Identification and investigation of genes involved in *Wolbachia*-host interaction and genes for control of pest insects

Elizabeth Sutton

Magdalen College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy (Zoology)

Michaelmas 2015

Acknowledgements

This DPhil has quite literally taken blood, sweat and tears to complete. Blood from numerous mosquito bites, sweat from working in a 28 °C insectary, and tears of frustration at the times it didn't quite go to plan. I would like to thank everyone who has helped and supported me during these four years.

Firstly thanks must go to my supervisor, Professor Steven Sinkins, who continued to supervise me even when he'd moved to Lancaster University and didn't have to any more. I am also very grateful to Dr Seb Shimeld, who took me on as his student halfway through my DPhil, and has provided invaluable guidance and support. Thanks also to my co-supervisor Professor Luke Alphey. Members of the Sinkins lab in both Oxford and Lancaster, as well as other groups in the Department of Zoology, have provided much assistance over the four years, both scientifically and otherwise. I would also like to thank everyone who helped me during my placement at Oxitec. My DPhil has involved several projects with external collaborators, and I am grateful to Dr Alistair Darby and his group at the University of Liverpool, Professor Helen White-Cooper at Cardiff University, Dr Paul Coupland at the Wellcome Trust Sanger Institute and Professor Julian Parkhill and his group at the Wellcome Trust Sanger Institute for their help and for hosting me at their institutions. I would also like to thank Magdalen College, the Department of Zoology and the BBSRC for giving me the opportunity to undertake this DPhil. Thanks also to the BBSRC PIPS scheme for providing me the opportunity to undertake internships during my DPhil, and to the organisations who hosted me – the Ebo Forest Research Project (even if I did nearly get shot, it does make a good story), and the Pathogen Genomics Group at the Wellcome Trust Sanger Institute, supervised by Professor Stephen Bentley, where my path towards becoming a bioinformatician really started.

Several people assisted with the scientific work performed here. Thank you to Dr Kirsty Stainton for assisting with the plasmid construction and injections in Chapter 2, Dr Simon Harris and Dr Paul Coupland for assisting with the computational analysis in Chapter 2, Dr Paul Coupland for assisting with the library preparation in Chapter 2, Dr Vincent Geoghegan for rearing the flies used in Chapter 3, Dr Caroline Harris for performing the mosquito rearing, dissections and extractions in Chapter 3, Dr Yachuan Yu, Dr Kelly Matzen, Dr Jenny Molloy, Dr Romisa Asadi, Ryan Turkel and Peter Elphick for rearing the inects used in Chapter 4, Dr Kelly Matzen, Dr Ian Goodhead, Dr Yachuan Yu and Dr Romisa Asadi for performing the dissections and extractions for sequencing in Chapter 4, and Dr

Ian Goodhead, Frances Blow and Dr Yachuan Yu for assisting with or performing the library preparation in Chapter 4,

On a personal level, I would like to thank my parents, and all my friends who have been there for me. Thanks to Chris, Jodi and Seb for providing me with places to stay, and Kumail, Ryan, and Kimberley for offering. I am lucky to have such good friends. Particular thanks to Mike for all his help, support, encouragement and patience.

Abstract

Insects pose large problems for health and agriculture; several major diseases are transmitted by insect vectors, and huge losses in food production occur due to insect pests. Current strategies for insect control are inadequate, with several disadvantages. Various novel approaches are being developed, aimed at either suppressing wild populations, or replacing them with insects refractory to disease-causing pathogens.

Some possible strategies utilise *Wolbachia*, a genus of maternally inherited intracellular bacteria. *Wolbachia* can manipulate host reproduction through a phenomenon known as cytoplasmic incompatibility (CI), which leads to spread of *Wolbachia* through populations. CI could potentially be used as a gene drive system for spread of pathogen-resistance genes, or for population suppression via the release of incompatible males. Furthermore, some strains of *Wolbachia* have been found to inhibit virus transmission by mosquitoes; there have been successful field trials of population replacement exploiting this inhibition.

At present, little is known about the mechanisms of CI, or viral inhibition by *Wolbachia*. Discussed in Chapter 2, candidate genes for CI involvement were identified by sequencing the genome of the non-CI strain *w*Au and comparing it to that of the CI strain *w*Mel. One candidate was used to transform *Drosophila melanogaster*, but it did not induce a CI phenotype. Chapter 3 discusses further identification and investigation of candidate genes for CI involvement by comparing RNA-seq data from *w*Au and the CI strain *w*Ri, and from *D. simulans* either uninfected or infected with one of these two strains. Similarly, candidate genes for viral inhibition were identified and investigated by comparing RNA-seq data from *Aedes albopictus* mosquitoes that were uninfected, naturally infected with *w*Alb, or transinfected with *w*Mel, which inhibits dengue in this species.

Other potential insect control strategies involve the release of modified insects with a genetic system that leads to population suppression. Discussed in Chapter 4, attempts were made using comparative genomics to identify the sex-determining locus in the mosquito *Ae. aegypti* , which could be utilised in genetic control systems. Some possible regions were identified, but the locus was discovered by another group before the work was completed. Chapter 4 also discusses the identification using RNA-seq of germline-specifically expressed or spliced genes in *Ae. aegypti* and the agricultural pest *Ceratitis capitata*; these could be used to control gene expression in various genetic insect control systems.

Overall this thesis demonstrates the use of high-throughput sequencing analyses to identify and investigate candidate genes of importance to insect control. It will hopefully serve as a platform for further research in this important field.

Contents

1 Introduction

1.1. Pest insects

1.1.1. Disease vectors

Insect vectors transmit several major global diseases, causing a large negative impact on global health due to high morbidity and mortality. Many of these diseases are increasing in prevalence, distribution and severity[1-3]. Mosquitoes are the main vectors. Malaria, caused by the *Plasmodium* parasite, transmitted by mosquitoes in the *Anopheles* genus, causes the most fatalities; in 2015 there were an estimated 214 million cases annually and 438,000 deaths[4]. However, numerous other diseases transmitted by mosquitoes and other insect species also impose a significant global disease burden.

1.1.1.1. Aedes aegypti and Aedes albopictus

The disease vectors used in this work were two mosquitoes in the *Aedes* genus, *Ae. aegypti* and *Ae. albopictus* (Figure 1.1). These are both vectors of dengue fever, the most prevalent mosquito-borne viral disease[5]. Dengue fever is caused by an RNA flavivirus with four serotypes (DENV1-4)[6]. The mortality rate is 1-5 %, but secondary infections with a different serotype can lead to dengue haemorrhagic fever, a more severe form of the disease with a mortality rate of around 26 %[7]. Over half the world's population has been estimated to be at risk of dengue fever[8], with an estimated 390 million cases annually, 96 million showing clinical symptoms[9], and 20,000 fatalities[10]. Trials of a candidate vaccine have shown partial effectiveness[11,12] but currently there is no licensed vaccine and no specific treatment[6]. *Ae. aegypti* and *Ae. albopictus* also vector other diseases, including chikungunya, another widespread viral disease for which there is no vaccine or cure[13], and yellow fever, for which there is a vaccine but which still causes an estimated 30,000 deaths annually[14]. *Ae. aegypti* is the primary vector of the above diseases, but *Ae. albopictus* is a

secondary vector of increasing importance given that its cold tolerance[15-17] combined with increased international travel[18] has allowed it to spread beyond the range of *Ae. aegypti* to North America[19] and Europe[20].

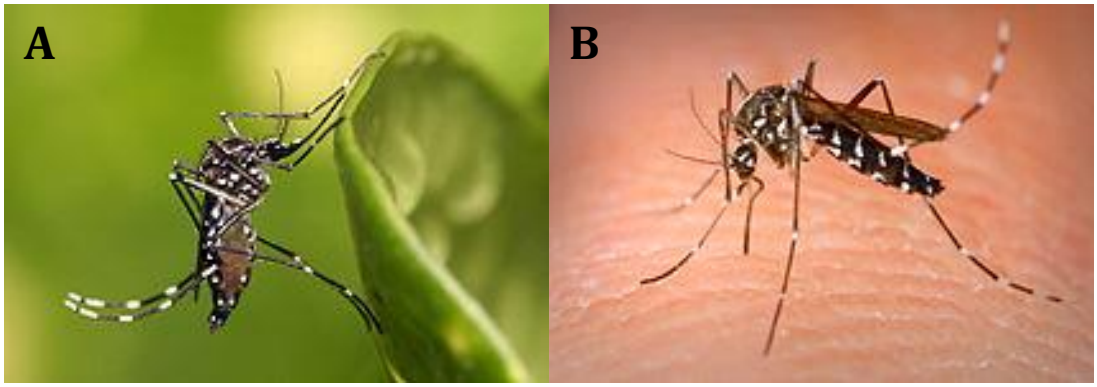

Figure 1.1. *Ae. aegypti* (A) and *Ae. albopictus* (B). Images from Wikipedia.

1.1.2. Agricultural pests

In addition to their role in disease transmission, there are also many insect species that act as agricultural pests, causing huge losses in food production and costing the agricultural industry billions of pounds. It is estimated that insects cause a loss of between 20 % and 40 % of potential food production, with damage occurring during growing, storage, and in consumers' homes[21]. Pest insects cause damage both by feeding on crops, and by spreading plant diseases.

1.1.2.1. Ceratitis capitata

The agricultural pest insect used in this work was the Mediterranean fruitfly (also known as medfly), *Ceratitis capitata* (Figure 1.2). *C. capitata* is one of the most destructive agricultural pests, affecting more than 250 types of crop and being present across the world[22]. Female *C. capitata* lay eggs in fruit; aside from the damage caused by the resulting larvae feeding on the fruit, this leaves holes that make the fruit more vulnerable to infection and rot[22].

Figure 1.2. *C. capitata*. Image from Wikipedia.

1.2. Control of pest insects

Given that there is no vaccine or cure for many insect-borne diseases, control of the insect vectors is currently the only option for disease control. Even if vaccines or specific treatments were available, vector management might be a more cost-effective strategy. Control of pest insects is also of huge importance to the agricultural industry, especially given the increasing demands on food production due to a growing world population.

1.2.1. Current control methods

One of the most widely used methods for the control of both disease vectors and agricultural pests is the use of insecticides. This can be effective in some cases, but its implementation can be logistically difficult and expensive[23], particularly for species such as *Ae. aegypti* that live closely with humans, even breeding inside houses[24]. Population re-expansion or re-invasion can occur rapidly without continued application[25], and development of insecticide resistance is also a serious problem[26]. *Ae. aegypti*, for example, has developed resistance to all common chemical classes of insecticides, such as organophosphates, pyrethroids, carbamates and organocholrines[27]. In addition to problems with implementation and effectiveness, insecticides can have a number of undesirable side effects. They can affect non-target insect species, which may result in

undesirable consequences, and there may also be general environmental toxicity. Furthermore, there are concerns about risks to human health, particularly in an agricultural setting, as insecticides may leave a residue on crops.

Alternative approaches that have been used for mosquitoes include removal of breeding sites[28,29], use of odour-baited traps[30,31], use of biological toxins[32], introduction of mosquito predators[28,33-35] and personal protection such as bednets[36]. However, these strategies can also be expensive, often require continual intervention, can be logistically difficult in urban areas, may be subject to development of resistance, and can have environmental effects. In addition, bednets are ineffective against day-biting species such as *Ae. aegypti* and *Ae. albopictus*.

A genetic approach that has seen some success with agricultural pests, although less so with disease vectors, is the sterile insect technique (SIT). This involves releasing males of the target species that have been sterilised by irradiation. When these males mate with wild females, no viable offspring are produced, leading to a decline in population size[37]. In the case of mosquitoes, it is important that only males are released, as males do not feed on humans and so do not transmit disease, whereas females do. In the case of agricultural pests such as *C. capitata*, male-only release is also advantageous, as it is females that cause crop damage. Release of males only also removes the possibility that the modified males will mate with modified rather than wild-type females[38].

The SIT was used to successfully eradicate the livestock pest New World screwworm[39], and has also been used to control *C. capitata* and other tephritid fruit flies[40-42]. In mosquitoes, it has been successful against *Culex quinquefasciatus*[43] and *Anopheles albimanus*[44], and recently there have been promising results of pilot studies in *Ae. albopictus*[45,46], but it was unsuccessful in *Ae. aegypti*[47].

The SIT has a number of disadvantages, a major problem being damaging effects of the irradiation on the males leading to reduced fitness and mating competitiveness[48], particularly in mosquitoes[49]. Furthermore, levels of sterility were relatively low in some cases, as not all mutations caused by irradiation result in complete sterility[50]. Aside from problems with effectiveness, there are also logistical issues with the SIT in some cases. Sex separation is required prior to release; this can be achieved by pupal sorting in *Aedes* and *Culex* mosquitoes as female pupae are larger than males, but in *Anopheles* mosquitoes there is little pupal size discrepancy, making sex separation difficult[51]. A transgenic line of *An. stephensi* mosquitoes expressing enhanced green fluorescent protein (EGFP) only in males has been developed[52], allowing sex separation with a fluorescence detection device, but this would be expensive to implement on a large scale, and still requires potentially damaging handling of the insects, as does pupal sorting.

1.2.2. Novel control methods

Given the impact of pest insects on public health and agriculture, and the inadequacies of current control methods described above, novel control methods for both disease vectors and agricultural pests are clearly needed. With this in mind, various new strategies are being developed which aim to overcome the problems of current control methods. Like the SIT, these methods use the target insect itself as the control agent, involving release of modified insects that mate with wild insects. They are therefore species-specific like the SIT and so remove the off-target effects of other methods. However, they lack the issues associated with irradiation in the SIT, and have other advantages, such as enabling genetic sex separation, and in some cases are self-sustaining. These methods can be broadly divided into two categories – population suppression and population replacement.

1.2.2.1. Population replacement

Population replacement strategies can be used for disease vectors, and aim to replace the natural population with insects that do not transmit disease, or at least with reduced vectorial capacity. Mosquitoes are the main targets of these strategies.

1.2.2.1.1. Using transgenic methods

One strategy for population replacement involves the introduction of transgenes that render mosquitoes refractory to the disease-causing pathogens. To date, *Ae. aegypti* has been engineered with transgenes conferring resistance to dengue virus[53-55] and *An. stephensi*[56-64] and *An. gambiae*[65,66] have been engineered with transgenes conferring resistance to *Plasmodium*.

1.2.2.1.2. Using Wolbachia

Certain strains of the endosymbiotic bacterium *Wolbachia*, which can infect mosquitoes, display traits that could lead to reduced disease transmission, most notably inhibitory effects against disease-causing pathogens. *Wolbachia* thus have the potential to be used as an alternative to refractory transgenes. This is discussed in more detail later.

1.2.2.1.3. Gene drive

Generation of mosquitoes with reduced ability to transmit disease is only part of a successful population replacement strategy. There must be a way to effect replacement of the wild population with these modified mosquitoes. The transgenes used in transgenic methods are unlikely to confer a strong enough advantage to be spread to fixation within an appropriate timescale, and transgenesis may even be associated with a fitness cost, as found in some studies[61,67,68]. It has been theorised that population replacement with a transgenic strain showing Mendelian inheritance of the transgene would require release of the modified mosquitoes on a scale that would be impractical given the wide

geographical distribution of disease-carrying mosquitoes[69]. Although recent modelling has suggested otherwise[70], strategies would certainly be more effective if they implemented a way to spread the transgenes through the population via super-Mendelian inheritance; this is known as gene drive. Gene drive mechanisms could also be useful in population suppression strategies (see below) in which not all progeny are killed, such as sex distortion, female-specific lethality, or decreasing female fertility; for example gene drive has been used to spread genes disrupting female reproduction in *An. stephensi*[71]. In the case of sex distortion systems targeting the X chromosome or other female-specific sequences, such as that developed in *An. gambiae*, gene drive could actually be achieved simply by inserting the effector gene on the Y-chromosome, or linked to the male-determining locus in species lacking sex chromosomes. Other strategies, however, require more complex mechanisms.

There are several selfish genetic elements that display super-Mendelian inheritance and have been theorised as possible gene drive mechanisms[72]. Initially efforts were made to use transposable elements for gene drive in mosquitoes, but these were unsuccessful[73]. Maternal-effect dominant embryonic arrest (Medea) is a gene drive system found in the beetle *Tribolium castaneum*, in which mothers carrying the *Medea* allele express a toxin in their germline that kills her progeny unless they also carry the *Medea* allele, thus promoting the spread of this allele[74]. A synthetic *Medea* system has been shown to drive population replacement in *Drosophila melanogaster*[75], but to date no system has been developed in mosquitoes. Underdominance is another technique that has been successfully demonstrated in *D. melanogaster*[76,77]. This is a system in which individuals that are heterozygous for a certain gene are less fit than homozygous individuals, resulting in spread of the dominant allele to fixation. Engineered underdominance systems use two genes encoding toxins each linked to a gene encoding an antidote for the other toxin; individuals that do not possess both alleles are inviable. Homing endonuclease genes

(HEGs), which encode endonucleases that cut DNA and cause a copy of the HEG to be inserted by homologous recombination, have been successfully used in *An. gambiae*[78]. More recently, the CRISPR/Cas9 genome editing system has been used for gene drive in *D. melanogaster*[79] and to spread anti-*Plasmodium* genes in *An. stephensi*[80] and genes causing female sterility in *An. gambiae*[71].

In addition to individual genetic elements, some endosymbiotic microorganisms display super-Mendelian inheritance, including *Wolbachia*. Given that certain strains of *Wolbachia* also display anti-pathogen inhibition, as mentioned above, this makes it an attractive overall possibility for population replacement. It has also been theorised that pathogen resistance transgenes could be spread utilising *Wolbachia*, or components of the cytoplasmic incompatibility (CI) mechanism it employs to promote its spread in mosquitoes[69,81,82]. This is discussed in more detail later.

1.2.2.2. Population suppression

Population suppression strategies aim to reduce the numbers of pest insects. For disease vectors, population suppression could be used alone or in combination with population replacement. For agricultural pests that directly damage crops it is the only option.

1.2.2.2.1. Using transgenic methods

Several population suppression strategies in development are analogous to the SIT, but use synthetic transgenic constructs to induce sterility rather than irradiation, which can improve male fitness and allows more flexibility. Prevention of sperm development is one such approach, and has been achieved in *An. gambiae*[83]. Generation of modified sperm that are incapable of viable fertilisation is another approach, and has also been achieved in *An. gambiae*[84].

Another promising strategy under development in both mosquitoes and agricultural pests is Release of Insects carrying a Dominant Lethal (RIDL). There are two versions of RIDL, known as bisex RIDL and female-specific RIDL. Bisex RIDL is similar in concept to the SIT but with transgene expression occurring during development of the offspring of modified males rather than in the germline of the modified males. In this system, modified males carry a dominant lethal gene that is passed onto offspring from matings with wild females, resulting in their death. Unlike the suppression methods above, this provides the opportunity for transgenes to interact with genes from the wild population, meaning that it may be easier for resistance to develop. However, it has the advantage that expression timing of the transgene can be varied. Death at late larval or pupal stages has the advantage that offspring of modified males will contribute to larval competition, which may enhance population suppression[85,86]. However, for agricultural pests death at a late stage is less desirable, as larval stages cause crop damage.

Expression of the lethal gene must be repressible to allow rearing of the line. This is currently achieved using the TetOff system[87,88], developed from components of the gene expression response to the antibiotic tetracycline in bacteria. In the two component version of this system (Figure 1.3), the lethal gene is placed under the control of a minimal promoter and the regulatory sequence tetO. Transcription of the gene occurs when tetO is bound by the tetracycline-repressible transactivator (tTA), a protein generated by fusion of the tetracycline repressor found in bacteria to the viral transcriptional activation VP16 domain[89]. This binding is blocked by tetracycline. Thus, insects carrying the gene can be reared successfully using a diet supplemented with tetracycline. A simplified one component version of this system has been developed in which tTA itself acts as the lethal gene[90] (Figure 1.4). In the absence of tetracycline, the minimal promoter produces a basal amount of tTA, which then promotes further tTA expression by binding to tetO, resulting in a positive feedback loop. Death is thought to occur by transcriptional squelching[91,92].

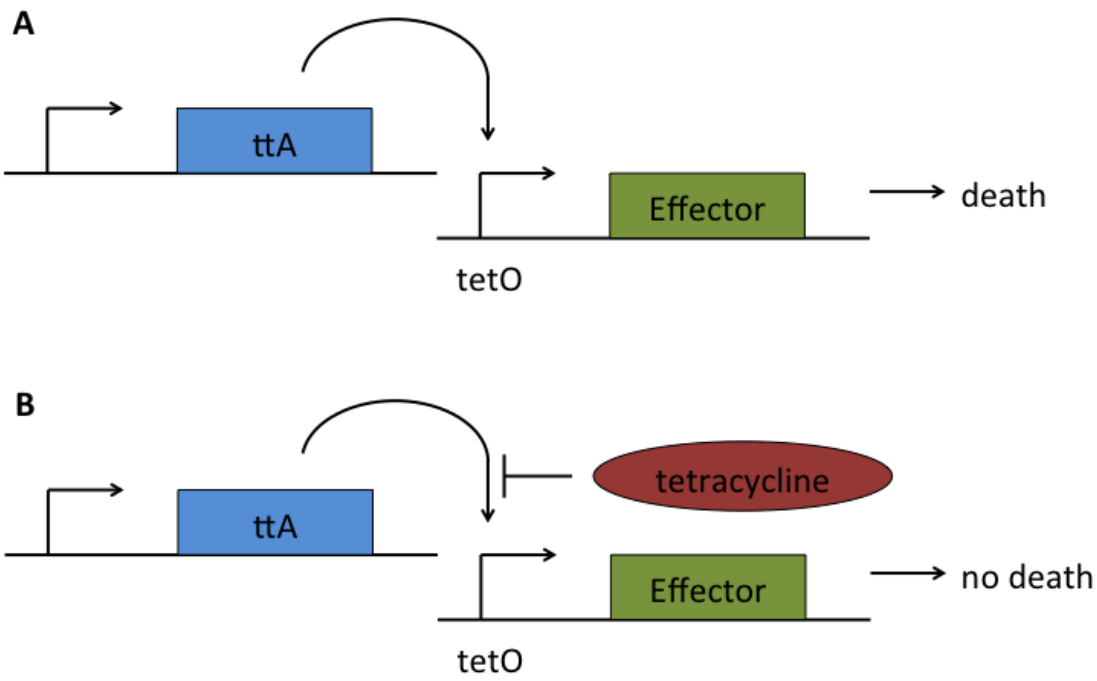

Figure 1.3. Two component RIDL. (A) In the absence of tetracycline, tTA is expressed and binds to tetO, driving expression of the lethal effector gene. (B) In the presence of tetracycline, tTA is expressed, but binding of tTA to tetO is blocked, preventing expression of the lethal effector gene.

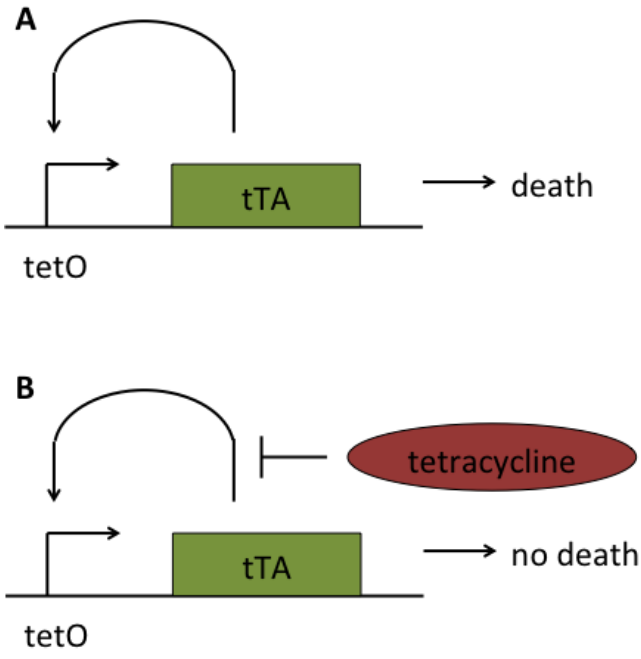

Figure 1.4. One component RIDL. (A) In the absence of tetracycline, basal tTA expression is driven by a minimal promoter. This tTA binds to tetO, driving its oxn expression. This positive feedback cycle results in lethal amounts of tTA. (B) In the presence of tetracycline, basal levels of tTA are expressed, but binding of tTA to tetO is blocked, preventing the positive feedback cycle.

Bisex RIDL has been implemented in *Ae. aegypti*[86], and has shown success in suppressing wild populations in field releases in Grand Cayman[93,94], Malaysia[95] and Brazil[96].

Female-specific RIDL functions as described above for bisex RIDL, except with an effector gene affecting only females. This has the advantage of allowing genetic sexing – removal of tetracycline in the last generation prior to release results in the generation of only viable males. This overcomes the problems with other methods of sex separation discussed above. Female-specific RIDL has been implemented in *Ae. aegypti*[97], *Ae. albopictus*[98], *C. capitata*[99], *Bactrocera oleae* (olive fly)[100], *Plutella xylostella (*diamondback moth)[101] and *Pectinophora gossypiella* (pink bollworm)[101]. In the mosquito lines, females are rendered flightless rather than being directly killed; however, this phenotype incapacitates the females by restricting access to sugar resources and the ability to escape from predation to such an extent that the effects are equivalent to lethality[98]. In addition, flightless females are effectively sterile as courtship and mating require wing oscillations[102], and flight is also required to acquire a blood meal[98].

Other systems that are conceptually similar to female killing are sex distortion systems, which result in the production of all or mostly male offspring from crosses of wild females with modified males. These systems either involve the prevention of development of female offspring, or conversion of genetic females to phenotypic males. Both result in a reduction in the number of females, leading to fewer matings, and thus the population declines. The former has been achieved in *An. gambiae*, using endonucleases that shred the X chromosome in sperm, and population suppression demonstrated[103]. Conversion of genetic females to phenotypic males has recently been achieved in *Ae. aegypti*[104], but is a way from application. As for female-specific RIDL, conversion of genetic females to phenotypic males would eliminate the need for sex separation. It would also increase the efficiency of mass rearing; current female-specific RIDL systems are late-acting, so

resources are still spent rearing females in the last generation before release that serve no purpose. If these females were converted to males the number of males for release would in theory be doubled for the same resources.

1.2.2.2.2. Using Wolbachia

In addition to transgenic synthetic constructs, population suppression could potentially be achieved by a method known as the Incompatible Insect Technique (IIT). This involves releasing male insects infected with *Wolbachia*, which produce inviable offspring upon mating with wild uninfected females or females infected with a different strain. This is due to CI, and is discussed in more detail below.

1.3. *Wolbachia*

Wolbachia is a genus of maternally inherited intracellular bacteria that were first identified in *Culex* mosquitoes[105]. They are very widespread, with estimates of the proportion of insect species infected as high as 66 %[106]; other arthropods and nematodes are also infected[107]. Mosquito genera infected include *Culex* and *Aedes*[108-110]; no natural infections had been reported in the *Anopheles* genus until a recent finding of *Wolbachia* in *An. gambiae*[111]. Several artificial transinfections have been generated, including in *An. stephensi*, which is naturally uninfected[112], *Ae. aegypti*[113-116], which is naturally uninfected, and *Ae. albopictus*[117-123] which has a natural superinfection of two strains[124,125].

Wolbachia display characteristics that give them potential to be used in both population suppression and population replacement strategies, mentioned above. Firstly, *Wolbachia* can induce reproductive manipulations in their hosts that confer a selective advantage on infected females, thus promoting the spread of the bacteria via maternal inheritance[126]. These manipulations include parthenogenesis, in which infected females produce only daughters, feminisation, in which genotypic males are transformed into phenotypic

females, and male killing, in which infected male embryos die while females develop normally[127]. The most common manipulation, and the only reproductive phenotype observed in mosquitoes, is CI[107,127], discussed further below. In addition to reproductive manipulation, *Wolbachia* can also display pathogen inhibition, also discussed further below.

1.3.1. Cytoplasmic incompatibility

In its simplest form, known as unidirectional, CI is a phenomenon in which uninfected females produce inviable offspring when mated with *Wolbachia*-infected males, while *Wolbachia*-infected females produce viable offspring when mated with both infected and uninfected males (Figure 1.5). Infected females therefore produce a greater number of offspring, and because *Wolbachia* are maternally inherited, in a previously naïve population the frequency of *Wolbachia* increases with each generation until it reaches an equilibrium, which can be close to fixation[106].

CI can also occur between hosts infected with different strains of *Wolbachia*. In this case, the CI can be either unidirectional or bidirectional. When it is unidirectional, females infected with one of the strains are compatible only with males infected with the same strain, but females infected with the other strain are compatible with any males. This is equivalent to the unidirectional CI described above in which only one strain of *Wolbachia* is present. In bidirectional CI, females infected with either strain are only compatible with males infected with the same strain (Figure 1.6). In a freely interbreeding population, females infected with the most prevalent *Wolbachia* strain have a reproductive advantage and this strain should spread[128,129], though differences in fitness effects, speciation resulting from the incompatibility and stochastic effects may impact this.

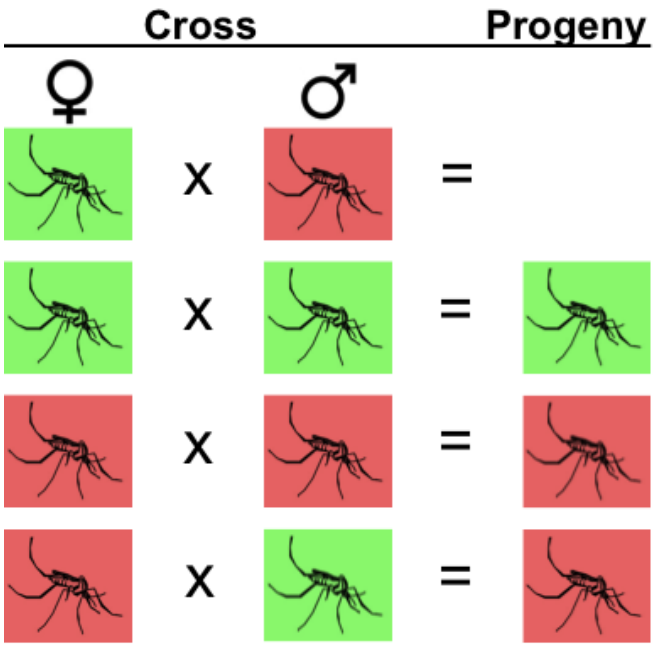

Figure 1.5. Unidirectional cytoplasmic incompatibility. Green = uninfected; red = *Wolbachia*-infected. Matings between uninfected females and infected males are incompatible, producing inviable offspring. Uninfected females are only compatible with uninfected males. In contrast, infected females are compatible with both infected and uninfected males, producing infected offspring in both cases.

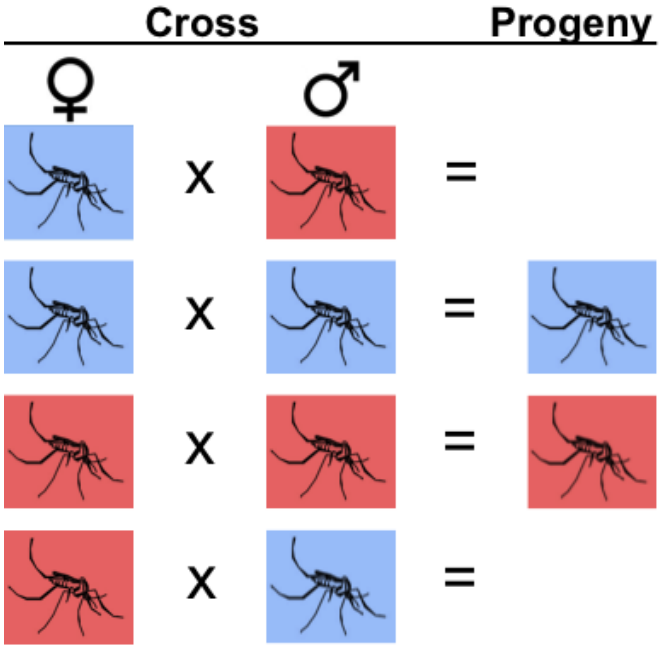

Figure 1.6. Biidirectional cytoplasmic incompatibility. Blue = *Wolbachia* strain 1; red = *Wolbachia* strain 2. Matings between males and females with different *Wolbachia* strains are incompatible, producing inviable offspring. Matings between males and females with the same *Wolbachia* strain are compatible, producing infected offspring in both cases.

The penetrance of CI may be complete, i.e. no embryos from incompatible crosses develop, or partial, i.e. a percentage of embryos from incompatible crosses develop normally. In some insect species, such as *Drosophila*, CI penetrance of natural infections is often relatively low[130-132], but incompatible crosses with naturally infected mosquitoes often show complete sterility[133,134]. Penetrance in transinfections is variable. Most show complete or almost complete penetrance, but some display lower penetrance than that for the natural host of the transinfected *Wolbachia* and/or the natural *Wolbachia* infection in the transinfected species[114,119].

1.3.1.1. Molecular mechanisms of CI

The exact molecular mechanisms of CI have not been elucidated, but are thought to fit what is known as the 'modification – rescue' ('mod-res') model[135]. According to this model, *Wolbachia* infection in males leads to a modification in sperm that prevents normal development of embryos arising from matings with these males, unless the female is infected with a compatible *Wolbachia* strain, in which case 'rescue' occurs in the fertilised ovum. *Wolbachia* can be classified based on their ability to perform the modification and rescue functions. Strains that can both induce and rescue CI, termed mod+/res+, are the most common[136], but all combinations of presence and absence of modification and rescue have been observed[135,137-139] . The fact that CI can occur in matings between males and females infected with different *Wolbachia* strains suggests that the precise nature of modification and/or rescue functions varies between *Wolbachia* strains[140]. In addition, *Cx. pipiens* embryos from different incompatible crosses were shown to develop to varying degrees[141], again suggesting that modification and rescue are likely to be a complex set of functions rather than single straightforward effects.

1.3.1.1.1. Mechanisms of modification

The modification in sperm results in severe disruption of the first mitotic division following fertilisation[136]. Paternal chromosomes in embryos from incompatible crosses show defects in condensation[142], and in *D. simulans* the maternal chromosomes were observed to enter the metaphase/anaphase transition while the paternal chromosomes did not segregate and were arrested in anaphase[143].

In prophase, prior to condensation defects, a delay in nuclear envelope breakdown and cyclin-dependent kinase 1 (Cdk1) activation has been observed in the male pronucleus in *Wolbachia* infection[144]. As mutations in cell cycle checkpoint proteins can lead to impaired chromosome condensation[145], this suggests that disruption of the cell cycle may contribute to the defects in paternal chromosome condensation observed in *Wolbachia* infection. This disruption could be caused either by direct interference with cell cycle proteins, or it could be caused indirectly via DNA damage, which leads to arrest at cell cycle checkpoints[146]. Consistent with the latter hypothesis, numerous studies have found markers of DNA damage in *Wolbachia* infection. For example, paternal chromosomes have been found to become fragmented during the first mitotic division[147,148]. In *Drosophila* studies have found chromosome bridging during anaphase[143,148,149], which indicates damaged or incompletely replicated DNA. One study found that levels of a marker of oxidative DNA damage were higher in *Wolbachia*-infected *Ae. albopictus* cells than in uninfected cells, and that there are an increased number of DNA strand breaks in meiotic spermatocytes in *Wolbachia*-infected *D. simulans*[150].

Another study has suggested that the defects in chromosome condensation and delay in Cdk1 activation are due to impaired deposition of histones H3.3 and H4 in the male pronucleus in interphase[151]. This study also found that in the pronuclei of sperm from *Wolbachia*-infected males there is prolonged retention of the replication factor

Proliferating Cell Nuclear Antigen (PCNA) into metaphase, indicating progression to mitosis with incompletely replicated DNA.

In addition to the characteristics discussed above, *Wolbachia* infection has been found to result in an excess of centrosomes in the fertilised egg that are not associated with the male and female pronuclei[143,149]. One cause of this phenotype is the absence of mitotic kinases which co-ordinate the centrosome replication cycle[152], suggesting that interference with these kinases may play a role in the CI modification. In some cases of *Wolbachia* infection, absence of centrosomes has been observed[136]. These findings suggest that *Wolbachia* may disrupt the association of male pronuclei with centrosomes[136]. Interference with this association has been shown to result from defects in chromosome condensation and cell cycle delays[149], thus the centrosome abnormalities observed in *Wolbachia* infection may be a consequence of the condensation and cell cycle defects described above.

It is of note that the embryonic defects associated with CI are not caused directly by *Wolbachia*, as the bacteria are not present in mature sperm[153,154]; rather *Wolbachia* must modify the sperm during development in a way that leads to effects after fertilisation. Such a modification could in theory be mediated via a diffusible factor secreted by *Wolbachia* into the sperm, or alternatively via a molecular imprint on the paternal chromatin[136]. In a study using gynogenetic *Drosophila* females[155], which produce diploid embryos that require extranuclear paternal factors for development but not a male pronucleus, crosses between uninfected gynogenetic females and *Wolbachia*-infected males resulted in viable progeny, suggesting that CI is due to an effect on the male pronucleus, consistent with the observed nuclear defects discussed above.

1.3.1.1.2. Mechanisms of rescue

If a female mating with a *Wolbachia*-infected male is infected with a compatible strain, rescue occurs and viable progeny are produced. On a cellular level, the paternal chromosomes condense and segregate normally during anaphase[143,144,147,149]. As for modification, the molecular mechanisms of rescue are poorly understood. However, a number of models have been proposed, which fall into three main classes[136].

In the first class of models, rescue is caused by a direct effect on the male pronucleus. An example is the 'lock and key' model; this hypothesises that *Wolbachia* in sperm produce a factor that binds to the male pronucleus (the 'lock'), which is removed by a factor produced by maternal *Wolbachia* (the 'key')[135,156,157]. In contrast, the 'titration-restitution' model proposes that paternal *Wolbachia* remove an essential factor from the male pronucleus which is restored by maternal *Wolbachia*[135,157,158].

In the second class of models, rescue is caused by compensatory modification of the female pronucleus. An example is the 'mistiming' model, which proposes that embryonic lethality is due to an effect on the male pronucleus causing asynchrony between the male and female pronuclei, and rescue is achieved via a compensatory change in timing of the female pronucleus[143,144,159].

The third class of models postulates that maternal *Wolbachia* induce rescue by affecting cytoplasmic events. A modified version of the 'mistiming' model is an example in this category, synchrony being restored in this case not by effects on the female pronucleus but by effects on cytoplasmic processes such as spindle assembly[136].

Cytological studies have been performed in attempts to determine the mechanisms of rescue, but so far there is little evidence to suggest which one of the three classes of model

is correct, let alone the molecular mechanisms involved. It has been demonstrated that rescue restores the synchrony of nuclear envelope breakdown and Cdk1 activation[144], but this could be due to a restoration of normal timing in the male pronucleus or a compensatory delay in the female pronucleus and so this finding cannot distinguish between the first two classes of models described above. An experimental distinction can be made between the 'lock and key' and 'titration' models and the 'mistiming' models, as the former suggests that the male pronucleus is incapable of normal development without rescue, while the latter suggests that the pronucleus could develop normally, it is just not synchronised with the female pronucleus. To distinguish between these models the number of androgenetic progeny (those that develop solely from the male pronucleus) arising from *Drosophila* crosses with disrupted pronuclear fusion was compared between compatible and incompatible crosses[160]. It was found that CI significantly reduced the frequency of androgenetic progeny, suggesting that in CI the male nucleus is rendered incapable of normal development without rescue.

1.3.1.2. Genes involved in CI

Several studies have attempted to identify genes involved in CI, using genomic, transcriptomic and proteomic analyses.

1.3.1.2.1. Genomic analysis

Genomic analysis has been facilitated by the genome sequencing of several strains of *Wolbachia*, including *w*Mel from *D. melanogaster*[161], *w*Ri[162], *w*Ha[163] and *w*No[163] from *D. simulans*, *w*Rec from *D. recens*[164], *w*Pip from *Cx. pipiens*[165,166], *w*AlbB from *Ae. albopictus*[167], *w*Bm from *Brugia malayi*[168], *w*Di from *Diaphorina citri*[169], and *w*DacA and *w*DacB from *Dactylopius coccus*[170]. Analysis of these genomes has revealed candidates for genes involved in CI.

Wolbachia genomes contain a large number of genes encoding proteins with ankyrin repeat (ANK) domains. ANK domains mediate protein-protein interactions[171], and while common in eukaryotes they are relatively rare in bacteria[172]. This suggests that they may play a role in *Wolbachia*-specific behaviour such as CI. Further evidence for a role of ANK proteins in CI is that of the *Wolbachia* genomes sequenced, the largest number of ANK genes was found in the genome of *w*Pip[165], the strain that results in the most complex CI crossing patterns observed[173,174]. In addition, variation in ANK genes was found between incompatible strains of *Wolbachia* in *Cx. pipiens*, along with differential expression in male and female mosquitoes[175,176], and ANK genes were found to differ between *w*Mel and *w*Au, a closely related *Wolbachia* strain that does not induce CI[177]. However, a large study in *Cx. pipiens* failed to reveal a correlation between ANK gene variability and CI crossing type[174], and transformation of *D. melanogaster* with various ANK genes did not induce or modify CI phenotypes[178], failing to support the hypothesis that ANK genes are involved in CI.

Bacteriophage-related genes have also been proposed as CI candidates. Several prophage regions have been identified in the sequenced genomes of *Wolbachia* strains that infect arthropods[161,162,165], with phages in the WO-B family being active. In contrast, the genome of the *Wolbachia* strain *w*Bm, which infects the nematode *B. malayi* and does not cause CI, contains no prophage regions[168], supporting a role of phage genes in CI. In addition, it was found that CI characteristics could be transferred to uninfected hosts by injecting egg cytoplasm from infected hosts after filtering through a membrane with a pore size that would exclude *Wolbachia*, suggesting that a virus particle can carry the factors responsible for CI[179]. However, several studies found no correlation between WO phylogeny and CI characteristics in *Cx. pipiens*[180-184] and transformation of *D. melanogaster* with several phage genes did not induce or modify CI phenotypes[178]. Genome sequencing showed that the *w*Rec strain in *D. recens*, which does cause CI, lacks some WO genes, and thus at least in this interaction the absent phage genes can be excluded as CI candidates[164].

1.3.1.2.2. Transcriptomic analysis

It is possible that in addition to *Wolbachia* genes, host genes may also be involved in CI. Several studies provide evidence for this hypothesis. For example, the *Wolbachia* strain *w*CauA induces CI in the host *Cadra cautella*, the almond moth, but when transferred into *Ephestia kuehniella*, the Mediterranean flour moth, a male killing phenotype is induced[185]. Similarly, the strains *w*MelPop and *w*Tei induce CI when transferred to *D. simulans*, but not in their native hosts[139,186]. In mosquitoes, bidirectional CI in two strains of *Cx. pipiens* was reduced by backcrossing females infected with one *Wolbachia* strain so that their nuclear genomes were almost identical to those of males infected with the other *Wolbachia* strain[175].

Potential host genes involved in CI can be identified by comparing gene expression in uninfected and *Wolbachia*-infected hosts, or between hosts infected with incompatible strains of *Wolbachia*. Such comparisons have been made in *Drosophila*, firstly in an *in vitro* system, by performing microarrays on *Wolbachia*-infected and uninfected *Drosophila* cell cultures, and subsequently *in vivo* by analysing expression of the genes found to be differentially expressed *in vitro*[187]. This study identified the gene for Angiotensin Converting Enzyme (*Ance*) as being differentially expressed in infected and uninfected hosts, with differential expression also observed between male and female flies. In addition, mutating the gene affected CI. This gene is required for spermatogenesis in *Drosophila*[188]; the fact that it has a reproductive function, combined with the above experimental results, suggest that *Ance* is involved in CI. *Wolbachia* infection in *D. simulans* was found to upregulate the non-muscle myosin II gene *zipper*, and when this gene was overexpressed in uninfected *D. melanogaster* an effect mimicking CI was produced[189]. The *Drosophila* gene *Hira* was found to be downregulated in *Wolbachia* infection, with the degree of downregulation correlated with the strength of CI, and mutation of this gene produced a CI-like effect[190]. Another study identified a number of candidates differentially

expressed between *Wolbachia* and uninfected testes and involved in spermatogenesis, including *Mst484Db* and *Juvenile hormone-inducible protein 26* (*JhI-26*)[191]. A CI-like phenotype was induced upon overexpression of *JhI-26* and the gene *CG10433*, encoding a male accessory protein whose expression is upregulated by JhI-26[192]. Further work and similar studies in mosquitoes will hopefully lead to the identification of more host genes involved in CI.

Potential *Wolbachia* genes involved in CI could also be identified using transcriptomic analyses, but prior to the work here no such analysis had been performed, likely due to difficulties in obtaining sufficient pure *Wolbachia* RNA.

1.3.1.2.3. Proteomic analysis

In addition to genomic and transcriptomic analyses, proteomics studies have revealed candidates for involvement in CI. *Wolbachia* proteins WPIP0282[193] and Hu beta[194], a DNA-binding protein, were identified as candidates based on detection in mosquito spermathecae and gonads. From the host perspective, a recent study suggests that CI modification may involve changes in seminal fluid proteins (Sfps) rather than or in addition to changes in sperm[195]. Sfps were found to be differentially expressed between extracts from spermathecae and seminal receptacles of *D. melanogaster* females mated with *w*Mel-infected or uninfected males, and knocking down two Sfp genes resulted in a partial CI phenotype in uninfected males.

1.3.1.3. Use of CI in insect control

As mentioned previously, CI could be utilised in different ways for insect control, for both population replacement and population suppression strategies. In population replacement strategies, it has been proposed as a gene drive mechanism to spread pathogen resistance

transgenes[69,81,82]. It also provides a convenient way to spread *Wolbachia* themselves in strategies making use of the disease refractory properties of *Wolbachia*.

There are in theory a variety of possible approaches to using CI to spread refractory transgenes, each with strengths and weaknesses. One theoretical strategy would be to genetically transform *Wolbachia* with pathogen resistance transgenes, then release mosquitoes carrying the transformed *Wolbachia* into the wild. Mathematical modelling indicates that this strategy would require a relatively small release of transgenic individuals[81]. *Wolbachia* have been successfully introduced into both infected and naïve mosquito species, as discussed previously, and have been shown to spread in laboratory, semi-field, and field populations[113,116,196]. However, *Wolbachia* genetic transformation is technically challenging, complicated by the inability to culture *Wolbachia* outside its host. Although methods for transformation have been suggested[197,198], it has not to date been accomplished. In addition, although cultured cell[199] and tissue[200] and transient somatic[201,202] infections have been achieved, and a natural infection has recently been found[111], it has so far not been possible to establish a stable *Wolbachia* transinfection in *An. gambiae*, which as the main vector of malaria is one of the most important species for vector control. Given the difficulties, this strategy seems to have little promise for application.

A similar strategy but one that overcomes the difficulties with *Wolbachia* transformation is to transform another maternally inherited cytoplasmic element, such as another obligate endosymbiont or a virus, and release individuals of the target species co-infected with this element and *Wolbachia*. The transformed cytoplasmic element should invade the target population together with the *Wolbachia*. As with direct transformation of *Wolbachia*, modelling suggests that this approach would require a relatively small release size[81]. Obligate endosymbionts have been successfully transformed[203], and evidence for

the feasibility of this approach for spreading transgenes is the concurrent spread of an mtDNA variant and the *w*Ri *Wolbachia* infection in *D. simulans* in California[204]. However this approach still requires generation of *Wolbachia* infections in host populations, which as discussed above may not be feasible in all cases.

An alternative strategy that eliminates the requirement for *Wolbachia* altogether is to transform the host nuclear genome with both a transgene for pathogen resistance and the gene(s) responsible for CI. Importantly this method would allow control of *An. gambiae*. However, it would require a large release of transgenic individuals[81] that would be impractical in most cases[205]. A variation of this approach which might reduce the release size required has been suggested for populations already containing *Wolbachia*, in which the pathogen resistance transgene would be linked to the gene(s) that restore compatibility of females with infected males[82]. While the germline transformation of several mosquito species has been accomplished[206-208] and is now commonly performed, these nuclear transformation strategies require the identification of the gene(s) responsible for CI, which so far has not been achieved.

With regards to population suppression strategies, CI also underpins the IIT, mentioned previously. This involves release of *Wolbachia*-infected males that are incompatible with the wild females; it is effectively a *Wolbachia*-mediated version of the SIT. This method has shown success in field trials with *Cx. quinquefasciatus*[128,209] and *Ae. polynesiensis*[210,211], but has yet to be fully explored in other species. Using CI to induce sterility has the advantage over traditional SIT that no irradiation is required. However, for naturally *Wolbachia*-free target species the potential for inadvertent release of females is a major drawback, as it could lead to invasion of the target population by *Wolbachia*, meaning that the released males would no longer be incompatible with wild females, and there is a possibility that the new infected population could be more competent vectors[212]. It may be possible to

combine this method with genetic sexing, discussed previously. Alternatively irradiation may be used to sterilise females[213-215]; while this may have some detrimental impact on fitness, females require less irradiation for sterilisation than males, so the effects are not as large as in traditional SIT.

For strategies in which population replacement or suppression is dependent on CI, knowledge of the mechanism(s) of CI is of central importance, and in some cases a requirement. In addition, population replacement strategies based on the disease-refractory properties of *Wolbachia* may require successive population invasions with different, incompatible *Wolbachia* strains or strain combinations, so improved understanding of CI would also be beneficial to these approaches.

1.3.2. Pathogen inhibition

Wolbachia has been found to directly induce resistance to a variety of pathogens, which gives it potential for use in population replacement strategies. In *Drosophila*, native *Wolbachia* strains confer resistance to *Drosophila* C virus, cricket paralysis virus, Flock house virus and Nora virus[216-218] and the fungus *Beauveria bassiana*[219]. In mosquitoes, native *Wolbachia* strains have been found to confer resistance to RNA viruses, such as West Nile virus (WNV) in *Cx. quinquefasciatus* and *D.melanogaster*[220], and to a degree, dengue virus in *Ae. albopictus*[221].

