sufficient to induce a high level of cutting. Once this threshold has been passed excess protein production could be negative if it leads to high levels of deposition or fitness costs.

Neither line showed restriction of expression to the germline. However, the impact of this somatic expression is not possible to determine from the qPCR data alone. qPCR data may not be compared directly between tissue types if the reference gene is not consistently expressed across all tissues. However, finding such stably expressed reference genes can be a challenge. For example, *RPL19* is a commonly used reference gene. However, in this work a significant difference in expression across tissues was observed. A second reference gene (*RPE*) was used in this study. However, it could not be used for standardisation as it had very low expression in comparison to *RPL19* and thus would require a large number of samples to ascertain signal over background variation when using this reference gene. This exercise was useful in validating the difference in expression across tissue types between reference genes as it also showed higher expression in the ovaries. This implied that the consistently higher expression in the ovaries had been caused by a higher amount of starting material rather than a true difference in expression between tissues. More confidence with these results could be gained with the use of more reference genes. *RPL19* was ranked as highly stable across different tissue types in *Ceratitis* and was identified as one of the best for usage (Sagri et al. 2017). However, this ranking did change across different tissue types, implying variance in gene expression. This makes direct comparison of expression between tissue types challenging, although the results can still be used for confirmation of Cas9 expression in different tissue types and allow us to see which lines are most effective.

The difficulty in comparing between tissues with confidence is clearly an issue in qPCR experiments. Further validation of reference gene expression would be needed to confirm if they are appropriate to compare between tissues. The use of multiple reference genes could help to mitigate this issue if the expression differences between tissues is not consistent between reference genes (such as one having higher expression in ovaries and another having higher expression in somatic tissue). If expression is consistently higher in one tissue in all reference genes this would not resolve this issue. Expanding the number of reference genes used can also be difficult as it increases the size of the experiment which introduces further variation as more plates have to be used and more batches have to be run, increasing the likelihood of batch variation. The ideal solution is to find a reference gene with stable expression across all tissue types. To do this, expression levels of common reference genes could be tested in *Ceratitis* to determine the tissue specific variation. This would allow researchers to make an informed choice when selecting reference genes if they wish to make comparisons between tissues.

The informativeness of qPCR data in determining which promotors will be most effective is doubtful. qPCR assays do allow for the exclusion of any promotors which are not expressing Cas9, as no expression will be detected. Low levels of somatic expression and whether they are sufficiently low is much more difficult to ascertain. The qPCR data shown in Chapter 3 showed a low level of somatic expression in *vasa* both in comparison to the other tested genes and its own germline expression levels. When tested as part of a homing drive there was, however, a high level of somatic expression. In work in other species *nanos* has consistently been shown to have a lower level of somatic expression than *vasa* (Du et al. 2023; Champer et al. 2018). Whether this is true species difference or the qPCR results do not accurately represent the level of somatic expression is unclear as somatic expression was an issue when the *vasa*-Cas9 construct was used in the homing drive (Meccariello et al. 2023).

This unpredictability of Cas9 promotor expression once again highlights the need for expanding the number of promoters available to us. To best determine how these Cas9 promotors truly behave, improvements to the qPCR protocol are necessary, without which we are still reliant on testing Cas9 promotors by crossing with guides and measuring cutting. This is time consuming and, given the difference in promotor activity between species (Du et al. 2023), would need repeating in every new species we wish to use the technology in. This is further compounded by the difference in requirements of Cas9 promotors for different designs.

The Cas9 promotors designed in Chapter 2 were envisioned for usage in toxin/antidote drives (Oberhofer et al. 2019; Champer et al. 2020a) which require slightly less tightly controlled expression. The use of essential genes reduces the risk of resistance allele formation therefore a higher level of somatic expression can be tolerated and Cas9 deposition into the embryos can be beneficial for drive spread (Champer et al. 2020a; 2021). The promotors in Chapter 3 were designed for use in a homing drive and required more tightly controlled expression than was needed in the research described in Chapter 2. For use in toxin/antidote drives all genes tested had potential for applied use. The eventual focus on *mei-W68* and *bgcn* was because these were technically the most achievable. *bgcn* was also used for homing drive designs as it showed the highest level of germline restriction through RT-PCR. *vasa* was selected despite the technical requirement of shortening the 5' UTR (Champer et al. 2017; Feng et al. 2021; Adolfi et al. 2020). The major difference between the two designs was in the 3'UTR. The use of native 3' UTR was only deemed necessary for the homing drives as these require tightly controlled expression associated with the use of native 3'UTRs (Ebron and Shukla 2016). Having a bgcn-Cas9 construct with and without a native 3' UTR provided an interesting test of the necessity of using a native 3'UTR. Although the bgcn-Cas9 homing drive was not established as a line in this work, this is now being undertaken due to the promising expression pattern obtained in Chapter 2. The results can be directly compared to the activity of the bgcn-Cas9 promotor observed in Chapter 2.