Wolbachia in artificial transinfections have also been shown to inhibit pathogens. The *Ae. albopictus* strain *w*Alb confers some resistance to dengue virus in *Ae. aegypti*[222], and to *P. falciparum* in *An. stephensi*[112]. The virulent *D. melanogaster* strain *w*MelPop was shown to inhibit dengue virus, chikungunya, *Plasmodium*[223], WNV[224] and the filarial nematode *B. pahangi*[225] in *Ae. aegypti*, and *P. gallinaceum*[202] and *P. falciparum*[226] in *An. gambiae*. Dengue resistance in *Ae. aegypti* has also been shown to be induced by the less virulent *Wolbachia*

strain *w*Mel[116], which also inhibits chikungunya and yellow fever viruses[227]. As well as reducing infection frequency and virus titres in *Ae. aegypti*, this transinfection also lengthens the extrinsic incubation period, the time taken for virus to reach the saliva following an infected bloodmeal, which reduces disease transmission as the mosquito is infectious for less of its lifespan[228]. This transinfection is the most advanced in development as a population replacement method, having demonstrated stability and success in reducing disease transmission in field trials[196,229,230].

Pathogen inhibition is dependent on the *Wolbachia*-host combination. For example, *w*MelPop but not *w*Mel was found to inhibit WNV in *Ae. aegypti*[224], and *w*AlbB inhibits dengue virus to a much higher degree in *Ae. aegypti* than in its native host *Ae. albopictus*[231]. Some *Wolbachia*-host combinations may actually lead to increased susceptibility to pathogens of *Wolbachia*-infected hosts relative to uninfected hosts; for example, the natural *w*Pip infection in *Cx. pipiens* may slightly increase susceptibility to *P. relictum*[232], though the lines compared were not properly backcrossed to eliminate host effects. Somatic infections of *w*AlbB, which has inhibitory effects in other hosts, can slightly enhance susceptibility to WNV in *Cx. tarsalis*[233] and to *P. berghei* in *An. gambiae*[234], though the relevance of these data to stable inherited infections is unknown.

Several studies suggest that density is another factor affecting pathogen inhibition. The degree of antiviral protection conferred by different *Wolbachia* strains in *Drosophila* positively correlated with their density, and reducing the abundance of the high density *w*Au strain removed its protective effects[235]. Similarly, antiviral resistance in *Drosophila* correlated with *Wolbachia* density in variants of the single *Wolbachia* strain *w*Mel[236]. Resistance to WNV was found to correlate with *Wolbachia* density in different populations of *Cx. quinquefasciatus* and *Cx. pipiens*[237]. The greater inhibition of dengue by *w*AlbB in *Ae. aegypti* than *Ae. albopictus* was suggested to be a result of the lower density of wAlbB in

Ae. albopictus – more inhibition was found at higher densities in *Ae. albopictus* cells[231]. As well as density, tissue distribution of *Wolbachia* may play a role in pathogen inhibition. For example, the natural infections in *Ae. albopictus* and *Ae. fluviatilis*, which both transmit viruses, are more restricted in their distribution than inhibitory *Wolbachia* such as *w*MelPop[238].

1.3.2.1. Molecular mechanisms of pathogen inhibition

The molecular mechanisms of *Wolbachia*-mediated pathogen inhibition are currently unclear. Several hypotheses have been proposed. Experimental studies have yielded conflicting evidence, and it may be that different mechanisms function in different *Wolbachia*-host combinations. The theories are not mutually exclusive, and there may be multiple mechanisms involved.

1.3.2.1.1. Immune upregulation

One mechanism that has been proposed for pathogen inhibition is upregulation of the host immune system by *Wolbachia*. Several studies have yielded evidence consistent with this. For example, expression of several immune genes was found to be upregulated by *w*MelPop in *Ae. aegypti*[223,225] and in a somatic *An. gambiae* infection[202]. It might be expected that *Wolbachia* would activate antibacterial rather than antiviral pathways, but some antimicrobial peptides are common to both pathways, such as cecropins, some of which were highly upregulated by *w*MelPop[225].

In addition to antimicrobial peptides, reactive oxygen species (ROS) also function as insect immune effectors[239], and have been suggested to play a role in *Wolbachia*-mediated pathogen inhibition. ROS are produced during oxidative phosphorylation, and can cause damage to proteins, lipids, and nucleic acids, known as oxidative stress[240]. They have been shown to confer resistance to various bacteria, fungi and *Plasmodium*[240]. Less is known on

whether ROS have a role in insect antiviral immunity, though increased oxidative stress was observed following viral infection in two lepidopteran insect pest cell lines[241], and they have been implicated as a viral defence in other species[242].

Upregulation of ROS by *Wolbachia* was found in naturally infected *Ae. albopictus* cell line[243], and in a transinfection of *w*Alb in *Ae. aegypti*[244]. In the latter study, increased ROS production was linked to activation of the Toll pathway and its activation of cecropins and defensins, another class of antimicrobial peptide. RNAi knockdown of these peptides reduced inhibition of dengue virus, directly suggesting a role for ROS in *Wolbachia*-mediated pathogen inhibition. Consistent with this, a recent study showed a positive correlation of oxidative stress and antiviral protection in *Wolbachia*-infected *Drosophila*[245].

In addition to effects on the humoral immune response, *w*Mel and *w*MelPop in *D. melanogaster*, *D. simulans* and *Ae. aegypti* were found to upregulate melanisation[246], a component of the cellular immune response.

These studies provide evidence for a role of immune upregulation in *Wolbachia*-mediated pathogen inhibition. However, immune upregulation has mostly been observed in artificial transinfections, with the exception of elevated ROS production in *Drosophila* infected with protective *Wolbachia* strains[247]. In naturally *w*Mel-infected *D. melanogaster*, another study found no upregulation of any of the three humoral response pathways or their associated antimicrobial peptides, but antiviral protection was still observed[248]. Antibacterial immune pathways were activated by transinfections of *w*MelPop and *w*Mel in *Ae. aegypti*[249], but not by the natural *w*Mel infection in *D. melanogaster*[245]. More generally *w*Mel and *w*MelPop were found to induce immune upregulation in *Ae. aegypti* but not *D. melanogaster*, despite conferring resistance to dengue virus in both hosts[250]. Furthermore,

immune upregulation induced by transinfected *w*Mel was much lower in *Ae. albopictus* than *Ae. aegypti*, despite both transinfections showing viral inhibition, with findings suggesting this was due to immunotolerance of the naturally infected *Ae. albopictus* to *Wolbachia*[122]. Taken together, these studies suggest that immune upregulation may not be a result of *Wolbachia* presence per se, but rather the introduction of a novel microbe. This suggests that other mechanisms are responsible at least in some cases.

1.3.2.1.2. Resource competition

The other main theory for the mechanism of *Wolbchia*-mediated pathogen inhibition is competition for resources between the pathogen and *Wolbachia*. Consistent with this is the strong inhibition shown by *w*MelPop; as it replicates more than other strains it is likely to consume more resources. Similarly, the correlation of *Wolbachia* density with pathogen inhibition, discussed above, is also consistent with the resource competition hypothesis. Resources proposed to be subject to competition include cholesterol and fatty acids. Cholesterol is required for *Wolbachia* replication[251] and possibly survival[252], as well as for several functions in insects[253-255]. Both insects and *Wolbachia* lack the biosynthetic pathways to produce cholesterol, so both depend on and compete for dietary cholesterol. Host cholesterol is also required for cell entry and replication of many viruses, including WNV[256] and dengue virus[257] in mosquitoes. Consistent with a role for cholesterol competition in *Wolbchia*-mediated pathogen inhibition, cholesterol-enriched diets reduced antiviral protection in *Wolbachia*-infected *D. melanogaster*[258]. Other studies have suggested that fatty acids are sequestered by *Wolbachia*[223,238,259,260]; these are thought to be essential for viral replication at least in dengue virus[261] and so competition between *Wolbachia* and viruses could lead to viral inhibition.

Wolbachia-mediated pathogen inhibition may also involve indirect competition for resources, through the modulation of autophagy. This is a pathway responsible for

degradation and recycling of cellular components. It has been shown to be required for replication of dengue virus, by allowing degradation of lipid droplets to fatty acids, which are thought to be used for energy by the viral replication machinery[259,260] and for virus particle assembly[262]. Replication of chikungunya virus has also been found to be promoted by autophagy[263]. *Wolbachia* in filarial nematodes and *Drosophila* was found to activate autophagy, and activation of autophagy by rapamycin in *Ae. albopictus* cells reduced *Wolbachia* density[264]. These findings are consistent with a possible competition between *Wolbachia* and viruses in the activation of autophagy and the resources resulting from this. Also consistent with this theory is the lack of inhibition by *Wolbachia* of the DNA virus Insect Iridovirus VI (IIV-6) in *Drosophila*, despite inhibition of a range of RNA viruses[217]. Unlike RNA viruses, replication of most DNA viruses occurs in the nucleus, so is less likely to require autophagy.

1.3.2.1.3. Other mechanisms

Other theories for the mechanism pathogen inhibition by *Wolbachia* include the hypothesis that *Wolbachia* sterically hinders pathogens, by blocking reconfiguration of cellular components required for viral replication[265,266]. This theory would be consistent with the density dependence of pathogen inhibition described above.

Manipulation of host microRNAs may play a role in pathogen inhibition. In *Ae. aegypti* cells, *w*MelPop was found to downregulate expression of the DNA methyltransferase gene *AaDnmt2* via upregulation of the microRNA aae-miR-2940; experimental upregulation of the *AaDnmt2* gene resulted in enhanced dengue virus replication, suggesting a link between this *Wolbachia* phenotype and dengue viral inhibition[267].

1.4. Aims

Given the impact of pest insects and the inadequacies of current control methods, novel control methods are required. The aims of this thesis are to identify and investigate genes of potential interest to novel control methods. Chapter 2 and Chapter 3 aim to identify and investigate *Wolbachia* and host genes involved in *Wolbachia*-host interactions. The focus of Chapter 2 is the identification and investigation of *Wolbachia* genes involved in the CI phenotype, using comparative genomics and *D. melanogaster* transformation for functional investigation. An improved understanding of CI would be beneficial for its use in insect control, and is also of interest from a basic biology perspective. Chapter 3 uses comparative transcriptomics to further identify and investigate candidate *Wolbachia* and host CI genes. Comparative transcriptomics is also used to identify and investigate host genes that may be involved in viral inhibition, which also has potential applications in novel control strategies. Again an improved understanding of this phenotype would be beneficial for these applications, as well as from a basic biology perspective. Chapter 4 aims to identify genes for use in transgenic control constructs. Firstly attempts are made using genomic analysis to identify the sex-determining locus in *Ae. aegypti*, then using transcriptomic analysis to identify genes that could be used to restrict expression to the male germline in transgenic constructs. Overall it is hoped that the identification and investigation of genes here will provide important mechanistic insights and platforms for further development of methods for control of pest insects.

2 Identification and investigation of candidate genes involved in CI using comparative genomics

2.1. Introduction

2.1.1. *Wolbachia* comparative genomics

One approach to identifying genes involved in CI is genomic analysis of *Wolbachia*, comparing closely related incompatible strains with each other, or comparing CI-inducing strains with related strains that do not induce CI. In *D. simulans* the *Wolbachia* strain *w*Au expresses neither sperm modification in males, nor rescue of CI in females[132,268,269,270,271]; in other words *w*Au is mod- resc-, while CI-inducing strains are designated mod+ resc+[126]. The *w*Mel strain, from *D. melanogaster*, is most closely related to *w*Au and does induce CI; thus genes differing between these strains are candidates for involvement in CI[177]. The genome sequence of *w*Mel has already been reported[161], so here *w*Au genome sequencing was undertaken to enable a comparative analysis. This work has been published - see Sutton *et al*. (2014)[272].

Although the genomes of several *Wolbachia* strains have been sequenced[161-170,273,274], acquisition of sequence data has been limited by the difficulty in obtaining a sufficient quantity and purity of *Wolbachia* genomic DNA (gDNA). *Wolbachia* are obligate endosymbionts that cannot be cultured outside their hosts, and are often present in relatively low abundance. Obtaining enough gDNA has thus required time-consuming amplification and purification protocols to minimize contamination with host gDNA. In addition, assembly has been complicated by numerous repeated sequences. Here the Pacific Biosciences (PacBio) RS II platform was utilised for sequencing; the long reads generated by this technology facilitate assembly through genomic repeats.

Previous comparative analysis of the genomes of mutual incompatibility-generating *Wolbachia w*Pip sub-strains infecting *Culex pipiens* mosquitoes revealed highly similar genomes with a small number of whole gene differences[165,166]. Most notably this included a transcriptional regulator gene designated *wtrM* identified in *w*Pip from *Cx. molestus* (*w*PipMol) but absent in *w*Pip from *Cx. quinquefasciatus* Pel and JHB (wPipPel and *w*PipJHB), which are bidirectionally incompatible with *Cx. molestus*[166]. Transfection of *Cx. quinquefasciatus* females with *wtrM* resulted in significant upregulation of *CPIJ005623*, a host gene implicated in CI based on knockdown studies[166]. Eight paralogous putative transcriptional regulator genes are present in *w*Mel (*WD0254*, *WD0255*, *WD0296*, *WD0508*, *WD0622*, *WD0623* and *WD0626* and *WD0627*). A specific comparison of these transcriptional regulator genes in *w*Mel and their homologues in *w*Au was therefore conducted to further investigate the hypothesis that disruptions to these genes could be responsible for the different CI phenotypes of these strains.

2.1.2. Candidate CI gene functional testing

Computational analysis can identify candidate genes that may be involved in a phenotype, but validation requires functional testing – manipulating the expression of the candidate genes and assessing the resulting phenotype. Due to the current inability to culture *Wolbachia* outside cells, direct manipulation of *Wolbachia* genes is not possible. Expression vectors have been used for transfection of *Wolbachia* genes and shown to change transcription levels of a host cell cycle regulator, but their tissue distribution is uneven, limiting their use for examining whether a CI-like phenotype can be induced[166]. The best approach may be to transform hosts with target *Wolbachia* genes and assess whether a CI phenotype can be induced, as has been performed for ankyrin repeat-encoding genes previously[178]. Of the host species in which CI occurs, *D. melanogaster* is the most amenable to transformation. Thus, *D. melanogaster* was transformed with a candidate CI gene, *WD0626*, chosen as it is the closest *w*Mel orthologue to the *wtrM* gene

discussed above. Crossing experiments were then performed to determine whether expression of this gene has an effect on CI phenotype.

2.2. Methods

2.2.1. *Wolbachia* comparative genomics

2.2.1.1. D. simulans

D. simulans flies infected with *w*Au from Coffs Harbour, Australia, were maintained at 25 °C with a 12:12 hour light: dark cycle, in plastic bottles containing 25 ml food (see Recipe 1 in Appendix 1), and transferred to fresh containers when necessary.

2.2.1.2. Cell culture

Ae. albopictus Aα23 cells transinfected with *w*Au, provided by Dr Sofia Pinto, were maintained using standard cell culture techniques. Cells were maintained in an incubator at 28 °C, in flasks containing 10 ml Schneider's medium with 10 % fetal bovine serum, penicillin and streptomycin, and passaged when required.

2.2.1.3. gDNA extraction

gDNA for genome sequencing was extracted using two different methods. To extract gDNA from whole flies, a slightly modified version of the protocol used by Iturbe-Ormaetxe *et al*.[275] was used. Approximately 10–25 ml flies were collected, sterilised in 50 % bleach for 3 minutes, rinsed in filter-sterilised dH_2O, then further sterilised in 70 % ethanol and rinsed again in filter-sterilised dH_2O. The flies were then homogenised in cold SPG buffer (3.8 mM KH_2PO_4, 4.9 mM L-glutamate, 7.2 mM K_2HPO_4, 218 mM sucrose) using a Polytron homogeniser (Kinematica, Switzerland). After homogenisation the sample was centrifuged at 3,200 g for 15 minutes. The supernatant was collected and the centrifugation repeated. The supernatant from the second centrifugation was sequentially filtered through 5 µm, 2.7 µm and 1.2 µm syringe filters. The filtrate was centrifuged at 18,000 g for 20 minutes

to pellet *Wolbachia*, which were resuspended in cold SPG buffer. The suspension was then incubated with 600 ng of DNase I (Roche, UK) at 37 °C for 30 minutes and subsequently with 5 µl of RNase A (Fermentas, UK) at 37 °C for 15 minutes to remove host DNA and RNA contamination. Cells were then lysed by incubation with 200 µg of proteinase K (Sigma-Aldrich, UK) at 56 °C. gDNA was purified using two phenol/chloroform/isoamyl alcohol extractions and one chloroform/isoamyl alcohol extraction.

To extract gDNA from cells, *Wolbachia* were first purified from the cells. Cells were dislodged from flasks by pipetting and scraping, and lysed by vortexing with borosilicate beads. The lysate was centrifuged at 2,500 g for 10 minutes at 4 °C, then filtered sequentially through 5 µm and 0.2 µm filters. Sucrose gradient centrifugation was performed at 18,500 g for 10–20 minutes at 4 °C to pellet the *Wolbachia*. gDNA was purified using the method described by Livak[276]. The Livak method was also used to extract gDNA from adult *w*Au-infected flies for polymerase chain reaction (PCR).

gDNA concentration and quality was assessed using a NanoDrop 1000 spectrophotometer (Thermo Scientific, USA).

2.2.1.4. gDNA purity assessment

Extracted *w*Au gDNA was analysed for contamination with host gDNA using quantitative PCR (qPCR). qPCR was performed on a PTC-200 Thermo Cycler with a Chromo4 continuous fluorescence detector (MJ Research, USA) using iQ™ SYBR® Green Supermix (Bio-Rad, UK) according to the manufacturer's instructions. The reaction parameters were 95 °C for 15 minutes, followed by 40 cycles of 95 °C for 15 seconds, primer-specific annealing temperature for 15 seconds and 60 °C for 15 seconds. Reactions were performed on five serial dilutions of the extracts, using primers specific for a *w*Au gene (*wsp*), a host nuclear gene (*RpL32* for *D. simulans* and *hth* for *Ae. albopictus*) and a host

mitochondrial sequence (mitochondrial rRNA). Primers were designed using Primer-BLAST[277]. Primer sequences are listed in Appendix 2.

Primer efficiency values were calculated from the C(t) values of the five dilutions. At each dilution, the relative C(t) values for each primer pair, corrected for differences in primer efficiencies, were calculated. Taking into account the different sizes of the genomes, these values were used to calculate ratios of the amount of *w*Au gDNA to host nuclear and mitochondrial gDNA. The mean ratios were determined and combined to give an overall ratio of *w*Au gDNA to host gDNA, then the figures in this ratio were converted into percentages.

2.2.1.5. Sequencing

*w*Au gDNA extracted from whole files was sequenced using the Illumina HiSeq 2000 platform. Library preparation and sequencing were performed at the Wellcome Trust Sanger Institute (Hinxton, UK). A ~200-300 base pair (bp) paired end library was constructed following the methods described by Quail *et al*.[278,279], using Kapa HiFi polymerase for PCR to reduce GC bias[280]. The library was given a unique index and sequenced as part of a lane with other samples.

*w*Au gDNA extracted from cells was sequenced using the PacBio RS II platform. Library preparation was performed with the assistance of Dr Paul Coupland at the Wellcome Trust Sanger Institute (Hinxton, UK). A ~10 Kbp library was constructed following standard protocols using a PacBio DNA Template Prep Kit. Sequencing was performed at the Wellcome Trust Sanger Institute (Hinxton, UK). Three SMRT® cells from this library were sequenced, with a movie length of 2 hours.

2.2.1.6. Genome assembly

Assembly of the Illumina data was performed with the assistance of Dr Simon Harris in the Pathogen Genomics Group at the Wellcome Trust Sanger Institute (Hinxton, UK). Assembly was performed using Velvet[281], with the *w*Mel genome as a reference.

Assembly of the PacBio data was performed by Dr Paul Coupland at the Wellcome Trust Sanger Institute (Hinxton, UK). Filters were set to exclude reads of quality <0.8, polymerase read length <500 nucleotides (nt) and sub-read length <500 nt. A *de novo* assembly was generated from the resulting sub-reads using the Hierarchical Genome Assembly Process (HGAP) version 1.0[282], with the genome size parameter set to 1.2 Mb (the approximate size of the *w*Mel genome).

Errors in the PacBio assembly were corrected using the data from the Illumina sequencing, with the assistance of Dr Simon Harris at the Wellcome Trust Sanger Institute (Hinxton, UK). The Illumina reads were mapped to the assembly using SMALT[283], then the assembly sequence was modified based on the mapped reads using Iterative Correction of Reference Nucleotides (iCORN)[284] with four iterations, resulting in the correction of 1 SNP and 88 indels.

2.2.1.7. Genome annotation

The final genome assembly was annotated by Dr Simon Harris at the Wellcome Trust Sanger Institute (Hinxton, UK), using the Automated Annotation Pipeline at the Wellcome Trust Sanger Institute, with the software Prokka[285]. Infernal[286] was used to identify RNA structures, followed by ARAGORN[287], RNammer[288] and Prodigal[289] to identify transfer RNAs (tRNAs) and transfer messenger RNAs (tmRNAs), ribosomal RNAs (rRNAs) and proteins, respectively. The predicted genes were compared against *Wolbachia* sequences

from RefSeq[290], using CD-HIT[291] to create a non-redundant protein database, then against UniProtKB/SwissProt[292]. Some annotations were edited manually.

2.2.1.8. Comparative analysis

The *w*Mel and *w*Au genomes were compared using Artemis Comparison Tool (ACT)[293]. Sequences of individual genes of interest were extracted and alignments and translations generated using Geneious (7.0.5)[294]. Predicted protein domains were identified using InterPro[295]. SNP analysis was performed with the aid of Synonymous Non-synonymous Analysis Program (SNAP) v1.1.1[296,297]. The Gene Ontology (GO) project[298] was used to aid categorisation of genes. The dot plot comparing *w*Au and *w*Mel genomes was generated using Dotter[299] with default parameters.

2.2.1.9. Statistical analysis

Statistical analysis was performed using R.

2.2.2. Candidate CI gene functional testing

2.2.2.1. D. melanogaster

*w*Mel-infected *D. melanogaster* of the line yw^{67c23} and uninfected *D. melanogaster* of the line w^{1118} and the GAL4 driver lines *yw*;ActGAL4/Cyo, *w;*BamGAL4, and *yw*;NosGAL4 were reared as described above. Flies were transferred to fresh containers fortnightly or when necessary.

2.2.2.2. Plasmid construction

Plasmid construction was performed with the assistance of Dr Kirsty Stainton.

2.2.2.2.1. Plasmid design

The plasmid used for transformation was designated pUAST-WD0626, and consisted of the *WD0626* gene ligated into the vector pUAST. This vector contains 5 repeats of the Upstream Activation Sequence (UAS), a regulatory sequence that promotes gene expression through binding of the protein GAL4[300]. Genes inserted into the pUAST vector are integrated into host genomes via P element-mediated transformation; pUAST contains P element inverted repeats flanking the gene of interest, which are recognised by the P element transposase, encoded on a separate helper plasmid which is injected simultaneously. pUAST also contains the marker gene *white*, which produces a dominant red-eyed phenotype. Figure 2.1 illustrates the main features of the plasmid pUAST-WD0626.

2.2.2.2.2. Gene amplification

The *WD0626* gene sequence was amplified by PCR from gDNA extracted from *w*Mel-infected flies using the Livak method[276]. PCR was performed on a 2720 Thermal Cycler (Applied Biosystems, UK) using standard PCR conditions, except for use of a Phusion Hot Start high-fidelity polymerase. The primers used had 5' tails containing recognition sites for the restriction enzymes *Eco*RI (forward primer) and *Kpn*I (reverse primer).

2.2.2.2.3. PCR product purification

PCR products were run on a 0.5 % agarose gel made with low melting point agarose, visualised using a transilluminator, excised using a clean scalpel, and placed into microcentrifuge tubes. DNA was then purified using a QIAquick Gel Extraction Kit (Qiagen, UK), according to the manufacturer's instructions. DNA concentration and quality was assessed using a NanoDrop 1000 spectrophotometer (Thermo Scientific, USA).

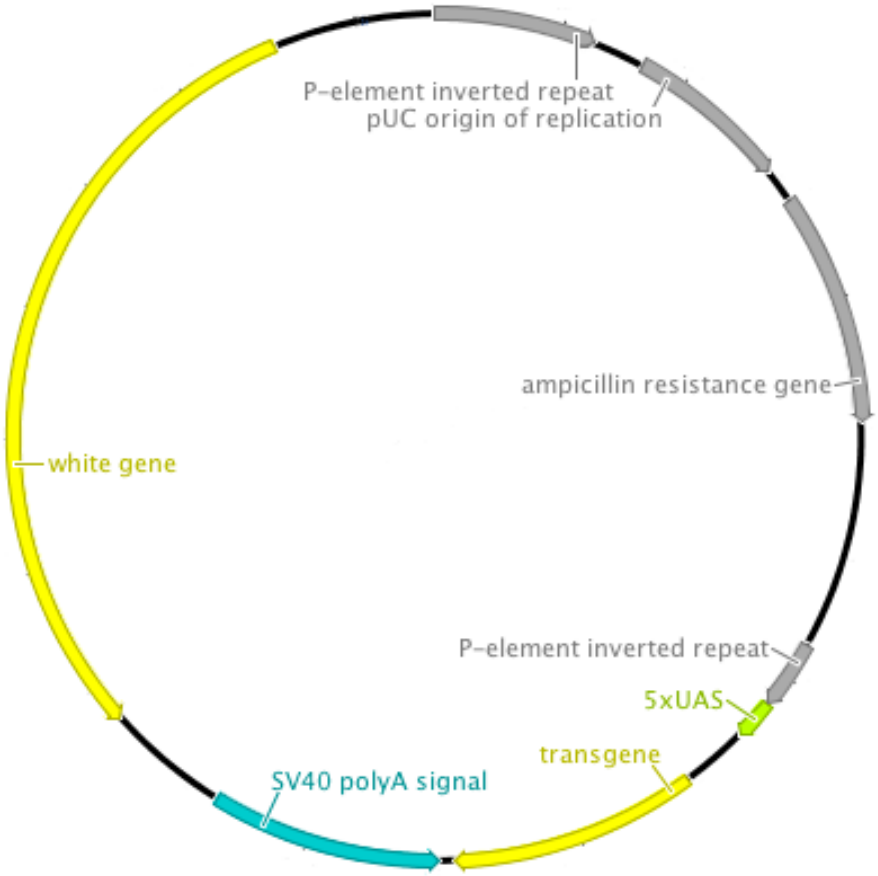

Figure 2.1. Features of pUAST-WD0626 plasmid. The ampicillin resistance gene allows selection of bacterial colonies containing the plasmid when cultured on LB-agar plates containing ampicillin. The P-element inverted repeats allow transposition of the transgene and associated sequences into the host genome, upon injection with a helper plasmid containing a transposase. The five copies of the regulatory element UAS promote expression of the transgene only when GAL4 is present, allowing control of expression. The SV40 polyA signal promotes polyadenylation of the transgene transcript. The *white* gene leads to a dominant red-eyed phenotype and is used to detect transformants.

2.2.2.2.4. Digestion

Purified PCR products and the pUAST vector were digested with *Eco*RI and *Kpn*I (New England Biolabs, UK). The reactions were performed as double digests in a total volume of 50 µl, containing 1 µl of each enzyme, 0.5 µl of bovine serum albumin (BSA), 5 µl of 10x NEBuffer 1, and up to 1 µg of DNA. Reactions were incubated at 37 °C for 1 hour, then at 65 °C at 20 minutes for restriction enzyme inactivation.

2.2.2.2.5. Ligation

Digested pUAST and PCR products were ligated together in 20 µl reactions containing 2 µl of T4 DNA ligase (New England Biolabs, UK), 2 µl of 10x T4 DNA ligase buffer, 100 ng of digested pUAST, and quantities of digested PCR products resulting in molar insert:vector ratios of between approximately 2:1 and 6:1. Reactions were incubated at 16 °C for 16 hours.

2.2.2.2.6. Transformation

Ligation mixtures were used to transform competent *E. coli*. 5 µl of ligation mixture was added to 25 µl or 50 µl of cells in 15 ml Falcon tubes and the mixture incubated on ice for 30 minutes. Cells were then heat-shocked at 42 °C for 30 seconds, followed by another 2 minute incubation on ice. 950 µl of SOC medium was added and the cells incubated at 37 °C for 1 hour with shaking at 200 rpm. 100 µl or 200 µl of the mixtures was spread on an LB-agar plate containing 100 µg/ml ampicillin, and the plates incubated at 37 °C overnight.

2.2.2.2.7. Colony screening

Colonies were screened the day following transformation for the presence of the desired plasmid, using PCR with primers annealing to pUAST either side of the insertion site, followed by analysis of the PCR products by gel electrophoresis. PCR was performed on a 2720 Thermal Cycler (Applied Biosystems, UK) using standard PCR conditions and with bacterial colonies used directly as templates. The primers used enabled the desired plasmids to be distinguished from undigested or self-ligated pUAST, which resulted in a much smaller product.

2.2.2.2.8. Plasmid purification

Colonies containing the desired plasmid were inoculated into 5 ml of LB broth containing 100 µg/ml ampicillin and incubated at 37 °C for 12-16 hours with shaking at 200 rpm. Plasmid DNA was then extracted using a QIAprep Spin Miniprep Kit (Qiagen, UK) according to the manufacturer's instructions. Plasmids were sent for sequencing by GATC Biotech (Cologne, Germany) and results analysed for the presence of errors.
Plasmids with no errors were used to re-transform *E. coli* using the method described above. Resulting colonies were inoculated into 5 ml of LB broth containing 100 µg/ml ampicillin and incubated at 37 °C for 8 hours with shaking at 200 rpm. 500 µl of these cultures was then inoculated into 250 ml of LB broth containing 100 µg/ml ampicillin, followed by incubation at 37 °C for 12-16 hours with shaking at 200 rpm. Plasmid DNA was then extracted using an EndoFree Plasmid Maxi Kit (Qiagen, UK) according to the manufacturer's instructions.

2.2.2.3. Embryo microinjection

2.2.2.3.1. Injection mix preparation

Volumes of DNA solutions containing 12 µg of the plasmid containing the gene of interest and 6 µg of the helper plasmid were added to a 1.5 ml microcentrifuge tube, along with dH_2O to a volume of 100 µl, followed by 200 µl of 100 % ethanol and 10 µl of 3M sodium acetate. The mixture was incubated on ice for 30 minutes, then centrifuged at maximum speed for 5 minutes. The supernatant was discarded and the pellet rinsed with 100 µl of 70 % ethanol and then resuspended in 20 µl of filter-sterilised injection buffer (0.1 mM Na_3PO_4 pH 6.8, 5 mM KCl). The mixes were left at room temperature for approximately one hour to aid dissolution of the DNA, then stored at -20 °C until use.

2.2.2.3.2. Embryo preparation

A minimum of one day prior to embryo collection flies were transferred from rearing bottles to empty plastic bottles with the bottoms replaced with mesh, which were placed inside sealed plastic containers. A food source was provided in the form of yeast paste on the cotton wool sealing the bottle, and a moistened paper towel was placed in the container to provide a suitable humidity level. Fresh yeast paste was provided daily.

For embryo collection the cotton wool sealing the bottle was replaced with a small pot or Petri dish of agar containing 30 % fruit juice, with yeast and a drop of acetic acid added on top to encourage oviposition. The first embryos of the day to be laid were discarded due to the possibility of overly mature embryos being present. After this, embryos laid within periods of approximately one hour were collected by rinsing into a small sieve.

Embryos were transferred to a strip of double-sided tape on a microscope slide, then chorions detached using another slide with double-sided tape. Embryos of a suitable age (judged by colour and general appearance) were transferred to a coverslip coated with a glue-like substance made by soaking double-sided tape in heptane overnight. All embryos were aligned in the same orientation. When embryos were suitably desiccated (judged by appearance and resistance), they were covered with Voltalef 10S oil (VWR International, UK).

2.2.2.3.3. Embryo microinjection

Embryos were injected with injection mix under 20x magnification using a FemtoJet microinjector system (Eppendorf, UK) with type II Femtotip microinjection needles (Eppendorf, UK). Injection was performed into the posterior end of the embryo, the region that eventually gives rise to gametes in adults. Dr Kirsty Stainton assisted with injections.

2.2.2.4. Rearing of injected flies

After injection embryos were incubated at 37 °C for one hour on at least one occasion to activate expression of the P element transposase, which was under the control of an hsp70 promoter. Coverslips of embryos were transferred to plates of agar containing 30 % fruit juice and smeared with yeast paste. Plates were placed inside sealed plastic containers with a moistened paper towel and were subject to the same rearing conditions described previously. From one day after injection plates were inspected under a microscope for the presence of larvae. Any larvae present were transferred to vials of *Drosophila* food, with a maximum density of approximately 30 larvae per vial.

2.2.2.5. Screening for transformants

Adult male and female injection survivors were collected shortly after eclosion and crossed individually to flies of the opposite sex from the w^{1118} strain. Virgin w^{1118} females were used. Virgin injection survivor females were used when possible, and if not a note was made of which males the female may have mated with. The progeny of these crosses were screened for red-eyed flies. Any red-eyed males were crossed to w^{1118} virgin females. The ratio of red-eyed flies to white-eyed flies in the progeny of these crosses was assessed. A ratio much greater than 1:1 suggests multiple insertions of the transgene, while a ratio much lower than 1:1 suggests that the insertion is detrimental. Thus for any crosses in which ratios varying greatly from 1:1 were observed the transformed parents were discarded. The transformed parents were also discarded in any crosses in which all the progeny of one sex were white-eyed and all the progeny of the other sex were red-eyed, as this suggests that the insertion is on the X chromosome (if all females red-eyed) or Y chromosome (if all males red-eyed), which would be problematic for future crosses. In all other cases the transformed parents were used to generate balanced transformed lines.

2.2.2.6. Insert location determination

Insert location determination was determined using a molecular approach based on an adapter ligation-mediated PCR method developed for mapping T-DNA inserts in *Arabidopsis*[301]. The molecular details of this strategy are illustrated in Figure 2.2 and adapter sequences listed in Table 2.1.

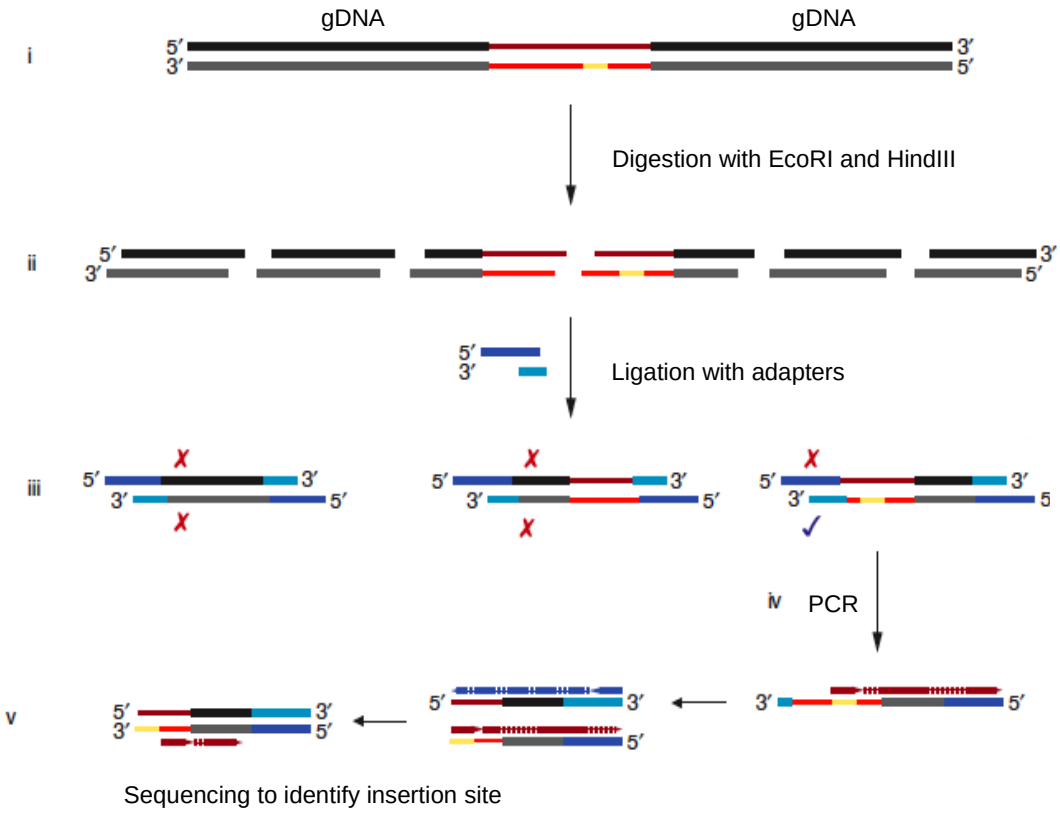

Figure 2.2. Adapter ligation-mediated PCR for insert location determination. In this method, gDNA containing the insert (i) was digested into fragments with *Eco*RI and *Hin*dIII (ii). The fragments were ligated to short double-stranded DNA molecules with 5' overhangs known as adapters, creating adapter-flanked templates (iii). The longer arm of the adapters (dark blue) contained a sequence identical to that of a primer known as an adapter primer, but there were no binding sites for this primer (crosses) unless synthesis of the complementary sequence of the 5' overhang occurred. This was achieved in fragments containing the insertion (tick) by the use of a primer (known as an insertion primer) complementary to part of the insertion sequence (yellow). The short strand of the adapter had a 3' C7 amino modification to prevent polymerase extension of this strand, which would create binding sites for the adapter primer in all fragments and allow pairs of adapter primers to amplify the sequence between them. Thus when PCR was performed using the adapter and insertion primers, DNA was amplified selectively in fragments containing the insertion (iv). The amplified sequence spanned the junction of the insertion and host gDNA. This junction was sequenced using a primer complementary to the insertion sequence (v), and the results analysed to determine on which chromosome the insert was located. Adapted from O'Malley *et al.* (2007)[301].

Oligonucleotide	Sequence
Adapter long strand	GTAATACGACTCACTATAGGGCACGCGTGGTCGACGGCC**CGGGCTGC**
*Eco*RI adapter short strand	5'-phosphate-AATT**GCAGCCCG**-amino C7-3'
*Hin*dIII adapter short strand	5'-phosphate-AGCT**GCAGCCCG**-amino C7-3'

Table 2.1. Sequences of adapters used for adapter ligation-mediated PCR. Dotted underscore indicates the sequence complementary to the binding site for adapter primer 1 (see main text). Solid underscore indicates the sequence complementary to the binding site for adapter primer 2 (see main text). Bold text indicates the complementary regions between long and short adapter strands that anneal to form the double-stranded adapters.

DNA was extracted from transformants using the Livak method[276]. DNA was then digested with *Eco*RI and *Hin*dIII and ligated to *Eco*RI and *Hin*dIII adapters in 20 µl reactions containing 6 µl of DNA solution, 0.1 µl of T4 DNA ligase, 2 µl of T4 DNA ligase buffer, 0.5 µl of each adapter, 0.2 µl of each enzyme, and 0.1 µl of ATP. Adapters were prepared prior to this by adding 20 µl of the adapter long strand and 20 µl of the adapter short strand to 1.21 ml of 1mM Tris (pH 8.3) in a 1.5 ml microcentrifuge tube, vortexing, placing the tube on a wet heat block at 96 °C for 2 minutes, then turning off the heat and leaving the block to cool to room temperature to allow the long and short strands to anneal to each other. After digestion and ligation, PCR was performed on the DNA using primers specific for the adapter (adapter primer 1) and insertion sequence. A second nested PCR reaction using further primers specific for the adapter (adapter primer 2) and insertion sequence was performed to enrich for the desired product. PCR products were purified using a QIAquick PCR Purification Kit (Qiagen, UK) and sent for sequencing by GATC Biotech (Cologne, Germany). Results were used as queries in a search of *D. melanogaster* chromosome sequences with the BLAST program. PCR mixtures were also visualised on an agarose gel, made with TAE buffer (64 mM EDTA, 1 M acetic acid, 2 M Tris base) containing 1 µl ethidium bromide per 100 ml buffer.

2.2.2.7. Generation of balanced lines

Balanced transformed lines were generated to prevent recombination of the insertion, which could lead to separation of the *white* gene marker from the gene of interest, or separation of the UAS from the gene of interest, for example. Balanced lines were created by crossing transformed males to virgin females with a balancer on the relevant chromosome. Red-eyed progeny with the correct phenotype were crossed to each other; offspring from these crosses will either contain the transgene balanced with a marker gene, or will be homozygous for the transgene (if it is homozygous viable). The lines were self-maintaining from this point.

2.2.2.8. Generation of GAL4/UAS flies

Genes under the control of UAS are usually not expressed in the absence of GAL4, which is not endogenous to *Drosophila*. Thus to generate transformed flies expressing the gene of interest, transformed males were crossed to virgin females from lines expressing GAL4. Because *Wolbachia* exert their effects in the germline, it was desired to express the genes of interest in these tissues. For expression in testes the GAL4 line *w:*BamGAL4 was used, which expresses GAL4 under the control of the germline-specific Bam promoter. For expression in ovaries initially the GAL4 line *yw*;NosGAL4 was used, which expresses GAL4 under the control of the ovary-specific Nanos (Nos) promoter. This line was chosen because the Nos promoter is known to drive high levels of expression and at the early stages most likely to be important for CI[302,303]. However, after the first set of crossing experiments the *yw*;NosGAL4 line was found to have become infected with *Wolbachia*, and so as the Bam promoter is active in the female as well as the male germline[304] the *w*;BamGAL4 line was subsequently used to drive expression of the transgene in ovaries.

Crosses were performed by transferring approximately 30 males and 30 females to empty plastic bottles with the bottoms replaced with mesh, which were placed inside sealed

plastic containers. A food source was provided in the form of yeast paste on the cotton wool sealing the bottle, and a moistened paper towel was placed in the container to provide a suitable humidity level.

After flies had been left for one day to mate, embryos were collected by replacing the cotton wool sealing the bottle with a small Petri dish of agar containing 30 % fruit juice, with yeast and a drop of acetic acid added on top to encourage oviposition. Embryos were transferred to bottles containing 25 ml of *Drosophila* food, at a maximum density of 125 embryos per bottle. Bottles were maintained under standard rearing conditions.

2.2.2.9. Confirmation of transgene presence

gDNA was extracted from flies used in crossing experiments using the Livak method[276]. PCR was performed as described previously, using primers specific for *WD0626*, and PCR products were visualised on an agarose gel, as described previously. Primers were designed using Primer-BLAST[277]. Primer sequences are listed in Appendix 2.

2.2.2.10. Screening for Wolbachia infection

gDNA was extracted from flies using the Livak method[276]. PCR was performed as described previously, using primers specific for the *Wolbachia* gene for *Wolbachia* surface protein (*wsp),* and PCR products were visualised on an agarose gel, as described previously. Primers were designed using Primer-BLAST[277]. Primer sequences are listed in Appendix 2.

2.2.2.11. Expression analysis

2.2.2.11.1. Dissection and RNA extraction

Virgin flies were used for dissection of abdomens at 8-10 hours after eclosion. Total RNA was extracted using a protocol based on chloroform extraction and isopropanol precipitation. Samples were placed in 1.5 ml microcentrifuge tubes and homogenised with

a sterile pestle in 200 µl of TRI Reagent (Sigma-Aldrich, UK). 40 µl of chloroform was added to each sample, followed by centrifugation at 20,000 g at 4 °C for 15 minutes. The resulting aqueous phases were transferred to new microcentrifuge tubes, 80 µl of isopropanol was added to each tube, and samples were centrifuged at 20,000 g at 4 °C for 30 minutes to pellet the RNA. The supernatants were removed and the pellets washed with 200 µl of 75 % ethanol (in RNase-free filter-sterilised dH_2O, prepared by autoclaving dH_2O with 0.01 % diethylpyrocarbonate), air-dried, then resuspended in 30 µl of RNase-free filter-sterilised dH_2O. The samples were then treated to remove DNA contamination. 1 U of DNase I was added to each sample, along with 10x DNase I buffer to a final concentration of 1x. Samples were incubated at 37 °C for 10-15 minutes to allow degradation of any DNA present, then at 65 °C for 10 minutes to inactivate the DNase I. Abdomens were pooled in groups of 6 for males and 4-5 for females.

2.2.2.11.2. cDNA synthesis

Complementary DNA (cDNA) was synthesised from RNA using an iScript cDNA Synthesis Kit (BioRad, UK) according to the manufacturer's instructions. cDNA concentration and quality was assessed using a NanoDrop spectrophotometer (Thermo Scientific, USA).

2.2.2.11.3. qRT-PCR

Quantitative reverse transcriptase PCR (qRT-PCR) was performed as described previously for qPCR. Primers were designed using Primer-BLAST[277]. Primer sequences are listed in Appendix 2.

2.2.2.12. Crossing experiments

Two sets of crossing experiments were performed, largely according to the method used by Yamada *et al.*[178,305]. Adult GAL4/UAS flies produced as described above were collected and crossed to flies of the opposite sex from the strains w^{1118} and *w*Mel, and GAL4/UAS

flies of the opposite sex from the same transformed line. Because CI penetrance in *D. melanogaster* decreases rapidly with increasing male age[130,132,306] and with increasing male development time[305], males were collected and crossed to females within 24 hours of eclosion, and only males eclosing on the first or second day of eclosion were used. In the second set of experiments males were collected and crossed to females within eight hours of eclosion in an effort to optimise conditions for CI, due to CI in control crosses being weaker than expected in the first set of experiments. Females were used at 2-5 days old. It was aimed to perform at least 20 replicate crosses of each type; when there were insufficient flies for this as many crosses as possible were performed.

To perform the crosses, for the first set of crossing experiments, single pairs of flies were placed in vials containing *Drosophila* food and left to mate for approximately 24 hours. Females were then transferred to compartments of plates containing agar with 30 % fruit juice, with yeast paste smeared on top. The plates were sealed with parafilm in which small holes were made for ventilation and maintained under standard rearing conditions. Females were transferred to new compartments approximately every 24 hours for two days. For the second set of crossing experiments, crosses were performed in vials containing agar with 30 % fruit juice rather than in plates.

After use, male and female GAL4/UAS flies were frozen at -20 °C to use later for confirmation of *WD0626* transgene presence. Approximately two days after removal of females from compartments or vials the number of larvae and the number of unhatched eggs in each compartment or vial were counted. Any crosses in which fewer than 25 eggs were laid in total were discounted from analysis. For females in crosses in which more than 25 eggs were laid but none hatched, spermathecae were dissected and studied under a microscope for the presence of sperm to check that the females had mated. Mean hatch

rates (the number of larvae as a proportion of the number of larvae plus the number of unhatched eggs) were calculated for each cross.

2.2.2.13. Statistical analysis

Statistical analysis was performed using online Wilcoxon[307] and Kruskal-Wallis[308] tests.

2.3. Results

2.3.1. *Wolbachia* comparative genomics

2.3.1.1. gDNA purity

Approximate calculations based on qPCR C(t) values for *w*Au and host genes were performed to estimate the degree of contamination with host gDNA in *w*Au gDNA samples extracted from cultured cells and whole adult flies. The estimated purity of *w*Au gDNA was ~60 % for the extract from cultured cells, and > 90 % for the extract from whole adult flies.

2.3.1.2. Sequencing and assembly

DNA sequencing was initially performed using the Illumina platform. However, the resulting assembly was fragmented in the regions of most interest, with scaffold positions uncertain. A second round of sequencing was therefore performed using the PacBio RS II system to obtain longer reads in an attempt to improve the assembly.

DNA sequencing using the Illumina HiSeq 2000 platform was performed on gDNA extracted from whole adult flies. In total 3,565,172 reads of 100 nt were generated, corresponding to 357 Mb of data. An assembly was generated from these reads with Velvet[281], using the *w*Mel genome as a reference. The final assembly comprised 77 contigs, with an N50 of 29.5 Kbp, a total size of 1.22 Mbp, and a mean coverage of 283X.

DNA sequencing using the PacBio RS II platform was performed on gDNA extracted from cells. In total 75,456 sub-reads (from 39,514 polymerase reads) were generated, with a mean sub-read length of 1,847 nt, corresponding to 139 Mb of data. A de novo assembly was generated from these sub-reads using HGAP (version 1.0)[282]. The final assembly comprised a single contig of 1.27 Mbp, with a mean coverage of 62X. Errors in the assembly were corrected using the Illumina data; this resulted in the correction of one single nucleotide polymorphism (SNP) and 88 indels.

2.3.1.3. w*Au genome features*

The wAu genome is a single circular chromosome of 1,268,461 bp. It has 1,266 predicted genes, corresponding to a coding content of 84 %. The major features of the genome, along with those of the *w*Mel genome, are shown in Table 2.2. Overall, the *w*Au and *w*Mel genomes are similar, but with a significant amount of rearrangement (Figure 2.3).

	*w*Au	*w*Mel
Genome size (bp)	1,268,461	1,267,782
G + C content (%)	35.22	35.23
Predicted CDSs	1,266	1,195
Coding density (%)	83.9	80.2
Mean gene size (bp)	840	850
Transfer RNAs	34	34
Ribosomal RNAs	1 of each	1 of each
Prophage regions	3	3

Table 2.2. General features of *w*Au and *w*Mel genomes.

Like *w*Mel, there is a large amount of DNA corresponding to mobile genetic elements in the *w*Au genome, including numerous insertion sequence (IS) elements. For example, 27 putative IS5 elements were identified, although most are likely to be inactive due to mutations or frameshifts. Some elements appear to have been active since the divergence of *w*Mel and *w*Au, as their locations in the genome differ between the two strains. In some cases their movement has resulted in disruption of genes in one strain. In many cases

where there are structural differences between the genomes of the strains, it seems that mobile elements have provided a mechanism for the rearrangement.

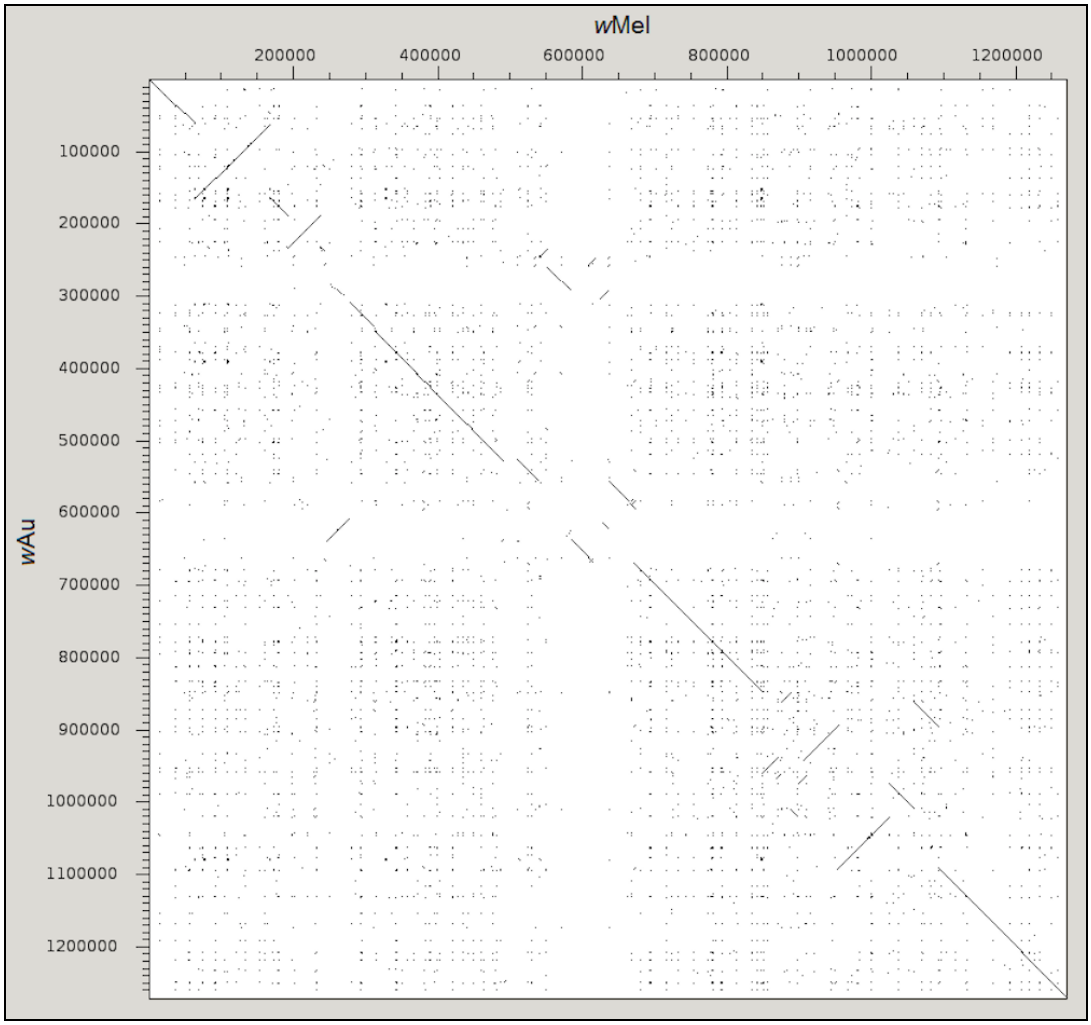

Figure 2.3. Dot plot illustrating the similarity between *w*Au and *w*Mel genomes. Axes show position in the genome (bp). Lines indicate regions of similarity. Lines on the diagonal from top left to bottom right indicate regions with the same location and arrangement in both genomes. Lines in the same orientation as this diagonal but located elsewhere indicate regions that are translocated in one genome relative to the other. Lines at right angles to the diagonal indicate regions that are inverted in one genome relative to the other. Parallel lines indicate repeated or similar regions.

2.3.1.4. Genes potentially inactive in w*Au*

Various other *w*Mel genes were also identified with large differences potentially causing inactivation in *w*Au, or which are absent from *w*Au entirely, as listed in Table 2.2. Any genes with differences between CI and non-CI strains are possible candidates for

involvement in the process, but it seems reasonable to focus more attention on genes that are entirely absent or potentially inactivated. As has previously been reported[177], a region corresponding to genes *WD0506* to *WD0518* is absent in *w*Au. No other indel of comparable size to the *WD0506* to *WD0518* segment was found to be absent in the *w*Au genome. However, of note was the absence in *w*Au of two genes, *WD0631* and *WD0632*, that have recently been identified as CI candidates due to the detection in *Cx. pipiens* spermathecae of a protein corresponding to the *WD0631* orthologue in the *w*Pip strain[193]; the *WD0631* and *WD0632* orthologues are transcribed as an operon[193]. Various other genes are potentially inactivated in *w*Au relative to *w*Mel, by truncation, frameshift, mobile element insertion, nonsense mutation or start codon mutation, as shown in Table 2.3.

2.3.1.5. Comparison of prophage regions

There are three prophage regions in the *w*Au genome, as for *w*Mel, although the location and structure of these regions differs between the two strains (Figure 2.4, Figure 2.5, Figure 2.6 and Figure 2.7). The prophage region designated WO-A (Figure 2.4) in *w*Mel (spanning *WD0259* – *WD0294*) is inverted in *w*Au relative to *w*Mel, and is further from the origin of replication (spanning *WPWAU0631* – *WPWAU0666*). Several genes in this region differ in one strain relative to the other beyond SNPs (Figure 2.4). Four genes are disrupted in *w*Au relative to *w*Mel, due to truncation (a shortened gene sequence due to partial deletion or genome rearrangement), frameshift, nonsense mutation, or start codon mutation, two are disrupted in *w*Mel relative to *w*Au, and two contain small in frame indels.

The region designated WO-B (Figure 2.5) in *w*Mel (spanning *WD0582* – *WD0644*) is closer to the origin of replication in *w*Au (spanning *WPWAU0282* – *WPWAU0318*). This region contains two segments, one closely related to P2 phage and the other to lambdoid phage[158]. In *w*Au the P2-like segment is inverted relative to its orientation in *w*Mel. In

addition, in *w*Mel there are intervening genes between the lambdoid-like block and P2-like block, whereas in *w*Au the two blocks are contiguous. These differences mean that unlike *w*Mel WO-B, the gene order of *w*Au WO-B is highly conserved with that of WO phage in

***w*Mel gene**	**Function**	**Matching *w*Au gene(s)**	**Difference in *w*Au**
WD0092	DNA processing chain A	*WPWAU0139/WPWAU0140*	Frameshift
WD0139	Transcriptional activator, tenA family, putative	*WPWAU0095*	Start codon mutation
WD0196	Hypothetical protein	No match	Truncation
WD0254	Transcriptional regulator, putative	*WPWAU0256/WPWAU0257/ WPWAU0686/WPWAU0687*	Frameshift
WD0274	Conserved hypothetical protein	*WPWAU0651*	Start codon mutation
WD0284	Conserved hypothetical protein	*WPWAU0640/WPWAU0641*	Frameshift
WD0288	Prophage LambdaW1, site-specific recombinase, resolvase family	*WPWAU0636*	Nonsense mutation
WD0294	Ankyrin repeat domain protein	*WPWAU0631*	Truncation
WD0295	Hypothetical protein	*WPWAU0322/WPWAU0323*	Frameshift
WD0382	Conserved hypothetical protein	*WPWAU0417/WPWAU0418/ WPWAU0419*	Frameshift
WD0383	Hypothetical protein	*WPWAU0420/WPWAU0421*	Frameshift
WD0385	Ankyrin repeat domain protein	*WPWAU0423/WPWAU0426/ WPWAU0427*	Mobile element insertion
WD0446	Hypothetical protein	*WPWAU0481/ WPWAU0482*	Frameshift
WD0462	Hypothetical protein	*WPWAU0494/WPWAU0495*	Frameshift
WD0463	ATPase, AAA family	*WPWAU0496*	Mobile element insertion
WD0472	ATPase, AAA family	*WPWAU0507/WPWAU0508*	Nonsense mutation
WD0507	DNA repair protein RadC, truncation	No match	Absent
WD0508	Transcriptional regulator, putative	No match	Absent
WD0509	DNA mismatch repair protein MutL-2	No match	Absent
WD0511	Conserved hypothetical protein	No match	Absent
WD0512	Hypothetical protein	No match	Absent
WD0513	Hypothetical protein	No match	Absent

WD0514	Ankyrin repeat domain protein	No match	Absent
WD0548	Hypothetical protein	*WPWAU0565*	Frameshift
WD0572	Conserved hypothetical protein	*WPWAU0271*	Frameshift
WD0582	Regulatory protein RepA, putative	*WPWAU0282*/*WPWAU0283*/ *WPWAU0284*	Frameshift
WD0591	Conserved hypothetical protein	*WPWAU0289*	Frameshift
WD0594	Prophage LambdaW4, DNA methylase	*WPWAU0291*/*WPWAU0292*	Frameshift
WD0609	Regulatory protein RepA, putative	*WPWAU0669*/*WPWAU0670*	Frameshift
WD0610	Helicase, SNF2 family	*WPWAU0671*/*WPWAU0672*	Nonsense mutation
WD0619	GlpT/PgpT/UhpT transporter family protein	*WPWAU0681*	Frameshift
WD0622	Transcriptional regulator, putative	*WPWAU0256*/*WPWAU0257*/ *WPWAU0686*/*WPWAU0687*	Frameshift
WD0630	Hypothetical protein	*WPWAU0249*	Truncation
WD0631	Hypothetical protein	No match	Absent
WD0632	Hypothetical protein	No match	Absent
WD0636	Prophage LambdaW5, ankyrin repeat domain protein	*WPWAU0314*/*WPWAU0315*	Frameshift
WD0638	Conserved hypothetical protein	*WPWAU0311*/*WPWAU0312*	Frameshift
WD0682	Ribosomal protein S10	*WPWAU0607*	Nonsense mutation
WD0686	Conserved domain protein	*WPWAU0613*/*WPWAU0614*/ *WPWAU0618*	Mobile element insertion
WD0696	Hypothetical protein	*WPWAU0698*	Nonsense mutation
WD0766	Ankyrin repeat domain protein	*WPWAU0768*	Nonsense mutation
WD1041	Surface protein-related protein	*WPWAU1092*/*WPWAU1093*/ *WPWAU1094*	Frameshift
WD1111	Hypothetical protein	*WPWAU0887*/*WPWAU0888*	Frameshift
WD1180	Recombination protein RecR	*WPWAU1182*	Nonsense mutation
WD1187	Hypothetical protein	*WPWAU1187*	Frameshift
WD1320	Multidrug resistance protein D	*WPWAU1315*	Frameshift

Table 2.3. *w*Mel genes potentially inactive in *w*Au. Truncation refers to shortening of the gene sequence due to partial deletion or genome rearrangement. Small hypothetical genes with multiple matches are not included, as these are likely to be remnants of mobile elements. Genes annotated as truncations in *w*Mel are not included, as these are likely to be inactive in *w*Mel.

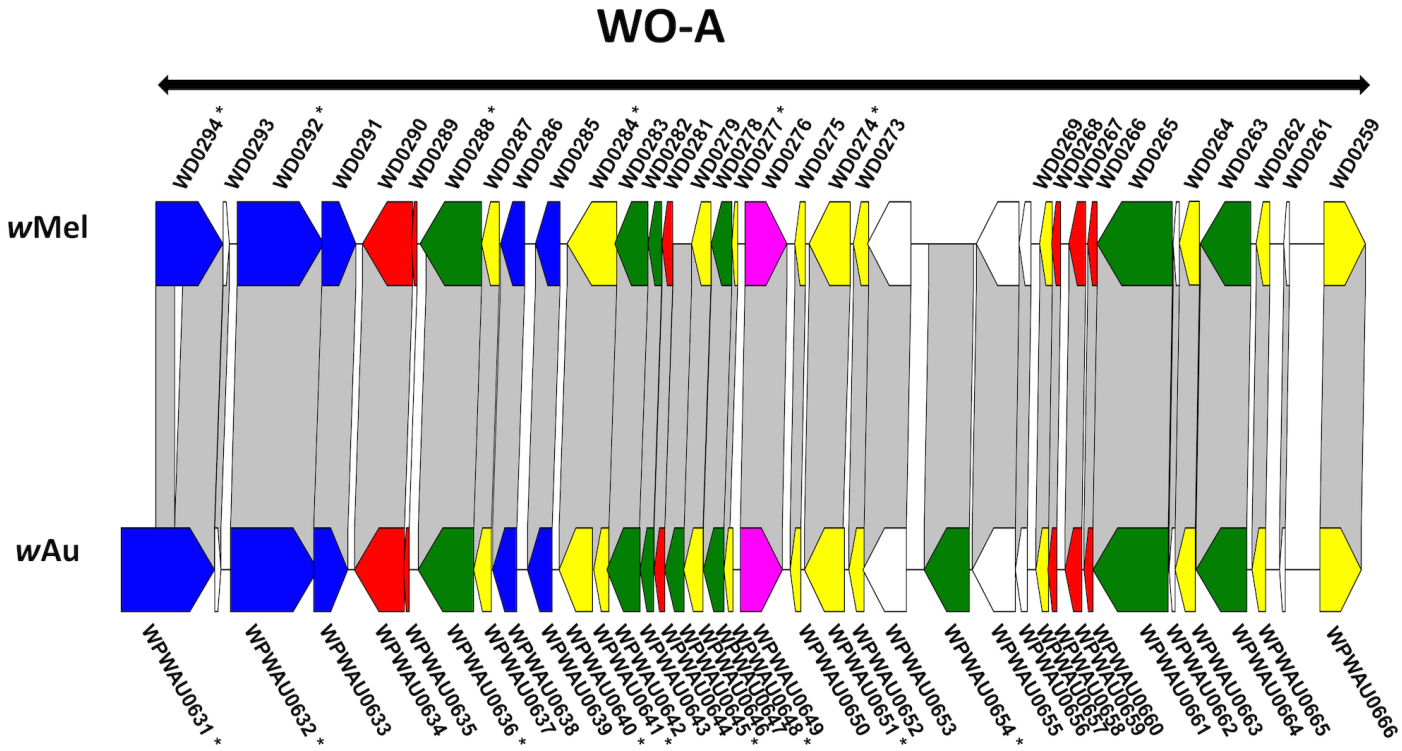

Figure 2.4. Comparison of WO-A in *w*Au and *w*Mel. Alignment of the WO-A prophage region between *w*Au and *w*Mel. Matching sequences corresponding to predicted CDSs, identified using ACT and Geneious alignments, are connected by grey blocks. Genes whose sequences differ between strains such that a CDS is not predicted in one strain are not represented in the strain lacking the predicted CDS, but their corresponding sequences are still connected to the CDSs in the other strain by grey blocks. The double-headed black arrow indicate regions that have been drawn inverted relative to their orientation in the genome, for clarity of alignment visualisation. Asterisks indicate genes with differences other than SNPs between *w*Au and *w*Mel. Internal indels less than 20 bp in size are not shown. Predicted CDSs are colour coded as follows: green, phage structural or replication genes; yellow, conserved hypotheticals; red, hypotheticals; blue, ankyrin repeat genes; magenta, transposases or reverse transcriptases. White arrows indicate sequences that are not annotated in one of the two strains and are probable pseudogenes or mis-annotations. Predicted CDSs that result from interruptions, frameshifts or nonsense mutations, which are combined into a single CDS in the other strain, are coloured the same as the CDS from which they are derived, even though they may also be pseudogenes.

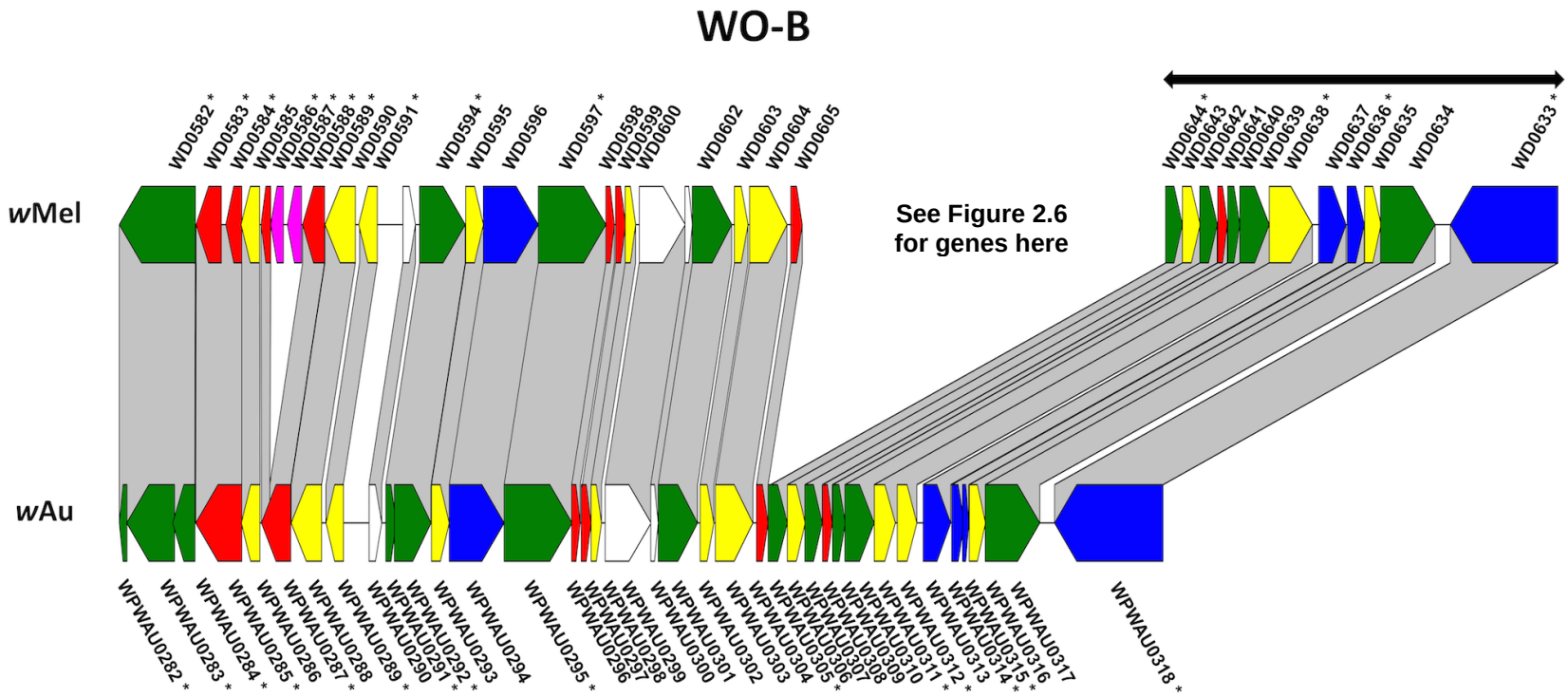

Figure 2.5. Comparison of WO-B in *w*Au and *w*Mel. Alignment of the WO-B prophage region between *w*Au and *w*Mel, constructed and presented as for Figure 2.4, with the same CDS colour coding, namely: green, phage structural or replication genes; yellow, conserved hypotheticals; red, hypotheticals; blue, ankyrin repeat genes; magenta, transposases or reverse transcriptases.

Figure 2.6. Comparison of genes inserted within *w*Mel WO-B in *w*Au and *w*Mel. Alignment of genes inserted within *w*Mel WO-B between *w*Au and *w*Mel, constructed and presented as for Figure 2.4, with the CDS colour coding: green, phage structural or replication genes; yellow, conserved hypotheticals; red, hypotheticals; blue, ankyrin repeat genes; magenta, transposases or reverse transcriptases; orange, enzyme genes; brown, membrane protein genes; purple, radC; black, transcriptional regulator genes; grey, others.

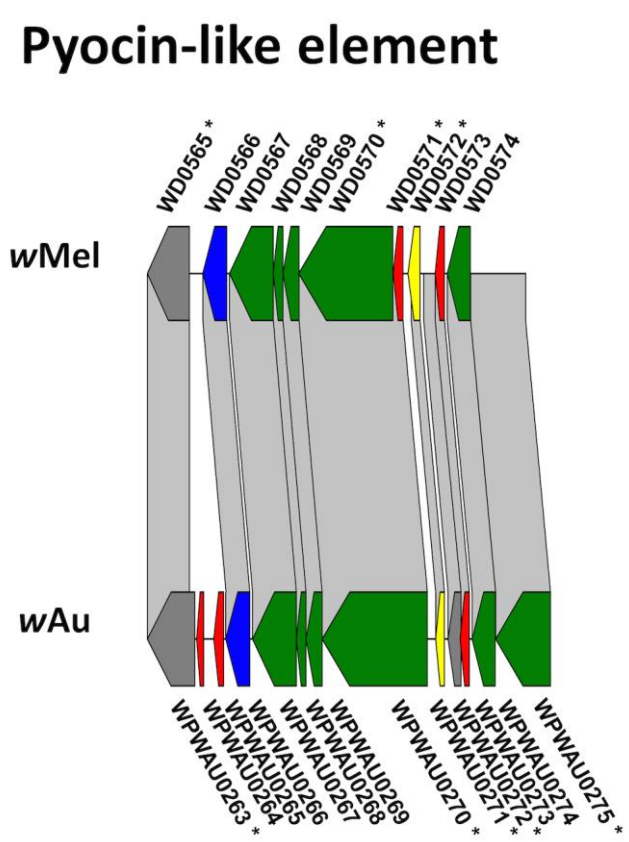

Figure 2.7. Comparison of pyocin-like element in *w*Au and *w*Mel. Alignment of the pyocin-like prophage region between *w*Au and *w*Mel, constructed and presented as for Figure 2.4, with the CDS colour coding: green, phage structural or replication genes; yellow, conserved hypotheticals; red, hypotheticals; blue, ankyrin repeat genes, grey, others.

*w*Kue[309], from which WO-A and WO-B were named[161]. As in WO-A, a high proportion of the genes in WO-B are disrupted in one of the two strains. Five genes are disrupted in *w*Au relative to *w*Mel, due to truncation or frameshift; another five are disrupted in *w*Mel relative to *w*Au, due to frameshift, IS element insertion or start codon mutation. Several of the genes between the lambdoid-like and P2-like blocks in *w*Mel, which include three of the transcriptional regulator genes discussed below, also differ significantly between *w*Mel and *w*Au (Figure 2.6). In addition to a higher than average frequency of SNPs between the two strains, five genes are disrupted in *w*Au relative to *w*Mel, due to truncation, frameshift or nonsense mutation, while two are disrupted in *w*Mel relative to *w*Au. Two genes contain small indels, and two more (*WD0631* and *WD0632*) are absent in *w*Au, discussed further below.

The third prophage region, a small pyocin-like element comprising *WD0565 – WD0574* in *w*Mel (Figure 2.7), is closer to the origin of replication in *w*Au (comprising *WPWAU0263 – WPWAU0275*). One gene is disrupted in *w*Au relative to *w*Mel, by a nonsense mutation, and four genes are disrupted in *w*Mel relative to *w*Au, by frameshift, nonsense mutation or truncation. In addition, there is an insertion after the first gene in *w*Au relative to *w*Mel, in which two additional genes are annotated. Overall the level of difference between *w*Mel and *w*Au in these three prophage regions is much higher than elsewhere in the genome.

2.3.1.6. Comparison of transcriptional regulator genes

Following a prior comparative genomic study of substrains of *w*Pip *Wolbachia* from *Cx. pipiens* mosquitoes, which implicated a transcriptional regular gene designated *wtrM* in CI[166], a comparison of the family of transcriptional regulator genes between *w*Au and *w*Mel constituted a focus of this study. These transcriptional regulator genes were found to differ in both organisation in the genome and sequence. Figure 2.8 illustrates the homology between *w*Au and *w*Mel transcriptional regulator genes; also shown is a comparison between *w*Mel and *w*Ri, a CI-inducing strain found in *D. simulans*. Figure 2.9 illustrates the differences between proteins that would be produced from the *w*Mel transcriptional regulator genes and their corresponding sequences in *w*Au. Of particular note is that the orthologue of *WD0622* is highly disrupted by a frameshift in *w*Au. The two identical sequences corresponding to *WD0622* (*WPWAU0256* and *WPWAU0687*) have a 1 bp insertion in *w*Au relative to *w*Mel, after bp 212, which causes a frameshift that would lead to premature termination of translation after 85 amino acids. Genes corresponding to the last 82 amino acids of *WD0622* (*WPWAU0257* and *WPWAU0686*) are also predicted since a substitution has produced a new start codon, although these may be mispredictions, as numerous SNPs and deletions have accumulated relative to *WD0622*. None of these genes include the DNA binding domains present in WD0622, suggesting that even if they are functional their activity is likely to differ significantly from that of their

counterpart in *w*Mel. Furthermore, there are IS5 elements inserted immediately downstream of the *w*Au genes corresponding to *WD0622*, so it may be that the regulation of the genes differs between the two strains. Genes such as this, which are conserved between multiple CI-inducing strains (*WD0622* has two orthologues in the CI-inducing *w*Ri strain) but disrupted in a non-CI strain, are prime candidates for involvement in CI.

In the *w*Mel genome, *WD0254*, *WD0255*, *WD0622*, *WD0623* and *WD0626* are in two regions that appear to be paralogous; *WD0254* is a truncated paralogue of *WD0622* (with the truncation appearing to be due to a transposase insertion) and *WD0255* a paralogue of *WD0623*. In the *w*Au genome, there also appear to be two genome segments containing paralogous sequences at approximately the same genomic positions as in *w*Mel. However, the segment at a similar position to *WD0254* and *WD0255* is inverted relative to its orientation in *w*Mel and is adjacent to genes that match *WD0628*-*WD0630* in *w*Mel, suggesting that one or more translocations have occurred, encompassing *WD0622* to *WD0630*; flanking IS elements provide a putative mechanism for its translocation and inversion. In *w*Mel there are sequence differences between the paralogues in the two paralogous transcriptional regulator regions. In *w*Au, however, *WPWAU0687*, *WPWAU0686* and *WPWAU0688* are identical to *WPWAU0256*, *WPWAU0257* and *WPWAU0255* respectively, while *WPWAU0691* is identical to *WPWAU0252* after the first 84 bp; this suggests that replacement by intragenomic recombination has occurred. *WD0296* appears to have an orthologue in *w*Au, *WPWAU0324*, in a similar genomic position, while there is no orthologue of *WD0508* present in *w*Au, consistent with a previous study that found that the region spanning *WD0506* to *WD0518* in *w*Mel is absent from *w*Au[177]. This gene is also absent in several other CI-inducing *Wolbachia* strains[177].

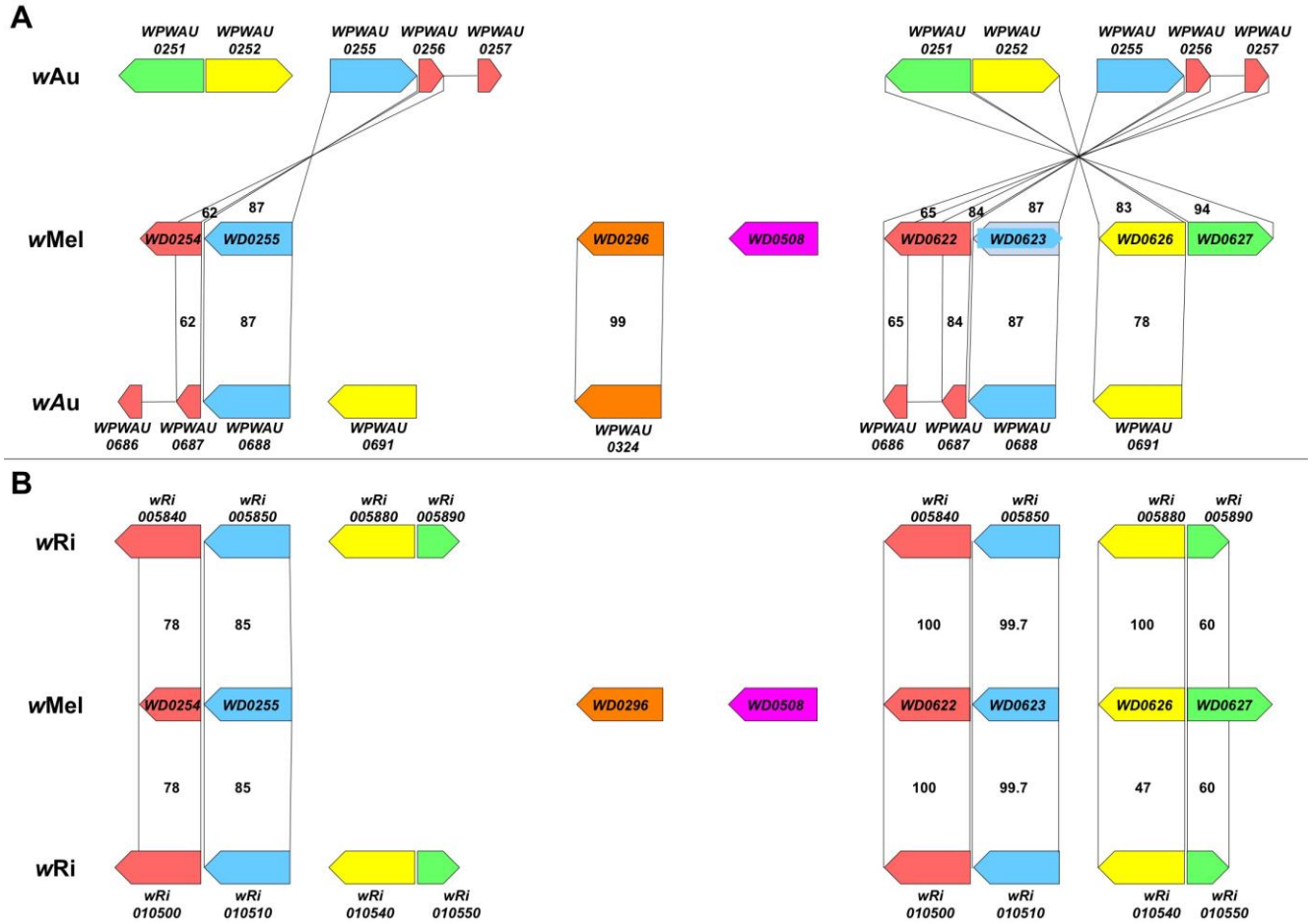

Figure 2.8. Homology between transcriptional regulator genes. An illustration of the *w*Mel transcriptional regulator genes and their homologues in *w*Au (A) and *w*Ri (B), identified using ACT and Geneious alignments. Genes depicted in the same colour are thought to be paralogous within an individual strain, and either orthologous or paralogous between strains. Numbers indicate the percentage amino acid similarity between any proteins produced from these genes.

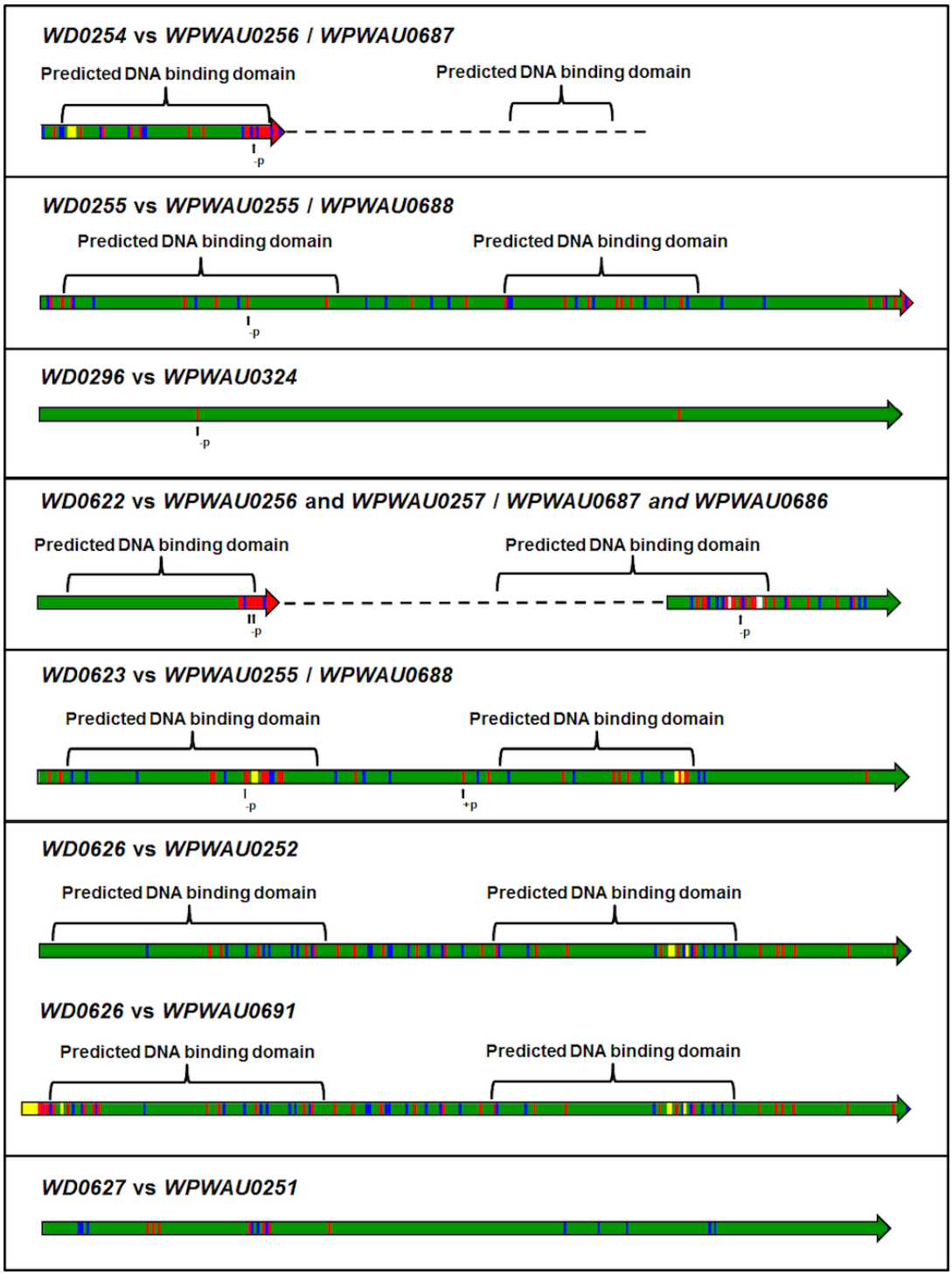

Figure 2.9. Differences in proteins from transcriptional regulator genes between *w*Au and *w*Mel. An illustration of the differences between proteins that would be produced from the *w*Mel transcriptional regulator genes and their corresponding sequences in *w*Au. Proteins predicted from the *w*Au sequences are shown. Green indicates identity with the *w*Mel translated sequence. Blue indicates substitution with a similar amino acid. Red indicates substitution with a dissimilar amino acid. Yellow indicates inserted amino acids that are present in the *w*Au protein but not the *w*Mel protein. White indicates deleted amino acids that are present in the *w*Mel protein but not the *w*Au protein. The locations of DNA binding domains predicted from the translated *w*Mel sequences are shown. Substitutions involving proline residues are indicated with arrows. – p: loss of a proline residue; + p: gain of a proline residue.

Analysis of the other transcriptional regulator gene sequences indicates that protein products from these genes would be different between *w*Au and *w*Mel (Figure 2.9), particularly *WD0623* and *WD0626*. The two identical *w*Au genes corresponding to *WD0623* (*WPWAU0255* and *WPWAU0688*) both have three small insertions in the putative DNA binding domains that result in an extra four amino acids and one amino acid substitution; there are 37 further amino acid substitutions, 23 of which are located in the putative DNA binding domains. There are substitutions involving proline residues; due to the unique conformational rigidity of the proline side chain, this could have a large impact on the secondary structure of any protein produced, and thus probably also its function. There are also SNPs and a 6 bp insertion in 5′ upstream regions where promoter elements have been shown to occur in prokaryotes, centred at −45 and −52[310]. In the *w*Au homologue of *WD0626*, *WPWAU0252*, two small insertions in the putative DNA binding domains would result in an extra three amino acids and one amino acid substitution. In addition there are 48 amino acid substitutions, 27 of which are located in the putative DNA binding domains. There are also SNPs and a 16 bp deletion in the 5′ upstream region spanning the −45 and −52 positions. Another factor that may affect expression is their different genomic location; in *w*Mel, *WD0622*, *WD0623* and *WD0626* are located within a prophage region, whereas in *w*Au this is not the case. The precise impact of all these differences described on protein function is hard to predict, but given in particular the changes in DNA binding domains it seems highly likely that their activity will be affected.

The total SNP density and the density of non-synonymous SNPs (dN) for *WD0623* and *WD0626* relative to their *w*Au orthologues are at the extreme end of the distributions of these measures over the genome (Figure 2.10). Comparing these measures between gene categories shows that the transcriptional regulator genes as a whole have a much higher density of total SNPs and non-synonymous SNPs than all other categories, as well as a higher density of synonymous SNPs (dS) (Figure 2.11). A Kruskal-Wallis test showed a

statistically significant difference between groups for total SNP density, dN and dS ($P < 0.05$ in all cases). dS was not significantly different between transcriptional regulator genes and other groups, except the structural protein and hypothetical protein groups. These findings suggest that the transcriptional regulator genes may be under positive selection.

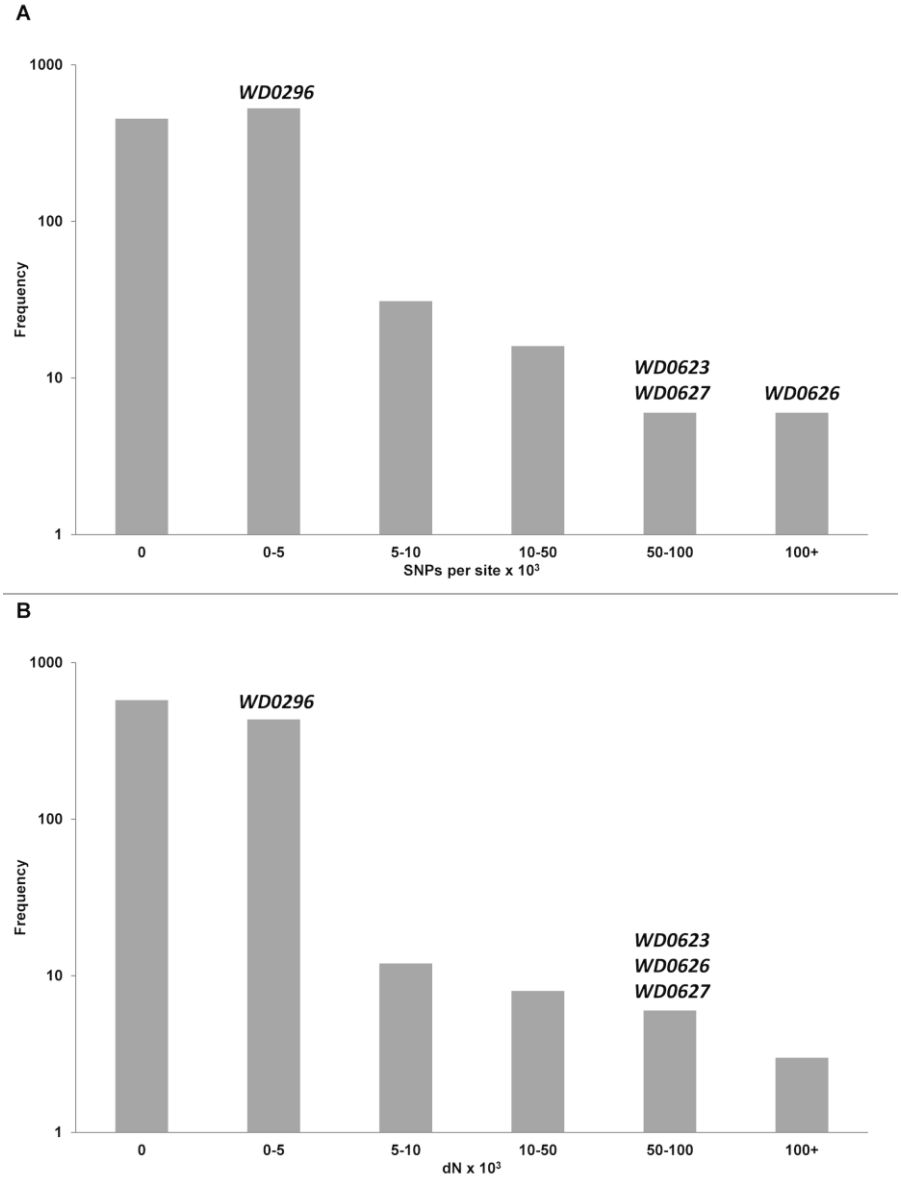

Figure 2.10. Frequency distribution of SNP density and dN. Graph showing the number of protein-coding genes in the *w*Mel genome within each SNP density (A) and dN (B) category. The frequency axis is drawn on a log scale. The bins containing the transcriptional regulator genes included in the analysis are indicated. Pseudogenes, genes that are potentially inactivated in *w*Au, IS elements and other genes with multiple ambiguous matches are excluded. The transcriptional regulator gene *WD0255* is excluded as its closest *w*Au sequence contains a frameshift and is more similar to *WD0623*. dN, number of non-synonymous SNPs per potential site.

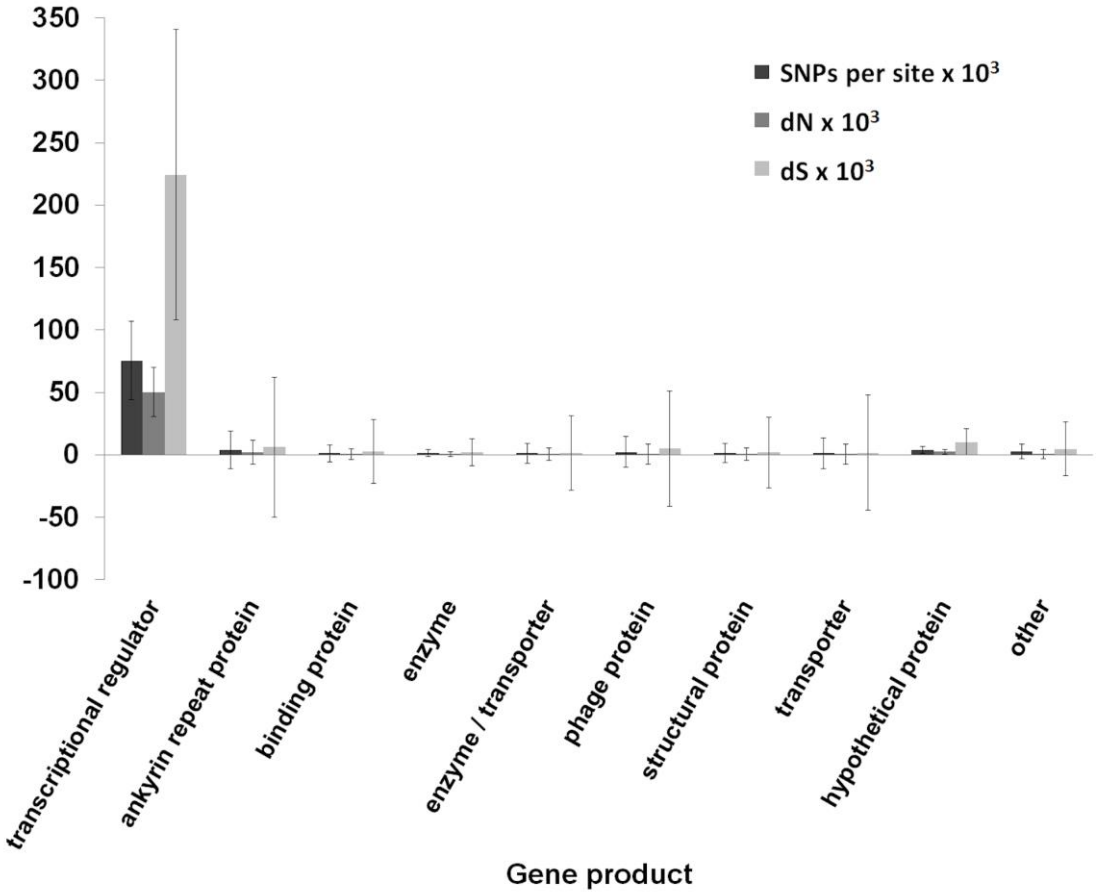

Figure 2.11. Mean SNP density, dN and dS by gene category. Graph showing the mean SNP density, dN, and dS for different categories of protein-coding gene in the *w*Mel genome. Error bars represent the standard error of the mean. Pseudogenes, genes that are potentially inactivated in *w*Au, IS elements and other genes with multiple ambiguous matches are excluded. The transcriptional regulator gene *WD0255* is excluded as its closest *w*Au sequence contains a frameshift and is more similar to *WD0623*. dN, number of non-synonymous SNPs per potential site; dS, number of synonymous SNPs per potential site.

2.3.2. Candidate CI gene functional testing

2.3.2.1. Transformation

Two lines with independent insertions of the *WD0626* construct were produced (designated lines 1 and 2).

2.3.2.2. Insert location determination

In order to generate balanced transformed lines it was necessary to know into which chromosome the transgene had been inserted. Insert location was determined using

adapter ligation-mediated PCR. For both lines there was only one PCR product (Figure 2.12), indicating a single insertion of the transgene. Using the sequences of these PCR products as queries in a BLAST search of *D. melanogaster* chromosome sequences revealed that for both lines the insertion was on chromosome 2.

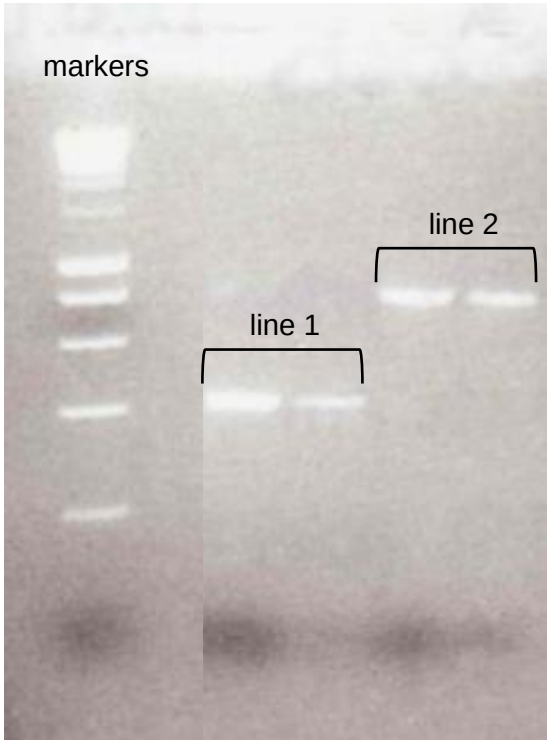

Figure 2.12. Gel showing products of adapter ligation-mediated PCR for determination of insert location.

2.3.2.3. WD0626 expression

qRT-PCR was performed in an attempt to confirm expression of *WD0626* in transformants, and determine its level relative to that in the natural *w*Mel infection. However, the primers used to amplify *WD0626* did not function as expected, producing poor standard curves, large variation between both biological and technical replicates, and showing amplification in control w^{1118} samples, in which there should be no *WD0626* expressed. In some cases the amplification in the w^{1118} samples was greater than in the *w*Mel-infected samples. Unfortunately it is therefore not possible to have confidence in the results of this analysis.

2.3.2.4. Crossing experiments

Two sets of crossing experiments were performed. In the first set males were used in crosses up to 24 hours after eclosion. In this set of experiments, control crosses of uninfected females with *w*Mel–infected males displayed weaker CI than expected, so in the second set of experiments males were used in crosses within eight hours of eclosion, as CI penetrance is stronger in younger males[130,132,306].

Figure 2.13 shows the results of the first set of crossing experiments for male BamGAL4/UAS-WD0626 flies. Female NosGAL4/UAS-WD0626 flies were found to be infected with *Wolbachia*, so crosses with these flies were excluded from analysis. Figure 2.14 shows the results of the second set of crossing experiments for male (A) and female (B) BamGAL4/UAS-WD0626 flies. In the second set of experiments the numbers of crosses in which mating occurred and sufficient eggs were laid were insufficient for analysis for BamGAL4/UAS flies from line 2 and for female BamGAL4/UAS flies from line 1 crossed to w^{1118} males, so these crosses were excluded from analysis. A Kruskal-Wallis test revealed no significant difference between the different crosses in the first set of experiments ($P > 0.05$ in all cases). There were, however, significant differences between the different crosses in the second set of experiments. For crosses involving transformed males, a Kruskal-Wallis test revealed a significant difference between crosses ($P = 6.6 \times 10^{-4}$). Post-hoc Wilcoxon tests with Bonferroni correction showed a significant difference between crosses with *w*Mel-infected females and BamGAL4/UAS-WD0626-1 males and w^{1118} females and *w*Mel-infected males ($P = 1.4 \times 10^{-3}$). For crosses involving transformed females, a Wilcoxon test revealed a significant difference between crosses with BamGAL4/UAS-WD0626-1 females and *w*Mel males and *w*Mel females and *w*Mel males ($P = 2.9 \times 10^{-3}$).

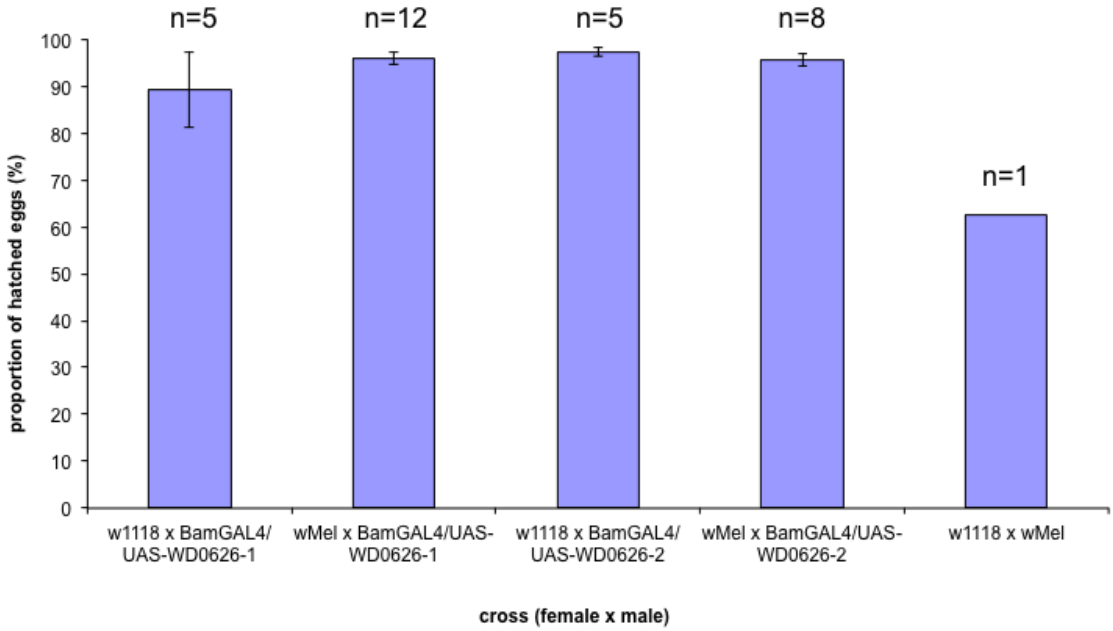

Figure 2.13. Hatch rates of crosses of male BamGAL4/UAS-WD0626, w^{1118} and *w*Mel flies from first set of crossing experiments. Error bars show standard error of the mean (SEM).