Due to the lack of success of the *vasa* as a Cas9 promotor in a homing drive, this work is being continued with the *bgcn* promotor. The potential promise of a homing drive in *Ceratitis* was clear from the results obtained in Meccariello et al. 2023. High cutting and homing rates have been achieved and the targeting of *tra* is a promising method of sex conversion. If an effective Cas9 promotor can be found, population homing drive should be achievable. The promotors in Chapter 2 could also have potential use for this purpose as well as potential utility in toxin/antidote drives and genetic sexing strains. However, for either purpose they must be validated.

I attempted to validate the expression of the Cas9 line in the work outlined in Chapter 5. If time allowed these would have been validated by crossing to a guide with an easily identifiable phenotype such as *white eye* for which a tested guide in exists (Meccariello et al. 2023). Due to time constraints it was not possible during this project to do this. Instead, validation of these lines was attempted by crossing them to an untested guide targeting *tra*. These crosses did not generate any sex conversion. However, due to the qPCR data obtained

I was relatively certain this was not due to an issue with the guide lines and it was hoped that the use of a tested *tra* guide would resolve this issue for future work. It was subsequently realised that there was an issue with the original guide design in that a PAM sequence was present in the final guide rendering it non-functional. This confirmed that the reason for failure was not with the Cas9 which can now be validated with crosses to functional guides in future work.

The design work completed in Chapter 5 was useful despite the error in the guide design reported above. The novel guide cassette I designed for the targeting of *dsx* has been fixed and can be pursed in subsequent research. *Tra* is a very promising target for editing as it allows for the full sex conversion of XX individuals into fertile pseudomales (Primo et al. 2020). The assembly of the *tra* cassette revealed several avenues of research to investigate. The difficulties in maintaining the guide multiplexing due to recombination highlighted a key issue in this field as multiplexed guides are a necessity for homing drives to tackle resistance allele formation (Champer et al. 2020b). To avoid this issue more research must be done to identify more U6 and 7sk sequences native to *Ceratitis*. The UEA_010 and UEA_011 plasmids I tested can be easily modified to remove the PAM sequence introduced in error. These new constructs can then be microinjected, and the resulting strains retested for their effectiveness. If additional U6 or 7sk sequences are identified, these guides can also be reintroduced into a single guide cassette. The parts designed in Chapter 5 clearly have future potential for use in genetic modification in medfly and contribute to the aim of expanding the genetic toolbox for this species.

The targeting of sex determination genes was a key part of Chapter 5. However, the primary signal of sex determination in medfly was focused on in Chapter 4. *Maleness-on-the-Y* (*MoY*) (Meccariello et al. 2019) has several potential uses in genetic pest control strategies in medfly. If it can be harnessed, a dominant, masculinising element (or its knockout) would allow us to develop sex conversion with ease. Original experiments examining the function of *MoY*, induced masculinisation through the injection of *MoY* protein or plasmid DNA. However, for use in pest control a stable insertion would need to be established. This has proven difficult, with embryonic lethality being the result of the creation of such strains (Meccariello, personal communication). I attempted to restrict expression of MoY to a specific developmental timeframe in order to prevent overexpression lethality. I placed *MoY* under the control of the male germline specific ß2-tubulin gene. If successful this would not only prevent the lethality

but also create an interesting potential genetic pest control, provided males could be produced that deposited *MoY* through sperm, inducing male development in their offspring.

The same post injection embryonic lethality was seen when attempting to integrate this construct into the medfly genome. The sex of the injection survivors was checked to see if there was any sex bias. However, no such bias was observed, which was unexpected. As ß2-tubulin is not expressed in females, if only male lethality was seen this would mean the expected expression was causing lethality. However, the female lethality implies there is either leaky expression from ß2-tubulin in females, or that off plasmid expression is sufficient to induce lethality. The results show that even a very small amount of *MoY* expression can induce lethality. With such a high sensitivity to mistimed or overexpression of *MoY* there is clearly a need for further research into the native expression pattern of *MoY*. The use of the *Cre-Lox* system I designed would hopefully allow for the integration of a *MoY* transgenic without lethality during microinjection. This would then allow more precise expression timing as the *MoY* would not be produced until the *Cre* and *Lox* systems meet.