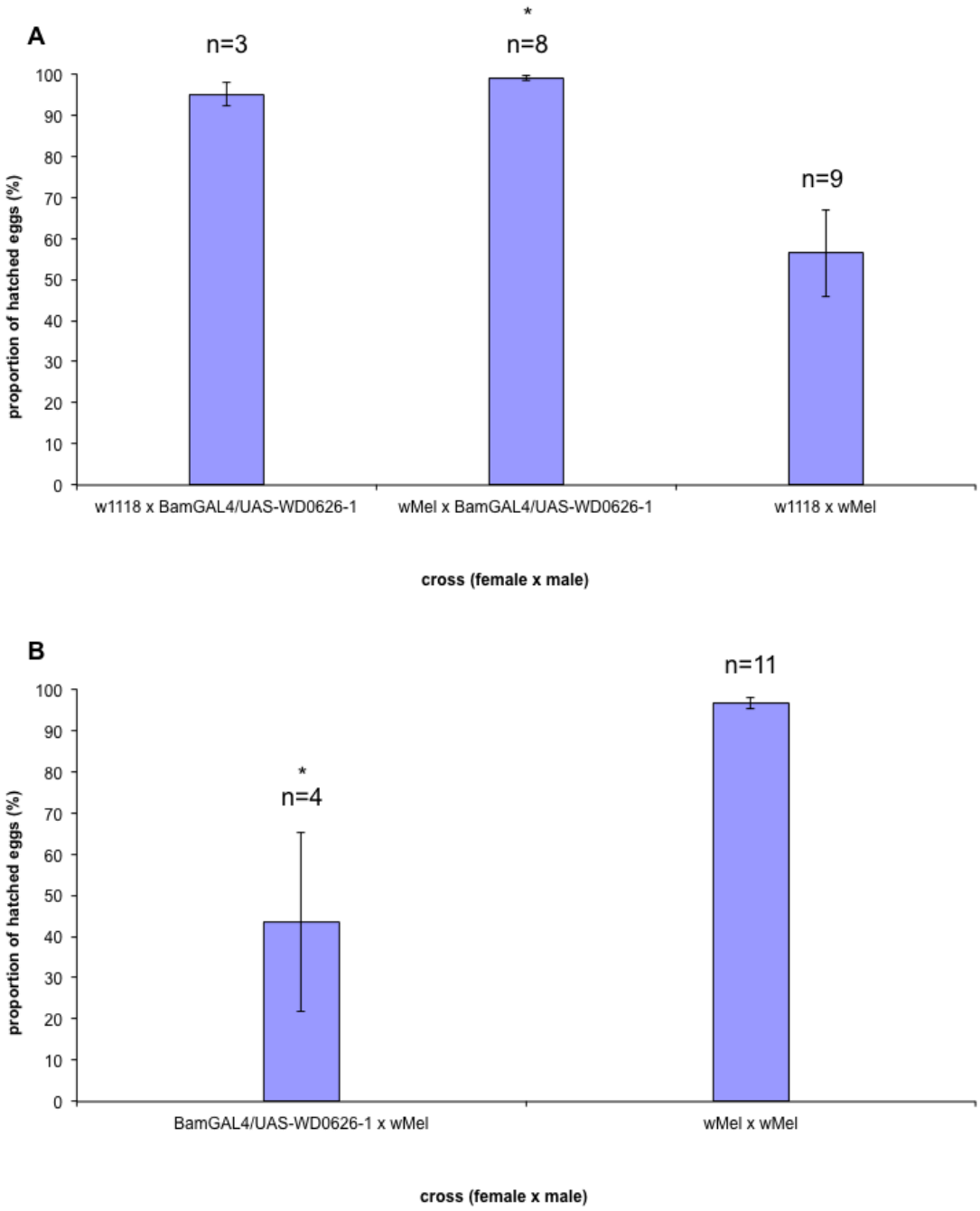

Figure 2.14. Hatch rates of crosses of BamGAL4/UAS-WD0626, w^{1118} and *w*Mel flies from second set of crossing experiments. (A) Hatch rates for crosses with male BamGAL4/UAS-WD0626 flies. (B) Hatch rates for crosses with female BamGAL4/UAS-WD0626 flies. Error bars show SEM. * $P < 0.05$ versus control cross (w^{1118} x *w*Mel or *w*Mel x *w*Mel).

2.4. Discussion

Here comparative genomics has been used to identify and investigate *Wolbachia* genes potentially involved in CI, followed by functional testing of one candidate gene. Previous work has compared the genomes of *w*Mel and *w*Bm, which infects the nematode *Brugia malayi*, to identify genes involved in host-specific adaptations[311], and some *w*Mel and *w*Au genes have been compared on an individual basis[178], but this is the first work known to compare whole *Wolbachia* genomes to identify candidate CI genes. There are also other noteworthy aspects to the methodology used, in addition to interesting findings.

2.4.1. *Wolbachia* comparative genomics

2.4.1.1. gDNA extraction

For the second round of sequencing, *Wolbachia* gDNA was extracted from infected cell lines, which represents a methodological departure from most previous studies, and is more convenient than techniques in previous *Wolbachia* genome sequencing projects, all of which except the sequencing of wBol1[274] have used *Wolbachia* gDNA extracted directly from their native hosts. It is a time-consuming and often laborious process to rear sufficient numbers of host insects for *Wolbachia* gDNA extraction, particularly for species with demanding rearing requirements. In contrast, transinfection of cells with *Wolbachia* is fairly easy to achieve, and amplification of cells to a suitable number is easier, quicker, and requires less space than whole organisms; this study used 24 flasks of cells, which were generated from a single flask in a few weeks. There were concerns that the sequence of the *w*Au from cultured cells might have accumulated differences compared to the *w*Au genome found in flies, due to a relaxation in cell lines of the selective pressures that apply in its native host. Even in native hosts, *Wolbachia* genomes can change in a short space of time; a recent study found a high level of variation in copy number of a region of eight genes known as 'Octomom' in *w*MelPop, and a line was identified in which a copy had been lost over the course of the study[312]. However, concerns over the fidelity of the *w*Au

genome sequence produced were alleviated by the observation of only one SNP between the sequence obtained using *w*Au from cultured cells and that using *w*Au from its native host. It is possible that after a longer period of time more differences from *Wolbachia in vivo* would accumulate, so use of recently generated *Wolbachia*-infected cell lines, as employed here, is advisable.

At ~60 %, the estimated purity of *w*Au versus host gDNA for the extract from cultured cells was lower than the >90 % for the extract from whole adult flies, which is comparable to the figure of up to 97 % reported previously using the same extraction method[275]. One explanation for the lower purity could be that *Wolbachia* densities may be lower within cultured cells than *in vivo*. Consistent with this, another study in which *Wolbachia* gDNA was purified from cells reported a higher level of host contamination[274], though at <20 % it was less than half that found here. A different cell line and different *Wolbachia* strain were used in this study; *Wolbachia* densities are known to vary *in vivo* between strains[235], and it could be that when *Wolbachia* is present in cell lines its density also varies depending on the cells used, which could explain the discrepancy in purities.

2.4.1.2. Sequencing and assembly

The achievement of a single contig assembly for *w*Au demonstrates that the PacBio RS II sequencing platform can be a very useful tool for rapid generation of complete bacterial assemblies. Furthermore, the generation of this single contig from a very small amount of DNA (approximately 2 ng), containing a substantial amount of host DNA contamination (~40 %), suggests that PacBio is well suited to use in cases where it is hard to obtain a large amount of gDNA, including obligate endosymbionts, like *Wolbachia*, that cannot be cultured outside host cells.

The sequence generated was largely consistent with data produced using the Illumina platform, with only one SNP between the two datasets. There were 88 indels relative to Illumina data; these were mostly single nucleotide insertions in the PacBio sequence, and were located in homopolymeric tracts, regions that are known to be prone to insertion errors in PacBio sequencing[313,314]. These were corrected after mapping the Illumina reads to the PacBio assembly. Combining the PacBio reads with the shorter but more accurate Illumina reads was found to be a very useful approach, consistent with other findings[315].

2.4.1.3. Comparative analysis

Comparison of the sequenced *w*Au genome to that of *w*Mel, which does induce CI, revealed a number of differences that could be involved in CI. Firstly, there were significant structural differences in the prophage regions, including disruption of a number of genes in both strains. Previous work has yielded conflicting evidence for involvement of prophage genes in CI. In contrast to the genomes of CI strains, the genome of the nematode-infecting non-CI strain *w*Bm contains no prophage regions[168], supporting a role for phage genes in CI. In addition, CI characteristics were transferred to uninfected hosts by injecting egg cytoplasm from infected hosts after filtering through a membrane with a pore size that would exclude *Wolbachia*, suggesting that a virus particle can carry the factors responsible for CI[179]. However, several studies found no correlation between WO phylogeny and CI characteristics in *Cx. pipiens*[180-183], the CI-inducing *w*Rec strain in *D. recens* lacks phage genes[164], and transformation of *D. melanogaster* with several phage genes did not induce or modify CI phenotypes[178].

Comparative analysis also suggested loss or potential inactivation of a number of genes outside the prophage regions in *w*Au. Transcriptional regulator genes in particular displayed considerable differences between *w*Au and *w*Mel, both in terms of genomic location and sequence; of these the *w*Mel gene *WD0622* may be the most promising to

examine as a CI candidate. Given that a transcriptional regulator gene has previously been implicated in CI[166], these represent important targets for further functional studies on the mechanism of CI. The *w*Mel genome region containing the transcriptional regulator genes *WD0622*, *WD0633*, *WD0626* and *WD0627* is also of particular interest with respect to CI given the proximity of *WD0631-2*, absent in *w*Au, because a *WD0631* orthologue protein was recently identified in mosquito spermathecae[193].

To confirm whether any of these transcriptional regulator genes are involved in CI, functional studies are required, as was performed for *WD0626* here. As more *Wolbachia* genomes are sequenced, further comparison of these genes between different strains will also be useful.

2.4.2. Candidate CI gene functional testing

Further investigation was performed for one member of the family of transcriptional regulator genes discussed above, *WD0626*. Results of the crossing experiments with *WD0626* transformants do not support a role for *WD0626* in CI. For *WD0626*-transformed males, there were no significant differences in hatch rate between crosses with wMel-infected or uninfected females, suggesting that *WD0626* does not induce modification. The mean hatch rates were lower for uninfected females though, and in one cross the hatch rate was lower (57 %) than that of the control incompatible cross of *w*Mel-infected males with uninfected females. The sample sizes of suitable crosses for analysis was fairly small, and it would be interesting to see if a significant difference between crosses of transformed males with *w*Mel-infected or uninfected females could be observed with a larger sample size. Interestingly, the hatch rates for crosses of *WD0626*-transformed males with uninfected females were more variable than those for crosses with *w*Mel-infected females. It is possible that other factors such as male age were impacting on the ability of

WD0626 to cause sterility, and that penetrance of sterility is more sensitive to this than in control crosses.

The hatch rate of crosses with *w*Mel-infected males was significantly lower for *WD0626*-transformed females than for *w*Mel-infected females, and was actually lower on average than that of incompatible crosses of uninfected females with *w*Mel-infected males, suggesting that *WD0626* does not induce rescue. However, as for crosses with *WD0626*-transformed males and uninfected females, the hatch rate was variable, and in one case was 100 %. Again the sample size of suitable crosses for analysis was fairly small, and it would be interesting to see if any rescue effect could be observed with a larger sample size.

The results here do not rule out a role of *WD0626* in CI, for a number of reasons. Firstly, it should be borne in mind that the messenger RNA (mRNA) expression levels of *WD0626* in the transformants could not be determined, and even if this had been successful, the level of protein expression would still have been uncertain; mRNA levels do not always correlate with protein levels, and this correlation can be weakened in transgenic systems due to different codon usage between the source of the transgene and the host, and the fact that transgene products might be less stable in the host cell. Proper protein folding may also have been impaired in *D. melanogaster* due to the absence of required co-factors present in *Wolbachia*. It could be that functional *WD0626* protein was expressed at a low level relative to that in the natural *w*Mel infection, and that at higher levels an effect on CI would have been observed. In addition to expression levels, expression timing and location, both at the organism level and cellular level, may differ for *WD0626* in the transformants compared to in the natural *w*Mel infection, and these could all affect CI. The different genomic context of *WD0626* in the transformants may affect expression and function characteristics, in addition to potentially unknown factors.

It is also worth noting that hatch rates are not the only measure of an effect on CI. It could be that *WD0626* was exerting an effect which was not manifest in hatch rates but which nonetheless affects embryo development. Consistent with this, work in *Cx. quinquefasciatus* found that knocking down the host candidate CI gene *CPIJ005623* did not affect hatch rates but did significantly increase the proportion of embryos surviving until later stages in incompatible crosses[166].

A major limitation of this work is the study of a single gene in isolation. CI is likely to be a complex process involving many genes, both *Wolbachia* and host. It could be that *WD0626* does have a role in CI, but works in combination with other genes. When Yamada *et al*. found no effect of expressing *w*Mel candidate CI genes individually, they tried expressing them in a *w*Au-infected background[178]. Similarly, it would be interesting to cross the *WD0626* transgene into a *Wolbachia*-infected background to investigate whether expression of *WD0626* could modify CI induction or rescue in the presence of other *Wolbachia* genes.

2.4.3. Conclusions

Here novel methodology has been used for *Wolbachia* genome sequencing, followed by the first comparison of whole genomes to identify genes potentially involved in CI. A number of candidate genes were identified, and although functional testing of one candidate failed to provide evidence for a role in CI, the work provides a useful platform for further investigation. Overall it contributes to the important aim of gaining an improved understanding of the molecular basis of CI.

3 Identification and investigation of candidate genes involved in *Wolbachia*-host interactions using comparative transcriptomics

3.1. Introduction

While comparative genomics is useful for identifying genes differing in presence/absence, sequence or genomic position, comparative transcriptomics identifies genes differing in mRNA expression (and thus, in most cases, levels of protein expression). This is also a useful way to identify candidate genes involved in a phenotype, as well as confirming transcription of genes identified by genomic analysis. Here, it was applied using RNA-seq to further identify and investigate candidate genes for CI involvement, as well as candidate genes for viral inhibition by *Wolbachia*.

3.1.1. Identification and investigation of candidate host CI genes

In addition to *Wolbachia* genes, host genes are likely to play a role in CI, as discussed in Chapter 1. Several studies have identified candidate host CI genes by comparing gene expression between *Wolbachia*-infected and uninfected hosts[187,189-191]. However, while functional analysis of these genes was consistent with a role in CI, in general genes identified by this methodology could have altered expression due to some other aspect of *Wolbachia*-host interaction, rather than being involved in CI. Comparing gene expression between a host infected with a CI strain and the same host infected with a non-CI strain would narrow down identified genes to those most likely to be involved in CI. No such comparison is known to have been performed previously; this analysis was therefore undertaken here, with gene expression in the absence of *Wolbachia* also measured to provide an extra control.

As described in Chapter 2, *w*Au is a non-CI strain, and the sequencing of its genome described in Chapter 2 made it more amenable to transcriptomic analysis. While the CI strain *w*Mel is the most closely related *Wolbachia* strain to *w*Au, hence why it was chosen for the comparative genomics discussed in Chapter 2, the two strains are not naturally present in the same host species – *w*Mel infects *D. melanogaster* while *w*Au infects *D. simulans*. However, other CI strains are found in *D. simulans*, one of the most common being *w*Ri[316]. Thus, to identify and investigate candidate host CI genes, RNA-seq was performed on *D. simulans* gonads that were uninfected, infected with *w*Ri, or infected with *w*Au. RNA was extracted from gonads rather than whole flies, as this is where CI-related effects are most likely to occur. Separate samples were taken comprising adult male testes and adult female ovaries, as *Wolbachia* is likely to have differing effects on gene expression in these tissues.

3.1.2. Identification and investigation of candidate *Wolbachia* CI genes

Comparing expression of orthologous *Wolbachia* genes between *w*Au and *w*Ri would allow further identification and investigation of *Wolbachia* candidate CI genes. There are relatively few studies that have analysed *Wolbachia* gene expression, and none known that have compared expression between a CI and a non-CI strain. Difficulty in obtaining sufficient suitable *Wolbachia* RNA is a factor in this, given it has so far proven impossible to culture *Wolbachia* outside a host, *Wolbachia* RNA comprises only a small fraction of the total RNA present in host cells, and bacterial RNA degrades much faster than eukaryotic RNA. Some *in vitro* studies have performed RNA-seq on purified *Wolbachia*[317,318], but non-native cell lines were used, which may change gene expression, as may the stressful purification protocol. Other studies have performed RNA-seq on *Wolbachia* from cultured cells without purification[319,320], and Woolfit *et al*. also analysed *w*MelPop and *w*MelCS strains in native host tissues (*D. melanogaster* heads), but only small non-coding RNAs were sequenced, and differential analysis was not performed[320].

There were therefore concerns that it may be difficult to obtain enough *Wolbachia*-derived reads at an affordable sequencing depth. However, differential expression analysis with *Wolbachia* RNA-seq data from native hosts has been successfully performed previously without purification and at a reasonable sequencing depth[321,322]. It was found that the *D. melanogaster* strain used to generate transcriptome data for the modENCODE project[323-325] was infected with *Wolbachia*, and the depth of sequencing here (calculated as maximum < 6 Gb per condition) was sufficient to perform differential expression analysis of *Wolbachia* genes[322]. Another study, on *w*Oo in the nematode *Onchocerca ochengi*, used an even lower sequencing depth, calculated as maximum <2 Gb per condition[321]. In this study, only 5 % of reads mapped to the *Wolbachia* genome, but this was still sufficient for almost 98 % of *Wolbachia* predicted gene models to be classified as transcribed. A more recent study of *w*Di in the nematode *Dirofilaria immitis* used a total sequencing depth of only 0.15 Gb over seven conditions, and still detected up to 97 % of predicted *w*Di transcripts[326].

The number of *Wolbachia* reads depends not just on sequencing depth, but also on *Wolbachia* density and the relative sizes of the *Wolbachia* and host transcriptome in the tissue used. *Wolbachia* density is generally higher in the reproductive tissues than elsewhere, and there is no reason to suppose the *D. simulans* transcriptome in these tissues is significantly larger or the *w*Au or *w*Ri transcriptomes significantly smaller than the corresponding transcriptomes in the successful studies. Thus, to identify and investigate candidate *Wolbachia* CI genes, RNA-seq was performed on *D. simulans* gonads infected with *w*Ri or *w*Au, as described above. To maximise the proportion of reads from *Wolbachia*, a novel protocol was used involving reverse polyA selection prior to ribodepletion, which should eliminate mature host mRNA as well as rRNA.

3.1.3. Identification and investigation of candidate host viral inhibition genes

As discussed in Chapter 1, in addition to causing CI, *Wolbachia* can also inhibit the transmission of pathogens, most notably RNA viruses. Previous studies have suggested various theories for this inhibition, such as immune stimulation[202,222,225,249,327], more specific elements of immune response such as increased concentration of ROS[243,244] and modulation of autophagy[328,329], competition for resources[223,258], and steric hindrance[265], but the exact mechanism(s) are unclear. Viral inhibition is dependent on both the host and *Wolbachia* strain. Thus a better understanding of the mechanisms of inhibition would allow a more informed choice before undertaking the time-consuming process of generating a transinfection, as well of being of general interest for use of such transinfections in insect control, and from a basic biology perspective. Genes potentially involved in the *Wolbachia* inhibitory phenotype may be identified through high-throughput transcriptomics comparing gene expression in hosts with inhibitory *Wolbachia* strains against that in uninfected hosts or those with non-inhibitory *Wolbachia* strains. Microarray experiments have been performed to this end[225,327] but no known RNA-seq studies.

One *Wolbachia* transinfection that has been found to inhibit virus transmission by the host is that of *w*Mel in the mosquito *Ae. albopictus*[122]. This mosquito is a major vector of dengue, chikungunya and other diseases[330]. All known *Ae. albopictus* populations carry a superinfection of the *Wolbachia* strains *w*AlbA and *w*AlbB[124,125]. An uninfected line of the Uju strain of *Ae. albopictus* has been produced using tetracycline curing[331]. This uninfected strain was used to generate the artificial transinfection with *w*Mel[122], which was found to completely abolish transmission of dengue[122] and chikungunya[332]. Thus, to identify and investigate genes that may be involved in viral inhibition, RNA-seq was performed on *Ae. albopictus* mosquitoes that were uninfected, infected with *w*Alb, or infected with *w*Mel.

3.2. Methods

3.2.1. Insects

3.2.1.1. D. simulans

D. simulans flies in a Brisbane genetic background were used. The maternal lineage of the uninfected and wAu-infected lines derive from several isofemale lines that were field collected from Perth in 2011/12. The maternal lineage of the wRi-infected line derives from isofemale lines collected from Brisbane in 2013. The uninfected and *w*Au-infected lines were outbred over several generations to a Brisbane genetic background prior to acquisition for this work. Once acquired, they were backcrossed for three generations (uninfected females crossed with the three different types of males) to further homogenise the genetic backgrounds. Flies were reared mainly by Dr Vincent Geoghegan in Prof. Steven Sinkins's group at Lancaster University (Lancaster, UK), as described in Chapter 2, though using a diet containing white sugar rather than oats and treacle (see Recipe 2 in Appendix 1). Flies used for experimental samples were reared at low density (100 eggs per bottle). Eggs were collected as described in Chapter 2. Flies used for experimental samples were collected within 2 hours of eclosion, and males and females separated.

3.2.1.2. Ae. albopictus

Ae. albopictus mosquitoes in an Indonesian background (Uju line) were used. Mosquitoes were reared by Dr Caroline Harris in Prof. Steven Sinkins's group at Lancaster University (Lancaster, UK), at 27 +/-2 °C and 70 +/-10 % relative humidity with a 12:12 hour light:dark cycle. Adults were kept in 30 cm^3 cages (BugDorm, Taiwan) and provided *ad libitum* access to water and sucrose in the form of a moistened ball of cotton wool and a sugar cube placed on top of the cage. Females were fed with defibrinated sheep blood approximately 5-10 days after eclosion, using Hemotek artificial membrane feeders

(Discovery Workshops, UK). Generation of the *w*Mel-infected and uninfected lines has been described previously[122].

3.2.2. Dissection and RNA extraction

3.2.2.1. D. simulans

For host RNA-seq, virgin males were used for dissection of testes at 24 +/- 1 hours after eclosion, and virgin females were used for dissection of ovaries at 3 days after eclosion. For *Wolbachia* RNA-seq, virgin males were used for dissection of testes at ~1 day after eclosion, and virgin females were used for dissection of ovaries at ~3 days after eclosion. Gonads were dissected in phosphate-buffered saline (PBS), transferred into lysis buffer from a Norgen Total RNA Purification Kit (Norgen Biotek Corp., Canada) containing 1 % β-mercaptoethanol, then flash frozen in liquid nitrogen. Samples were stored at -80 °C until RNA extraction. After thawing, samples were homogenised by vortexing with borosilicate beads. Total RNA was then extracted using a Norgen Total RNA Purification Kit (Norgen Biotek Corp., Canada), according to the manufacturer's instructions. Contaminating DNA was removed with Ambion Turbo DNase, according to the manufacturer's instructions. RNA concentration and quality was assessed using a NanoDrop spectrophotometer (Thermo Scientific, USA).

For each condition (*w*Au-infected, *w*Ri-infected and uninfected), three biological replicates were prepared. Each replicate for host RNA-seq comprised pooled gonads from five flies. Each replicate for *Wolbachia* RNA-seq comprised pooled ovaries from 10 females or pooled testes from 20 males.

3.2.2.2. Ae. albopictus

Dissection and RNA extraction was performed by Dr Caroline Harris at Lancaster University (Lancaster, UK). For each condition (*w*Alb-infected, *w*Mel-infected and

uninfected), three biological replicates were prepared, each consisting of a pool of seven mosquitoes.

3.2.3. Sequencing

3.2.3.1. D. simulans and Wolbachia

Library preparation and sequencing were performed at the Wellcome Trust Institute for Human Genetics (Oxford, UK). For the samples for host RNA-seq, stranded paired-end libraries were generated using standard polyA selection. For the samples for *Wolbachia* RNA-seq, stranded paired-end libraries were generated using a custom preparation comprising reverse polyA selectiol. In both cases ribodepletion was performed using a Ribo-Zero Gold rRNA Removal Kit (Epidemiology) (Illumina, UK), and libraries were completed using an Illumina TruSeq Stranded Total RNA Library Prep Kit (Illumina, UK). Sequencing was performed on the Illumina HiSeq 2000 platform, using a read length of 100 nt. *D. simulans* libraries were split across two lanes.

3.2.3.2. Ae. albopictus

Library preparation and sequencing were performed at Imperial College (London, UK). Unstranded paired-end libraries were generated using standard polyA selection. Libraries were sequenced on the Illumina HiSeq 1500 platform, using a read length of 101 nt. Libraries were split across multiple lanes.

3.2.4. Sequence data processing

3.2.4.1. D. simulans and Wolbachia

Sequence data processing up to and including alignment of reads to genomes was performed at the Wellcome Trust Institute for Human Genetics (Oxford, UK). Quality control was performed separately for separate lanes to identify any problematic technical variation between lanes. Reference indexes of the *D. simulans* genome assembly release

2.0.1 (downloaded from FlyBase[333]) and the *w*Au and *w*Ri genome assemblies (downloaded from GenBank[334]) were constructed using Bowtie2 (v2.1.0)[335]. Reads were aligned to the *D. simulans* index using TopHat2 (v2.0.12)[336], and to the *w*Au and *w*Ri indexes using Bowtie2 (v2.1.0)[335]. PCR and optical duplicate reads were removed using Picard Tools[337].

3.2.4.2. Ae. albopictus

The overall quality of the sequencing reads was assessed using FastQC (v0.11.2)[338]. Quality control was performed separately for separate lanes to identify any problematic technical variation between lanes. Raw reads were processed to remove adapter sequences and sequences of poor quality using cutadapt (v1.7.1)[339]. A reference index of the *Ae. albopictus* genome assembly, kindly provided by Jiabao Xu (University of California, Irvine, USA), was constructed using Bowtie2 (v2.2.1)[335]. Trimmed reads were aligned to this index using TopHat2 (v2.0.11)[336]. A transcript assembly was generated from the alignments using Cufflinks (v2.2.1)[340] with RABT methodology[341] followed by Cuffmerge (v1.0.0)[342]. In addition, a *de novo* assembly was generated using Trinity (version 2014/07/17)[343]. The assemblies were evaluated using the script contig_stats.pl[344] and the program RSEM-EVAL[345].

3.2.5. Expression analysis

Expression quantification and differential expression analysis was performed using Cuffdiff2 (v2.2.1)[346]. For *D. simulans* and *Wolbachia*, the alignment files provided by the Wellcome Trust Institute for Human Genetics were used, along with downloaded reference annotation files corresponding to the genome assemblies described above. For *Ae. albopictus*, the assembly and annotation produced by Cufflinks and Cuffmerge was used. Expression levels were quantified in units of fragments per kilobase of transcript per million mapped reads (FPKM); this accounts for the fact that longer transcripts will

produce more reads, and normalises expression levels to the amount of total input RNA. A false discovery rate (FDR) value of 0.05 was used as the threshold for significance. Global expression was visualised using CummeRbund (version 0.1.3)[347].

3.2.5.1. Identification of Wolbachia orthologues

To enable comparison of expression between *w*Au and *w*Ri, orthologues between the two strains were identified. BLAST searches were performed using *w*Au gene sequences as queries against *w*Ri gene sequences, and vice versa. Orthologues were defined as genes displaying reciprocal best hits.

3.2.6. Assignment of functional information

D. simulans GO annotations were already present in the downloaded reference annotation. *Wolbachia* and *Ae. albopictus* GO annotations were assigned using BLAST2GO (version 3.0.10 or 3.1.3, respectively)[348], based on BLASTx searches of predicted gene sequences. For *Wolbachia*, the search was performed against the nr database. For *Ae. albopictus*, it was performed against protein sequences from the *Ae. aegypti* AaegL3.3 geneset (downloaded from VectorBase). For *Wolbachia*, the May 2015 version of the gene ontology files was used. For *Ae. albopictus*, the September 2015 version was used. In both cases the InterProScan function was used to add annotations based on similarity to domains in InterPro[295]. Functional enrichment analysis using GOSeq (see below) required assignment of all ancestral GO terms as well as the most specific; this was performed in R using the Bioconductor package GO.db (version 3.0.0)[349]. Kyoto Encyclopaedia of Genes and Genomes (KEGG) pathway annotations were assigned using BLAST2GO, based on GO annotations.

3.2.6.1. Ae albopictus dengue virus host factors

The names of *Ae. aegypti* orthologues of *D. melanogaster* genes identified as putative dengue virus host factors (DVHFs) by Sessions *et al*.[350] were downloaded from FlyBase, and their protein sequences downloaded from VectorBase. BLAST searches were performed using these protein sequences as queries against predicted *Ae. albopictus* transcript sequences, and vice versa. Putative *Ae. albopictus* DVHFs were defined as genes displaying reciprocal best hits.

3.2.7. Functional enrichment analysis

Functional enrichment analysis was performed using GOSeq (version 1.20.0)[351]. A two-sided test with Benjamini-Hochberg multiple testing correction was used. An FDR value of 0.05 was used as the threshold for significance.

3.3. Results

3.3.1. Identification and investigation of candidate host CI genes

3.3.1.1. Sequencing and read alignment

RNA sequencing was performed on RNA extracted from separate ovaries and testes samples from *D. simulans* infected with *w*Ri, *w*Au or uninfected. Three replicates were sequenced for each sample type. In total 720,932,532 paired end reads of 100 nt were generated, corresponding to 72.1 Gb of data. 84.7 % of these reads mapped to the *D. simulans* genome using TopHat2 (v2.0.12)[336].

3.3.1.2. Global expression analysis

Following alignment, quantification was performed using Cuffdiff2 (v2.2.1)[346]. In total, 14,946 annotated genes were identified as being expressed in at least one of the samples. 13,711 of these were expressed in at least one ovary sample, and 14,850 in at least one testes sample.

Global results were visualised using CummeRbund (v0.1.3)[347]. A heatmap representing expression levels of all expressed genes in each of the six sample types is shown in Figure 3.1. At first glance this suggests that overall gene expression among the different *Wolbachia* infection statuses is fairly similar; there is a much greater overall difference between ovaries and testes, with a large number of genes expressed quite highly in testes but at very low levels in ovaries. The clustering in the heatmap shows that in terms of global expression, *w*Ri-infected and *w*Au-infected ovaries are more similar to each other than to uninfected ovaries, but *w*Ri-infected testes are more similar to uninfected than to *w*Au-infected testes. This is shown more clearly by the dendrogram in Figure 3.2.

Scatter plots and volcano plots between the samples show the direction of this pronounced difference in *w*Au-infected testes – there is a large subset of genes that show significant downregulation in *w*Au-infected versus both *w*Ri-infected and uninfected testes (Figure 3.3 and Figure 3.4). As in Figure 3.1, it is apparent there are much larger differences between testes and ovaries than between different *Wolbachia* infection statuses. In all comparisons between testes and ovaries there appears to be a large subset of genes upregulated in the testes compared to the ovaries.

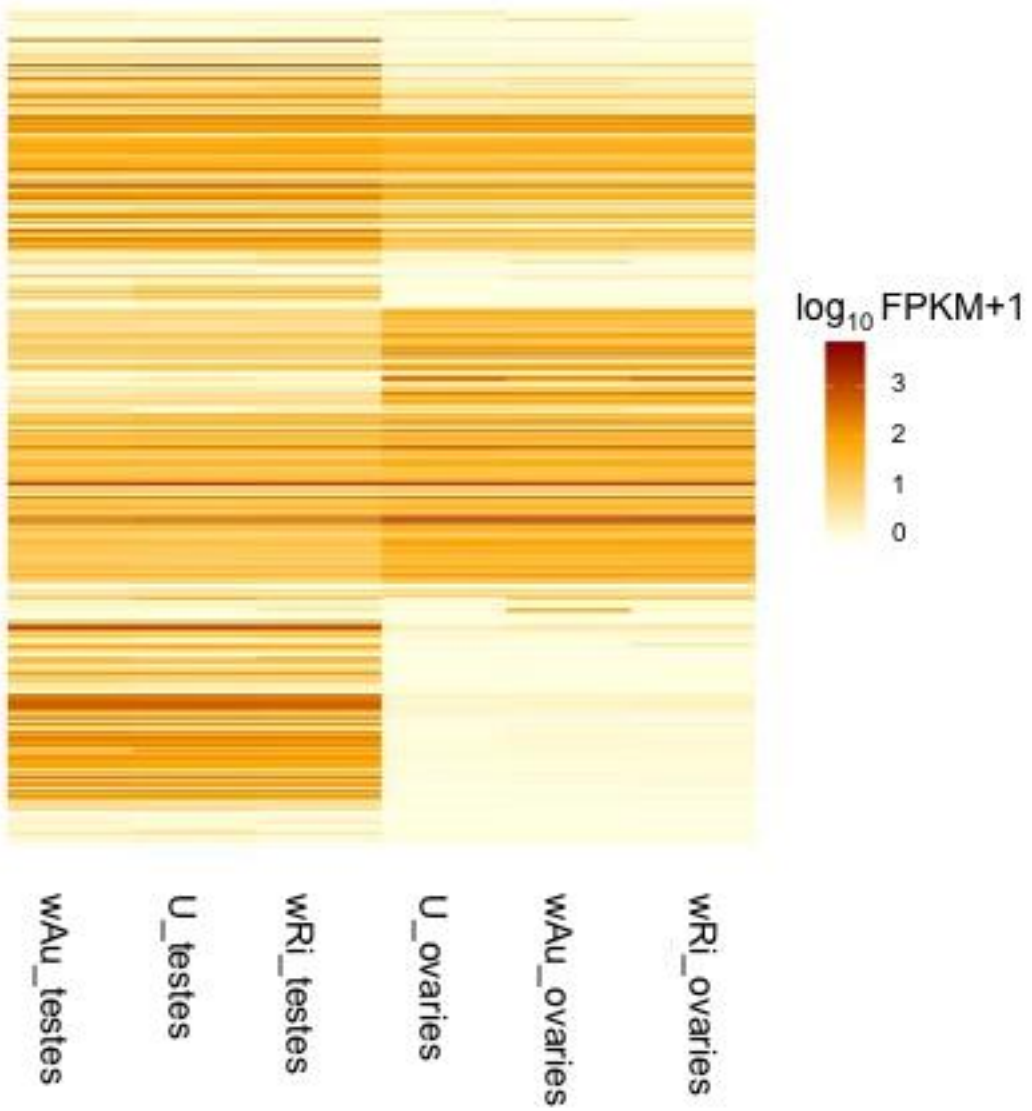

Figure 3.1. Heatmap illustrating expression levels for all expressed host genes in ovaries and testes from uninfected, *w*Ri-infected and *w*Au-infected *D. simulans.* U = uninfected. FPKM, fragments per kilobase of transcript per million mapped reads.

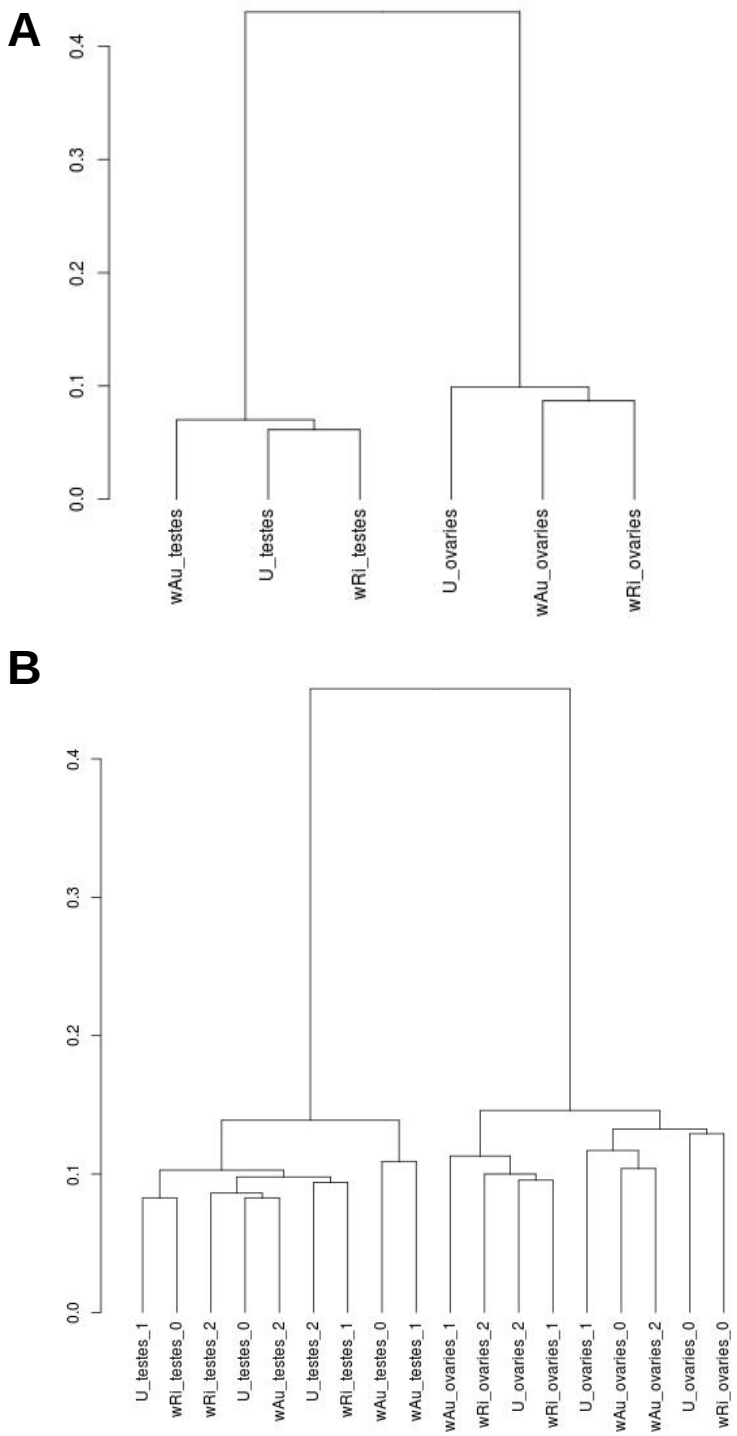

Figure 3.2. Dendrogram showing similarity in terms of host gene expression among ovaries and testes from uninfected, *w*Ri-infected and *w*Au-infected *D. simulans*, based on Jensen-Shannon distances, with biological replicates combined (A) or separate (B). U = uninfected.

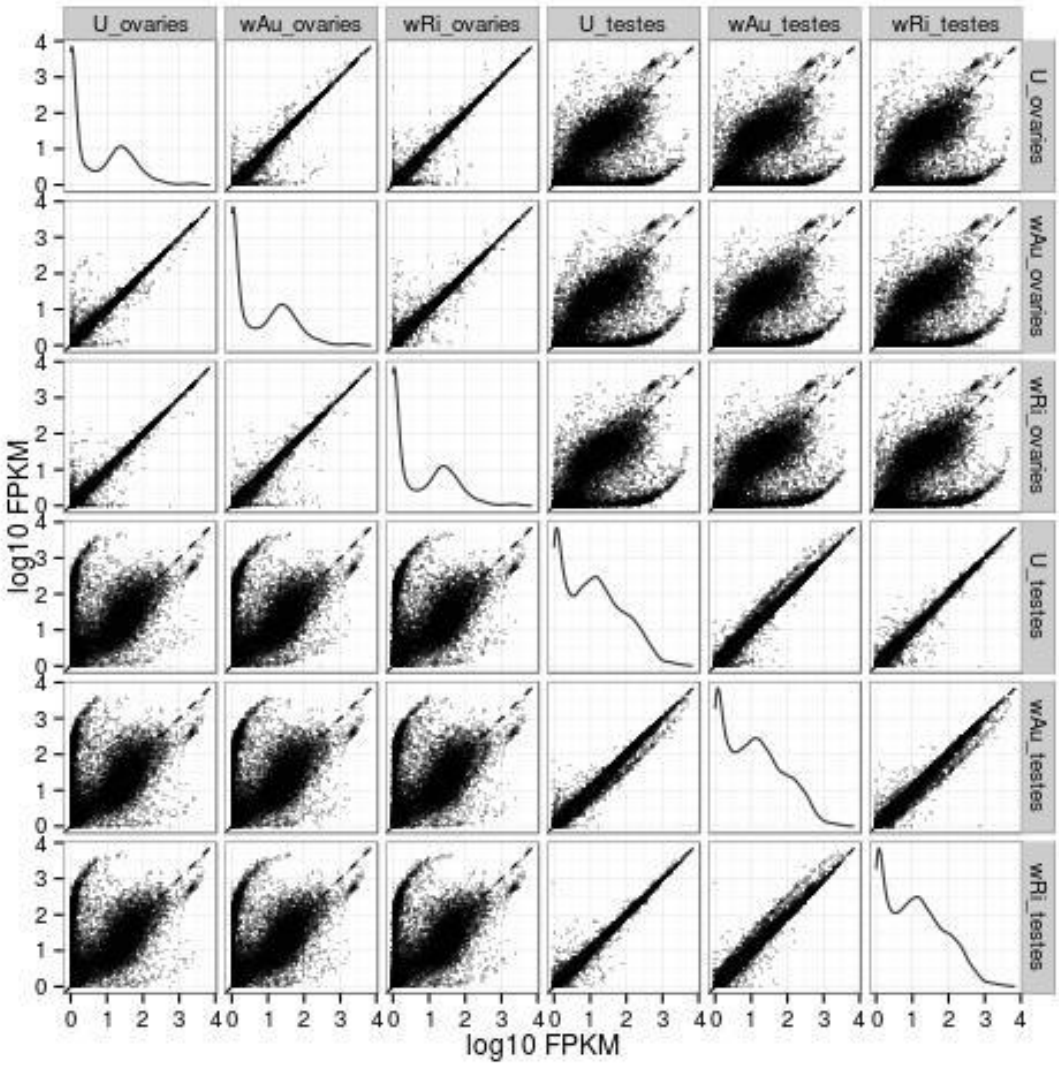

Figure 3.3. Scatter plots comparing host gene expression in ovaries and testes from uninfected, *w*Au-infected and *w*Ri-infected *D. simulans*. Plots between identical samples show the distribution of FPKM values for that sample. Dotted lines are identity lines. U = uninfected. FPKM, fragments per kilobase of transcript per million mapped reads.

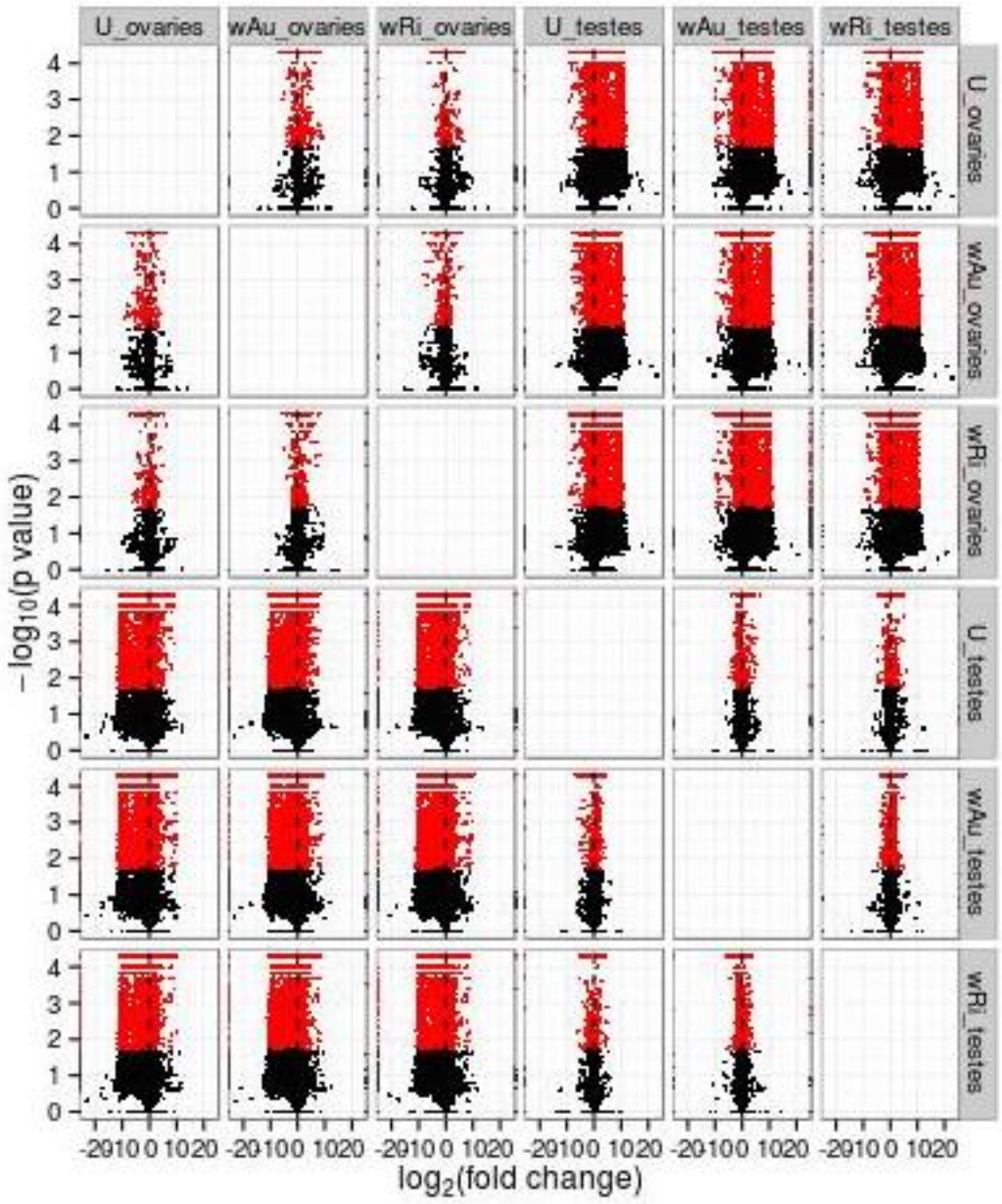

Figure 3.4. Volcano plots comparing host gene expression in ovaries and testes from uninfected, *w*Au-infected and *w*Ri-infected *D. simulans*. Red = significantly differentially expressed. Black = not significantly differentially expressed. U = uninfected.

3.3.1.3. Differential expression analysis

Differential expression analysis was performed using Cuffdiff2 (v2.2.1)[346]. In total 446 genes (of 13,711 expressed; 3.3 %) in ovaries and 1,320 (of 14,850 expressed; 8.9 %) in testes were identified as being significantly differentially expressed (FDR < 0.05) between at least two of the different *Wolbachia* infection statuses. The numbers of significantly upregulated and downregulated genes in each comparison are shown in Table 3.1 and

Table 3.2. Again it can be seen that there are a disproportionately large number of genes downregulated in *w*Au-infected versus *w*Ri-infected and uninfected testes. Additional comparisons were made between *w*Ri-infected versus uninfected and *w*Au-infected gonads combined, as these are the genes most likely to be involved in CI as opposed to some other *Wolbachia*-host interaction. Lists of these genes along with their functional annotations, if any, and expression information, are in Supplementary file 1, available in the Oxford University Research Archive (ORA) at http://ora.ox.ac.uk/objects/uuid:eb49f3cf-cfdf-40c9-a56c-f1a05985b4c0.

.

		Higher expression			
		Uninfected	*w*Au	Uninfected and *w*Au	*w*Ri
Lower expression	Uninfected	-	155	-	89
	*w*Au	106	-	-	91
	Uninfected and wAu	-	-	-	16
	*w*Ri	103	152	40	-

Table 3.1. Numbers of significantly differentially expressed host genes in ovaries from uninfected, *w*Au-infected and *w*Ri-infected *D. simulans.*

		Higher expression			
		Uninfected	*w*Au	Uninfected and *w*Au	*w*Ri
Lower expression	Uninfected	-	171	-	100
	*w*Au	634	-	-	732
	Uninfected and *w*Au	-	-	-	39
	*w*Ri	119	336	50	-

Table 3.2. Numbers of significantly differentially expressed host genes in testes from uninfected, *w*Au-infected and *w*Ri-infected *D. simulans.*

3.3.1.3.1. Genes differentially regulated by wRi in both ovaries and testes

Three genes were found to have significantly altered expression in *w*Ri-infected versus both *w*Au-infected and uninfected gonads for both ovaries and testes. These were *FBgn0190622*, *FBgn0194931*, and *FBgn0195967*. Their expression was lower in *w*Ri-infected versus both *w*Au-infected and uninfected gonads in all cases. *FBgn0195967* is annotated with cytochrome P450 domains. *FBgn0190622* and *FBgn0194931* are both annotated with a molecular function of fatty-acyl-CoA reductase (alcohol-forming) activity, and with a male sterility, NAD-binding domain. This represents the C-terminal NAD-binding region of the male sterility protein from *Arabidopsis* and *Drosophila*.

3.3.1.3.2. Genes exclusively expressed or unexpressed in wRi-infected gonads

Genes that are expressed exclusively in or have no expression exclusively in *w*Ri-infected gonads compared to both *w*Au-infected and uninfected gonads may be particularly good candidates for involvement in CI.

One gene, *FBgn0184169* was expressed in *w*Ri-infected but not in *w*Au-infected or uninfected ovaries. It was not annotated with any functional information. One gene, *FBgn0268675*, was expressed in *w*Ri-infected but not in *w*Au-infected or uninfected testes. Again it was not annotated with any functional information. Four genes were expressed in *w*Au-infected and uninfected but not *w*Ri-infected ovaries – *FBgn0188360*, *FBgn0182430*, *FBgn0183080*, and *FBgn0189900*. Three of these genes had annotated molecular functions - *FBgn0188360*, serine-type endopeptidase activity; *FBgn0182430*, phosphatase activity; *FBgn0189900*, transferase activity, transferring acyl groups other than amino-acyl groups. One gene, *FBgn0187164*, was expressed in *w*Au-infected and uninfected but not *w*Ri-infected testes. It was not annotated with any functional information. Even in the tissues in which they were expressed, all of these genes had relatively low (<10 FPKM) expression, so may not be that functionally important, or could even be false positives.

3.3.1.3.3. Genes identified as candidate CI genes in other studies

Other studies have identified candidate CI genes from comparative expression analyses[166,187,189-192]. The expression of these genes or their *D. simulans* orthologues was assessed to determine whether they were significantly differentially expressed among different *Wolbachia* statuses here (Table 3.3). One candidate gene, *Ance*, was significantly differentially expressed in wRi-infected versus both *w*Au-infected and uninfected ovaries. However, the direction of differential expression was the opposite to that previously observed in *D. melanogaster*, with lower expression in *w*Ri-infected than *w*Au-infected and uninfected ovaries. Two other candidate genes, *JhI-26* and *CG10433* were upregulated in *w*Ri-infected versus *w*Au-infected testes, but not versus uninfected testes. They also displayed higher expression in uninfected than *w*Au-infected testes, so it seems that the difference is likely due to a downregulation in the presence of *w*Au rather than an upregulation by *w*Ri.

3.3.1.4. Functional enrichment analysis

Functional enrichment analysis determines whether particular gene categories are over-represented among a set of genes relative to what would be expected by chance. Here, using GOSeq (version 1.20.0)[351], it was applied to two types of classification, GO terms and KEGG pathways, to determine whether any categories were over-represented among differentially expressed genes. All comparisons shown in Table 3.1 and Table 3.2 were tested.

3.3.1.4.1. GO enrichment analysis

GO terms are a set of descriptions of gene products developed by the Gene Ontology Consortium[298]. There are three branches, which comprise terms for describing gene products in terms of their associated biological process, molecular function, or cellular component, respectively. The numbers of significantly over-represented GO terms (FDR <

Gene	*D. simulans* gene ID	Reference	Differentially expressed between *Wolbachia* infection statuses?	Differentially expressed between testes and ovaries?
Ance	*FBgn0262093*	Xi *et al.* (2008)[187]	*w*Ri < uninfected ovaries *w*Ri < *w*Au ovaries	In all comparisons
zipper	*FBgn0196215*	Clark *et al.* (2006)[189]	No	In all comparisons
Hira	*FBgn0188437*	Zheng *et al.* (2011)[190]	No	In all comparisons
grau	*FBgn0196547*	Pinto *et al.* (2013)[166]	No	In all comparisons
JhI-26	*FBgn0182937*	Zheng *et al.* (2011)[191] Liu *et al.* (2014)[192]	*w*Ri > *w*Au testes uninfected > *w*Au testes	uninfected testes > ovaries wRi testes > ovaries
CG10433	*FBgn0196498*	Liu *et al.* (2014)[192]	*w*Ri > *w*Au testes uninfected > *w*Au testes	In all comparisons

Table 3.3. Differential expression status of candidate host CI genes identified by other studies, in ovaries and testes from uninfected, *w*Au-infected and *w*Ri-infected *D. simulans*.

0.05) in each comparison are shown in Table 3.4 and Table 3.5. Lists of the terms are in Supplementary file 2, available in the ORA at http://ora.ox.ac.uk/objects/uuid:eb49f3cf-cfdf-40c9-a56c-f1a05985b4c0.

		Higher expression			
		Uninfected	*w*Au	Uninfected and *w*Au	*w*Ri
Lower expression	Uninfected	-	16	-	6
	*w*Au	0	-	-	2
	Uninfected and *w*Au	-	-	-	0
	*w*Ri	11	9	0	-

Table 3.4. Numbers of significantly over-represented GO terms in host genes differentially expressed among ovaries from uninfected, *w*Au-infected and *w*Ri-infected *D. simulans.*

		Higher expression			
		Uninfected	*w*Au	Uninfected and *w*Au	*w*Ri
Lower expression	Uninfected	-	0	-	3
	*w*Au	68	-	-	79
	Uninfected and *w*Au	-	-	-	3
	*w*Ri	4	0	0	-

Table 3.5. Numbers of significantly over-represented GO terms in host genes differentially expressed among testes from uninfected, *w*Au-infected and *w*Ri-infected *D. simulans.*

The numbers of over-represented GO terms are largely consistent with the numbers of differentially expressed genes for each comparison, with the largest numbers of over-represented GO terms found among genes with significantly lower expression in *w*Au-infected than *w*Ri-infected or uninfected testes. Only three GO terms were found to be over-represented among genes differentially expressed between *w*Ri-infected and both *w*Au-infected and uninfected gonads. These were 'serine-type endopeptidase activity', 'serine-type peptidase activity', and 'serine hydrolase activity', and were over-represented among genes upregulated in *w*Ri-infected versus both *w*Au and uninfected testes. These terms are all related to each other, with 'serine-type endopeptidase activity' being a child

of 'serine-type peptidase activity', which is in turn a child of 'serine hydrolase activity', so it is probable that the latter two are over-represented due to their association with the first. These three terms (in the order stated above) were also the most significantly enriched terms among genes with higher expression in *w*Au-infected than uninfected ovaries, and genes with higher expression in *w*Au-infected than *w*Ri-infected ovaries.

3.3.1.4.2. KEGG pathway enrichment analysis

KEGG pathways are biological pathways defined by the Kyoto Encyclopaedia of Genes and Genomes, which maintains a database of the associations between genes and pathways[352,353]. No KEGG pathways were significantly enriched (FDR < 0.05) among genes differentially expressed between *w*Ri-infected and both *w*Au-infected and uninfected gonads. The only KEGG pathway that was significantly enriched in any comparison was 'arginine and proline metabolism', which was over-represented in genes downregulated in *w*Au-infected versus uninfected testes (FDR = 3.07×10^{-13}).

3.3.2. Identification and investigation of candidate *Wolbachia* CI genes

3.3.2.1. Sequencing and read alignment

RNA sequencing was performed on RNA extracted from separate ovaries and testes samples from *D. simulans* infected with *w*Ri or *w*Au. Three replicates were sequenced for each sample type. In total 353,615,930 paired end reads of 100 nt were generated, corresponding to 35.4 Gb of data. After deduplication there were 350,548,800 reads. Using Bowtie (v2.1.0) for genome alignment, 4.2 % of deduplicated *w*Ri-infected reads mapped to the *w*Ri genome, and 2.1 % of deduplicated *w*Au-infected reads mapped to the *w*Au genome.

3.3.2.2. Global expression analysis

Following alignment, quantification was performed using Cuffdiff2 (v2.2.1)[346]. Excluding tRNA and rRNA, 1,141 annotated *w*Ri genes were identified as being expressed in *w*Ri-infected gonads, 1,136 in ovaries and 1,135 in testes. Excluding tRNA and rRNA, 1,233 annotated *w*Au genes were identified as being expressed in *w*Au-infected gonads, 1,188 in ovaries and 1,221 in testes.

Global results were visualised using CummeRbund (v0.1.3)[347]. There are a number of issues with comparing gene expression across different species, discussed in more detail below, and it is not possible to perform such a comparison with Cuffdiff. Therefore, comparisons were made between testes and ovaries for each strain; analysis of differences between the results of these comparisons acts as a proxy for directly comparing gene expression between the two species. Heatmaps representing expression levels of all expressed genes in testes and ovaries for each strain are shown in Figure 3.5. These suggest that, unlike for the host genes, overall *Wolbachia* gene expression in ovaries and testes is quite similar, though one feature that particularly stands out is the upregulation of a small subset of genes in *w*Au-infected testes compared to *w*Au-infected ovaries.

Scatter plots show that in general expression is more dissimilar between ovaries and testes infected with *w*Au than those infected with *w*Ri, with the trend line for the *w*Au plot being further from the line of identity (Figure 3.6). The plots also illustrate that there are relatively large number of genes expressed in *w*Au-infected testes but not at all in *w*Au-infected ovaries; on the other hand, there are only a few genes with exclusive expression in *w*Ri-infected testes.

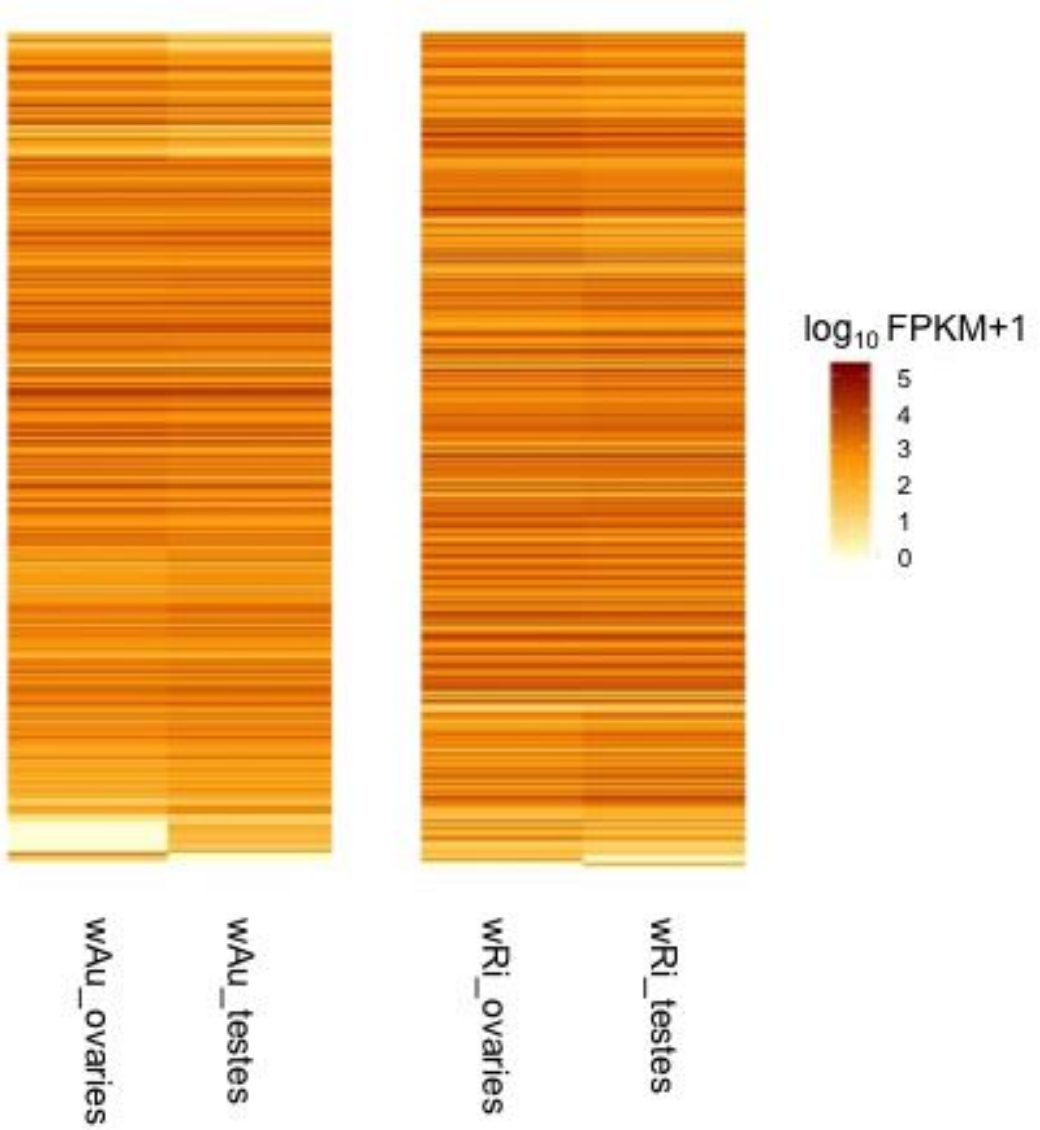

Figure 3.5. Heatmap illustrating expression levels for all expressed *Wolbachia* genes in ovaries and testes from *w*Au-infected and *w*Ri-infected *D. simulans.* FPKM, fragments per kilobase of transcript per million mapped reads.

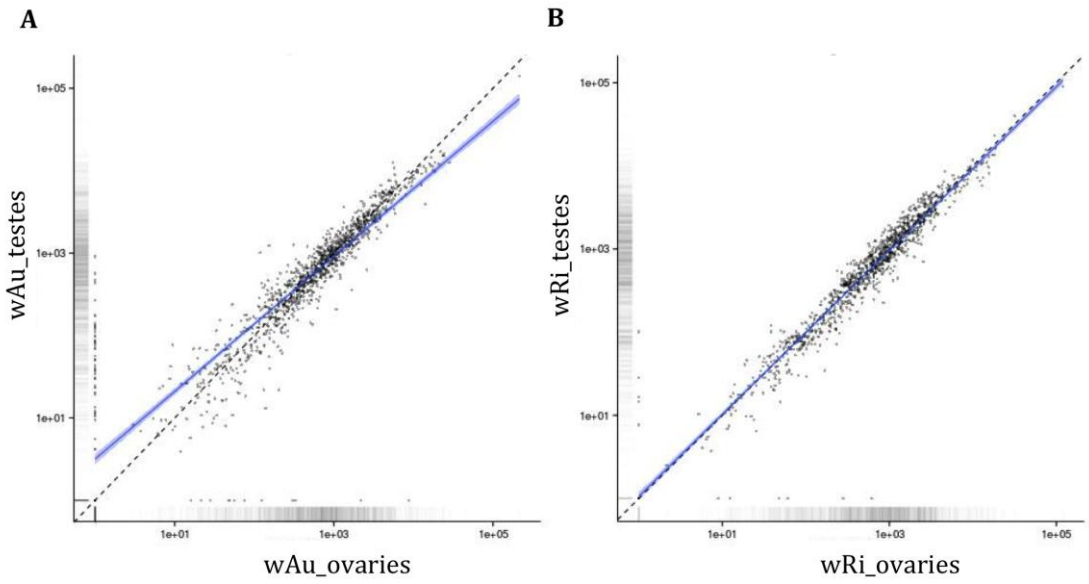

Figure 3.6. Scatter plots comparing *Wolbachia* gene expression in ovaries and testes from *w*Au-infected (A) and *w*Ri-infected (B) *D. simulans*. Bars adjacent to axes show the distribution of FPKM values for that sample. Dotted lines are identity lines. Blue lines are trend lines.

3.3.2.3. Differential expression analysis

In theory, differential expression analysis between species could be performed by directly comparing expression values of orthologues, accounting for interspecies differences such as gene length and GC content. However, as stated above, there are issues with this approach, and it seems it is rarely used in practice. The main problem is that differences in mappability between species, for example due to the presence of paralogues that cause reads to multiply map in one species but not the other, can make genes appear to be falsely differentially expressed between species[354,355]. Therefore differential expression analysis was performed between testes and ovaries in each strain separately. Orthologues were identified by reciprocal best hit BLAST analysis, and the lists of differentially expressed genes were compared to identify genes that were differentially expressed between ovaries and testes in one strain but whose orthologues were not differentially expressed in the other strain. Genes involved in CI may not necessarily be differentially expressed between testes and ovaries and so this approach may lead to false negatives,

but this was deemed to be preferable to the false positives that would likely arise from direct interspecies comparison.

Differential expression analysis was performed using Cuffdiff2 (v2.2.1)[346]. In total, 289 genes (of 1,141 expressed; 25 %) were identified as being significantly differentially expressed between *w*Ri-infected testes and ovaries, and 132 genes (of 1,233 expressed; 11 %) were identified as being significantly differentially expressed (FDR < 0.05) between *w*Au-infected testes and ovaries. The differential expression did not seem to be biased in a particular direction for either strain; in *w*Ri 134 genes showed higher expression in ovaries and 155 showed higher expression in testes, while in *w*Au 65 genes showed higher expression in ovaries and 67 showed higher expression in testes. Lists of these genes along with their functional annotations, if any, and expression information, are in Supplementary file 3, available in the ORA at http://ora.ox.ac.uk/objects/uuid:eb49f3cf-cfdf-40c9-a56c-f1a05985b4c0.

Reciprocal best hit BLAST analysis identified 924 pairs of putative orthologues between *w*Au and *w*Ri (excluding rRNA and tRNA), 921 of which were expressed in at least one of the strains. 61 (46 %) of the 132 *w*Au genes differentially expressed between ovaries and testes had putative *w*Ri orthologues that were not differentially expressed in the same manner between ovaries and testes (Figure 3.7). Of these 61 genes, 33 had higher expression in *w*Au-infected ovaries and 28 had higher expression in *w*Au-infected testes. 196 (68 %) of the 289 *w*Ri genes differentially expressed between ovaries and testes had putative *w*Au orthologues that were not differentially expressed in the same manner between ovaries and testes (Figure 3.7). Of these 196 genes, 90 had higher expression in *w*Ri-infected ovaries and 106 with higher expression in *w*Ri-infected testes. Differentially expressed genes with orthologues that were not differentially expressed in the other *Wolbachia* strain are highlighted in red in the lists of genes in Supplementary file 3.

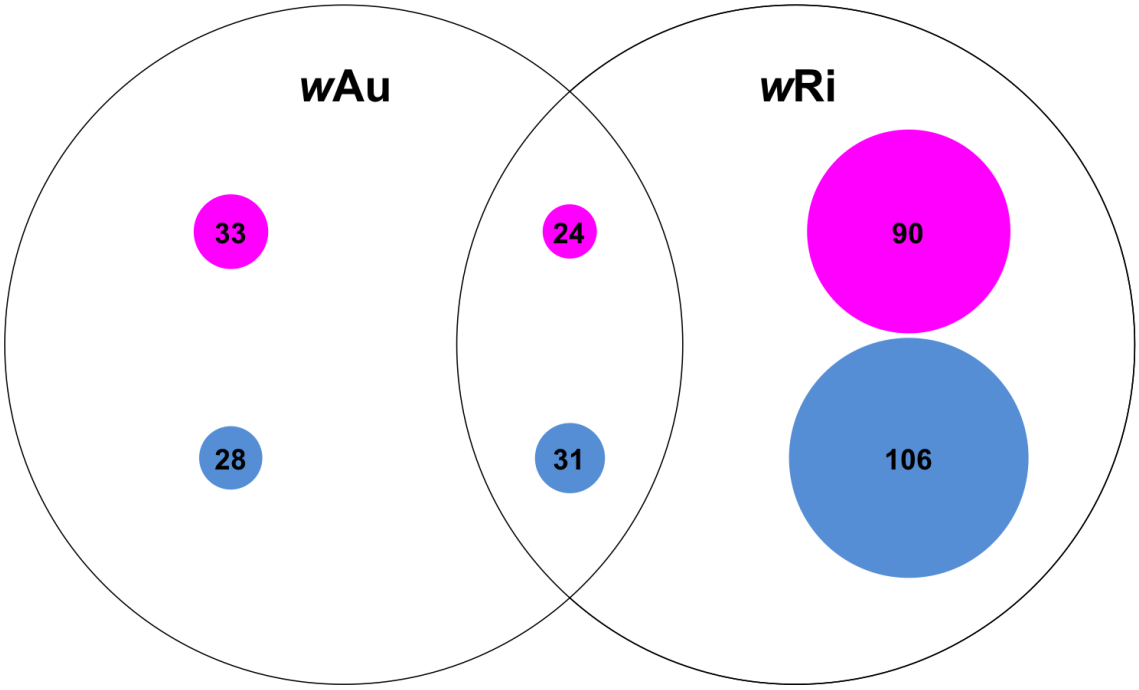

Figure 3.7. Orthologues differentially expressed between testes and ovaries in wAu and wRi. Pink indicates genes upregulated in ovaries versus testes. Blue indicates genes upregulated in testes versus ovaries. The size of the coloured circles is proportional to the number of genes they represent.

3.3.2.4. Functional enrichment analysis

Functional enrichment analysis was performed using GOSeq (version 1.20.0)[351] as described above for host genes, except unlike for *D. simulans*, the *Wolbachia* genome annotations do not contain GO information, so this was obtained based on BLAST hits using BLAST2GO (version 3.1.3)[348]. Genes differentially expressed between ovaries and testes in each *Wolbachia* strain were tested (upregulated and downregulated separately), as well as genes from these lists with orthologues that were not differentially expressed in the same way in the other *Wolbachia* strain.

3.3.2.4.1. GO enrichment analysis

No GO terms were found to be significantly over-represented (FDR < 0.05) in any set of genes tested.

3.3.2.4.2. KEGG pathway enrichment analysis

No KEGG pathways were found to be significantly over-represented (FDR < 0.05) in any set of genes tested.

3.3.2.4.3. Other enrichment analysis

Further analysis was also performed to determine whether gene categories with previous evidence for involvement in CI – transcriptional regulators, phage genes, and ankyrin repeat genes – were enriched in any of the gene sets described above. None of the gene categories were found to be significantly over-represented (FDR < 0.05) in any set of genes tested.

3.3.2.5. Analysis of transcriptional regulator gene expression

As discussed in Chapter 2, the family of transcriptional regulator genes in *Wolbachia* are promising candidates for CI. One of these genes is present in the CI-inducing strain *w*Mel but not in *w*Au (a non-CI strain)[177,272], while another varies in presence between incompatible strains of *Cx. pipiens*[166], and others, as demonstrated in Chapter 2, show large differences in sequence and organisation between *w*Mel and *w*Au[272]. Here, the expression levels of these transcriptional regulator genes in *w*Au and *w*Ri were analysed; firstly to confirm whether or not they are being expressed, and how much, and secondly to perform a crude comparison of expression levels between these two strains.

The results show varying levels of expression among the transcriptional regulator genes, with some very highly expressed (Table 3.6 and Table 3.7). Three *w*Ri transcriptional regulator genes – *wRi005890*, *wRi015000* and *wRi010550* – are significantly differentially expressed between ovaries and testes, with the paralogous genes *wRi005890* and *wRi010550* displaying higher expression in testes, and *wRi015000* displaying higher

expression in ovaries. No *w*Au transcriptional regulator genes are significantly differentially expressed between testes and ovaries.

Gene	Expression in ovaries (percentile)	Expression in testes (percentile)	Log2(fold change)	FDR
WPWAU_0251	86	90	0.26	0.26
WPWAU_0252	6	8	-0.03	1.00
WPWAU_0255	10	10	-0.32	0.69
WPWAU_0256	32	19	-1.19	0.34
WPWAU_0257	8	0	-inf	1.00
WPWAU_0324	91	91	0.02	0.95
WPWAU_0686	7	0	-inf	1.00
WPWAU_0687	37	23	-0.94	0.39
WPWAU_0688	10	11	-0.16	0.83
WPWAU_0691	3	5	0.23	1.00

Table 3.6. Expression levels of *w*Au transcriptional regulator genes. Expression levels are given as percentile of overall expression of all genes in that tissue. Log2(fold change) and FDR values are for FPKM values in testes relative to those in ovaries. Genes are colour-coded as in Figure 2.8, with genes thought to be paralogous within strains and paralogous or orthologous between strains depicted in the same colour.

Gene	Expression in ovaries (percentile)	Expression in testes (percentile)	Log2(fold change)	FDR
wRi005840	19	16	-0.46	0.14
wRi005850	11	8	-0.45	0.28
wRi005880	14	16	0.36	0.34
wRi005890	79	89	0.60	**0.00**
wRi010500	20	15	-0.70	**0.02**
wRi010510	14	9	-0.69	0.07
wRi010540	13	16	0.42	0.24
wRi010550	77	88	0.69	**0.00**

Table 3.7. Expression levels of *w*Ri transcriptional regulator genes. Expression levels are given as percentile of overall expression of all genes in that tissue. Log2(fold change) and FDR values are for FPKM values in testes relative to those in ovaries. Significant FDR values are highlighted in bold. Genes are colour-coded as in Figure 2.8, with genes thought to be paralogous within strains and paralogous or orthologous between strains depicted in the same colour.

The genes colour-coded in red in Table 3.6 and Table 3.7 were identified in Chapter 2 as particularly promising candidates for CI. In *w*Au, these genes are predicted as coding

sequences (CDSs) even though there has been a frameshift relative to the orthologous genes in *w*Ri that would cause premature termination. The *w*Au paralogues *WPWAU_0257* and *WPWAU_0686* are predicted CDSs downstream of the frameshift. These genes had relatively low expression (8th and 7th percentile) in ovaries and no expression in testes. The *w*Au paralogues *WPWAU_0256* and *WPWAU_0687* are predicted CDSs upstream of the frameshift, and appeared to have a reasonable level of expression (32nd-38th percentile). These results suggest that these genes may still be active, but they likely have different functions to their counterparts in *w*Ri due to their truncation. The *w*Ri counterparts of this gene, *wRi005840* and *wRi01550* are also expressed (16th-20th percentile), and interestingly *wRi015500* has significantly higher expression in ovaries than testes.

There also appear to be differences between strains in the most highly expressed transcriptional regulator genes. The most highly expressed of the transcriptional regulator genes in *w*Au, *WPWAU_0324* (91st percentile in both testes and ovaries) lacks an orthologue in *w*Ri. *WPWAU_0251* is also highly expressed in *w*Au; its two orthologues *wRi005890* and *wRi010550*, are also highly expressed in *w*Ri, but unlike in *w*Au, in *w*Ri the expression is significantly higher in testes for both orthologues.

3.3.3. Identification and investigation of candidate host viral inhibition genes

3.3.3.1. Sequencing and read alignment

RNA sequencing was performed on RNA extracted from whole adult *Ae. albopictus* mosquitoes. Three replicates were sequenced for each sample type. In total 904,411,738 reads of 101 nt were generated, corresponding to 91.3 Gb of data. After trimming and quality filtering a total of 896,792,764 paired end reads were used for genome alignment with TopHat2 (v2.0.11). 67.9 % of these mapped to the *Ae. albopictus* genome.

3.3.3.2. Transcript assembly

Unlike the completed *D. simulans* and *Wolbachia* genome assemblies used above, the genome of *Ae. albopictus* is relatively poorly assembled, having only been recently sequenced, and comprises 401,027 scaffolds[356]. The genome annotation was completed at the same time, and is also likely to be of lower quality than those of organisms that have better genome assemblies and/or are more well studied. For these reasons, it was decided to generate a transcript assembly for expression quantification, rather than simply quantifying expression of pre-annotated transcripts as for *D. simulans* and *Wolbachia* above. Due to the very fragmented genome assembly, it was decided to generate a *de novo* assembly with Trinity, in addition to an assembly based on alignment to the genome, and compare the assemblies to decide which to take forward. The *de novo* assembly was generated using Trinity (version 2014/07/17)[343]. The alignment-based assembly was generated using Cufflinks (v2.2.1)[341] with RABT methodology[341], followed by merging of assemblies from different samples with Cuffmerge (v1.0.0)[342]. Basic assembly statistics were evaluated, in addition to an assembly score using RSEM-EVAL[345]. The results are shown in Table 3.8, along with assembly statistics for the *Ae. aegypti* AaegL3.3 transcriptome assembly from VectorBase; as the two species are closely related, *Ae. albopictus* transcriptome statistics should be similar to those of *Ae. aegypti*.

Assembly	**Assembly size (nt)**	**Number of genes**	**Number of transcripts**	**Mean transcript length (nt)**	**N50 transcript length (nt)**	**RSEM-EVAL score (x10^9)**
Alignment-based assembly	129,369,955	40,568	67,586	1,914	3,067	-48.4
De novo assembly	46,521,043	51,757	74,934	1,090	2,071	-35.1
Ae. aegypti transcriptome assembly AaegL3.3	32,498,104	17,478	18,840	1,725	2,392	-

Table 3.8. Assembly statistics for *Ae. albopictus de novo* and alignment-based transcriptomes and *Ae. aegypti* transcriptome. nt, nucleotide.

The alignment-based assembly was taken forward for further analysis, based on these results and for ease of downstream analysis – the *de novo* assembly contains only newly predicted transcripts and genes with no associated information, whereas the alignment-based assembly contains many transcripts and genes derived from the pre-existing annotation, which is useful for determining information about the transcripts, and reciprocally for adding information to the geneset that will be used as a reference by others in future.

3.3.3.3. Global expression analysis

Following alignment, quantification was performed using Cuffdiff2 (v2.2.1)[346]. Global results were visualized using CummeRbund (v0.1.3)[347]. A heatmap representing expression levels of all expressed genes in each of the three sample types is shown in Figure 3.7. At first glance this suggests that overall gene expression among the different *Wolbachia* infection statuses is fairly similar, particularly between *w*Mel-infected and uninfected mosquitoes. The dendogram in Figure 3.8 shows more clearly that these are closer to each other than to *w*Alb-infected mosquitoes in terms of global gene expression.

Scatter plots and volcano plots between the samples do not reveal a pronounced direction for the differences (Figure 3.9 and Figure 3.10), indicating both significant upregulation and downregulation in all comparisons.

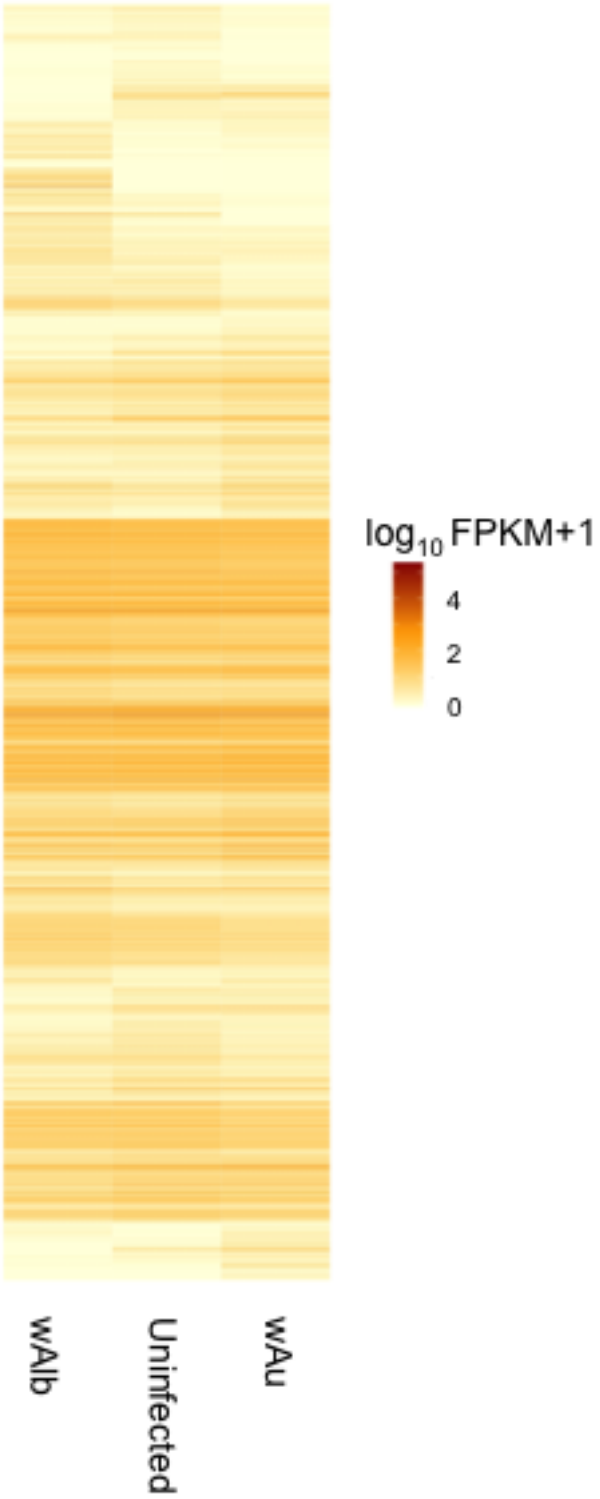

Figure 3.7. Heatmap illustrating expression levels for all expressed host genes in uninfected, *w*Alb-infected and *w*Mel-infected *Ae. albopictus*. FPKM, fragments per kilobase of transcript per million mapped reads.

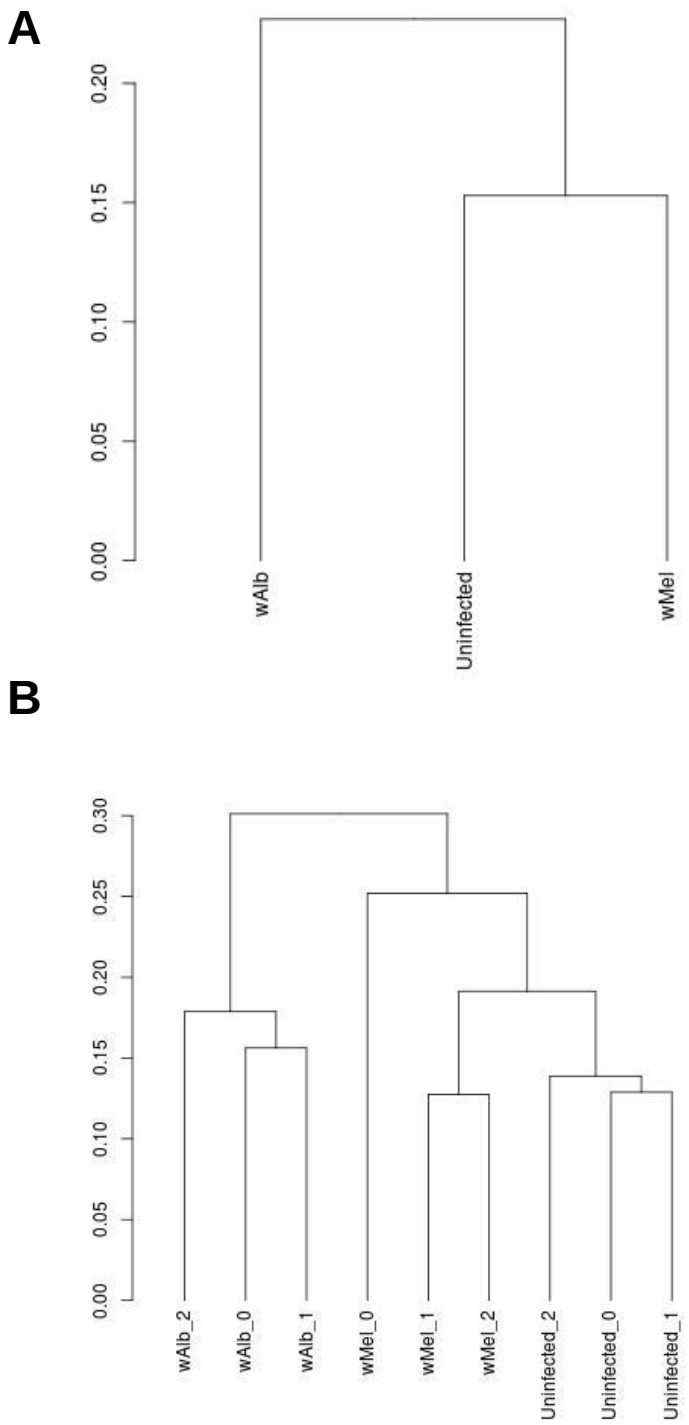

Figure 3.8. Dendrogram showing similarity in terms of host gene expression among uninfected, *w*Alb-infected, and *w*Mel-infected *Ae. albopictus,* based on Jensen-Shannon distances, with biological replicates combined (A) or separate (B).

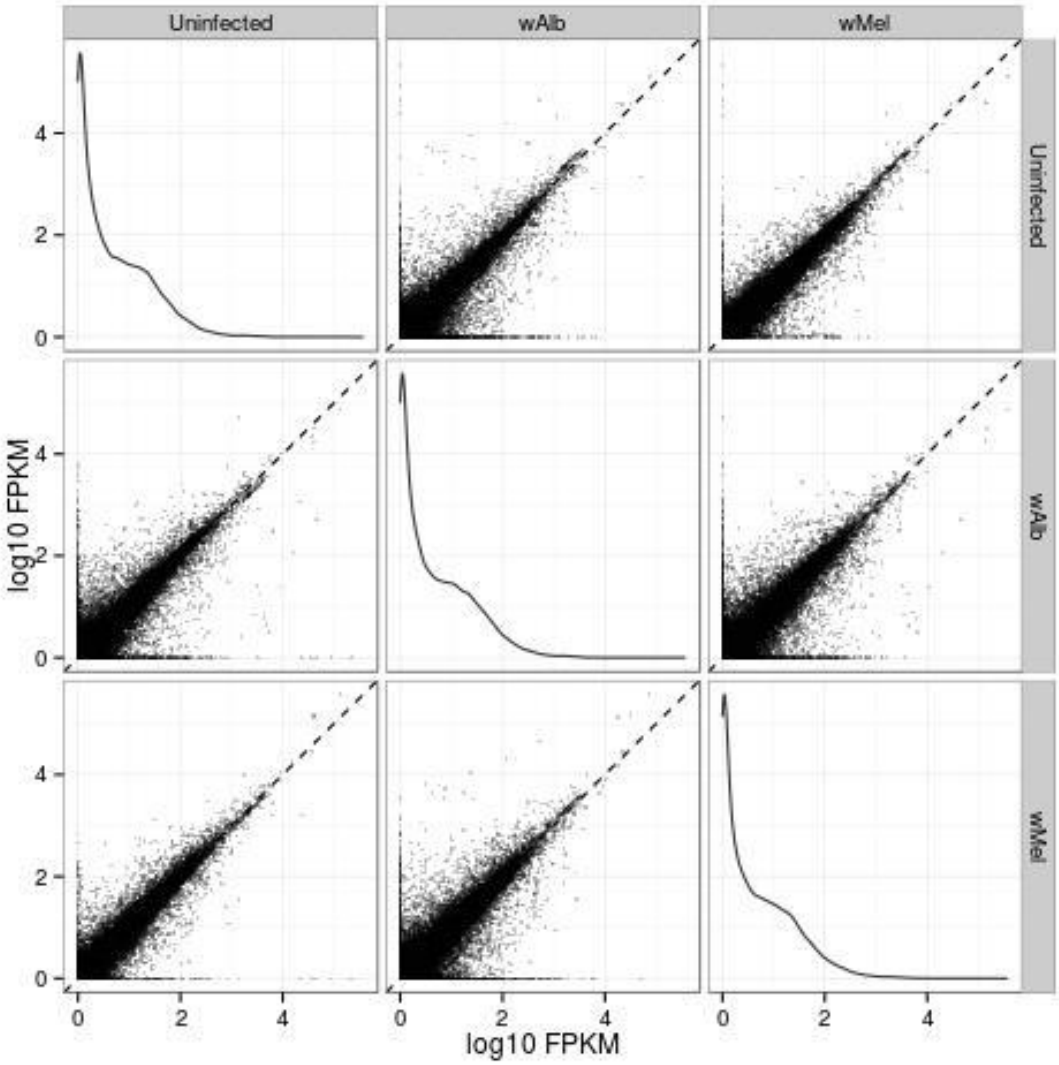

Figure 3.9. Scatter plots comparing host gene expression in uninfected, *w*Alb-infected and *w*Mel-infected *Ae. albopictus*. Plots between identical samples show the distribution of FPKM values for that sample. Dotted lines are identity lines. FPKM, fragments per kilobase of transcript per million mapped reads.

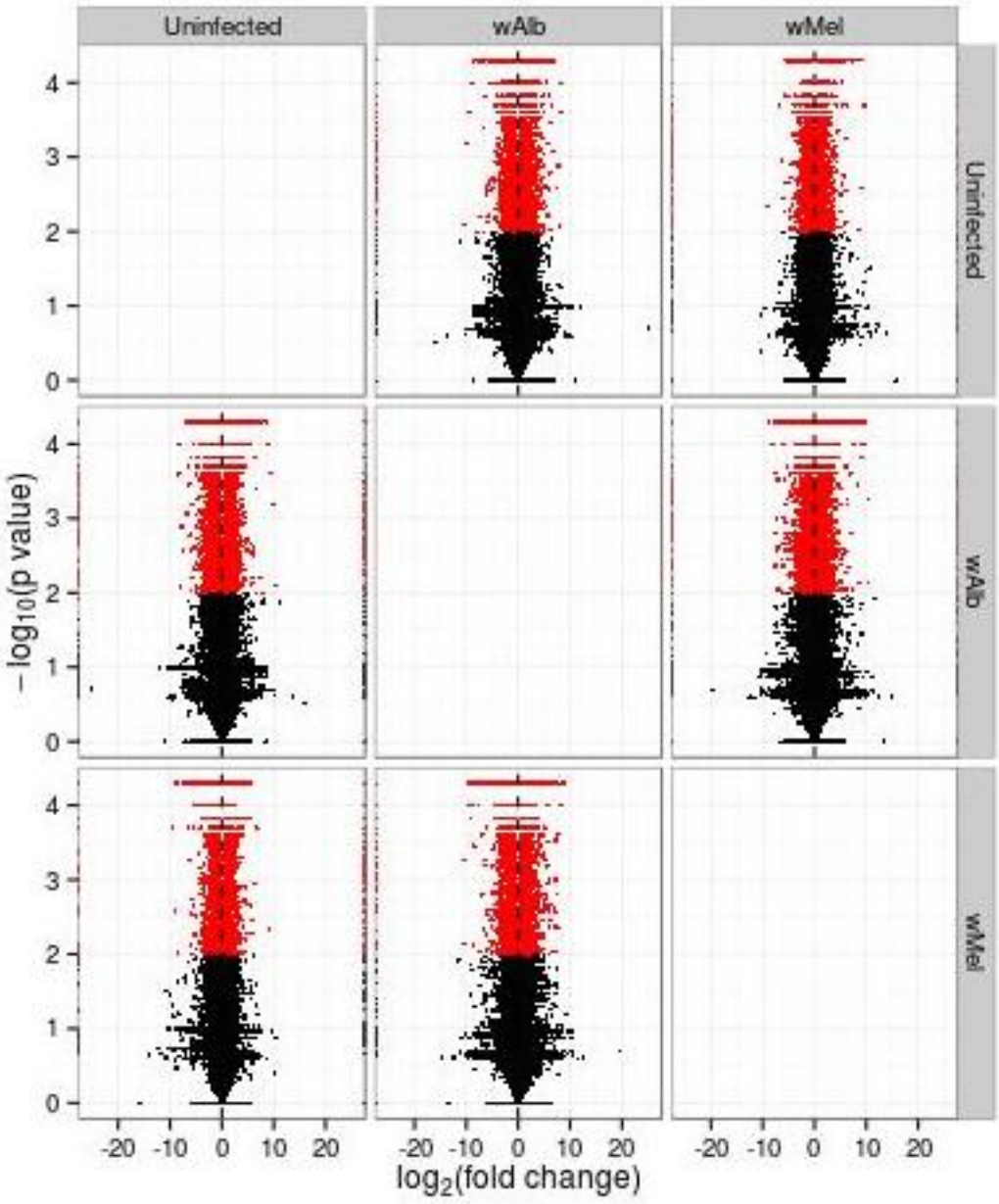

Figure 3.10. Volcano plots comparing host gene expression in ovaries and testes from uninfected, *w*Alb-infected and *w*Mel-infected *Ae. albopictus*. Red = significantly differentially expressed. Black = not significantly differentially expressed.

3.3.3.4. Differential expression analysis

Differential expression analysis was performed using Cuffdiff2 (v2.2.1)[346]. In total, 11,901 genes (of 40,568 expressed; 29 %) were identified as being significantly differentially expressed between at least two of the different *Wolbachia* infection statuses. The numbers of significantly upregulated and downregulated genes in each comparison are shown in Table 3.9. Additional comparisons were made between *w*Mel-infected versus uninfected

and *w*Alb-infected mosquitoes combined, as these are the genes most likely to be involved in viral inhibition as opposed to some other *Wolbachia*-host interaction. Lists of these genes along with their functional annotations, if any, and expression information, is available at https://www.dropbox.com/sh/3rpx4dsss6xnh46/AACHJYAWS1Be-ynz-X37NG8ua?dl=0.

		Higher expression			
		Uninfected	wAlb	Uninfected and wAlb	wMel
Lower expression	Uninfected	-	3,388	-	2,539
	wAlb	2,346	-	-	3,858
	Uninfected and wAlb	-	-	-	1,583
	wMel	2,450	4,912	1,381	-

Table 3.9. Numbers of significantly differentially expressed host genes in uninfected, *w*Alb-infected and *w*Mel-infected *Ae. albopictus.*

3.3.3.4.1. Genes exclusively expressed or unexpressed in w*Mel-infected mosquitoes*

Genes that are expressed exclusively or have no expression exclusively in *w*Mel-infected mosquitoes compared to both *w*Alb-infected and uninfected gonads may be particularly good candidates for involvement in viral inhibition. The numbers of exclusively expressed or unexpressed genes in each comparison are shown in Table 3.10. List of these genes along with their functional annotations, if any, and expression information, are in Supplementary file 4, available in the ORA at http://ora.ox.ac.uk/objects/uuid:eb49f3cf-cfdf-40c9-a56c-f1a05985b4c0.

Many of these genes had low expression levels even in the samples in which they were expressed, and so could represent false positives. Therefore the number of genes with a mean expression level above 10 FPKM in the sample types in which they were expressed

		Expression			
		Uninfected	*w*Alb	Uninfected and *w*Alb	*w*Mel
No expression	Uninfected	-	601 (171)	-	138 (34)
	*w*Alb	440 (128)	-	-	501 (157)
	Uninfected and *w*Alb	-	-	-	64 (19)
	*w*Mel	299 (68)	879 (262)	164 (35)	-

Table 3.10. Numbers of exclusively expressed or unexpressed host genes in uninfected, *w*Alb-infected and *w*Mel-infected *Ae. albopictus.* Numbers in brackets indicate the number of genes with a mean expression level above 10 FPKM in the sample types in which they are expressed.

was also determined. Most genes exclusively expressed in *w*Mel-infected mosquitoes were not annotated and could not be assigned anyfunctional information. However, seven of the nine genes that were exclusively expressed in *w*Mel-infected mosquitoes and were annotated with a molecular function were annotated with the function of odorant binding (a significantly over-represented category in genes upregulated in *w*Mel-infected versus both *w*Alb-infected and uninfected mosquitoes; see below).

3.3.3.5. Functional enrichment analysis

Functional enrichment analysis was performed using GOSeq (version 1.20.0)[351] as described above for investigation of candidate CI genes. The *Ae. albopictus* genome annotation does not contain GO information, so again this was obtained based on BLAST hits using BLAST2GO (version 3.1.3)[348].

3.3.3.5.1. GO enrichment analysis

For genes upregulated in *w*Mel-infected versus both *w*Alb-infected and uninfected mosquitoes, 43 GO terms were found to be significantly over-represented (FDR < 0.05) (Table 3.11). For genes downregulated in *w*Mel-infected versus both *w*Alb-infected and

uninfected mosquitoes, 33 GO terms were found to be significantly over-represented (FDR < 0.05) (Table 3.12).

GO ID	Term	Category	FDR
GO:0004252	serine-type endopeptidase activity	MF	6.57E-13
GO:0008236	serine-type peptidase activity	MF	1.41E-12
GO:0017171	serine hydrolase activity	MF	1.41E-12
GO:0016705	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen	MF	1.32E-10
GO:0020037	heme binding	MF	1.87E-10
GO:0046906	tetrapyrrole binding	MF	1.91E-10
GO:0004497	monooxygenase activity	MF	7.27E-10
GO:0005506	iron ion binding	MF	9.34E-09
GO:0005576	extracellular region	CC	9.34E-09
GO:0006508	proteolysis	BP	2.73E-08
GO:0003824	catalytic activity	MF	3.74E-08
GO:0055114	oxidation-reduction process	BP	3.91E-08
GO:0016491	oxidoreductase activity	MF	9.68E-08
GO:0005549	odorant binding	MF	2.29E-07
GO:0004175	endopeptidase activity	MF	1.07E-06
GO:0016787	hydrolase activity	MF	1.31E-06
GO:0008233	peptidase activity	MF	2.20E-06
GO:0070011	peptidase activity, acting on L-amino acid peptides	MF	7.07E-06
GO:0006040	amino sugar metabolic process	BP	1.83E-05
GO:0008061	chitin binding	MF	4.75E-05
GO:0006030	chitin metabolic process	BP	4.75E-05
GO:1901071	glucosamine-containing compound metabolic process	BP	6.52E-05
GO:0006022	aminoglycan metabolic process	BP	1.11E-04
GO:0006760	folic acid-containing compound metabolic process	BP	1.62E-04
GO:0042558	pteridine-containing compound metabolic process	BP	4.60E-04
GO:0042302	structural constituent of cuticle	MF	7.24E-04
GO:0005215	transporter activity	MF	7.95E-04
GO:0016798	hydrolase activity, acting on glycosyl bonds	MF	9.26E-04
GO:0009396	folic acid-containing compound biosynthetic process	BP	1.35E-03
GO:1901607	alpha-amino acid biosynthetic process	BP	1.88E-03
GO:0005615	extracellular space	CC	1.88E-03
GO:0004553	hydrolase activity, hydrolyzing O-glycosyl compounds	MF	2.15E-03
GO:0008237	metallopeptidase activity	MF	2.40E-03
GO:0042559	pteridine-containing compound biosynthetic process	BP	3.19E-03
GO:0016053	organic acid biosynthetic process	BP	3.19E-03
GO:0046394	carboxylic acid biosynthetic process	BP	3.19E-03
GO:0008652	cellular amino acid biosynthetic process	BP	8.49E-03
GO:0044421	extracellular region part	CC	1.12E-02
GO:0016646	oxidoreductase activity, acting on the CH-NH group of donors, NAD or NADP as acceptor	MF	1.44E-02
GO:0042398	cellular modified amino acid biosynthetic process	BP	1.91E-02
GO:0016701	oxidoreductase activity, acting on single donors with incorporation of molecular oxygen	MF	2.42E-02
GO:0044283	small molecule biosynthetic process	BP	2.57E-02
GO:0016998	cell wall macromolecule catabolic process	BP	3.18E-02

Table 3.11. GO terms significantly over-represented among genes upregulated in *w*Mel-infected versus both uninfected and *w*Alb-infected *Ae. albopictus*. BP = biological process. CC = cellular component. MF = molecular function

GO ID	Term	Category	FDR
GO:0005634	nucleus	CC	1.64E-07
GO:0016568	chromatin modification	BP	4.21E-04
GO:0003677	DNA binding	MF	5.50E-04
GO:0051276	chromosome organization	BP	6.12E-04
GO:0006325	chromatin organization	BP	6.12E-04
GO:0016569	covalent chromatin modification	BP	8.40E-04
GO:0016570	histone modification	BP	8.40E-04
GO:0005488	binding	MF	1.55E-03
GO:0000123	histone acetyltransferase complex	CC	1.55E-03
GO:0031248	protein acetyltransferase complex	CC	1.55E-03
GO:1902493	acetyltransferase complex	CC	1.55E-03
GO:0018205	peptidyl-lysine modification	BP	1.55E-03
GO:0043231	intracellular membrane-bounded organelle	CC	2.06E-03
GO:0043227	membrane-bounded organelle	CC	2.06E-03
GO:0004402	histone acetyltransferase activity	MF	7.79E-03
GO:0090595	acetyl-CoA:L-lysine N6-acetyltransferase	MF	7.79E-03
GO:0043189	H4/H2A histone acetyltransferase complex	CC	1.03E-02
GO:1902562	H4 histone acetyltransferase complex	CC	1.03E-02
GO:0016573	histone acetylation	BP	1.16E-02
GO:0003676	nucleic acid binding	MF	1.17E-02
GO:0006473	protein acetylation	BP	1.40E-02
GO:0006475	internal protein amino acid acetylation	BP	1.40E-02
GO:0018393	internal peptidyl-lysine acetylation	BP	1.40E-02
GO:0018394	peptidyl-lysine acetylation	BP	1.40E-02
GO:0005694	chromosome	CC	1.79E-02
GO:0006338	chromatin remodeling	BP	1.79E-02
GO:0043543	protein acylation	BP	1.84E-02
GO:0044428	nuclear part	CC	2.03E-02
GO:0005654	nucleoplasm	CC	2.09E-02
GO:0005657	replication fork	CC	2.14E-02
GO:0006260	DNA replication	BP	2.66E-02
GO:0046872	metal ion binding	MF	3.82E-02
GO:0044427	chromosomal part	CC	3.95E-02

Table 3.12. GO terms significantly over-represented among genes downregulated in *w*Mel-infected versus both uninfected and *w*Alb-infected *Ae. albopictus*. BP = biological process. CC = cellular component. MF = molecular function.