The final toolbox part I explored was the development of essential gene targets and rescue elements outlined in Chapter 6. Essential genes have several uses in genetic pest control strategies, the primary being for their use in toxin/antidote drives to prevent resistance allele formation (Oberhofer et al. 2019; Champer et al. 2020a). In this work I selected potential essential genes for targeting and sequenced these across two wildtype strains. It is expected that essential genes are highly conserved due to their necessity for survival, though this may not always be the case. My results highlighted the presence of population specific single nucleotide polymorphisms, indicating that these 'essential' genes are tolerant to some sequence variation (although it should be noted all SNPs detected were silent mutations). If checking only within a single population these mutations would not have been detected, and there could be additional mutations in wildtype populations.

When designing rescue elements there are two options in how the redesign can be completed. The entire gene can be recoded or the area of the gene which the guide is complementary to can be recoded. Recoding a smaller section of the gene has benefits in that native codon usage is often more efficient. However, this does run the risk of a recombination event uncoupling the rescue element from the rest of the drive, forming an essential gene resistance allele. As I envisioned these rescue elements for use in distant site rescue designs this risk was sufficiently mitigated by the distance between where the drive was inserted and the native gene. These essential gene designs have yet to be tested but can

be used for the development of toxin/antidote drives once efficient Cas9 promotors and gene of interest guides have been designed and validated.

In this thesis I have designed and constructed several elements which are of use in genetic modification of medfly. When beginning my PhD I had an aim to construct my own gene drives in *Ceratitis* and I designed the parts to achieve this. This was a complex task which required skills in molecular design, construction, medfly husbandry and microinjection. In the end I did not succeed in constructing a gene drive. However, I have contributed towards the development of a gene drive that was completed in the research described in Meccariello et al. 2023 and I also generated several parts which were necessary towards this goal. This work will be continued, and I believe I achieved my revised aim of contributing to the expansion of the toolbox in this important agricultural pest.

7.2 References

Adolfi, Adriana, Valentino M. Gantz, Nijole Jasinskiene, Hsu-Feng Lee, Kristy Hwang, Gerard Terradas, Emily A. Bulger, et al. 2020. 'Efficient Population Modification Gene-Drive Rescue System in the Malaria Mosquito *Anopheles stephensi*'. *Nature Communications* 11 (1): 5553. https://doi.org/10.1038/s41467-020-19426-0.

Bradshaw, Corey J. A., Boris Leroy, Céline Bellard, David Roiz, Céline Albert, Alice Fournier, Morgane Barbet-Massin, Jean-Michel Salles, Frédéric Simard, and Franck Courchamp. 2016. 'Massive yet Grossly Underestimated Global Costs of Invasive Insects'. *Nature Communications* 7 (1): 12986. https://doi.org/10.1038/ncomms12986.

Carrami, Eli M., Kolja N. Eckermann, Hassan M. M. Ahmed, Héctor M. Sánchez C., Stefan Dippel, John M. Marshall, and Ernst A. Wimmer. 2018. 'Consequences of Resistance Evolution in a Cas9-Based Sex Conversion-Suppression Gene Drive for Insect Pest Management'. *Proceedings of the National Academy of Sciences* 115 (24): 6189–94. https://doi.org/10.1073/pnas.1713825115.

Challinor, A. J., J. Watson, D. B. Lobell, S. M. Howden, D. R. Smith, and N. Chhetri. 2014. 'A Meta-Analysis of Crop Yield under Climate Change and Adaptation'. *Nature Climate Change* 4 (4): 287–91. https://doi.org/10.1038/nclimate2153.

Champer, Jackson, Samuel E Champer, Isabel K Kim, Andrew G Clark, and Philipp W Messer. 2021. 'Design and Analysis of CRISPR-Based Underdominance Toxin-Antidote Gene Drives'. *Evol. Appl.* 14 (4): 1052–69. https://doi.org/10.1111/eva.13180.