3.3.3.5.2. KEGG pathway enrichment analysis

No KEGG pathways were significantly enriched (FDR < 0.05) among genes differentially expressed between *w*Mel-infected and both *w*Alb-infected and uninfected mosquitoes. The 'oxidative phosphorylation' and 'taurine and hypotaurine metabolism' pathways were significantly over-represented in genes upregulated in *w*Mel-infected versus *w*Alb-infected mosquitoes (FDR = 2.02×10^{-5} and 2.62×10^{-2} respectively). The 'oxidative phosphorylation' pathway was also significantly over-represented in genes upregulated in uninfected versus *w*Alb-infected mosquitoes (FDR = 1.34×10^{-6}).

3.3.3.5.3. Dengue virus host factor enrichment analysis

Genome-wide screening has previously identified putative dengue virus host factors (DVHFs) implicated in dengue virus propagation in *D. melanogaster* cells[350]. To determine whether such DVHFs might be involved in dengue inhibition by *w*Mel, *Ae. albopictus* orthologues were determined, and enrichment analysis performed to establish whether these genes are over-represented among genes differentially expressed between uninfected, *w*Alb-infected and *w*Mel-infected *Ae. albopictus*.

With multiple testing correction, putative *Ae. albopictus* DVHFs are not significantly over-represented (FDR < 0.05) in any set of differentially expressed genes, though without multiple testing correction, DVHFs are significantly over-represented in genes upregulated in *w*Mel-infected versus uninfected *Ae. albopictus*.

3.4. Discussion

Here comparative transcriptomics has been used to identify and investigate host and *Wolbachia* genes potentially involved in CI, as well as candidate host genes potentially involved in viral inhibition by *Wolbachia*. This high-throughput, genome-wide approach represents the first known time that RNA-seq has been used for these purposes. Furthermore, while RNA-seq has been performed to identify genes involved in *Wolbachia*-host symbiosis in filarial nematodes[321,326,357] and *D. melanogaster*[322], this represents the first known transcriptomic analysis of any kind aiming to identify *Wolbachia* genes involved in viral inhibition and CI.

3.4.1. Identification and investigation of candidate host CI genes

Candidate host CI genes were identified and investigated by comparative transcriptomic analysis in *D. simulans* ovaries and testes infected with *w*Au or *w*Ri, or uninfected. Global expression analysis revealed that for neither testes nor ovaries was *w*Ri-infected tissue

more different to *w*Au-infected than to uninfected tissue, suggesting that most genes whose expression is affected by *Wolbachia* are not involved in CI, but another *Wolbachia*-host interaction. *w*Ri-infected testes were actually more similar to uninfected than to *w*Au-infected testes, with a large number of genes downregulated in *w*Au-infected testes. There were a wide variety of GO terms over-represented among these downregulated genes, several associated with metabolism but also some associated with mating and reproduction. At first glance this seems a surprising result; it might be expected that *w*Ri-infected testes would be the most different, as *w*Ri induces the additional phenotype of CI. Different densities of *w*Au and *w*Ri in the testes may play a role in this observation. *w*Au has been found to be at higher titres in various tissues, including testes (no information for ovaries), than a number of other *Wolbachia* strains, including *w*Ri[235]. It might be expected that higher densities of *Wolbachia* would have a greater effect on global gene expression. Interestingly, *w*Au has been found to confer greater protection against viruses than does *w*Ri[218]. The degree of protection was found to correlate with *w*Au density[235], which could be consistent with the idea of greater *Wolbachia* densities having a larger effect on host gene expression. It is also possible that there are other factors involved in the viral inhibition that may help to explain the more pronounced effect on host gene expression by *w*Au than *w*Ri.

Another factor that could affect regulation of host gene expression by *Wolbachia* is the age of the association between the *Wolbachia* and its host. It is difficult to comment further on this, however, as it is not clear which direction, if any, the effect would take, or which association is older, with some evidence suggesting *w*Au is a relatively recent acquisition by *D. simulans*[358], and other evidence suggesting it is ancient[272].

KEGG pathway enrichment analysis did not reveal any significantly over-represented pathways among host genes differentially expressed between *w*Ri-infected and both *w*Au-

infected and uninfected gonads. A low number of genes associated with a pathway was likely a factor in this. However, GO enrichment analysis revealed that serine-type endopeptidase activity (along with two of its ancestral terms) was enriched in genes upregulated in *w*Ri-infected versus both *w*Au-infected and uninfected testes, and was also the most significantly enriched term among genes upregulated in *w*Au-infected versus *w*Ri-infected or uninfected ovaries. One of the four genes with expression in *w*Au-infected and uninfected but not *w*Ri-infected ovaries was also annotated with this function.

Serine-type endopeptidases (also known as serine proteases) are enzymes that catalyse the cleavage of peptide bonds in proteins, using the amino acid serine at the active site[359]. They can be categorised as either trypsin-like or subtilisin-like[360]. InterPro annotations suggest that the products of all seven serine protease genes upregulated in *w*Ri-infected versus uninfected and *w*Au-infected testes are trypsin-like. Members of this family perform a wide range of functions[361-363], with evidence suggesting roles in immune response, development, and other processes in *Drosophila* (summarised by Shah *et al*.[364]). Consistent with their identification here as candidate CI genes, serine proteases are also heavily involved in reproduction in *Drosophila*. A high proportion of proteins in *Drosophila* seminal fluid are serine proteases or serine protease inhibitors of unknown function[365-367], like the predicted products of the genes found here to be upregulated in *w*Ri-infected versus uninfected and *w*Au-infected testes. Proteases expressed in *Drosophila* females are also thought to play a role in mating, by cleaving proteins transferred by the male[368-373]. Reproduction in other insects[374,375], as well as mammals[376], also involves proteases. There is insufficient evidence to conclude that the serine endopeptidase genes identified here have a role in CI, and enrichment of serine-type endopeptidase activity among genes upregulated in ovaries by *w*Au but not *w*Ri fails to support a role for these proteins in CI in this tissue. However, enrichment of this category in *w*Ri-infected testes combined with its

known involvement in reproduction make them promising candidates for further investigation.

Other promising host candidates include *FBgn0190622* and *FBgn0194931*, which are downregulated in both *w*Ri-infected testes and ovaries compared to their *w*Au-infected and uninfected counterparts, and are annotated with a domain associated with male sterility in *Arabidopsis* and *Drosophila*. In Arabidopsis, the protein with this domain, MALE STERILITY 2, is involved in male gametogenesis, and leads to abnormalities in pollen development when mutated[377]. This is consistent with the effective sterility of males with incompatible females in CI. It is interesting that *w*Ri appears to regulate all three of these genes in the same direction in both testes and ovaries. Other studies have found candidate CI genes that appear to be regulated in opposite directions in male and female hosts, for example *Ance*, which was upregulated in ovaries and downregulated in testes[187]. Such a pattern fits with a model of CI in which modification involves alteration of host gene expression in males and rescue involves a compensatory opposing alteration in females. However, it is also possible that the compensatory alteration could be in the same direction, to bring expression in females 'in line' with that in males. Consistent with this, the host gene *grauzone*, which affected CI when knocked down, was found to be upregulated by *Wolbachia* infection in both ovaries and testes in the mosquito *Cx. quinquefasciatus*[166].

FBgn0262093 (*Ance*), mentioned above, is another interesting candidate. This gene was previously identified as being differentially expressed in infected and uninfected *Drosophila*, with upregulation in *Wolbachia*-infected versus uninfected ovaries, and downregulation in *Wolbachia*-infected versus uninfected testes[187]. This was consistent for a number of *Wolbachia* strains, including *w*Ri. In addition, the penetrance of CI was reduced in incompatible crosses with females carrying a loss-of-function mutation in *Ance*.

This gene is required for spermiogenesis in *Drosophila*[188], consistent with a role in CI. However, the differential expression of *Ance* here was found to be in the opposite direction to that of the previous study, with significantly lower expression in *w*Ri-infected versus both uninfected and *w*Au-infected ovaries. There was no significant difference among testes. The inconsistency of the findings with those of the previous study could reflect differences between *D. melanogaster* and *D. simulans* associations with *Wolbachia*, or the older adult females used here, but downregulation by CI-inducing *Wolbachia* in ovaries would seem to be more consistent with the finding that loss of function of the gene partially rescues CI. Analysis of other candidate genes identified in other studies also revealed different results. *JhI-26* and *CG10433*, identified as being upregulated in *w*Mel-infected *D. melanogaster* testes[191,192], appeared to be downregulated by *w*Au, but were not upregulated in *w*Ri-infected versus uninfected testes. It should be borne in mind that the candidate genes identified in other studies were identified in other species, and there may be differences in the mechanisms of CI between different *Wolbachia* strains and different hosts.

3.4.2. Identification and investigation of candidate *Wolbachia* CI genes

Candidate *Wolbachia* CI genes were identified and investigated by comparative transcriptomic analysis in *D. simulans* ovaries and testes infected with *w*Ri or *w*Au. Global expression analysis suggested that there was more of an overall difference between gene expression in testes and ovaries for *w*Au. Possibly consistent with the effect on host gene expression in *w*Au-infected testes, a relatively large number of *w*Au genes were expressed in testes but not in ovaries.

In contrast to the results of the global analysis, the number of genes significantly differentially expressed between testes and ovaries was actually higher in *w*Ri. A high proportion of differentially expressed genes in both strains had orthologues in the other

strain that were not differentially expressed in the same manner, consistent with the differing phenotypes of the strains.

Functional enrichment analysis did not identify any significantly over-represented categories among the differentially expressed genes. Insufficient functional annotations were likely a factor in this. Analysis of transcriptional regulator gene expression, however, did yield interesting results. The most promising genes identified in Chapter 2, containing a frameshift that would lead to premature termination in *w*Au, appeared to be expressed in both strains. Some transcriptional regulator genes, including two with relatively high expression, were differentially expressed between testes and ovaries in *w*Ri, but none were in *w*Au. The most highly expressed transcriptional regulator gene in *w*Au is one that is absent in *w*Ri. Taken together, these findings lend support to the hypothesis that these genes are involved in CI.

3.4.3. Identification and investigation of candidate host viral inhibition genes

Candidate host genes involved in viral inhibition by *Wolbachia* were identified by comparative transcriptomic analysis in *Ae. albopictus* infected with its natural strains *w*AlbA and *w*AlbB, *w*Mel, which inhibits dengue and chikungunya viruses, or uninfected with *Wolbachia*. Global expression analysis revealed a greater difference between *w*Alb-infected and uninfected mosquitoes than between *w*Mel-infected and uninfected mosquitoes, suggesting that viral inhibition does not involve as many host genes as other *Wolbachia*-host interactions.

Nonetheless, large numbers of genes were found to be differentially expressed between *w*Mel-infected and both *w*Alb-infected and uninfected mosquitoes. Several gene categories were found to be enriched among these genes. For genes upregulated by *w*Mel, interestingly the top three over-represented GO terms were ‘serine-type endopeptidase

activity', 'serine-type peptidase activity' and 'serine-hydrolase activity', the same three terms found to be over-represented in *D. simulans* testes infected with the CI strain *w*Ri versus those infected with the non-CI strain *w*Au or uninfected, as discussed above. It is possible that the enrichment of these terms here also represents their involvement in CI; *w*Mel-infected *Ae. albopictus* displays unidirectional incompatibility with uninfected and bidirectional incompatibility with *w*Alb-infected *Ae. albopictus*[122]. Though as discussed previously, serine-type endopeptidases do have a wide variety of roles, so it is also possible that they are involved in the viral inhibition phenotype.

The other most significantly over-represented terms could be consistent with a role of ROS in viral inhibition by *Wolbachia*, as suggested by previous studies[243,244], although work on this particular transinfection failed to support this hypothesis[378]. ROS are produced as a by-product of oxidative phosphorylation or as a response to pathogens[379,380]. Consistent with a role of ROS in viral inhibition, genes involved in the KEGG pathway 'oxidative phosphorylation' were over-represented in *w*Mel-infected versus uninfected *Ae. albopictus*. This pathway was also strongly implicated in the response to dengue infection in *Ae. aegypti*[381]. In addition, several of the over-represented terms, including some of the most significant, are related to redox reactions, such as 'oxidoreductase activity' and 'oxidation-reduction process'. Heme, which is an iron-containing tetrapyrrole, is a pro-oxidant molecule that has shown to be involved in redox homeostasis in insects[382], and is also found in NADPH oxidase, one of the enzymes responsible for generation of ROS[383,384]. Consistent with this, 'heme binding', 'tetrapyrrole binding' and 'iron ion binding' were three of the most significantly over-represented terms. However, over-representation of these terms does not necessarily imply increased generation of ROS. Other studies have also supported an effect of *Wolbachia* on host iron metabolism, but in a way that suggests it protects hosts from oxidative stress[385,386]. Iron

metabolism is involved in many cellular processes, and its alteration by *Wolbachia* may reflect various host-symbiont interactions[387].

The terms mentioned above are also related to metabolism, as were many of the other over-represented terms associated with genes upregulated by *w*Mel. This could be consistent with the theory that pathogen inhibition by *Wolbachia* is mediated by competition for resources, such as cholesterol[223,258] with increased metabolic demands from *w*Mel compared to *w*Alb limiting resources available to viruses. Support for alteration of metabolism by *w*Mel in *Ae. albopictus* comes from work showing that *w*Mel-infected as well as *w*MelPop-infected *Ae. albopictus* cell lines displayed altered host lipid composition, with a reduced availability of lipids that may be important to virus propagation[388]. A possible explanation for higher metabolic demands of *w*Mel is that it is found at much higher titres in *Ae. albopictus* than is *w*Alb[122]. Other studies have also found protective *Wolbachia* to be at higher titres than *Wolbachia* that do not display pathogen inhibition[112,223,231,235-237].

Another quite significantly over-represented term was 'odorant binding', and this function was also associated with a high proportion of genes with functional information expressed exclusively in *w*Mel-infected mosquitoes. It is not certain why this should be the case, or whether it has a role in viral inhibition by *Wolbachia*. It is hard to envisage a mechanism by which this would be the case, though it is conceivable that *Wolbachia* might upregulate the feeding behavior of its hosts in order to meet increased metabolic demands, which might involve upregulation of odorant binding. Increased feeding and foraging behavior due to infection has been observed in other insects[389,390]. However, this is only speculation here, and in fact the *Wolbachia* strain *w*MelPop was found to decrease blood-feeding success in *Ae. aegypti*, albeit only in older mosquitoes[391]. Nonetheless, this finding does

raise an interesting ethical point regarding the use of *Wolbachia* transinfections for disease control, as increased feeding behavior would be an undesirable characteristic.

Interestingly, none of the over-represented GO terms were obviously immune-related, and only three of the 2,964 genes differentially expressed in *w*Mel-infected versus both *w*Alb-infected and uninfected *Ae. albopictus* had the word 'immune' in either their annotation or associated GO terms. Several other studies have suggested a role for an immune response in pathogen inhibition by *Wolbachia*[202,222,225,249,327]. However, previous work on this particular *w*Mel transinfection found that while upregulation of three immune genes was observed initially, there was no significant upregulation after ten generations[122], and further work on a wider range of immune genes also failed to reveal immune upregulation[388], consistent with the apparent lack of immune upregulation shown here. Furthermore, the scale of even the initial upregulation found by Blagrove *et al*.[122] was much lower than that observed for *w*MelPop in *Ae. aegypti*[225] and *An. gambiae*[202]. A transinfection of *Ae. aegypti* with *w*Mel also displayed lower immune upregulation than the *Ae. aegypti w*MelPop transinfection[250]. In its native host *D. melanogaster*, *w*Mel was found to confer viral resistance without any observed immune upregulation[250]. Along with other studies, these results suggest that the combination of host and *Wolbachia* affects whether immune upregulation occurs or not, and it has been proposed that chronic immune upregulation may only occur when a previously *Wolbachia*-naïve host is transinfected with *Wolbachia*. Consistent with this, the results here do not support immune upregulation as a mechanism for viral inhibition by *w*Mel in *Ae. albopictus*.

There was also enrichment of many GO terms in genes downregulated in *w*Mel-infected versus both *w*Alb-infected and uninfected *Ae. albopictus*. These mostly constituted terms associated with chromatin modification, particularly histone acetylation. This process relaxes the structure of chromatin, making the DNA more accessible to transcription

factors and increasing the level of transcription. Previous studies have also found epigenetic effects of *Wolbachia* – cytosine methylation, associated with decreased transcription, was altered by *w*MelPop in *Ae. aegypti*[392], and methylation profiles correlated with the *Wolbachia*-induced trait of feminisation in a leafhopper[393]. Furthermore, in *Ae. aegypti* cells, *w*MelPop was found to downregulate the *Ae. aegypti* methyltransferase responsible for cytosine methylation, and this was linked to dengue inhibition by *w*MelPop[267]. Our findings here provide supporting evidence for epigenetic regulation of gene expression by *Wolbachia*, though it is the first known time an effect on acetylation has been suggested.

Overall, our results do not support immune upregulation as a mechanism for *Wolbachia*-induced viral inhibition at least in this particular transinfection, but are consistent with other theories such as generation of ROS and effects on metabolism. Of course it is possible that several mechanisms may be involved, and different *Wolbachia*-host combinations may display different mechanisms. Transcriptomic analysis of *Wolbachia* genes involved would also be beneficial, as discussed above for investigation of CI.

3.4.4. Conclusions

These analyses represent the first known use of RNA-seq to identify and investigate candidate host and *Wolbachia* genes involved in CI, and host genes involved in viral inhibition by *Wolbachia*. It is the first known time that any transcriptomic analysis has been performed in *Wolbachia* to identify genes involved in these *Wolbachia*-host interactions.

It should be noted that there are limitations to the interpretation of the findings presented here. Firstly, while the host insects were backcrossed to homogenise the genetic background, there could still have been extraneous variables resulting from different host

strains that affected gene expression. For example, *Wolbachia* density has been shown to increase with lab adaptation[394,395], so variations in time since lab colonisation could have affected *Wolbachia* density, which may in turn have affected gene expression. In addition, due to the expense of RNA-seq, only one time point was taken for each sample type in all cases. It could be that there are genes involved in CI or viral inhibition that are active at different ages that have been missed here. Furthermore, developmental stage is likely to affect expression of genes involved in CI in germ cells. The samples here were timed to provide a mix of stages of germ cells, but this could obscure important differences in gene expression in individual stages. For the investigation into the mechanism of pathogen inhibition, whole mosquitoes were used for RNA-seq. Analysis of individual tissues, such as midgut, haemocel and salivary glands, may yield improved understanding of the pathogen inhibition mechanism.

In addition to limitations relating to the samples, there were also limitations to the computational analysis. *Wolbachia* genes were only classed as candidate CI genes if they were differentially expressed between ovaries and testes in one *Wolbachia* strain but not the other, which may exclude some valid candidates. In addition, the host transcriptomic analyses here have focused on protein-coding genes; it could be that non-coding RNAs play a role in *Wolbachia*-host interactions. Consistent with this, several studies have found differences in host microRNA expression in *Wolbachia* infection[319,396,397], and suggested a role for host microRNAs in dengue inhibition by *Wolbachia*[267,398].

It should also be noted that quantification by RNA-seq is not always accurate, and promising findings on individual genes would benefit from qRT-PCR validation. Furthermore, expression at the mRNA level does not necessarily correlate with expression at the protein level; proteomic analysis would be useful in all cases. Nevertheless, this

work has yielded interesting results on *Wolbachia*-host interactions and provided several candidates for further investigation.

4 Identification of candidate genes for transgenic control of pest insects

4.1. Introduction

An alternative strategy to the use of *Wolbachia* for insect control is the release of transgenic insects carrying synthetic constructs that lead to population suppression, as discussed in Chapter 1. This chapter focuses on the identification of components that could be used in such synthetic constructs.

4.1.1. Searching for *Ae. aegypti* sex locus

4.1.1.1. Sex determination in Ae. aegypti

Unlike many species, including most other insects, *Ae. aegypti* does not possess heteromorphic sex chromosomes, with sex determined instead by a region on a chromosome[399]. Males are the heterogametic sex, with a dominant allele designated M and a recessive allele designated m, while females have the genotype mm[400]. At the time this work was undertaken, linkage and physical mapping had revealed the sex locus to be close to the centromere of chromosome 1, in region 1q21[401-405], but its nature, mechanism, and sequence had not been determined. Y chromosomes are highly repetitive and heterochromatic, leading to difficulties in sequencing and assembly[406]. It is likely that discovery of the sex locus has been impeded for the same reasons; consistent with this, intense staining of region 1q21 suggests that it is also heterochromatic[405]. Since this work was performed, a gene designated *Nix* has since been determined to be the male-determining M factor[104], but the efforts here to identify the sex locus prior to this discovery will be discussed nonetheless.

4.1.1.2. Potential uses of sex locus for transgenic control

There are various ways in which the sex locus could be used for transgenic control. Most obviously, it could be used in sex distortion systems. Expressing the male-determining M factor in females during development could potentially turn them into genetic males, resulting in a biased sex ratio that would lead to population suppression; ectopic expression of *Nix* was indeed found to have this effect[104]. A sex-distorting system could also potentially be created by targeting the recessive m allele, with an effector such as an engineered site-specific nuclease. The effector could be active during development, or alternatively in spermatogenesis, so that only functional sperm containing the M allele would be produced. As mentioned in Chapter 1, an equivalent to the latter system has been produced in *An. gambiae*, using a homing endonuclease that targets the X chromosome in sperm[103]. Similar to the former system, siRNA silencing of *doublesex*, a downstream gene involved in sex development, resulted in interference with female-specific traits in *Ae. aegypti*[407]. However, in another study RNAi-mediated knockdown of *Ae. aegypti doublesex* was found to prevent the development of females altogether[408]. Thus it is possible that targeting the m allele during development would lead to the death of females rather than turning them into phenotypic males, but such an effect would still be useful for population suppression.

These applications all require knowledge of the sequence of the sex locus M or m alleles. Thus, given that this was unknown at the time, genome sequencing of multiple male and female *Ae. aegypti* was performed and the sequences compared to identify differences that could potentially represent the sex locus.

4.1.2. Identification of *Ae. aegypti* and *C. capitata* testis-specifically expressed or spliced genes

Transgenic control systems require expression of an effector transgene in a particular tissue and/or at a particular developmental stage, and usually require that the transgene not be expressed elsewhere or at another time. For some insect control strategies, germline-specific transgene expression is required, male germline-specific expression being of particular interest. These include sex distortion systems targeting sperm that would result in female offspring[103], prevention of sperm development[83], or systems in which embryos fertilised by sperm from modified males are inviable[84].

Expression can be controlled with tissue- or stage-specific regulatory regions (promoters and/or untranslated regions (UTRs)), such as those of the *Actin-4* gene, which have been used for female-specific transgene expression in *Ae. aegypti*[97], *Ae. albopictus*[98] and *An. stephensi*[409]. In addition, alternative splicing may be exploited; for example sex-specific introns have been utilised to add additional specificity to the *Actin-4* regulatory regions in the species above, and to achieve female-specific expression in *C. capitata*[99], *Bactrocera oleae* (olive fly)[100], *Plutella xylostella* (diamondback moth)[101] and *Pectinophora gossypiella* (pink bollworm)[101]. Analogous components to drive germline-specific expression, particularly in males, would be useful for the applications described above. Conserved components that could be used across multiple species would be particularly useful.

Several insect genes with testis-specific expression have been identified, often first in *D. melanogaster*, for example *β2-tubulin*[410]. Homologues of *β2-tubulin* have been identified and the promoters found to drive testis-specific expression in other species, including *An. gambiae*[53], *Ae. aegypti*[411] and *C. capitata*[412]. However, studies on *D. melanogaster* suggest that expression timing in male germline cells must be taken into consideration when building synthetic constructs. In *D. melanogaster*, transcription is repressed with the onset

of the meiotic divisions[413,414]. Barring a few exceptions[415-417], genes whose product is required after this transcriptional repression are transcribed in primary spermatocytes, before the meiotic divisions – the transcripts are then stored and translated as required[418]. To achieve conditional transgene expression, synthetic constructs often utilise bipartite expression systems, such as the GAL4-UAS[300] or the TetOff[87,88] system used in RIDL, discussed in Chapter 1. In these systems, expression of the effector transgene is driven by binding of a transcription factor, and it is this transcription factor rather than the transgene directly that is under the control of the regulatory elements restricting expression. If testis-specific regulatory regions were to be used in such systems, they would need to drive pre-meiotic protein expression, otherwise the transcription factor would not be translated early enough to drive transcription of its target.

High-throughput transcriptional profiling[419] and subtractive hybridisation[408] studies have recently yielded a number of potential germline-specific transcripts in *Ae. aegypti*, and a number of microarray experiments have yielded candidates in other species such as *An gambiae*[420,421] and *D. melanogaster*[422-424]. However, no known studies have been performed with sufficient time resolution to determine the activity of regulatory regions at different stages of spermatogenesis. Information on insect testis-specific splicing is even more sparse; testis-specific splice forms of the genes *achi* and *vis* have been discovered in *D. melanogaster*[425], but no testis-specific splice forms are known to have been identified in *Ae. aegypti*, *C. capitata*, or any other pest insect.

Thus, here RNA-seq was performed on staged testis samples from *Ae. aegypti* and *C. capitata*, to identify genes with testis-specific expression peaking early in spermatogenesis, whose regulatory regions are therefore likely to be suitable for pre-meiotic protein expression. Using a novel computational pipeline, testis-specific splice forms were also identified. Data from *D. melanogaster* was used as a control, and to

facilitate interspecies comparison to identify conserved components that may function in constructs across multiple species.

4.2. Methods

4.2.1. Searching for *Ae. aegypti* sex locus

4.2.1.1. Ae. aegypti

Ae. aegypti of the Asian wild-type strain (originating from Malaysia, colonised by the Institute of Medical Research (Kuala Lumpur) in 1977) were reared by Dr Kelly Matzen at Oxitec Ltd (Abingdon, UK), at 27 +/-2 °C and 70 +/-10 % relative humidity with a 12:12 hour light:dark cycle. Larvae were reared in trays and fed with Tetramin® (Tetra GmbH, Germany). Males and females for experimental samples were separated as pupae. Adults were maintained in cages with *ad libitum* access to a 10 % sucrose solution supplemented with 14 U mL^{-1} penicillin and 14 μg mL^{-1} streptomycin (Sigma-Aldrich, UK). Adult females for both colony maintenance and experimental work were fed defibrinated horse blood (TCS Biosciences Ltd, UK) 3-5 days after eclosion. *Ae. aegypti* used for gDNA extraction were inbred for four generations, using a single male and a single female for each generation.

4.2.1.2. gDNA extraction

gDNA extraction was performed by Dr Ian Goodhead at the University of Liverpool (Liverpool, UK) using the CTAB method[426]. 12 individual male and 12 individual female samples were prepared.

4.2.1.3. Sequencing

Library preparation was performed with the assistance of Dr Ian Goodhead and Frances Blow in Dr Alistair Darby's group at the University of Liverpool (Liverpool, UK). Unstranded paired-end libraries were prepared using a NEBNext Ultra DNA Library Prep

Kit (New England Biolabs, UK), according to the manufacturer's instructions. Fragmentation was performed using a Covaris S2 SonoLAB Single machine (Covaris Inc., USA). Libraries were evaluated using an Agilent 2100 Bioanalyzer (Agilent, UK), according to the manufacturer's instructions. Sequencing was performed at the Centre for Genomic Research at the University of Liverpool (Liverpool, UK) on the Illumina HiSeq 2500 platform, using a read length of 150 nt.

4.2.1.4. Sequence data processing

Raw reads were processed to remove adapter sequences and sequences of poor quality using cutadapt[339]. A reference index of the *Ae. aegypti* genome assembly AaegL2 (downloaded from VectorBase[427]) was constructed and reads aligned to this using Bowtie2 (v2.1.0)[335]. Reads were also aligned to a reference index generated from contigs rather than scaffolds, using the same tools.

4.2.1.5. Coverage analysis

Coverage for each scaffold or contig was computed using BEDTools (v2.18.2)[428]. Scaffolds and contigs with significantly different coverage between males and females were identified using edgeR[429]. An FDR value of 0.05 was used as the threshold for significance. A BLASTn search of male-biased scaffolds against female-biased scaffolds was performed using the BLAST+ package[430].

4.2.1.6. Variation analysis

Alignment files were processed prior to variant calling using PicardTools (v1.98)[337] to add read group information, remove duplicate reads and create a sequence dictionary for indel realignment. Indel realignment and variant calling were performed using GATK (2.8.1)[431]. Variant sites with two haplotypes in all male samples and one haplotype in all female samples were extracted using vcftools (version 0.1.9)[432] and awk.

4.2.1.7. Assembly analysis

De novo assemblies were generated using Velvet (v 1.2.07)[281] with VelvetOptimiser. Male and female sequencing reads were aligned to assemblies using Bowtie2 (v2.1.0)[335]. The male-exclusive assembly was aligned with the reference assembly AaegL2 using Mugsy (v1r2.3)[433].

4.2.1.8. Statistical analysis

Statistical analysis was performed using R, and an online calculator for Chi-squared tests[434].

4.2.1.9. Searching for Nix

BLAST databases were generated from the male assembly and male-exclusive assembly and tBLASTx searches of the gene sequence of *Nix* performed using the BLAST+ package[430].

4.2.2. Identification of *Ae. aegypti* and *C. capitata* testis-specifically expressed or spliced genes

4.2.2.1. Insects

4.2.2.1.1. Ae. aegypti

Ae. aegypti of the Asian wild-type strain were reared by Dr Yachuan Yu at Cardiff University (Cardiff, UK) and Dr Jenny Molloy at Oxitec Ltd (Abingdon, UK), as described above.

4.2.2.1.2. C. capitata

C. capitata of the Toliman wild-type strain (originating from Guatemala, colonised in 1990) were reared by Dr Romisa Asadi, Ryan Turkel and Peter Elphick at Oxitec Ltd (Abingdon, UK), at 26 +/- 1 °C and 65 +/- 10 % relative humidity with a 12:12 hour

light:dark cycle. Larvae and adults were kept in plastic containers with *ad libitum* access to a diet containing maize meal, sucrose and yeast. Pupae were allowed to eclose in a Petri dish containing sand. Males and females for experimental samples were separated shortly after eclosion, before mating.

4.2.2.1.3. D. melanogaster

D. melanogaster of the w^{1118} strain were reared by Dr Yachuan Yu at Cardiff University (Cardiff, UK), at 25 +/- 1 °C with a 12:12 hour light:dark cycle. All stages were kept in plastic containers with *ad libitum* access to a diet containing cornmeal, dextrose and yeast. Males and females used for experimental samples were separated shortly after eclosion, before mating.

4.2.2.2. Dissection and RNA extraction

4.2.2.2.1. For sequencing

Dissection and RNA extraction for RNA-seq was performed by Dr Yachuan Yu and Dr Romisa Asadi in Prof. Helen White-Cooper's group at Cardiff University (Cardiff, UK). Virgin *Ae. aegypti*, *C. capitata* and *D. melanogaster* males were used for dissection of staged testis and gonadectomised carcass at three days after eclosion. For *Ae. aegypti*, two staged testis samples (referred to as 'early' and 'late') were prepared by bisecting testes; the apical region contains cysts of male germline cells in earlier stages of development, up to late spermatocytes, and the basal region contains spermatid cysts in later stages of development. Both of these samples also contained somatic cells from the testis sheath. For *C. capitata*, four staged testis samples were prepared – early spermatocytes, late spermatocytes, round spermatids and elongated spermatids – by spilling cysts out of the testes and examining isolated cysts with a Nikon Eclipse Ti-S inverted microscope (Nikon, Surrey, UK). Cysts at specific stages were identified based on cell size and morphology, and collected manually with a pulled-out Pasteur pipette. A *C. capitata* ovary sample was also

prepared from virgin females five days after eclosion. For *D. melanogaster*, three staged testis samples were prepared – early spermatocytes, late spermatocytes and spermatids – using the same method as for *C. capitata*. All tissues were dissected into PBS. Total RNA was extracted using TRIzol® (Life Technologies Ltd, Paisley, UK), according to the manufacturer's instructions. A single replicate was prepared for each sample type. Tissues from multiple individuals were pooled for each sample.

4.2.2.2.2. For RT-PCR

Virgin *Ae. aegypti* and *C. capitata* males were used for dissection of testis and gonadectomised carcass at 0-3 days after eclosion. Virgin *Ae. aegypti* and *C. capitata* females were used for dissection of ovary and gonadectomised carcass at 4-6 days after eclosion and ~24 hours post-blood meal (PBM). For qRT-PCR, staged testis samples were also prepared as described above, except in this instance only two samples – spermatocytes and spermatids – were prepared for *C. capitata*. All tissues were dissected into PBS, then either stored in RNALater (Qiagen, Manchester, UK) or lysis buffer (Life Technologies Ltd, Paisley, UK or Norgen Bioteck Corp., Ontario, Canada) at -20 °C until RNA extraction, or used immediately for RNA extraction. Total RNA was extracted using either a Norgen Total RNA Purification Kit (Norgen Biotek Corp., Ontario, Canada) or an Ambion RNAqueous Kit (Life Technologies Ltd, Paisley, UK), according to the manufacturer's instructions. Three replicates were prepared for each sample type. Tissues from multiple individuals were pooled for each sample.

4.2.2.3. cDNA synthesis

cDNA for RT-PCR was synthesised using a RevertAid First Strand cDNA Synthesis Kit (Thermo Scientific, Pittsburgh, USA) with random hexamer primers, according to the manufacturer's instructions. cDNA concentration and quality was assessed using a NanoDrop spectrophotometer (Thermo Scientific, Pittsburgh, USA).

4.2.2.4. Sequencing

Library preparation was performed by Dr Yachuan Yu in Prof. Helen White-Cooper's group at Cardiff University (Cardiff, UK). Unstranded single end libraries were prepared using an Illumina TruSeq RNA Sample Preparation Kit (Illumina, UK), according to the manufacturer's instructions. Sequencing was performed at the NGS facility at Glasgow Polyomics (University of Glasgow, Glasgow, UK) on the Illumina Genome Analyzer II platform, using a read length of 73 nt.

4.2.2.5. Data from other studies

RNA-seq data from other studies were downloaded from the SRA[435] (Table 4.1).

Species	Sample	Data number(s)	Sample number	Experiment number	Reference
D. melanogaster	female (1 day, virgin)	SRR023595 and SRR035393	SRS004689	SRX008013	Graveley *et al.* (2011)[314]
		SRR023680 and SRR023702	SRS004689	SRX008188	
	ovaries (4 days, virgin)	SRR070396 and SRR070417	SRS118276	SRX029395	Unavailable
	male head (1 day)	SRR070432 and SRR070433	SRS118262	SRX029371	Unavailable
		SRR100280	SRS118262	SRX042021	Unavailable
Ae. aegypti	female carcass (4-6 days, 24h PBM)	SRR923829	SRS453851	SRX316548	Akbari *et al.* (2013)[410]
Ae. aegypti	ovaries (4-6 days, 24h PBM)	SRR923830	SRS453856	SRX316548	
C. capitata	(0-6 days)	SRR836189	SRS417380	Unavailable	Unavailable

Table 4.1. RNA-seq data from other studies downloaded from the SRA. PBM, post-bloodmeal.

4.2.2.6. Sequence data processing

The overall quality of the sequencing reads was assessed using FastQC (v0.10.1)[338]. Raw reads were processed to remove adapter sequence using FASTA/Q Clipper from the FASTX_Toolkit[436] and sequences of poor quality using Sickle[437]. Reference indexes of the *Ae. aegypti* genome assembly AaegL2 (obtained from VectorBase[427]), the *C. capitata* genome assembly Ccap_1.0 (obtained from NCBI[438]) and the *D. melanogaster* genome assembly r5.54 (obtained from FlyBase[333]) were constructed using Bowtie2 (v2.1.0)[335]. Trimmed reads were aligned to these indexes using TopHat2 (v2.0.9)[336]. Transcript assemblies were generated from the alignments using Cufflinks (v2.1.1)[340] with reference annotation based transcript (RABT) methodology[341] followed by Cuffmerge (v1.0.0)[342]. Transcript expression in each sample was quantified using Cuffdiff2 (v2.1.1)[346].

4.2.2.7. Identification of candidate testis-specifically expressed or spliced genes

Candidate testis-specifically expressed or spliced genes were identified from the output of Cuffdiff2, and their sequences obtained, using custom written Python scripts (available at github.com/ElizabethSutton) in combination with BEDTools (version 2.16.2)[428]. Filtering steps were applied as described in the results section. For steps requiring alignment of sequences, Geneious (7.0.5)[293] was used.

For the candidate testis-specifically expressed genes, only genes with higher expression in the samples from early stages of spermatogenesis (the early sample for *Ae. aegypti*; the early spermatocytes sample for *C. capitata*) than in the samples from the later stages (the late sample for *Ae. aegypti*; the mean expression in the late spermatocytes, round spermatids and elongated spermatids samples for *C. capitata*) were taken forward.

4.2.2.8. Interspecies comparison

Sequences of all transcripts predicted by Cufflinks were extracted using a custom Python script (available at github.com/ElizabethSutton) in combination with BEDTools (version 2.16.2)[428]. tBLASTx searches of the transcript sequences from one species against the transcript sequences from another were performed using the BLAST+ package[430].

4.2.2.9. Experimental testing

4.2.2.9.1. RT-PCR

RT-PCR primers were designed using Primer-BLAST[277]. Primer sequences are listed in Appendix 2. RT-PCR was performed on a TGradient thermocycler (Biometra, Goettingen, Germany) using a PCRBIO kit (PCR Biosystems Ltd, London, UK), according to the manufacturer's instructions. The reaction parameters were: 95 °C for 30 seconds, 2 cycles of 95 °C for 15 seconds, 55 °C for 30 seconds and 72 °C for 2 minutes, 33 cycles of 95 °C for 15 seconds, 55 °C for 15 seconds and 72 °C for 30 seconds, and finally 72 °C for 1 minute. RT-PCR products were visualized on 1.5-2 % agarose gels, as described previously.

4.2.2.9.2. qRT-PCR

qRT-PCR was performed on an Mx3500P instrument (Stratagene, La Jolla, USA) using iQ™ SYBR® Green Supermix (Bio-Rad, Hemel Hempstead, UK) according to the manufacturer's instructions. The reaction parameters were: 95 °C for 5 minutes, followed by 40 cycles of 95 °C for 15 seconds, 55 °C for 15 seconds and 60 °C for 15 seconds. Reactions were performed with serial dilutions to determine primer efficiency. Reactions with primers targeting alpha-tubulin were performed for normalisation. Two technical replicates were performed and the mean expression value taken.

4.2.2.9.3. Purification and sequencing of PCR products

To confirm that the PCR results reflected the predicted candidates, PCR products were purified using a QIAquick PCR Purification Kit (Qiagen, Manchester, UK) and sent for sequencing by GATC Biotech (Cologne, Germany).

4.2.2.10. Promoter prediction

4.2.2.10.1. Identification of motifs conserved between species

Experimentally validated candidate testis-specifically expressed genes were used as queries in a tBLASTx search of the nr database to identify homologous genes. The upstream sequences of the candidate genes and their top two matching genes from other insect species were obtained (2,500 Kb of upstream sequence, or however much sequence was present before the next gene or end of the scaffold, if less than 2,500 Kb). These sequences were input into PROMO[439], rVista 2.0[440] and Melina II[441] to identify conserved motifs.

4.2.2.10.2. Identification of motifs conserved within species

2,500 Kb of upstream sequence was obtained for each *Ae. aegypti* candidate testis-specifically expressed gene that corresponded to an annotated gene with a 5'UTR. These sequences were input into PROMO[439] and Melina II[441] to identify conserved motifs.

4.3. Results

4.3.1. Searching for *Ae. aegypti* sex locus

To attempt to identify the sex locus, the genomes of 12 male and 12 female *Ae. aegypti* were sequenced and compared. Three main approaches were taken towards identifying candidates from the sequencing data – analysis of coverage differences between males and females, analysis of sequence variation between males and females, and analysis of separate male and female assemblies. Use of different approaches would enable greater

confidence in candidates identified by more than one approach, and was necessary as the nature of the sex locus was unknown at the time. The M and m alleles could have been sufficiently different to be represented as distinct haplotypes in the current genome assembly, in which case they would likely display differing coverage between males and females, and the M allele would likely be present in an assembly constructed from male reads but not an assembly constructed from female reads. They might not be identified by searching for regions with sequence variation between males and females, as reads from the two different alleles may not map to the same haplotype. On the other hand, if the M and m alleles were sufficiently similar to be collapsed into one haplotype in the current assembly, analysis of sequence variation would be the best way to identify them; such a region is less likely to show differential coverage, or obvious differences between male and female assemblies.

4.3.1.1. Sequencing and read alignment

In total 376,122,418 reads of 150 nt were generated, corresponding to 56.4 Gb of data. After trimming and quality filtering a total of 359,689,764 reads were used for genome alignment with Bowtie2 (v2.1.0)[335]. 86.2 % of these mapped to the *Ae. aegypti* genome, corresponding to a mean coverage of approximately 41X.

4.3.1.2. Coverage analysis

The coverage of male reads and female reads was computed for each scaffold and each contig in the latest available version of the *Ae. aegypti* genome assembly (AaegL2) using BEDTools (v2.18.2)[428]. To assess whether scaffolds or contigs with differing male and female coverage were promising candidates for further investigation, the correlation between coverage ratios for mapped scaffolds or contigs and their position was assessed. Mapping data were obtained from Nene *et al*. (2007)[442], Timoshevskiy *et al*. (2013)[405] and Juneja *et al*. (2014)[443]. As the approximate location of the sex locus was known to be near

the centromere of chromosome 1, a greater difference between male and female coverage for scaffolds or contigs mapped to this chromosome or to this region might be expected if analysis of coverage differences were a suitable method for identification of the sex locus.

Statistical analysis revealed no significant difference in the magnitudes of the log2 male to female scaffold coverage ratios between any of the chromosomes ($P = 7.3x10^{-2}$, Kruskal-Wallis test). However, there were significant differences between scaffolds mapped in the approximate location of the sex locus (defined as the region between (and including) the bins containing the closest mapped scaffolds known to be on different sides of the sex locus) and both scaffolds mapped to chromosomes 2 and 3 ($P = 5.3x10^{-3}$, Wilcoxon test) and scaffolds mapped elsewhere on chromosome 1 ($P = 4.1x10^{-2}$, Wilcoxon test). The same analysis performed for contigs revealed a significant difference between chromosomes ($P = 5.59x10^{-16}$, Kruskal-Wallis test). Using post-hoc Wilcoxon tests with Bonferroni corrections, a significant difference was found for all pairwise comparisons of chromosomes (chr1 vs chr2, $P = 2.4x10^{-2}$; chr1 vs chr3, $P = 1.7x10^{-14}$; chr2 vs chr3, $P = 1.5x10^{-10}$). It is unclear why there is a significant difference between chromosomes 2 and 3. As for scaffolds, there were also significant differences between the magnitudes of the log2 coverage ratios for contigs mapped near the sex locus and contigs mapped to chromosomes 2 and 3 ($P = 1.4x10^{-8}$, Wilcoxon test), and contigs mapped elsewhere in chromosome 1 ($P = 1.2x10^{-3}$, Wilcoxon test). Scatter plots illustrating female to male coverage ratios for contigs at different chromosomal positions (based on data from Juneja *et al*. (2014)[443]) are shown in Figure 4.1.

These findings suggest that the sex locus may be contained on a scaffold or contig with differing male and female coverage. Therefore, scaffolds and contigs with significantly different male and female coverage were identified using edgeR[429]. 573 scaffolds had significantly different coverage between males and females (FDR < 0.05), 352 with higher

coverage in males and 221 with higher coverage in females. 2,485 contigs had significantly different coverage between males and females (FDR < 0.05), 1,176 with higher coverage in males and 1,309 with higher coverage in females. Scaffolds or contigs that have been physically mapped to a location other than approximately where the sex locus is known to be are unlikely to contain the sex locus, unless the mapping is incorrect, but the majority of these scaffolds and contigs could not be excluded on this basis. Most were unmapped, with only 37 of the 573 scaffolds mapped to an unsuitable location for the sex locus.

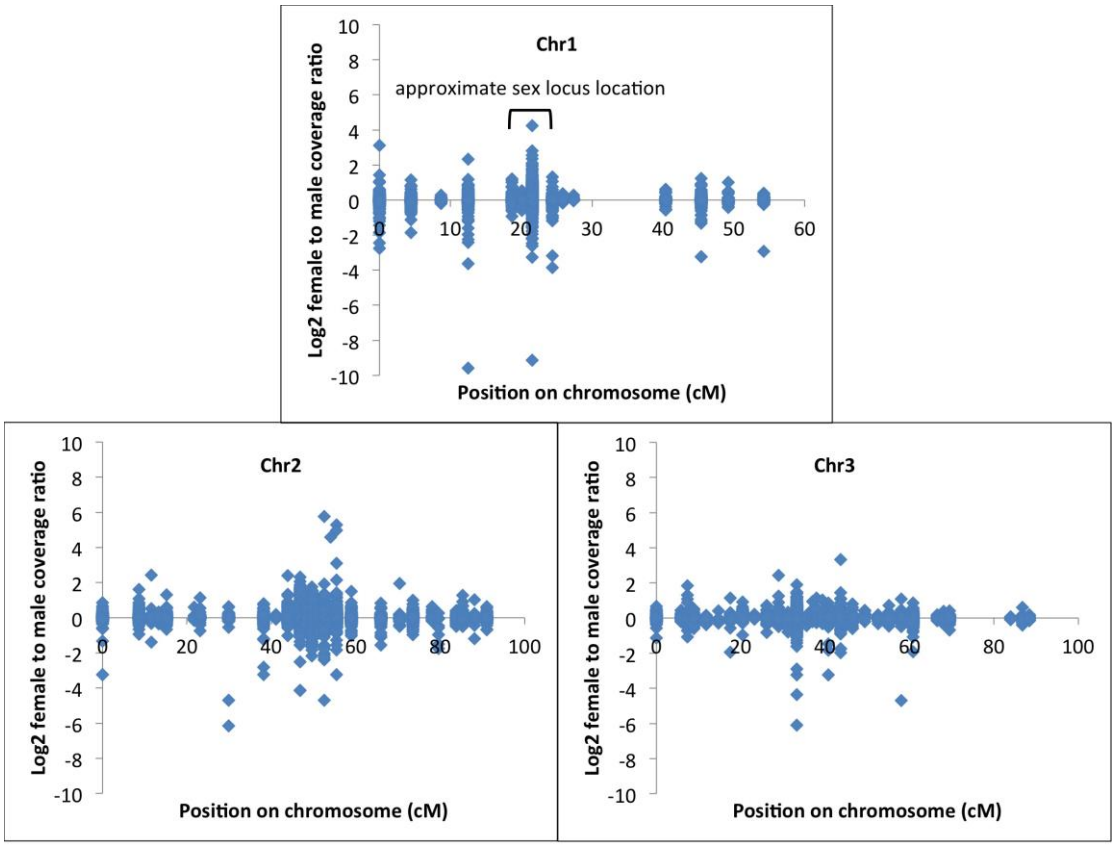

Figure 4.1. Scatter plots illustrating female to male coverage ratios at different chromosomal positions. The approximate sex locus location is defined as the region between (and including) the positions at which the closest mapped contigs known to be on different sides of the sex locus have been mapped.

If the M and m alleles represent different haplotypes that have been assembled on separate scaffolds or contigs, their sequences might be sufficiently similar to yield a significant hit in a BLAST search. The sequences of male-biased scaffolds were therefore

used as queries in a BLAST search against scaffolds with female-biased scaffolds. This yielded 18,589 hits, too many to feasibly analyse, so this line of enquiry was not pursued further.

It seemed likely that the sex locus would constitute a coding region, so the presence of annotated genes on scaffolds with differential coverage was then analysed. There were 347 annotated genes on male-biased scaffolds and 936 annotated genes on female-biased scaffolds. It is plausible that a gene representing the sex locus will show differential expression between males and females, and so filtering these lists based on differential expression between the sexes could be a way to narrow them down to the most promising candidates. However, genes found to be more highly expressed ($P < 0.05$) in males or in females by Harker *et al.*[444] or Dissanayake *et al*.[445] were not significantly over-represented on male-biased scaffolds or female-biased scaffolds respectively ($P > 0.05$, Chi-squared test). Likewise, genes found to be male- or female-specifically expressed by Akbari *et al*.[419] were not significantly over-represented on male-biased scaffolds or female-biased scaffolds respectively ($P > 0.05$, Chi-squared test). Furthermore, there were 574 annotated genes with suggested sex-differential or specific expression on sex-biased scaffolds. 220 of these were on scaffolds mapped to an unsuitable location for the sex locus, but the remaining 354 was still too large a set for experimental testing. This line of enquiry was therefore not pursued further.