Champer, Jackson, Esther Lee, Emily Yang, Chen Liu, Andrew G Clark, and Philipp W Messer. 2020a. 'A Toxin-Antidote CRISPR Gene Drive System for Regional Population Modification'. *Nat. Commun.* 11 (1): 1082. https://doi.org/10.1038/s41467-020-14960-3.

Champer, Jackson, Jingxian Liu, Suh Yeon Oh, Riona Reeves, Anisha Luthra, Nathan Oakes, Andrew G. Clark, and Philipp W. Messer. 2018. 'Reducing Resistance Allele Formation in CRISPR Gene Drive'. *Proceedings of the National Academy of Sciences* 115 (21): 5522–27. https://doi.org/10.1073/pnas.1720354115.

Champer, Jackson, Riona Reeves, Suh Yeon Oh, Chen Liu, Jingxian Liu, Andrew G. Clark, and Philipp W. Messer. 2017. 'Novel CRISPR/Cas9 Gene Drive Constructs Reveal Insights into Mechanisms of Resistance Allele Formation and Drive Efficiency in Genetically Diverse Populations'. Edited by Harmit S. Malik. *PLOS Genetics* 13 (7): e1006796. https://doi.org/10.1371/journal.pgen.1006796.

Champer, Samuel E, Suh Yeon Oh, Chen Liu, Zhaoxin Wen, Andrew G Clark, Philipp W Messer, and Jackson Champer. 2020b. 'Computational and Experimental Performance of CRISPR Homing Gene Drive Strategies with Multiplexed gRNAs'. *Sci Adv* 6 (10): eaaz0525. https://doi.org/10.1126/sciadv.aaz0525.

Chan, Yuk-Sang, David S. Huen, Ruth Glauert, Eleanor Whiteway, and Steven Russell. 2013. 'Optimising Homing Endonuclease Gene Drive Performance in a Semi-Refractory Species: The *Drosophila melanogaster* Experience'. Edited by Shree Ram Singh. *PLoS ONE* 8 (1): e54130. https://doi.org/10.1371/journal.pone.0054130.

Chen, Lei, Gui Wang, Ya-Nan Zhu, Hui Xiang, and Wen Wang. 2016. 'Advances and Perspectives in the Application of CRISPR/Cas9 in Insects'. *Dong Wu Xue Yan Jiu = Zoological Research* 37 (4): 220–28. https://doi.org/10.13918/j.issn.2095-8137.2016.4.220.

Deutsch, Curtis A, Joshua J Tewksbury, Michelle Tigchelaar, David S Battisti, Scott C Merrill, Raymond B Huey, and Rosamond L Naylor. 2018. 'Increase in Crop Losses to Insect Pests in a Warming Climate'. *Science* 361 (6405): 916–19. https://doi.org/10.1126/science.aat3466.

Du, Jie, Weizhe Chen, Xihua Jia, Xuejiao Xu, Emily Yang, Ruizhi Zhou, Yuqi Zhang, Matt Metzloff, Philipp W. Messer, and Jackson Champer. 2023. 'New Germline Cas9 Promoters Show Improved Performance for Homing Gene Drive'. Preprint. Genetics. https://doi.org/10.1101/2023.07.16.549205.

Ebron, Jey Sabith, and Girish C. Shukla. 2016. 'Molecular Characterization of a Novel Androgen Receptor Transgene Responsive to MicroRNA Mediated Post-Transcriptional Control Exerted via 3'-Untranslated Region: Characterization of a Novel Androgen Receptor Reporter'. *The Prostate* 76 (9): 834–44. https://doi.org/10.1002/pros.23174.

Feng, Xuechun, Víctor López Del Amo, Enzo Mameli, Megan Lee, Alena L. Bishop, Norbert Perrimon, and Valentino M. Gantz. 2021. 'Optimized CRISPR Tools and Site-Directed Transgenesis towards Gene Drive Development in *Culex quinquefasciatus* Mosquitoes'. *Nature Communications* 12 (1): 2960. https://doi.org/10.1038/s41467-021-23239-0.

Gutierrez, Andrew Paul, Luigi Ponti, Markus Neteler, David Maxwell Suckling, and José Ricardo Cure. 2021. 'Invasive Potential of Tropical Fruit Flies in Temperate Regions under Climate Change'. *Communications Biology* 4 (1): 1141. https://doi.org/10.1038/s42003-021-02599-9.