4.3.1.3. Variation analysis

Sites with variation in the male and female sequence data relative to the reference genome assembly were determined with GATK (2.8.1) HaplotypeCaller[431]. To assess whether variant sites were promising candidates for further investigation, the correlation between variant site frequency for mapped scaffolds and their position was assessed, using the same mapping data as previously described. A greater level of variation from the reference

genome for scaffolds mapped to chromosome 1 or to the region of the sex locus might be expected if analysis of sequence variation were a suitable method for identification of the sex locus.

Statistical analysis revealed no significant difference in overall variant site frequency between any of the chromosomes ($P > 0.05$, Kruskal-Wallis test), but a significant difference in indel frequency between chromosomes ($P = 1.9 \times 10^{-2}$), although no pairwise comparisons were significant ($P > 0.05$, post-hoc Wilcoxon test with Bonferroni correction). However, chromosome 1 displayed the lowest indel frequency, which fails to support the hypothesis that this variation might be associated with the sex locus. Wilcoxon tests showed a significant difference in overall variant site frequency between scaffolds mapped near the sex locus and scaffolds mapped on chromosome 2 or 3 ($P = 6.8 \times 10^{-3}$), or elsewhere on chromosome 1 ($P = 8.3 \times 10^{-4}$). Both of these comparisons were also significant when considering only non-indel variants ($P = 2.1 \times 10^{-3}$ and $P = 1.4 \times 10^{-3}$), and the comparison of scaffolds near the sex locus versus elsewhere on chromosome 1 was significant for indels ($P = 6.9 \times 10^{-3}$). However, overall the level of variation seemed to be greatest for chromosome 3 (Figure 4.2).

Variant sites that could potentially represent the sex locus would be expected to display two haplotypes in males and one haplotype in females. In total there were 1,049 variant sites displaying two haplotypes in all males and one in all females, 131 indels and 918 non-indel variants such as SNPs or multi-nucleotide polymorphisms (MNPs). 582 of these variant sites were on scaffolds mapped to unsuitable locations for the sex locus, but the remaining 467 is still too large a set for experimental testing. In addition, the results above fail to provide strong support for an association between the sex locus and sequence variation from the reference genome. This line of enquiry was therefore not pursued further.

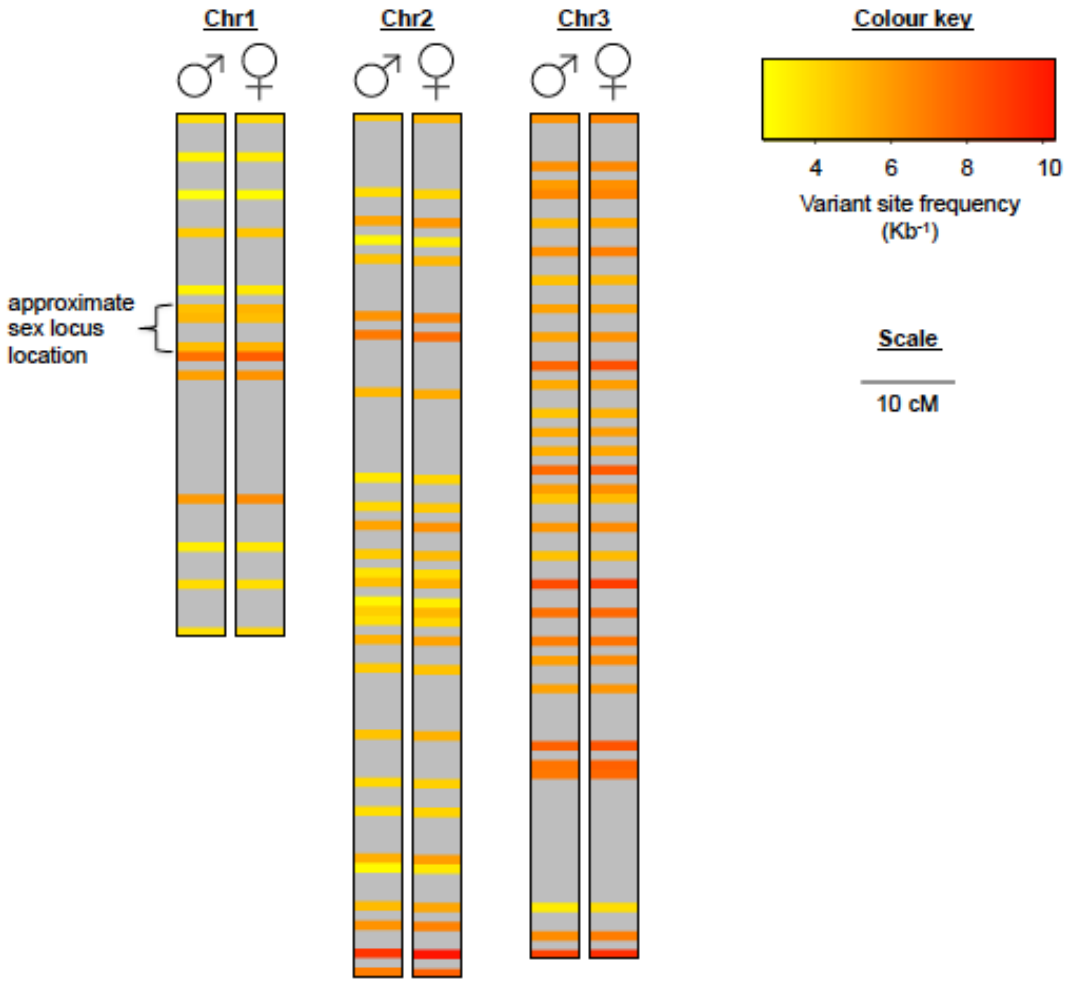

Figure 4.2. Heatmap illustrating variant site frequency in male and female samples at different chromosomal positions. Values illustrated are mean variant site frequency for scaffolds mapped within bins of 1 centimorgan. The approximate sex locus location is defined as the region between (and including) the bins containing the closest mapped scaffolds known to be on different sides of the sex locus.

4.3.1.4. Assembly analysis

Separate *de novo* assemblies from male reads and female reads were constructed using Velvet (v 1.2.07)[281]. Male reads that map to the male assembly but not to the female assembly were possible candidates for the M allele, so these reads were used to generate a male-exclusive assembly. This assembly comprised 11,291 contigs, with an N50 of 425 bp and a total size 4.94 Mbp. 1,317 scaffolds in the reference genome (AaegL2) produced alignments with this male-exclusive assembly using Mugsy (v1r2.3)[433]. These alignments overlapped with 864 annotated genes, although most of the overlaps were small. One of the most interesting candidates was *AAEL05440*. This gene was present on the scaffold

with the longest alignment to the male-exclusive assembly. This scaffold has been mapped to chromosome 1, although not at a position near the region in which the sex locus is known to be. Male and female reads mapping to this gene were visualised using Integrative Genomics Viewer[446] to further assess its suitability as a candidate for the sex locus (Figure 4.3). Some features of the read alignments were consistent with what would be expected for the sex locus. For example, there were regions of the gene lacking female coverage. In the final two exons, there appeared to be two haplotypes in males but one in females, consistent with what would be expected if M and m alleles showed sequence variation. However, in other areas of the gene there were two haplotypes in females. Furthermore, the female reads in different regions of the gene appeared not to be linked by pairing, inconsistent with lack of female coverage representing regions deleted in females.

4,515 (out of 11,291) contigs in the male-exclusive assembly did not align at all to the reference assembly. A BLASTx search of these contigs against nr produced 8,851 hits for 65 contigs, against 4,873 genes. Many hits were proteins related to mobile elements. There were also many hypothetical proteins. The sequences yielding these hits are possible candidates for the sex locus, but again there were too many for experimental testing.

4.3.1.5. Searching for Nix

Following the completion of the work described above, the sex locus was determined by another group to be the gene *Nix*[104]. This gene was not present in the published assembly, but it was possible that it might have been present in the male assembly or the male-exclusive assembly generated here. However, BLAST searches using the sequence of *Nix* as a query failed to reveal an alignment corresponding to the whole gene. There were some shorter significant alignments, but the longest of these was only 107 bp, a relatively small fraction of the full gene length of 985 bp.

Figure 4.3. Screenshot of *AAEL05440* visualised in Integrative Genomics Viewer. The top set of reads is from female samples; the bottom set of reads is from male samples. Coloured reads are improperly paired. Coloured lines in the overall coverage graphs indicate differences from the reference sequence. The structure of the gene is shown in the bottom panel. Thicker blocks represent exons.

4.3.2. Identification of *Ae. aegypti* and *C. capitata* testis-specifically expressed or spliced genes

4.3.2.1. Sequencing and read alignment

RNA sequencing was performed on RNA extracted from two *Ae. aegypti*, four *C. capitata* and three *D. melanogaster* staged testis samples, an *Ae. aegypti* gonadectomised male sample, and a *C. capitata* ovary sample. In total 420,411,150 reads of 73 nt were generated, corresponding to 30.7 Gb of data. After trimming and quality filtering a total of 356,634,803 reads were used for genome alignment with TopHat2 (v2.0.9)[336]. 89.1 % of the *Ae. aegypti* reads mapped to the *Ae. aegypti* genome, 89.8 % of the *C. capitata* reads mapped to the *C. capitata* genome, and 97.0 % of the *D. melanogaster* reads mapped to the *D. melanogaster* genome.

Data from *C. capitata* female, *Ae. aegypti* female and ovary, and *D. melanogaster* female, ovary and male head samples from other experiments were downloaded from the SRA[435]. The *Ae. aegypti* female sample was gonadectomised; an ovary sample was therefore used in addition so that data from all female tissues were present. The *C. capitata* and *D. melanogaster* female samples were not gonadectomised, but ovary samples were still included in the analysis, as many genes expressed in the testis could potentially also be expressed in the ovary, and their detection may be impeded by the large amount of other tissue in a whole female sample. The *Ae. aegypti* ovary and female samples were from recently fed females (~24 hours post-blood meal), as these should include transcripts expressed during oogenesis, thus enabling elimination of genes expressed in both male and female gametogenesis. As it was available, a *D. melanogaster* head sample rather than a gonadectomised male sample was included as there are some genes that are expressed exclusively in the head and testis, but few genes expressed in testis, elsewhere in males, but not in females, and head is likely to be a cleaner control sample than gonadectomised male, which may contain traces of testis (Prof. Helen White-Cooper, personal

communication). Of these reads, 95.0 % of the *Ae. aegypti* reads mapped to the *Ae. aegypti* genome, 69.2 % of the *C. capitata* reads mapped to the *C. capitata* genome, and 95.1 % of the *D. melanogaster* reads mapped to the *D. melanogaster* genome.

4.3.2.2. Identification of candidate testis-specifically expressed genes

Candidate testis-specifically expressed genes were identified from the total set of predicted genes by running a custom Python script on the output of the standard TopHat-Cufflinks-Cuffdiff RNA-seq analysis pipeline, and applying various filtering steps (described below) to maximise sensitivity whilst removing unsuitable genes and minimising false positives (Table 4.2).

Species	Predicted genes	Predicted testis-specifically expressed genes	Predicted testis-specifically expressed early genes
Ae. aegypti	26,401	388	57
C. capitata	24,733	667	68

Table 4.2. Identification of candidate testis-specifically expressed genes. The second column refers to all genes predicted by Cufflinks from all samples used in analysis, including non-coding RNA genes. The last two columns refer to genes identified after application of the filters described in the main text.

An expression level of 10 FPKM in the early sample for *Ae. aegypti* and the early spermatocytes sample for *C. capitata* was chosen as a threshold for candidates. A threshold was set as predicted genes with low expression are more likely to be false positives, and also regulatory elements associated with relatively strong expression are desired for use in synthetic constructs; 10 FPKM was chosen as it is the boundary between low and moderate expression for *D. melanogaster* RNA-seq data on FlyBase[333]. The threshold for expression in samples other than testis (gonadectomised male, ovary and female) was not set at zero, to allow for some noise in the data, but rather at 1 FPKM, based on quantification of the *D. melanogaster* genes *can*, *comr*, *nht* and *Taf12L*, which are supposedly testis-specific[323] (Table 4.3).

	Early spermatocytes	Late spermatocytes	Spermatids	Female	Ovary	Male head
can	44.22	38.78	27.98	0.03	0.01	0.02
comr	17.39	11.13	2.51	0.00	0.00	0.00
nht	86.97	55.93	9.32	0.00	0.00	0.00
Taf12L	329.92	223.67	53.68	0.00	0.97	0.79

Table 4.3. Expression levels (FPKM) of *D. melanogaster* supposedly testis-specifically expressed genes in testis and non-testis tissues.

Many potential candidates appeared to be short non-coding RNAs. Quantification of short non-coding RNAs is likely to be inaccurate in a protocol using polyA selection. Therefore the only genes taken forward for further analysis were those that either coincided with a locus already annotated as a protein-coding gene, or novel predicted genes that were over 1 Kb in length.

After application of the filtering steps above, predicted testis-specifically expressed genes with higher expression in early spermatogenesis than in late spermatogenesis were identified (Table 4.2). For each species, the top ten candidates in order of expression level in the earliest testis sample were taken forward for experimental testing. Genes encoding proteins associated with transposable elements were excluded, as there are likely to be multiple copies of these in the genome, and it would be difficult to design PCR primers that would target only one. For *Ae. aegypti*, one additional candidate was also taken forward, as a homologue of the gene was identified as a candidate in *C. capitata*; candidates that are conserved between species may simplify construct generation in different species. The candidate genes tested experimentally are shown in Table 4.4 and Table 4.5.

Annotated ID	Annotated product
Novel gene 1	
AAEL000463	hypothetical protein
AAEL001333	protein arginine n-methyltransferase 1
AAEL003021	hypothetical protein
AAEL004293	conserved hypothetical protein
AAEL006665	histone H1, putative
AAEL009267	hypothetical protein
AAEL010268	hypothetical protein
AAEL011838	conserved hypothetical protein
AAEL012065	cytoplasmic polyadenylation element binding protein (cpeb)
AAEL012239	testis development protein nyd-sp29

Table 4.4. Candidate *Ae. aegypti* testis-specifically expressed genes tested experimentally.

Annotated ID	Annotated product
LOC101449084	uncharacterised
LOC101449780	malate dehydrogenase, mitochondrial-like
LOC101451111	protein I'm not dead yet-like
LOC101451785	protein lifeguard 1-like
LOC101454909	bestrophin-2-like
LOC101455161	ATP synthase subunit beta, mitochondrial-like
LOC101457895	uncharacterised
LOC101459316	zinc finger protein 624-like
LOC101459689	phosphate carrier protein, mitochondrial-like
LOC101462854	malate dehydrogenase, cytoplasmic-like

Table 4.5. Candidate *C. capitata* testis-specifically expressed genes tested experimentally.

4.3.2.3. Experimental testing of candidate testis-specifically expressed genes

4.3.2.3.1. RT-PCR

Reverse transcriptase PCR (RT-PCR) for the selected candidates was performed on total RNA derived from testis, gonadectomised male, ovary, and gonadectomised female samples, to confirm that the candidates were testis-specifically expressed in adults. For some candidates, the RT-PCR results suggested that there was also expression of the gene in other tissues, mostly ovary, and one candidate failed to produce a positive result in the

testis sample. However, the results supported the prediction of testis-specific expression for several candidates (Figure 4.4 and Figure 4.5), discussed below.

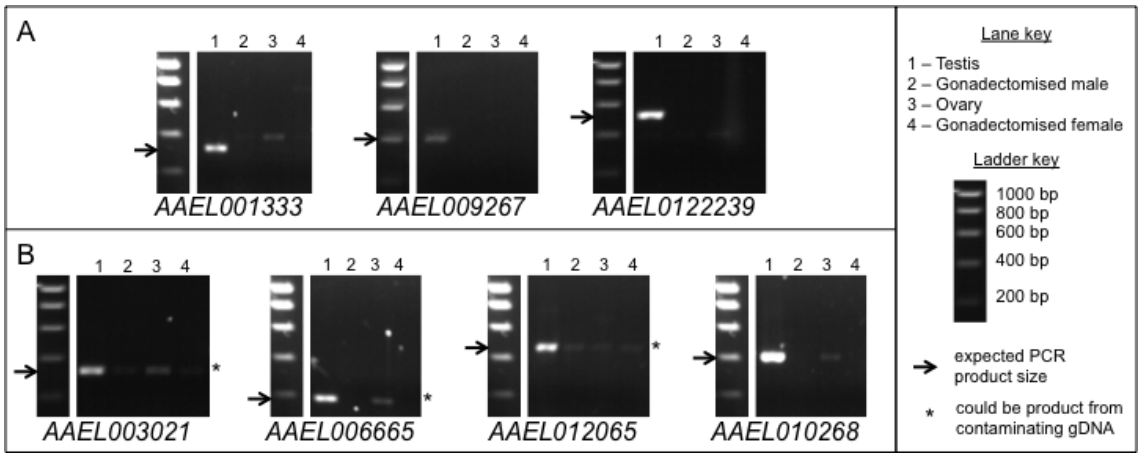

Figure 4.4. Gels showing PCR results for *Ae. aegypti* expression candidates. (A) Candidates for which no band of the expected size for the testis sample could be seen in non-testis samples. (B) Candidates for which a band of the expected size for the testis sample could be seen in a non-testis sample, but it was faint, and in the cases indicated by asterisks, could have resulted from contaminating gDNA. Expected PCR product sizes are indicated with arrows. In some cases bands of other sizes are of the expected size for products amplified from contaminating gDNA. Other bands of unexpected sizes may represent isoforms that were not predicted, or non-specific amplification.

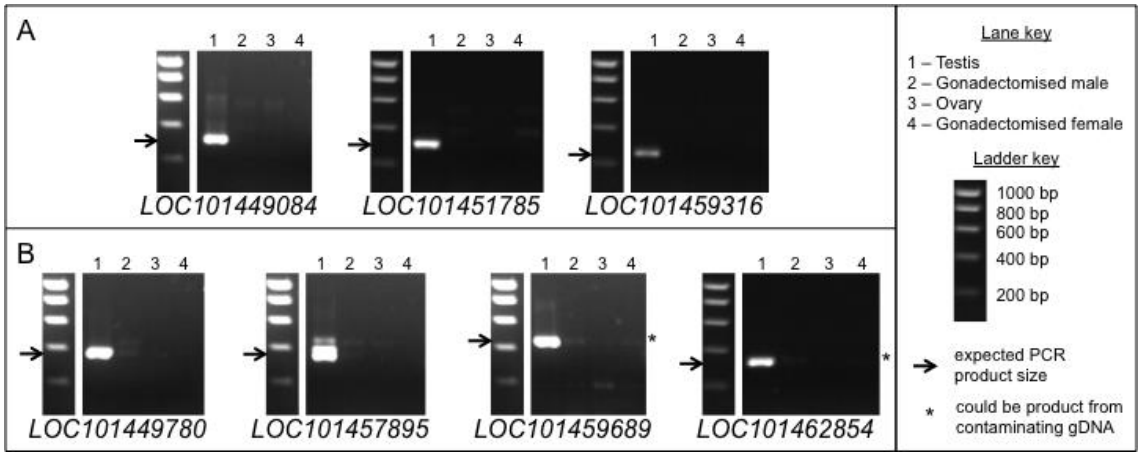

Figure 4.5. Gels showing PCR results for *C. capitata* expression candidates. (A) Candidates for which no band of the expected size for the testis sample could be seen in non-testis samples. (B) Candidates for which a band of the expected size for the testis sample could be seen in a non-testis sample, but it was faint, and in the cases indicated by asterisks, could have resulted from contaminating gDNA. Expected PCR product sizes are indicated with arrows. In some cases bands of other sizes are of the expected size for products amplified from contaminating gDNA. Other bands of unexpected sizes may represent isoforms that were not predicted, or non-specific amplification.

Three candidate *Ae. aegypti* genes (Figure 4.4A), corresponding to the annotated loci *AAEL001333*, *AAEL009267* and *AAEL0122239*, and three candidate *C. capitata* genes (Figure 4.5A), corresponding to the annotated loci *LOC101449084*, *LOC101451785* and *LOC101459316*, displayed the expected outcome of RT-PCR amplification from testis and no amplification from other samples, and were taken forward for further testing. Four additional candidate *Ae. aegypti* genes (Figure 4.4B), corresponding to the annotated loci *AAEL003021*, *AAEL006665*, *AAEL010265* and *AAEL010268*, and four additional candidate *C. capitata* genes (Figure 4.5B), corresponding to the annotated loci *LOC101449780*, *LOC101457895*, *LOC101459689* and *LOC101462854*, were also taken forward despite some amplification in non-testis samples. In these cases the quantity of product from the non-testis samples was low, and in some cases the product could have resulted from amplification of contaminating gDNA.

4.3.2.3.2. qRT-PCR

Quantitative RT-PCR (qRT-PCR) for the candidate genes taken forward for further testing was performed on staged testis samples (early and late samples for *Ae. aegypti*, spermatocytes and spermatids samples for *C. capitata*), to confirm that the candidate genes displayed the desired expression pattern of higher expression early in spermatogenesis (Figure 4.6 and Figure 4.7). Gonadectomised male, ovary, and gonadectomised female samples were also used in the qRT-PCR to quantify the level of expression in these tissues, if any. Candidates with a low level of non-testis expression may still be usable for restricting expression to the testis, particularly in combination with other strategies, such as use of testis-specific splicing.

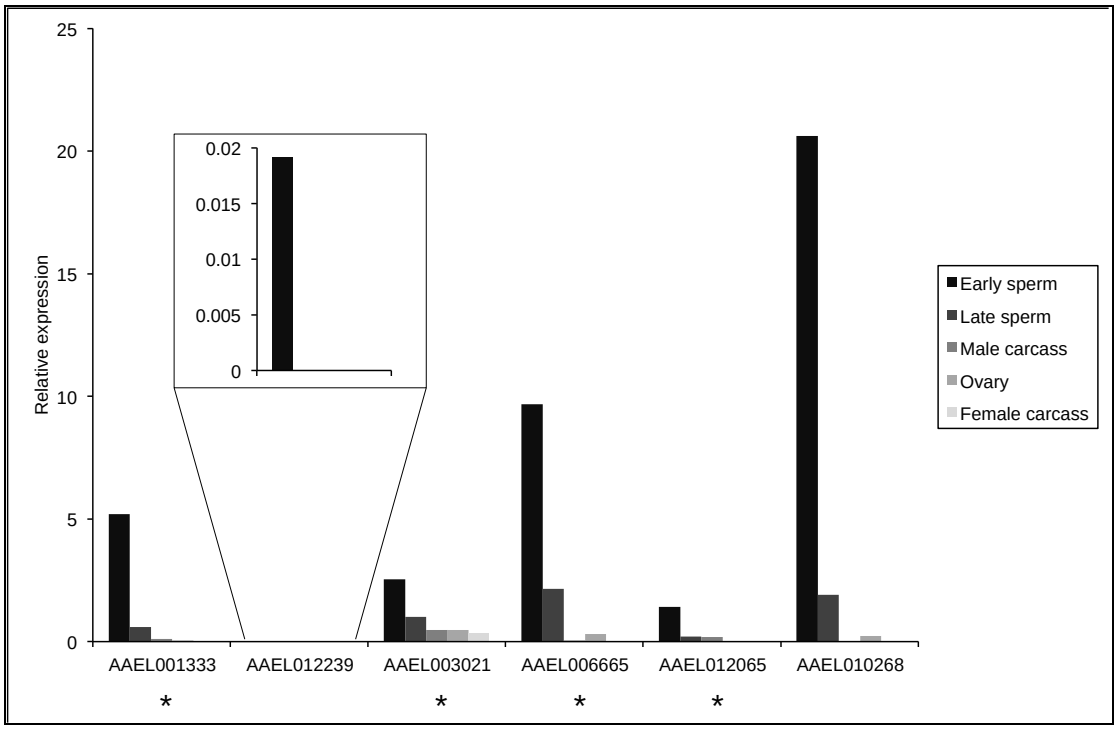

Figure 4.6. Relative expression levels in different tissues for *Ae. aegypti* expression candidates, determined using qRT-PCR. Results for *AAEL012239* are shown inset, as the expression level for this gene was too low to view at the same scale as for the other genes. * Primers could also have amplified from gDNA, so apparent low expression in non-testis tissues could be a result of gDNA contamination.

The timing was confirmed for all *Ae. aegypti* candidates (Figure 4.6) except *AAEL009267*, for which the qPCR failed, and for four of the *C. capitata* candidates (Figure 4.7). For the other three *C. capitata* candidates, *LOC101449084*, *LOC101457895* and *LOC101459689*, no expression was detected in spermatocytes. The results for all the *Ae. aegypti* candidates except *AAEL012239* suggested some expression in non-testis tissues, but this was at a low level compared to that in testis, and in four of the five cases amplification could have resulted from contaminating gDNA. For all the *C. capitata* candidates, no expression was detected in non-testis samples.

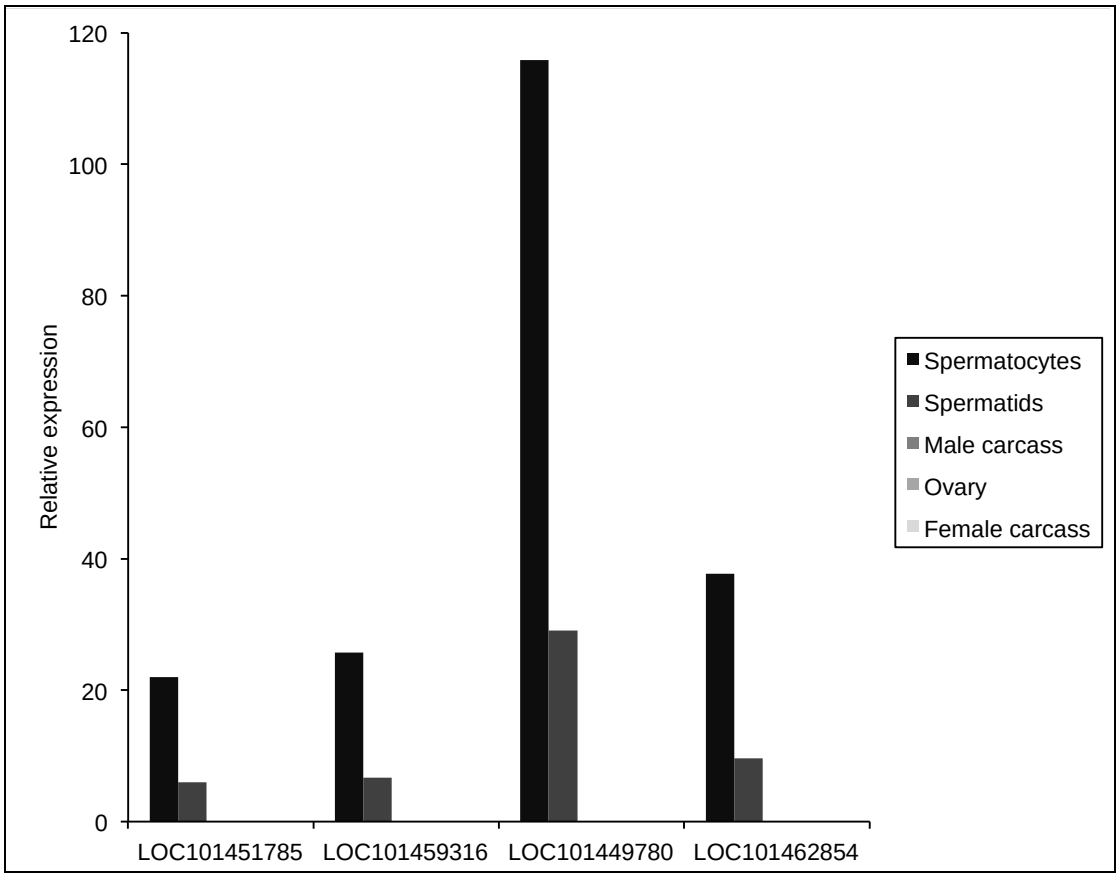

Figure 4.7. Expression levels in different tissues for *C. capitata* expression candidates, determined using qRT-PCR.

4.3.2.4. Identification of candidate testis-specifically spliced genes

Similarly to the candidate testis-specifically expressed genes, analysis was performed on RNA-seq data from two staged testis samples in *Ae. aegypti* and four staged testis samples in *C. capitata*, along with gonadectomised male, ovary and female samples to identify genes with testis-specific splice forms.

Candidate testis-specifically spliced genes were identified from the total set of predicted genes by running a custom Python script on the output of the standard TopHat-Cufflinks-Cuffdiff RNA-seq analysis pipeline, and applying various filtering steps (described below) to maximise sensitivity whilst removing unsuitable genes and minimising false positives (Table 4.6). These filtering steps may exclude some valid genes, but for the intended

downstream application it is not necessary to identify all testis-specifically spliced genes; it was more important to minimise false positives.

Species	Predicted genes	Predicted testis-specifically spliced genes
Ae. aegypti	26401	27
C. capitata	24733	33

Table 4.6. Identification of candidate testis-specifically spliced genes. The second column refers to all genes predicted by Cufflinks from all samples, including non-coding RNA genes. The last column refers to genes identified after application of the filters described in the text.

An expression level of 10 FPKM (in the early sample for *Ae. aegypti* and the early spermatocytes sample for *C. capitata*) was chosen as a threshold for the predicted testis-specific splice forms, using the same rationale as discussed for the candidate testis-specifically expressed genes. The threshold for expression of predicted testis-specific splice forms in tissues other than testis was not set at zero, to allow for some noise in the data, but rather at 0.4 FPKM, based on quantification of transcripts from the *D. melanogaster* genes *achi* and *vis* that are supposedly testis-specifically spliced[425]. It was also required that at least one other splice form of the gene was expressed in at least one other sample (gonadectomised male, ovary or female) at a level of 10 FPKM or above, to distinguish testis-specific splicing from testis-specific expression.

In addition to the above expression thresholds, a threshold for exon-exon junction coverage was set to minimise false positives; only introns with more than 10 reads spanning the exon-exon junction were taken forward. False positives may also arise due to low coverage in a particular sample, causing incorrect assembly of a transcript in this sample, for example with a few nucleotides missing at the end, and giving the appearance of alternative splicing. To minimise this source of error, the only introns taken forward were those differing by more than 20 bp at one end at least from introns in other

transcripts from the same gene. Finally, only candidates for which the predicted testis-specific intron was within an annotated gene were taken forward, to avoid false positives that are in fact intergenic regions but predicted as introns due to incorrect merging of transcripts during assembly.

Experimental validation of testis-specific splicing required distinguishing between splice forms using RT-PCR. For candidates in which the alternative splicing was such that primers could be designed common to both predicted testis-specific and other splice forms, there were sufficiently few candidates that all of them were taken forward for experimental testing. There were further candidates for which at least one primer to detect the predicted testis-specific splice form could not be designed outside a predicted testis-specific exon, thus requiring separate primers to confirm expression of other splice forms in other tissues; for each species the top five of these candidates in order of ascending intron size were taken forward for experimental validation. For *C. capitata*, one additional candidate was also taken forward, as a homologue of the gene was identified as a candidate in *Ae. aegypti*; as mentioned above, candidates that are conserved between species may simplify construct generation in different species. The candidate genes tested experimentally are shown in Table 4.7 and Table 4.8.

Annotated ID	Annotated product
AAEL000028	Clip-Domain Serine Protease family B.
AAEL001898	conserved hypothetical protein
AAEL001916	eukaryotic translation initiation factor 4e
AAEL002083	DEAD box ATP-dependent RNA helicase
AAEL005651	ethanolamine-phosphate cytidylyltransferase
AAEL005935	hypothetical protein
AAEL007258	hypothetical protein
AAEL008110 *	centrosomin (arrow protein)
AAEL008428	conserved hypothetical protein
AAEL010509	bridging integrator
AAEL011105	adducin
AAEL011153	hypothetical protein
AAEL011968	conserved hypothetical protein
AAEL012262	conserved hypothetical protein
AAEL013571	synaptobrevin
AAEL014957	conserved hypothetical protein
AAEL017395	
AAEL018039	
AAEL018211 *	
AAEL018350	

Table 4.7. Candidate *Ae. aegypti* testis-specifically spliced genes tested experimentally.
* These genes contained two candidate testis-specifically spliced introns.

Annotated ID	Annotated product
LOC101448609	transport and Golgi organization protein 11-like
LOC101449153 **	centrosomin-like
LOC101449503	cytochrome c oxidase subunit 7A-related protein, mitochondrial-like
LOC101450565	IQ domain-containing protein D-like
LOC101450641	actin, indirect flight muscle-like
LOC101452861	uncharacterized LOC101452861
LOC101452984	E3 ubiquitin-protein ligase RNF181 homolog
LOC101453609	eukaryotic translation initiation factor 5-like
LOC101457260	UPF0468 protein CG5343-like
LOC101459514 *	macrophage erythroblast attacher-like
LOC101461218	ADIPOR-like receptor CG5315-like

Table 4.8. Candidate *C. capitata* testis-specifically spliced genes tested experimentally.
* This gene contained two candidate testis-specifically spliced introns. ** This gene contained three candidate testis-specifically spliced introns.

4.3.2.5. Experimental testing of candidate testis-specifically spliced genes

4.3.2.5.1. RT-PCR

RT-PCR for the selected candidates was performed on testis, gonadectomised male, ovary, and gonadectomised female samples, to confirm that the candidates were testis-specifically spliced. For cases in which it was possible, primers were designed common to both predicted testis-specific and other splice forms, but with different expected product sizes to allow discrimination of splice forms. It is possible that RT-PCR with this type of primers will give a false negative result in non-testis tissues, as the predicted testis-specific splice form may evade detectable amplification if it is only a minor splice form and/or the amplicon is larger than those from other splice forms and so amplifies less efficiently. Therefore PCR was also performed with primers spanning the predicted testis-specific exon-exon junction, which should amplify only the predicted testis-specific splice form. For cases in which design of primers common to testis-specific and other splice forms was not possible, PCR with primers spanning the predicted testis-specific exon-exon junction was performed, in addition to PCR with additional primers designed to amplify other splice forms to confirm expression of the gene in other tissues.

The PCR results varied between candidate genes. For some candidates the predicted testis-specific splice form was not detected, for others it was detected in samples other than testis, and for others no splice forms at all were detected in non-testis samples, suggesting that the gene is testis-specifically expressed rather than differentially spliced. However, the results supported the prediction of testis-specific splicing for some candidates (Figure 4.8 and Figure 4.9), discussed below.

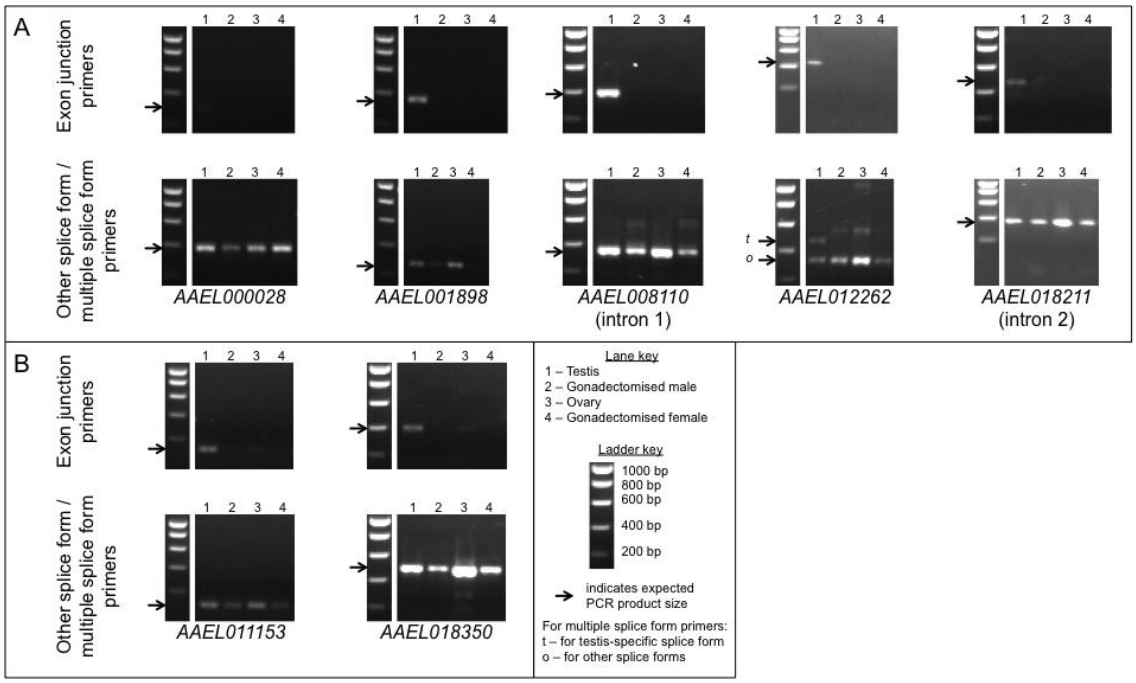

Figure 4.8. Gels showing PCR results for *Ae. aegypti* splicing candidates. (A) Candidates for which no band of the expected size for the predicted testis-specific splice form could be seen in non-testis samples. (B) Candidates for which a band of the expected size for the predicted testis-specific splice form could be seen in a non-testis sample, but it was only faint. For each candidate, the top image shows the results of PCR with primers designed to span the predicted testis-specific exon junction, to confirm the testis-specificity of the predicted testis-specific splice form. The bottom image shows the results of PCR with primers designed to amplify either another non-testis-specific splice form of the gene, or multiple splice forms including both the testis-specific splice form and non-testis-specific splice forms, to confirm expression of the gene in non-testis tissues. Expected PCR product sizes are indicated with arrows. Bands of unexpected sizes may represent other splice forms that were not predicted, or non-specific amplification.

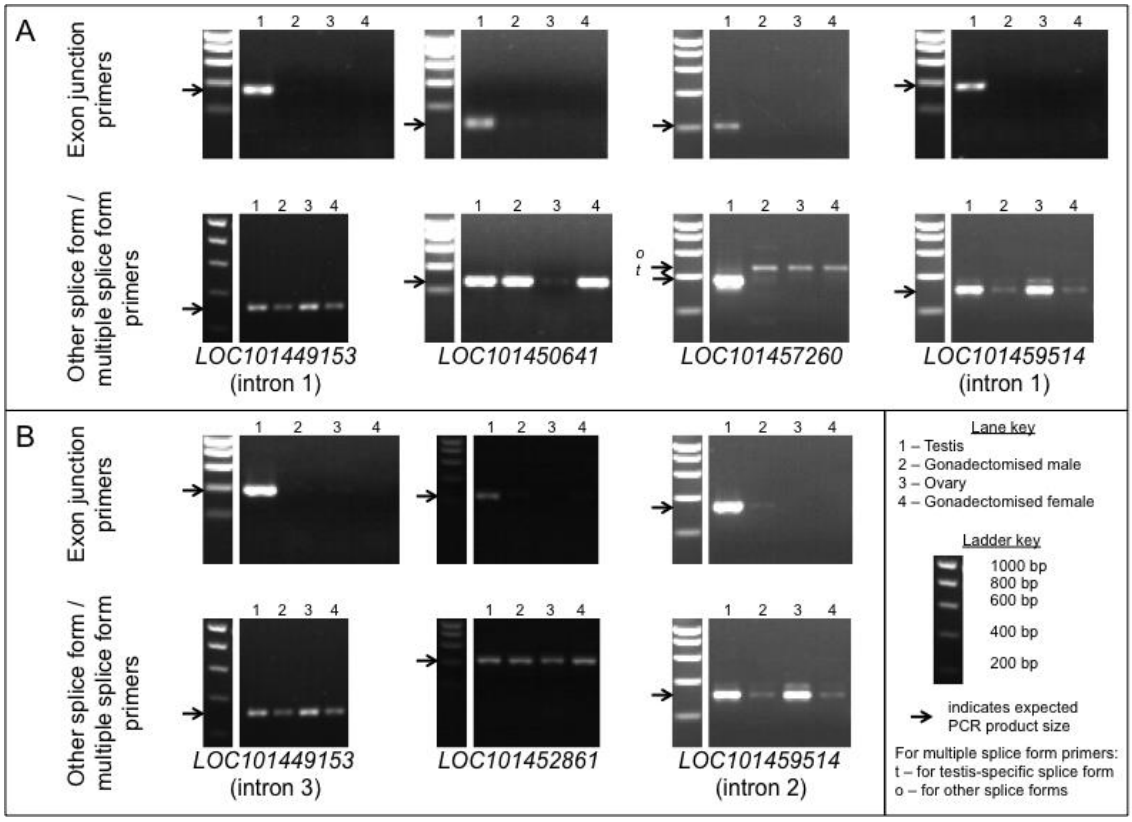

Figure 4.9. Gels showing PCR results for *C. capitata* splicing candidates. (A) Candidates for which no band of the expected size for the predicted testis-specific splice form could be seen in non-testis samples. (B) Candidates for which a band of the expected size for the predicted testis-specific splice form could be seen in a non-testis sample, but only a faint band. For each candidate, the top image shows the results of PCR with primers designed to span the predicted testis-specific exon junction, to confirm the testis-specificity of the predicted testis-specific splice form. The bottom image shows the results of PCR with primers designed to amplify either another non-testis-specific splice form of the gene, or multiple splice forms including both the testis-specific splice form and non-testis-specific splice forms, to confirm expression of the gene in non-testis tissues. Expected PCR product sizes are indicated with arrows. Bands of unexpected sizes may represent other splice forms that were not predicted, or non-specific amplification.

Five *Ae. aegypti* candidate introns (Figure 4.8A), within the annotated loci *AAEL000028*, *AAEL001898*, *AAEL008110*, *AAEL012262*, and *AAEL018211*, and four *C. capitata* candidate introns (Figure 4.9A), within the annotated loci *LOC101449153*, *LOC101450641*, *LOC101457260* and *LOC101459514*, displayed the expected outcome of a positive result for the predicted testis-specific splice form in testis, and a negative result for this splice form but positive result for other splice forms in other samples. These nine candidates were taken forward for further testing. Two additional *Ae. aegypti* candidate introns (Figure 4.8B), within the annotated loci *AAEL011153* and *AAEL018350*, and three additional *C.*

capitata candidate introns (Figure 4.9B), within the annotated loci *LOC101449153*, *LOC101452861* and *LOC101459514*, were also taken forward despite positive PCR results for the predicted testis-specific splice form in non-testis samples, as the quantity of product from the non-testis samples was low, and the product was only present for PCR with primers spanning the predicted testis-specific exon-exon junction. In these cases, expression of the predicted testis-specific splice form in non-testis tissues could be at a low level relative to the expression of other splice forms, and so the candidate could potentially still be useful for the intended application.

4.3.2.5.2. qRT-PCR

The suitability of a testis-specific intron for use in a synthetic construct as discussed above will be affected by the proportions of different splice forms for the corresponding gene in the testis. If used to direct testis-specific expression of a coding region, the higher the proportion of transcripts spliced in the testis-specific form, the more desired coding transcripts produced. In order to determine this proportion for the candidates taken forward for further testing, qRT- PCR was performed (Figure 4.10 and Figure 4.11). Gonadectomised male, ovary, and gonadectomised female samples were also used in the qRT-PCR to determine the expression level of the predicted testis-specific splice form in these tissues, if any, relative to the expression level of other splice forms. Candidates with a low level of non-testis expression of the predicted testis-specific splice form relative to other splice forms may still be usable for synthetic biology applications, particularly in combination with other strategies, such as use of testis-specific regulatory regions, for restricting expression to the testis.

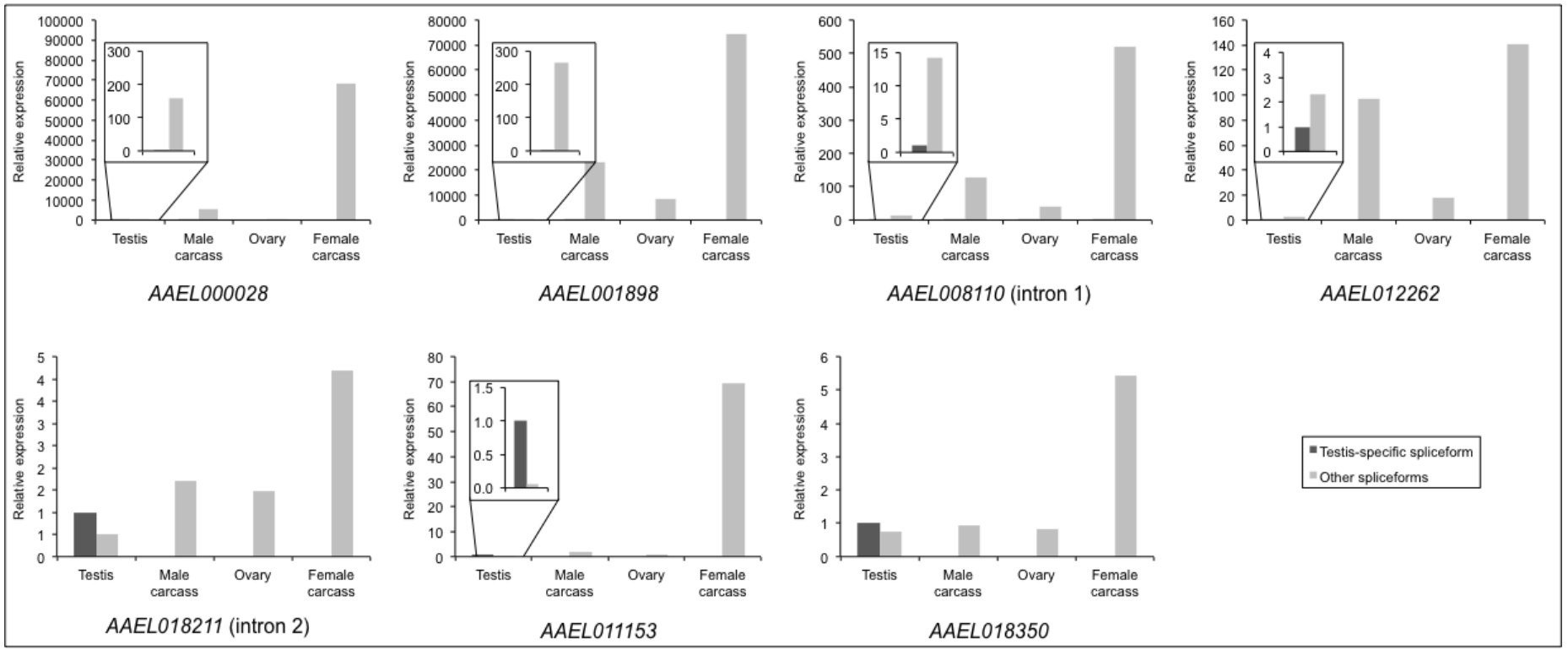

Figure 4.10. Relative expression levels in different tissues for predicted testis-specific and other splice forms of *Ae. aegypti* splicing candidates, determined using qRT-PCR. Where expression levels in the testis are too low to view at the same scale as for the other splice forms, results for testis are shown inset. The relative expression value of the testis-specific splice form is set at 1 in all cases.

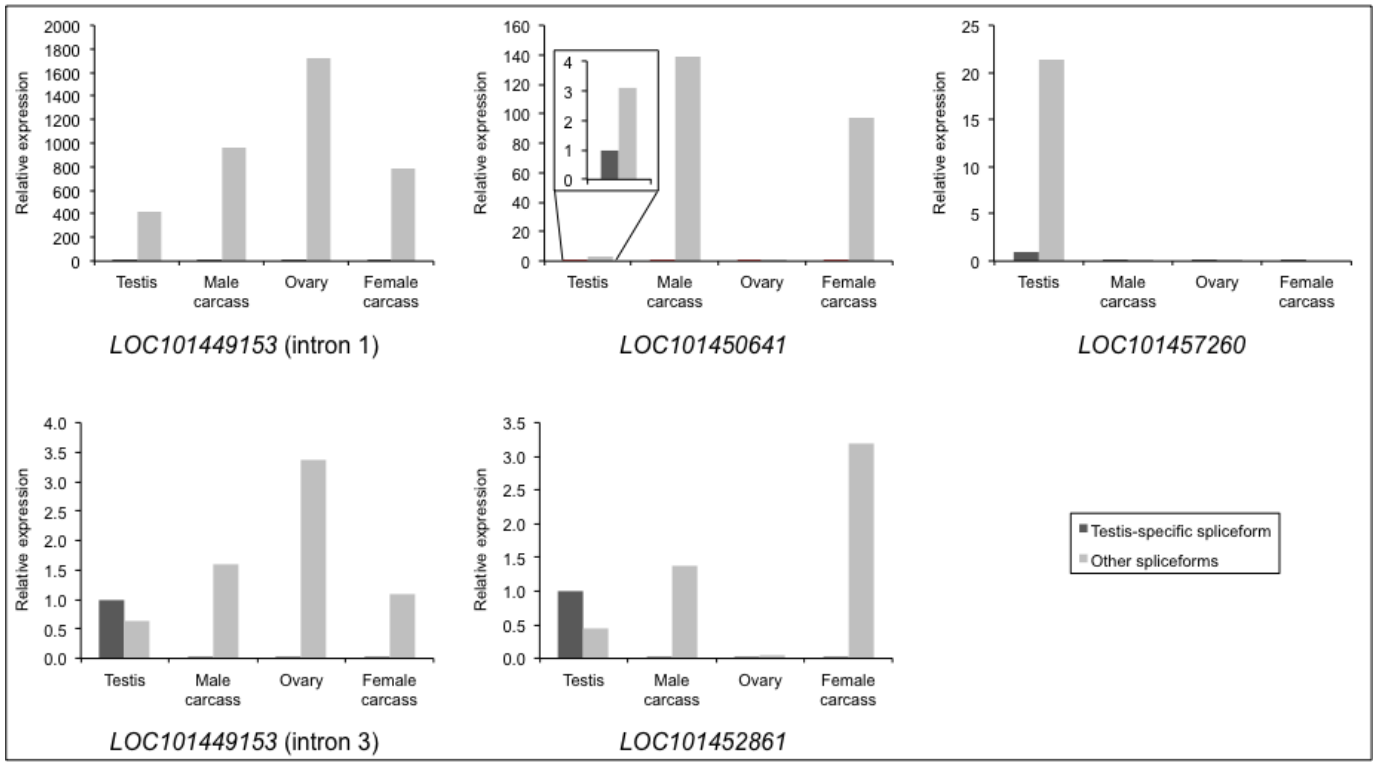

Figure 4.11. Relative expression levels in different tissues for predicted testis-specific and other splice forms of *C. capitata* splicing candidates, determined using qRT-PCR. Where expression levels in the testis are too low to view at the same scale as for the other splice forms, results for testis are shown inset. The relative expression value of the testis-specific splice form is set at 1 in all cases.