Liquido, NJ, McQuate, GT, Hanlin, MA, and Suiter, KA. 2017. 'Host Plants of the Mediterranean Fruit Fly, *Ceratitis capitata* (Wiedemann), Version 3.5'. *USDA Compendium of Fruit Fly Host Information*.

Lotze-Campen, Hermann. 2011. 'Climate Change, Population Growth, and Crop Production: An Overview'. In *Crop Adaptation to Climate Change*, edited by Shyam S. Yadav, Robert J. Redden, Jerry L. Hatfield, Hermann Lotze-Campen, and Anthony E. Hall, 1st ed., 1–11. Wiley. https://doi.org/10.1002/9780470960929.ch1.

McFarlane, Gus R, C Bruce A Whitelaw, and Simon G Lillico. 2018. 'CRISPR-Based Gene Drives for Pest Control'. *Trends Biotechnol.* 36 (2): 130–33. https://doi.org/10.1016/j.tibtech.2017.10.001.

Meccariello, Angela, Shibo Hou, Serafima Davydova, James Fawcett, Alexandra Siddall, Philip T. Leftwich, Flavia Krsticevic, Philippos Aris Papathanos, and Nikolai Windbichler. 2023. 'Gene Drive and Genetic Sex Conversion in the Global Agricultural Pest *Ceratitis capitata*'. Preprint. Synthetic Biology. https://doi.org/10.1101/2023.08.16.553191.

Meccariello, Angela, Marco Salvemini, Pasquale Primo, Brantley Hall, Panagiota Koskinioti, Martina Dalíková, Andrea Gravina, et al. 2019. '*Maleness-on-the-Y* (*MoY*)

Orchestrates Male Sex Determination in Major Agricultural Fruit Fly Pests'. *Science* 365 (6460): 1457–60. https://doi.org/10.1126/science.aax1318.

Nguyen, Thu N. M., Amanda Choo, and Simon W. Baxter. 2021. 'Lessons from Drosophila: Engineering Genetic Sexing Strains with Temperature-Sensitive Lethality for Sterile Insect Technique Applications'. *Insects* 12 (3): 243. https://doi.org/10.3390/insects12030243.

Oberhofer, Georg, Tobin Ivy, and Bruce A Hay. 2019. 'Cleave and Rescue, a Novel Selfish Genetic Element and General Strategy for Gene Drive'. *Proc. Natl. Acad. Sci. U. S. A.* 116 (13): 6250–59. https://doi.org/10.1073/pnas.1816928116.

Primo, Pasquale, Angela Meccariello, Maria Grazia Inghilterra, Andrea Gravina, Giuseppe Del Corsano, Gennaro Volpe, Germano Sollazzo, et al. 2020. 'Targeting the Autosomal *Ceratitis capitata Transformer* Gene Using Cas9 or dCas9 to Masculinize XX Individuals without Inducing Mutations'. *BMC Genetics* 21 (S2): 150. https://doi.org/10.1186/s12863-020-00941-4.

Sagri, Efthimia, Panagiota Koskinioti, Maria-Eleni Gregoriou, Konstantina T. Tsoumani, Yiannis C. Bassiakos, and Kostas D. Mathiopoulos. 2017. 'Housekeeping in Tephritid Insects: The Best Gene Choice for Expression Analyses in the Medfly and the Olive Fly'. *Scientific Reports* 7 (1): 45634. https://doi.org/10.1038/srep45634.

Sultana, Sabira, John B Baumgartner, Bernard C Dominiak, Jane E Royer, and Linda J Beaumont. 2017. 'Potential Impacts of Climate Change on Habitat Suitability for the Queensland Fruit Fly'. *Sci. Rep.* 7 (1): 13025. https://doi.org/10.1038/s41598-017-13307-1.

Thomas, Michael C., John B. Heppner, Robert E. Woodruff, Howard V. Weems, Jr., Gary J. Steck, and Thomas R. Fasulo. 2022. 'Mediterranean Fruit Fly, *Ceratitis capitata* (Wiedemann) (Insecta: Diptera: Tephritidae)'. *EDIS* 2004 (8). https://doi.org/10.32473/edis-in371-2001.

Weldon, C. 2022. '*Ceratitis capitata* (Mediterranean Fruit Fly)'. https://doi.org/10.1079/cabicompendium.12367.

White, Ian M., and Marlene M. Elson-Harris. 1992. *Fruit Flies of Economic Significance: Their Identification and Bionomics*. Wallingford, Oxon, UK: CAB International in association with ACIAR.