The qRT-PCR for the *C. capitata* candidate introns within the annotated locus *LOC101459514* failed to produce meaningful results, with calculations suggesting negative expression of splice forms, so these introns were excluded. Based on the qRT-PCR results for the other candidates, the estimated ratio of testis-specific splice form to other splice forms in the testis ranged from 0.4 % to 1,000 % in *Ae. aegypti* (Figure 4.10) and 5 % to 200 % in *C. capitata* (Figure 4.11). Candidates at the lower ends of these ranges are unlikely to be suitable for use in a synthetic construct. For example, the results suggest that for *AAEL001898*, only 0.4 % of mature transcripts in the testis would retain the intron, and thus only 0.4 % of transcripts would be of the desired form if this intron were used to direct testis-specific expression of a coding region. However, candidates at the higher ends of the ranges are more likely to be suitable, and will be taken forward for testing in synthetic constructs. In some cases the qRT-PCR results suggested expression of the testis-specific splice form in non-testis samples, but this was mostly at a very low level (< 1 %) relative to the expression of other splice forms in these tissues. For the *C. capitata* candidates *LOC10450641* and *LOC101457260* the results suggested that 17-100 % of the splice forms in non-testis samples were actually the predicted testis-specific splice form. However, the expression of all splice forms in these non-testis samples was low compared to expression in the testis (too low to be visible in the graphs in Figure 4.11), so relative errors in quantification are likely to be higher.

4.3.2.6. Interspecies comparison

To determine whether any of the candidates we identified were conserved between species, tBLASTx searches between *Ae. aegypti* and *C. capitata* candidates were performed, using the candidate sequences from one species as queries and all transcripts predicted by Cufflinks from the other species as a database. The *D. melanogaster* dataset was also used as a database, to provide further confidence in conservation, and also because more supporting information is available on *D. melanogaster* genes.

These BLAST searches revealed that the *Ae. aegypti* testis-specifically expressed candidate corresponding to the annotated locus *AAEL009267* and the *C. capitata* testis-specifically expressed candidate corresponding to the annotated locus *LOC101459316* are homologous, and furthermore both show homology to a *D. melanogaster* gene, *CG7691*, that was also identified as testis-specifically expressed, with higher expression early in spermatogenesis. The *Ae. aegypti* testis-specifically spliced candidate corresponding to the annotated locus *AAEL008110* (centrosomin) and the *C. capitata* testis-specifically spliced candidate corresponding to the annotated locus *LOC101449153* (centrosomin-like) are homologous, and both show homology to the *D. melanogaster* centrosomin gene, which is known to have a role in spermatogenesis and display testis-specific splicing in this species[447].

4.3.2.7. Promoter prediction

In theory it would be possible to build a construct utilising a promoter to restrict expression without exact knowledge of the span of the promoter, by just including a sufficient length of upstream sequence (likely around 2 Kb) to cover the promoter. However, the larger a construct the less the chance of a successful transformation, so it is desirable to include only the promoter if possible. Thus, promoter prediction was attempted for candidate testis-specifically expressed genes.

Two approaches were used. Firstly, important promoter elements should be conserved for homologous genes between species, so a comparison was made between upstream sequences of candidate genes and their top two matching genes from a BLAST search against other insect species. Secondly, important promoter elements may be conserved for genes with a similar expression profile within species, so upstream sequences for candidate genes within species were compared. For the latter approach, only annotated genes with 5'UTRs were used, as otherwise the supposed upstream sequence may include

5'UTR, making it more difficult to find conserved motifs in the actual upstream sequence. Only *Ae. aegypti* genes were used, as it was possible to extract multiple upstream sequences at once from VectorBase[427]; this is not possible for *C. capitata*.

Three programs were used to identify transcription factor binding sites – PROMO[439], rVista 2.0[440] and Melina II[441]. PROMO identifies known sites, and was set to identify only sites present in all sequences (but not necessarily in the same location), restricted to *Dipteran* factors and sites. rVista 2.0 also identifies known binding sites, but can be set to show only sites that are conserved in the same location between sequences; this setting was used, restricted to insect sites, for the target gene and its top BLAST hit from another species. Melina II can identify *de novo* sites, and was used as there may be sites in *Ae. aegypti* and *C. capitata* that are unknown.

Neither approach for promoter prediction revealed a clearly defined promoter region for any of the genes tested using any of the software. An example of an interspecies comparison using PROMO is shown in Figure 4.12, illustrating the lack of consensus between binding sites. On further analysis it was discovered that most of these genes did not have a 5'UTR predicted, and so the 'upstream' sequence likely also contains 5'UTR. Given the variability in 5'UTR length, the position of the transcription start site (TSS) is hard to determine computationally, which makes promoter identification difficult.

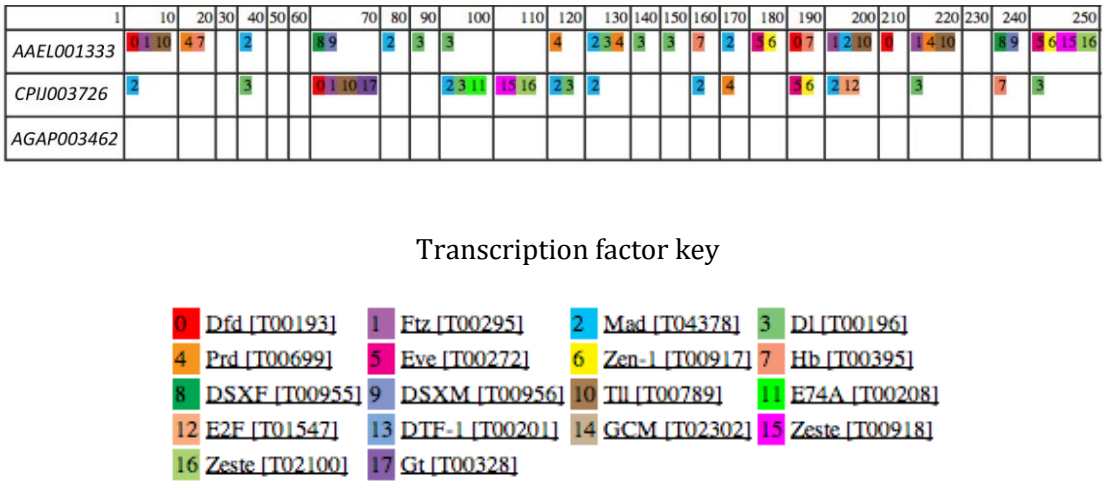

Figure 4.12. Transcription factor sites in upstream sequence of *AAEL001333* and homologues in other species. Sites were identified using PROMO[439].

4.4. Discussion

4.4.1. Searching for *Ae. aegypti* sex locus

Attempts were made to identify the *Ae. aegypti* sex locus by comparing the genome sequences of males and females. Unfortunately, the sex locus was not identified from this work; it has now been determined as the gene *Nix*[104].

Comparisons of coverage between male and female samples initially seemed a promising approach. Identification of Y-linked genes based on a reduced sequencing coverage in males relative to females has been proposed previously[406], and was adopted by Hall *et al.* in a strategy termed the chromosome quotient (CQ) method[448], which was used to identify novel Y-linked genes in *An. gambiae*[448], a male-biased gene in *Ae. aegypti*[449], and subsequently the male-determining M factor itself[104].

However, all approaches here identified too many candidates for functional testing. This is likely in part due to the use of inbred individuals from one strain of *Ae. aegypti*, meaning that the comparative approaches taken here would also identify regions linked to the sex locus in these individuals. Adding sequencing data from other strains would aid in

eliminating these linked regions from identification; differences between males and females conserved across different strains would have been much stronger candidates. Identification of the male-determining M factor utilised data from two different *Ae. aegypti* strains[104].

One way to narrow down candidates would have been to determine whether their physical location was consistent with the known region of the sex locus. Previous work used fluorescent in situ hybridisation (FISH) to confirm whether candidates mapped to the correct location[449,104]. Analysis of transcriptomic profiles for candidate sequences is another way to validate candidates – Hall *et al*. used RNA-seq data to restrict candidates for the M allele to those expressed in early embryos and not in females[104].

Another limitation with the methods here is that they focused too heavily on sequences and genes present in the published *Ae. aegypti* genome assembly (AaegL2). This assembly was generated from data derived from both male and female gDNA[442], meaning that the M allele will only have a quarter of the coverage of autosomal sequences. The overall coverage was only 7.63X, so the M allele is likely to have less than 2X coverage. The sequencing was performed using Sanger technology, which has a bias against heterochromatic DNA, so the actual coverage may be even less. *Nix*, the gene determined to be the male-determining factor, was not present in the published genome assembly[104].

However, even the new assemblies generated here did not contain the full *Nix* sequence. This could be due to insufficient coverage, and/or difficulties in assembly in the region of the sex locus. An improved assembly in this region may have aided in identification of the sex locus. To this end, after the work described here a bacterial artificial chromosome (BAC) library was generated by Amplicon Express (Washington, USA) with the aim of using PacBio sequencing of BAC clones to generate an improved assembly in this region,

and to provide more information on candidate sequences – knowledge of their flanking regions may enable them to be assigned to a physical location, for example. BAC clones for sequencing were to be identified by BAC walking starting from the nearest known mapped scaffolds, or probing with candidate sex locus sequences. However, the male-determining M factor was identified before this could be completed.

Although the work here was not fruitful, hopefully identification of the sex locus by others[104] will provide a basis from which to build transgenic insect control systems targeting sex development.

4.4.2. Identification of *Ae. aegypti* and *C. capitata* testis-specifically expressed or spliced genes

RNA-seq data were used in a high-throughput approach to identify testis-specifically expressed and spliced genes in the disease vector *Ae. aegypti* and agricultural pest *C. capitata*. This genome-wide approach represents an advance on previous efforts to find regions for use in insect control constructs, which attempted to identify candidates on an individual basis, often based on distant homology to *D. melanogaster* genes. Using staged testis samples allowed sufficient time resolution to select testis-specifically expressed genes with expression levels highest in early spermatogenesis; the regulatory sequences of these genes are more likely to be suitable for bipartite synthetic constructs than those from genes more highly expressed in later stages.

The findings complement those of Akbari *et al.*[450], who identified regulatory regions specific to the female germline in *Ae. aegypti*. These regions could be used to drive ovary-specific expression in strategies such as *Medea* and UD^{MEL}, which have been shown to result in population replacement in *Drosophila*[75,77,451], while the testis-specific elements identified here could be used in alternative population replacement strategies such as sex

distortion and prevention of sperm development. This could also be achieved with the testis-specific introns identified here. These could be used on their own or in combination with testis-specific regulatory elements, or even to achieve testis-specific expression with regulatory elements active in the testis but not testis-specific – this would allow a wider choice of regulatory elements.

The genes identified also provide a choice of expression levels in a synthetic construct, with varying expression levels for the testis-specifically expressed genes and varying splice form ratios for the testis-specifically spliced genes. This may be useful as different applications utilising testis-specific expression may require different expression levels. It may also be useful that some of the genes display testis-specific expression or splicing that is consistent between *Ae. aegypti*, *C. capitata* and *D. melanogaster*. Conserved genes such as this can simplify construct generation across different species, as it is possible that the same or similar sequences may be used for multiple species.

Unfortunately, it was not possible to identify promoter regions for the candidate-testis specifically expressed genes computationally; this was complicated by lack of knowledge of the 5'UTRs. To determine accurate TSS positions and thus enable promoter prediction to be focused on an appropriate region, 5'RACE could be performed.

In addition to the results of this work, the methodology is also noteworthy. To identify testis-specific splicing, a novel computational pipeline for this type of analysis was developed. Whilst the majority of the work is performed by the pre-existing Tuxedo suite of programs, these do not produce finished analyses with regards to alternative splicing, but rather an intermediate output that requires further computation to produce user-friendly candidate lists and sequences. The novel pipeline combines this software with custom-written Python scripts to achieve this. Unlike other methods for identifying

differential splicing from RNA-seq data[452], it identifies splice forms generated by all types of splice event, not only exon skipping. The outputs are particularly tailored for subsequent experimental testing, containing intron flanking sequences along with numbered exon junction positions to facilitate PCR primer design, and alignments for all transcripts of each gene. In addition to its application here to identify testis-specific splicing, the pipeline could be applied to other sample sets, for example to identify splice forms specific to other tissues, developmental stages, disease states or external conditions. A flow chart illustrating the pipeline is shown in Figure 4.13.

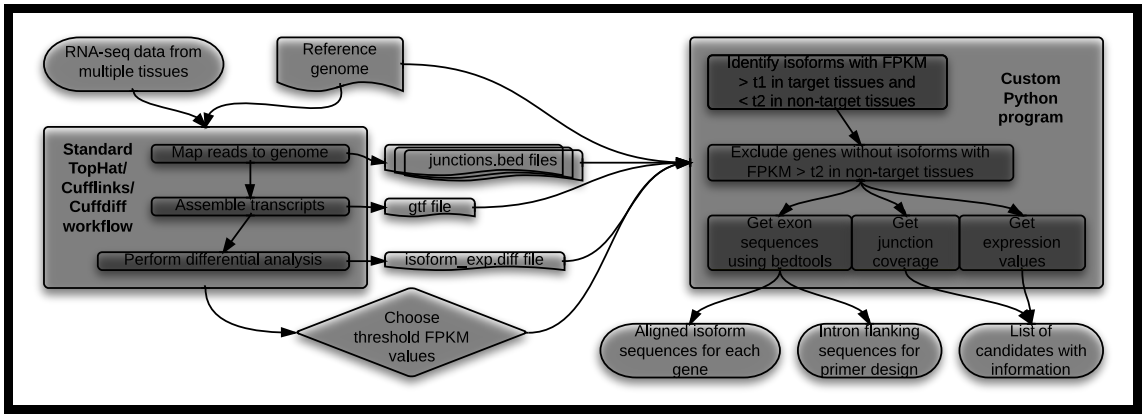

Figure 4.13. Novel computational pipeline to identify tissue-specifically expressed or spliced genes. Output of the Tuxedo suite of tools is processed with user-defined thresholds to output lists of candidates with information that allows further sorting and filtering, such as intron sizes, expression values and exon junction coverage. It also outputs sequences in a convenient format for primer design for subsequent experimental testing - intron-flanking sequences with numbered exon junctions, and alignments of all isoforms for each gene.

Returning to the results, it should be noted that are some limitations. Based on the RT-PCR and qRT-PCR results, it is possible that some of the candidates identified may not be completely testis-specifically expressed or spliced, although at the levels present it is hard to distinguish between genuine expression in other tissues and noise in the results. Depending on various factors such as the tissue, developmental stage, and effector gene action, it is possible that a small amount of off-target expression may be acceptable. Furthermore, multiple components for restricting expression to the testis could

potentially be combined, such as a testis-specific regulatory region and a testis-specific intron, resulting in greater specificity than that of the individual components.

There are also a number of other factors that could affect the success of a transgenic line. For example, evidence suggests that the size of a construct is inversely correlated with transformation efficiency using transposon-based integration systems such as *piggyBac*[453]. It may be necessary to shorten the sequences of candidate testis-specific regulatory regions or introns before their inclusion in synthetic constructs, using only the parts of core importance to the testis-specificity. Another potential issue is that candidates may be dependent on genomic context for their testis-specificity, in which case they may function suboptimally in a synthetic construct.

Furthermore, the timing of expression driven by the candidate testis-specific regulatory regions might be inappropriate. This work focused on identifying regulatory regions that are active early in spermatogenesis, as expression too late would definitely not be useful for the intended application. However, expression too early might also be a problem for some strategies, as it could for example result in the production of fewer or no sperm rather than rendering sperm non-functional. This would obviously be detrimental for a sex-distorting line, as fully functional sperm are still required in this strategy. It may also be undesirable in a strategy aiming to create sterile males, as fewer sperm may reduce the reproductive competitiveness of transgenic males relative to wild type males due to sperm competition, and their ability to suppress receptivity to further mating in wild type females. The role of sperm in suppressing female remating in *Ae. aegypti* is unclear, but in *An. gambiae*, spermless males were shown to elicit normal mating responses[83]. However, this may not translate to *Ae. aegypti*, as the mechanisms of mating responses in *An. gambiae* differ from those in other mosquitoes, including those in the *Aedes* genus[454]. In *C. capitata*, some studies found that spermless males[455], or sterile males[456,457], which transfer

fewer sperm[458,459], were less able to suppress female remating, suggesting that sperm abundance affects mating receptivity in this species. However, studies on other tetriphids have not found an affect of sperm abundance on female remating[460-462]. The authors of one of these studies[462] suggested that the mating status of the male may play a role, with sperm abundance being unimportant in multiply mated males. The applicability of this hypothesis to population control strategies involving sperm manipulation depends on the mating behavior of transgenic males in the field, currently an unknown. Sperm abundance cannot be ruled out as an important factor.

Another timing issue that may be affected by the regulatory regions used is the period for which the transgenic products persist. Initial attempts at developing a sex-distorting system in *An. gambiae* resulted instead in embryonic lethality due to persistence of the transgenic endonuclease[84]; this was addressed by engineering the endonuclease to reduce its thermostability and thus persistence[103].

Whether the candidate testis-specific regulatory regions and introns that have been identified display suitable timing profiles and other characteristics will be confirmed by functional testing in synthetic constructs. Candidates have been further tested since the work described here was carried out, and work is currently underway to generate synthetic constructs. Overall, the results here provide a useful platform from which to develop such constructs, in addition to furthering knowledge of basic insect biology.

4.4.3. Conclusions

These analyses, using genome resequencing and RNA-seq to identify candidate genes for transgenic control of pest insects, demonstrate the power of high-throughput sequencing approaches, and the ability to apply them for different goals. The aim of the genome resequencing was to identify the *Ae. aegypti* sex locus for use as a target or an effector in

synthetic constructs, while the aim of the RNA-seq was to identify testis-specifically expressed or spliced sequences for use in controlling expression of effectors in synthetic constructs.

The *Ae. aegypti* sex locus, while not identified by this work, has since been determined[104]. The work here did identify a number of testis-specifically expressed or spliced genes in *Ae. aegypti* and *C. capitata*, including some that were conserved between species. These findings provide additional tools in the rapidly developing field of insect synthetic biology, and will hopefully aid in developing novel control strategies for pest insects.

5 General discussion

5.1. Summary

Insects are responsible for transmission of several widespread and serious diseases, as well as costly agricultural losses. There are several inadequacies with current control methods, so novel techniques are being developed, in which a modified version of the insect itself acts as the control agent. The intracellular bacteria *Wolbachia* are of interest for several strategies due to their interactions with their hosts. Pathogen inhibition and CI phenotypes are of particular relevance to insect control, as well as being of interest from a basic biology perspective. In this thesis, genes that may be involved in these *Wolbachia*-host interactions have been identified and investigated, as well as genes with the potential for use in transgenic constructs for insect control.

Chapter 2 identified *Wolbachia* genes that may be involved in CI, by sequencing the genome of a *Wolbachia* strain that does not cause CI, *w*Au, and comparing it to that of *w*Mel, which does cause CI. To sequence the *w*Au genome, PacBio reads were generated using gDNA extracted from cultured cells, which proved to be a convenient novel method for obtaining *Wolbachia* gDNA for genome sequencing. The PacBio reads were assembled into a single contig, with errors corrected using Illumina data. Comparative genomics revealed a number of genes with large differences between the two strains, including some genes absent entirely from *w*Au. Several transcriptional regulator genes, a family previously implicated in CI, were found to display differences between the two strains, including frameshifts that would likely result in loss of function in *w*Au. Statistical analysis revealed a higher frequency of SNPs in these genes than genes in other categories and suggested that they may be under positive selection, supporting a possible role for these

genes in CI. One transcriptional regulator gene was used to transform *D. melanogaster* for functional testing, but no significant CI effect was observed.

Chapter 3 used comparative transcriptomics to further identify and investigate *Wolbachia* genes that may be involved in CI, as well as host genes potentially involved in CI and host genes potentially involved in *Wolbachia*-mediated pathogen inhibition. These analyses are thought to represent the first use of RNA-seq for these purposes. Candidate *Wolbachia* CI genes were identified by comparing *Wolbachia* RNA-seq data from *D. simulans* testes and ovaries infected with either *w*Au or the CI-causing strain *w*Ri. A number of genes were identified that were differentially expressed between testes and ovaries infected with *w*Ri whose orthologues were not differentially expressed between testes and ovaries infected with *w*Au; these are good candidates for involvement in CI. This included three members of the transcriptional regulator gene family. Other members of this family were confirmed as being expressed, and some seemed to display differing expression levels between the two strains, though expression levels were not directly compared due to issues with interspecies comparison.

Candidate host CI genes were identified by comparing host RNA-seq data from *D. simulans* testes and ovaries infected with *w*Au, *w*Ri or uninfected. This analysis revealed a number of differentially expressed host genes that may be involved in CI, including *Ance*, a gene identified as a CI candidate in a previous study. The function of serine-type endopeptidase activity was enriched among genes upregulated by *w*Ri in testes, though as this is associated with a wide variety of cellular processes it does not provide much insight into CI.

Host genes potentially involved in pathogen inhibition were identified by comparing RNA-seq data from *Ae. albopictus* mosquitoes transinfected with *w*Mel, which shows resistance

against dengue and chikungunya viruses[122,332] naturally infected with *w*AlbA/B, or uninfected. A number of candidate viral inhibition genes were identified, and a number of functional annotations were enriched among these candidate genes. Enriched functions among genes upregulated by *w*Mel were not consistent with a role of general immune activation in pathogen inhibition, but many were related to metabolism and so were consistent with resource competition playing a role.

Chapter 4 attempted to identify genes for potential use in transgenic constructs. Firstly, attempts were made to identify the sex locus in *Ae. aegypti* by comparing male and female genome sequences. Despite some approaches seeming promising, this work did not succeed in identifying a strong candidate before the locus was identified by another group[104]. Secondly, transcriptomic analysis was performed on RNA-seq data from *Ae. aegypti* and *C. capitata* to identify testis-specifically expressed or spliced genes, which could be used to restrict expression to the male germline in transgenic insects. Several candidate genes were identified.

5.2. Use of high-throughput sequencing

All chapters share the use of high throughput sequencing for genome-wide identification of genes. This methodology has seen explosive growth in recent years, with decreasing costs and technological improvements making it a viable and effective option for investigating many biological questions. Genome sequencing has been facilitated by the recent development of long-read sequencing technologies, such as PacBio, which was used here to create a finished *Wolbachia* genome assembly. Other long read technologies being brought to the market, such as Oxford Nanopore sequencing[463], may further advance this field. In the field of transcriptomics, RNA sequencing has significant advantages over previous array-based methods, such as the ability to detect previously unknown transcripts. This feature has been important in some of the applications here. *Ae.*

albopictus does not have a well characterised transcriptome, and several candidate pathogen inhibition genes identified in Chapter 3 were novel. While the *Ae. aegypti* and *C. capitata* testis-specifically expressed genes in Chapter 4 corresponded to annotated loci, some of the testis-specifically spliced transcripts had not previously been predicted.

5.3. Future work and considerations

While high-throughput sequencing is an excellent tool for rapidly identifying a number of candidate genes, confirmation of the nature of function of these genes requires experimental validation. Although some functional testing was performed in Chapter 1, further analysis is required to confirm the roles of the genes identified in Chapters 2 and 3 in *Wolbachia*-host interactions, and the suitability of the genes identified in Chapter 4 for use in transgenic constructs.

Beyond laboratory analysis, any application of insect control methods will require rigorous field testing. There are various issues that may affect the success of modified males in the field. Survival and mating competitiveness of modified males with wild males under natural conditions is of critical importance for all strategies in which the insect itself is the control agent. With regards to *Wolbachia*-based population replacement strategies, semi-field trials have shown *Wolbachia*-infected *Ae. aegypti*[464] and *Ae. polynesiensis*[210] males to be equally or almost equally as competitive as wild males, and *Ae. aegypti* bisex RIDL males also showed no significant difference to wild males in terms of mating competitiveness[465]. Environmental conditions may also affect success – for example *Wolbachia*-mediated *Plasmodium* resistance was found to be affected by temperature[466]. Field testing is also important for satisfying regulatory concerns. This is of particular importance for population replacement methods, which are not self-limiting. Ideally there should be a way to remove transgenes or *Wolbachia* from populations in case of unforeseen consequences. In *Wolbachia* strategies this could be achieved by releasing

insects with an incompatible strain. In the case of naturally *Wolbachia*-infected target species such as *Ae. albopictus*, reversion to the wild-type infection could be readily achieved.

Field trials have already shown success in a number of cases, such as the use of *w*Mel-transinfected *Ae. aegypti* for population replacement[196,229,230], and the use of *Ae. aegypti* bisex RIDL lines for population suppression[93-96]. However, it is still too early to know the long-term effectiveness of these strategies and whether they can be successful in other locations with different conditions, such as differing climate, host species or strain, host population dynamics and human population dynamics. Of relevance to *Wolbachia* strategies is the fact that *Wolbachia* density has been shown to decrease with time following introduction into a new host[467]; as several studies have shown that pathogen inhibition is density-dependent[231,235-237] this may reduce its effectiveness as a control strategy. Pathogens may also evolve resistance to inhibition by *Wolbachia* or refractory transgenes over time. Likewise, population suppression strategies must consider the possibility of the target insects developing resistance against the suppression mechanism.

While genetic control strategies have progressed to the field trial stage in some cases, they are still an area of growing research, and recent technological advances have the potential to improve transgenic strategies. Many transgenic insect lines have been developed using transformation techniques that result in random integration of the transgenic construct into the genome, but site-specific insertion can now be achieved using systems such as the phage phi C31 system[66,453,468]. This should prove useful in generating stable transgenic lines with desired characteristics. Furthermore, new technologies that allow site-directed mutagenesis, such as HEGs, zinc finger nucleases (ZFNs), transcription-activator like effector nucleases (TALENs) and the CRISPR/Cas9 system, also allow more precise

manipulation when engineering effects on insect survival, reproduction or disease susceptibility (e.g. reviewed by Gabrieli *et al*.[469]).

Each control strategy is associated with different advantages and disadvantages, and in reality the best solution may be to use a combination of strategies – for disease vectors this is known as integrated vector management (IVM). Broadly speaking, complete population suppression is difficult to achieve, so for disease vectors population replacement may be a more appropriate goal. However, it would benefit from additional control with suppression strategies – this would lower the number of modified mosquitoes required to exceed the initial frequency threshold for *Wolbachia* or transgenes to spread, and also reduce biting nuisance in addition to disease transmission.

5.4. Conclusions

In conclusion, novel methods using *Wolbachia* or transgenic constructs have great potential for the control of insect disease vectors and agricultural pests. This work has identified and investigated a number of candidate genes that may be involved in *Wolbachia*-host interactions and thus of interest to *Wolbachia*-based strategies, as well as genes for use in transgenic control constructs. Further work is required to validate the results, but the findings provide an important platform from which to develop control strategies for insect pests.

References

1. Gubler, D. J., The global emergence/resurgence of arboviral diseases as public health problems. *Arch Med Res* **2002,** *33* (4), 330-42.

2. Gould, E. A.; Solomon, T., Pathogenic flaviviruses. *Lancet* **2008,** *371* (9611), 500-9.

3. Chen, L. H.; Wilson, M. E., Dengue and chikungunya in travelers: recent updates. *Curr Opin Infect Dis* **2012,** *25* (5), 523-9.

4. WHO, World Malaria Report 2015. **2015**.

5. WHO, Insect-borne diseases.

6. WHO, Dengue and severe dengue - Fact sheet no. 117. **2015**.

7. Ranjit, S.; Kissoon, N.; Jayakumar, I., Aggressive management of dengue shock syndrome may decrease mortality rate: a suggested protocol. *Pediatr Crit Care Med* **2005,** *6* (4), 412-9.

8. Brady, O. J.; Gething, P. W.; Bhatt, S.; Messina, J. P.; Brownstein, J. S.; Hoen, A. G.; Moyes, C. L.; Farlow, A. W.; Scott, T. W.; Hay, S. I., Refining the global spatial limits of dengue virus transmission by evidence-based consensus. *PLoS Negl Trop Dis* **2012,** *6* (8), e1760.

9. Bhatt, S.; Gething, P. W.; Brady, O. J.; Messina, J. P.; Farlow, A. W.; Moyes, C. L.; Drake, J. M.; Brownstein, J. S.; Hoen, A. G.; Sankoh, O.; Myers, M. F.; George, D. B.; Jaenisch, T.; Wint, G. R.; Simmons, C. P.; Scott, T. W.; Farrar, J. J.; Hay, S. I., The global distribution and burden of dengue. *Nature* **2013,** *496* (7446), 504-7.

10. Hill, C. A.; Kafatos, F. C.; Stansfield, S. K.; Collins, F. H., Arthropod-borne diseases: vector control in the genomics era. *Nat Rev Microbiol* **2005,** *3* (3), 262-8.

11. Capeding, M. R.; Tran, N. H.; Hadinegoro, S. R.; Ismail, H. I.; Chotpitayasunondh, T.; Chua, M. N.; Luong, C. Q.; Rusmil, K.; Wirawan, D. N.; Nallusamy, R.; Pitisuttithum, P.; Thisyakorn, U.; Yoon, I. K.; van der Vliet, D.; Langevin, E.; Laot, T.; Hutagalung, Y.; Frago, C.; Boaz, M.; Wartel, T. A.; Tornieporth, N. G.; Saville, M.; Bouckenooghe, A.; Group, C. S., Clinical efficacy and safety of a novel tetravalent dengue vaccine in healthy children in Asia: a phase 3, randomised, observer-masked, placebo-controlled trial. *Lancet* **2014,** *384* (9951), 1358-65.

12. Villar, L.; Dayan, G. H.; Arredondo-García, J. L.; Rivera, D. M.; Cunha, R.; Deseda, C.; Reynales, H.; Costa, M. S.; Morales-Ramírez, J. O.; Carrasquilla, G.; Rey, L. C.; Dietze, R.; Luz, K.; Rivas, E.; Miranda Montoya, M. C.; Cortés Supelano, M.; Zambrano, B.; Langevin, E.; Boaz, M.; Tornieporth, N.; Saville, M.; Noriega, F.; Group, C. S., Efficacy of a tetravalent dengue vaccine in children in Latin America. *N Engl J Med* **2015,** *372* (2), 113-23.

13. WHO, Chikungunya - Fact sheet no. 327. **2015**.

14. WHO, Yellow fever - Fact sheet no. 100. **2014**.

15. Hawley, W. A.; Pumpuni, C. B.; Brady, R. H.; Craig, G. B., Overwintering survival of *Aedes albopictus* (*Diptera*: *Culicidae*) eggs in Indiana. *J Med Entomol* **1989,** *26* (2), 122-9.

16. Hanson, S. M.; Craig, G. B., *Aedes albopictus* (*Diptera*: *Culicidae*) eggs: field survivorship during northern Indiana winters. *J Med Entomol* **1995,** *32* (5), 599-604.

17. Romi, R.; Severini, F.; Toma, L., Cold acclimation and overwintering of female *Aedes albopictus* in Roma. *J Am Mosq Control Assoc* **2006,** *22* (1), 149-51.

18. Paupy, C.; Delatte, H.; Bagny, L.; Corbel, V.; Fontenille, D., *Aedes albopictus*, an arbovirus vector: from the darkness to the light. *Microbes Infect* **2009,** *11* (14-15), 1177-85.

19. Gratz, N. G., Critical review of the vector status of *Aedes albopictus*. *Med Vet Entomol* **2004,** *18* (3), 215-27.

20. Scholte, J.; Schaffner, F., Waiting for the tiger: establishment and spread of the *Aedes albopictus* mosquito in Europe. In *Emerging pests and vector-borne diseases in Europe* **2007,** pp241-60.

21. Oxitec Agriculture - The need. http://www.oxitec.com/agriculture/the need.

22. Oxitec Agriculture - Our products - Medfly. http://www.oxitec.com/agriculture/our-products/medfly.

23. Curtis, C. F.; Mnzava, A. E., Comparison of house spraying and insecticide-treated nets for malaria control. *Bull World Health Organ* **2000,** *78* (12), 1389-400.

24. Oxitec, Health - The need. http://www.oxitec.com/health/the need.

25. WWF, Hazards and exposures associated with DDT and synthetic pyrethroids used for vector control. **1999**.

26. Hemingway, J.; Field, L.; Vontas, J., An overview of insecticide resistance. *Science* **2002,** *298* (5591), 96-7.

27. Vontas, J.; Kioulous, E.; Pavlidi, N.; Morou, E.; della Torre, A.; Ranson, H., Insecticide resistance in the major dengue vectors *Aedes albopictus* and *Aedes aegypti*. *Pesticide Biochemistry and Physiology* **2012,** *104* (2), 126-31.

28. Walker, K.; Lynch, M., Contributions of *Anopheles* larval control to malaria suppression in tropical Africa: review of achievements and potential. *Med Vet Entomol* **2007,** *21* (1), 2-21.

29. Keiser, J.; Utzinger, J.; Singer, B. H., The potential of intermittent irrigation for increasing rice yields, lowering water consumption, reducing methane emissions, and controlling malaria in African rice fields. *J Am Mosq Control Assoc* **2002,** *18* (4), 329-40.

30. Kline, D. L., Traps and trapping techniques for adult mosquito control. *J Am Mosq Control Assoc* **2006,** *22* (3), 490-6.

31. Okumu, F. O.; Madumla, E. P.; John, A. N.; Lwetoijera, D. W.; Sumaye, R. D., Attracting, trapping and killing disease-transmitting mosquitoes using odor-baited stations - The Ifakara Odor-Baited Stations. *Parasit Vectors* **2010,** *3*, 12.

32. Lacey, L. A., *Bacillus thuringiensis* serovariety *israelensis* and *Bacillus sphaericus* for mosquito control. *J Am Mosq Control Assoc* **2007,** *23* (2 Suppl), 133-63.

33. Kay, B.; Vu, S. N., New strategy against *Aedes aegypti* in Vietnam. *Lancet* **2005,** *365* (9459), 613-7.

34. Soumare, M. K.; Cilek, J. E., The effectiveness of *Mesocyclops longisetus* (*Copepoda*) for the control of container-inhabiting mosquitoes in residential environments. *J Am Mosq Control Assoc* **2011,** *27* (4), 376-83.

35. Sinh Nam, V.; Thi Yen, N.; Minh Duc, H.; Cong Tu, T.; Trong Thang, V.; Hoang Le, N.; Hoang San, L.; Le Loan, L.; Que Huong, V. T.; Kim Khanh, L. H.; Thuy Trang, H. T.; Lam, L. Z.; Kutcher, S. C.; Aaskov, J. G.; Jeffery, J. A.; Ryan, P. A.; Kay, B. H., Community-based control of *Aedes aegypti* by using *Mesocyclops* in southern Vietnam. *Am J Trop Med Hyg* **2012,** *86* (5), 850-9.

36. Hawley, W. A.; ter Kuile, F. O.; Steketee, R. S.; Nahlen, B. L.; Terlouw, D. J.; Gimnig, J. E.; Shi, Y. P.; Vulule, J. M.; Alaii, J. A.; Hightower, A. W.; Kolczak, M. S.; Kariuki, S. K.; Phillips-Howard, P. A., Implications of the western Kenya permethrin-treated bed net study for policy, program implementation, and future research. *Am J Trop Med Hyg* **2003,** *68* (4 Suppl), 168-73.

37. Alphey, L.; Benedict, M.; Bellini, R.; Clark, G. G.; Dame, D. A.; Service, M. W.; Dobson, S. L., Sterile-insect methods for control of mosquito-borne diseases: an analysis. *Vector Borne Zoonotic Dis* **2010,** *10* (3), 295-311.

38. Rendón, P.; McInnis, D.; Lance, D.; Stewart, J., Medfly (*Diptera*: *Tephritidae*) genetic sexing: large-scale field comparison of males-only and bisexual sterile fly releases in Guatemala. *J Econ Entomol* **2004,** *97* (5), 1547-53.

39. Lindquist, D. A.; Abusowa, M.; Hall, M. J., The New World screwworm fly in Libya: a review of its introduction and eradication. *Med Vet Entomol* **1992,** *6* (1), 2-8.

40. Krafsur, E., Sterile insect technique for suppressing and eradicating insect populations: 55 years and counting. *J. Agric. Entomol.* **1998,** *15*, 303-317.

41. Keng-Hong, T. E., *Area-wide Control of Fruit Flies and Other Insect Pests*. Penerbit Universiti Sains Malaysia, Penang: **2000**.

42. Koyama, J.; Kakinohana, H.; Miyatake, T., Eradication of the melon fly, *Bactrocera cucurbitae*, in Japan: importance of behavior, ecology, genetics, and evolution. *Annu Rev Entomol* **2004,** *49*, 331-49.

43. Patterson, R. S.; Weidhaas, D. E.; Ford, H. R.; Lofgren, C. S., Suppression and elimination of an island population of *Culex pipiens quinquefasciatus* with sterile males. *Science* **1970,** *168* (3937), 1368-70.

44. Lofgren, C. S.; Dame, D. A.; Breeland, S. G.; Weidhaas, D. E.; Jeffery, G.; Kaiser, R.; Ford, H. R.; Boston, M. D.; Baldwin, K. F., Release of chemosterilized males for the control of *Anopheles albimanus* in El Salvador. 3. Field methods and population control. *Am J Trop Med Hyg* **1974,** *23* (2), 288-97.

45. Bellini, R.; Calvitti, M.; Medici, A.; Carrieri, M.; Celli, S.; Maini, S., Use of the sterile insect technique against *Aedes albopictus* in Italy: first results of a pilot trial. In *Area-wide control of insect pests: from research to field implementation* **2007,** pp 505-15.

46. Bellini, R.; Medici, A.; Puggioli, A.; Balestrino, F.; Carrieri, M., Pilot field trials with *Aedes albopictus* irradiated sterile males in Italian urban areas. *J Med Entomol* **2013,** *50* (2), 317-25.

47. McDonald, P. T.; Hausermann, W.; Lorimer, N., Sterility introduced by release of genetically altered males to a domestic population of *Aedes aegypti* at the Kenya coast. *Am J Trop Med Hyg* **1977,** *26* (3), 553-61.

48. Alphey, L., Re-engineering the sterile insect technique. *Insect Biochem Mol Biol* **2002,** *32* (10), 1243-7.

49. Dame, D. A.; Curtis, C. F.; Benedict, M. Q.; Robinson, A. S.; Knols, B. G., Historical applications of induced sterilisation in field populations of mosquitoes. *Malar J* **2009,** *8 Suppl 2*, S2.

50. Klassen, W.; Curtis, C., History of the sterile insect technique. In *Sterile Insect Technique* **2005,** pp3-36.

51. Papathanos, P. A.; Bossin, H. C.; Benedict, M. Q.; Catteruccia, F.; Malcolm, C. A.; Alphey, L.; Crisanti, A., Sex separation strategies: past experience and new approaches. *Malar J* **2009,** *8 Suppl 2*, S5.

52. Catteruccia, F.; Benton, J. P.; Crisanti, A., An *Anopheles* transgenic sexing strain for vector control. *Nat Biotechnol* **2005,** *23* (11), 1414-7.

53. Olson, K. E.; Higgs, S.; Gaines, P. J.; Powers, A. M.; Davis, B. S.; Kamrud, K. I.; Carlson, J. O.; Blair, C. D.; Beaty, B. J., Genetically engineered resistance to dengue-2 virus transmission in mosquitoes. *Science* **1996,** *272* (5263), 884-6.

54. Franz, A. W.; Sanchez-Vargas, I.; Adelman, Z. N.; Blair, C. D.; Beaty, B. J.; James, A. A.; Olson, K. E., Engineering RNA interference-based resistance to dengue virus type 2 in genetically modified *Aedes aegypti*. *Proc Natl Acad Sci U S A* **2006,** *103* (11), 4198-203.

55. Mathur, G.; Sanchez-Vargas, I.; Alvarez, D.; Olson, K. E.; Marinotti, O.; James, A. A., Transgene-mediated suppression of dengue viruses in the salivary glands of the yellow fever mosquito, *Aedes aegypti*. *Insect Mol Biol* **2010,** *19* (6), 753-63.

56. Isaacs, A. T.; Li, F.; Jasinskiene, N.; Chen, X.; Nirmala, X.; Marinotti, O.; Vinetz, J. M.; James, A. A., Engineered resistance to *Plasmodium falciparum* development in transgenic *Anopheles stephensi*. *PLoS Pathog* **2011,** *7* (4), e1002017.

57. Isaacs, A. T.; Jasinskiene, N.; Tretiakov, M.; Thiery, I.; Zettor, A.; Bourgouin, C.; James, A. A., Transgenic *Anopheles stephensi* coexpressing single-chain antibodies resist *Plasmodium falciparum* development. *Proc Natl Acad Sci U S A* **2012,** *109* (28), E1922-30.

58. Sumitani, M.; Kasashima, K.; Yamamoto, D. S.; Yagi, K.; Yuda, M.; Matsuoka, H.; Yoshida, S., Reduction of malaria transmission by transgenic mosquitoes expressing an antisporozoite antibody in their salivary glands. *Insect Mol Biol* **2013,** *22* (1), 41-51.

59. Ito, J.; Ghosh, A.; Moreira, L. A.; Wimmer, E. A.; Jacobs-Lorena, M., Transgenic anopheline mosquitoes impaired in transmission of a malaria parasite. *Nature* **2002,** *417* (6887), 452-5.

60. Moreira, L. A.; Ito, J.; Ghosh, A.; Devenport, M.; Zieler, H.; Abraham, E. G.; Crisanti, A.; Nolan, T.; Catteruccia, F.; Jacobs-Lorena, M., Bee venom phospholipase inhibits malaria parasite development in transgenic mosquitoes. *J Biol Chem* **2002,** *277* (43), 40839-43.

61. Luckhart, S.; Giulivi, C.; Drexler, A. L.; Antonova-Koch, Y.; Sakaguchi, D.; Napoli, E.; Wong, S.; Price, M. S.; Eigenheer, R.; Phinney, B. S.; Pakpour, N.; Pietri, J. E.; Cheung, K.; Georgis, M.; Riehle, M., Sustained activation of Akt elicits mitochondrial dysfunction to block *Plasmodium falciparum* infection in the mosquito host. *PLoS Pathog* **2013,** *9* (2), e1003180.

62. Hauck, E. S.; Antonova-Koch, Y.; Drexler, A.; Pietri, J.; Pakpour, N.; Liu, D.; Blacutt, J.; Riehle, M. A.; Luckhart, S., Overexpression of phosphatase and tensin homolog improves fitness and decreases *Plasmodium falciparum* development in *Anopheles stephensi*. *Microbes Infect* **2013,** *15* (12), 775-87.

63. Dong, Y.; Das, S.; Cirimotich, C.; Souza-Neto, J. A.; McLean, K. J.; Dimopoulos, G., Engineered *Anopheles* immunity to *Plasmodium* infection. *PLoS Pathog* **2011,** *7* (12), e1002458.

64. Corby-Harris, V.; Drexler, A.; Watkins de Jong, L.; Antonova, Y.; Pakpour, N.; Ziegler, R.; Ramberg, F.; Lewis, E. E.; Brown, J. M.; Luckhart, S.; Riehle, M. A., Correction: Activation of Akt signaling reduces the prevalence and intensity of malaria parasite infection and lifespan in *Anopheles stephensi* mosquitoes. *PLoS Pathog* **2010,** *6* (8).

65. Kim, W.; Koo, H.; Richman, A. M.; Seeley, D.; Vizioli, J.; Klocko, A. D.; O'Brochta, D. A., Ectopic expression of a cecropin transgene in the human malaria vector mosquito *Anopheles gambiae* (*Diptera*: *Culicidae*): effects on susceptibility to *Plasmodium*. *J Med Entomol* **2004,** *41* (3), 447-55.

66. Meredith, J. M.; Basu, S.; Nimmo, D. D.; Larget-Thiery, I.; Warr, E. L.; Underhill, A.; McArthur, C. C.; Carter, V.; Hurd, H.; Bourgouin, C.; Eggleston, P., Site-specific integration and expression of an anti-malarial gene in transgenic *Anopheles gambiae* significantly reduces *Plasmodium* infections. *PLoS One* **2011,** *6* (1), e14587.

67. Catteruccia, F.; Godfray, H. C.; Crisanti, A., Impact of genetic manipulation on the fitness of *Anopheles stephensi* mosquitoes. *Science* **2003,** *299* (5610), 1225-7.

68. Irvin, N.; Hoddle, M. S.; O'Brochta, D. A.; Carey, B.; Atkinson, P. W., Assessing fitness costs for transgenic *Aedes aegypti* expressing the GFP marker and transposase genes. *Proc Natl Acad Sci U S A* **2004,** *101* (3), 891-6.

69. Sinkins, S. P.; Gould, F., Gene drive systems for insect disease vectors. *Nat Rev Genet* **2006,** *7* (6), 427-35.

70. Okamoto, K. W.; Robert, M. A.; Gould, F.; Lloyd, A. L., Feasible Introgression of an Anti-pathogen Transgene into an Urban Mosquito Population without Using Gene-Drive. *PLoS Negl Trop Dis* **2014,** *8* (7), e2827.

71. Hammond, A.; Galizi, R.; Kyrou, K.; Simoni, A.; Siniscalchi, C.; Katsanos, D.; Gribble, M.; Baker, D.; Marois, E.; Russell, S.; Burt, A.; Windbichler, N.; Crisanti, A.; Nolan, T., A

CRISPR-Cas9 gene drive system targeting female reproduction in the malaria mosquito vector *Anopheles gambiae*. *Nat Biotechnol* **2016,** *34*, 78-83.

72. Hurst, G. D.; Werren, J. H., The role of selfish genetic elements in eukaryotic evolution. *Nat Rev Genet* **2001,** *2* (8), 597-606.

73. Sethuraman, N.; Fraser, M. J.; Eggleston, P.; O'Brochta, D. A., Post-integration stability of piggyBac in *Aedes aegypti*. *Insect Biochem Mol Biol* **2007,** *37* (9), 941-51.

74. Beeman, R. W.; Friesen, K. S.; Denell, R. E., Maternal-effect selfish genes in flour beetles. *Science* **1992,** *256* (5053), 89-92.

75. Chen, C. H.; Huang, H.; Ward, C. M.; Su, J. T.; Schaeffer, L. V.; Guo, M.; Hay, B. A., A synthetic maternal-effect selfish genetic element drives population replacement in *Drosophila*. *Science* **2007,** *316* (5824), 597-600.

76. Akbari, O. S.; Matzen, K. D.; Marshall, J. M.; Huang, H.; Ward, C. M.; Hay, B. A., A synthetic gene drive system for local, reversible modification and suppression of insect populations. *Curr Biol* **2013,** *23* (8), 671-7.

77. Reeves, R. G.; Bryk, J.; Altrock, P. M.; Denton, J. A.; Reed, F. A., First steps towards underdominant genetic transformation of insect populations. *PLoS One* **2014,** *9* (5), e97557.

78. Windbichler, N.; Menichelli, M.; Papathanos, P. A.; Thyme, S. B.; Li, H.; Ulge, U. Y.; Hovde, B. T.; Baker, D.; Monnat, R. J.; Burt, A.; Crisanti, A., A synthetic homing endonuclease-based gene drive system in the human malaria mosquito. *Nature* **2011,** *473* (7346), 212-5.

79. Gantz, V. M.; Bier, E., Genome editing. The mutagenic chain reaction: a method for converting heterozygous to homozygous mutations. *Science* **2015,** *348* (6233), 442-4.

80. Gantz, V. M.; Jasinskiene, N.; Tatarenkova, O.; Fazekas, A.; Macias, V. M.; Bier, E.; James, A. A., Highly efficient Cas9-mediated gene drive for population modification of the malaria vector mosquito *Anopheles stephensi*. *Proc Natl Acad Sci U S A* **2015,** *112* (49), e6736-43

81. Turelli, M.; Hoffmann, A. A., Microbe-induced cytoplasmic incompatibility as a mechanism for introducing transgenes into arthropod populations. *Insect Mol Biol* **1999,** *8* (2), 243-55.

82. Sinkins, S. P.; Godfray, H. C., Use of *Wolbachia* to drive nuclear transgenes through insect populations. *Proc Biol Sci* **2004,** *271* (1546), 1421-6.

83. Thailayil, J.; Magnusson, K.; Godfray, H. C.; Crisanti, A.; Catteruccia, F., Spermless males elicit large-scale female responses to mating in the malaria mosquito *Anopheles gambiae*. *Proc Natl Acad Sci U S A* **2011,** *108* (33), 13677-81.

84. Windbichler, N.; Papathanos, P. A.; Crisanti, A., Targeting the X chromosome during spermatogenesis induces Y chromosome transmission ratio distortion and early dominant embryo lethality in *Anopheles gambiae*. *PLoS Genet* **2008,** *4* (12), e1000291.

85. Atkinson, M. P.; Su, Z.; Alphey, N.; Alphey, L. S.; Coleman, P. G.; Wein, L. M., Analyzing the control of mosquito-borne diseases by a dominant lethal genetic system. *Proc Natl Acad Sci U S A* **2007,** *104* (22), 9540-5.

86. Phuc, H. K.; Andreasen, M. H.; Burton, R. S.; Vass, C.; Epton, M. J.; Pape, G.; Fu, G.; Condon, K. C.; Scaife, S.; Donnelly, C. A.; Coleman, P. G.; White-Cooper, H.; Alphey, L., Late-acting dominant lethal genetic systems and mosquito control. *BMC Biol* **2007,** *5*, 11.

87. Thomas, D. D.; Donnelly, C. A.; Wood, R. J.; Alphey, L. S., Insect population control using a dominant, repressible, lethal genetic system. *Science* **2000,** *287* (5462), 2474-6.

88. Heinrich, J. C.; Scott, M. J., A repressible female-specific lethal genetic system for making transgenic insect strains suitable for a sterile-release program. *Proc Natl Acad Sci U S A* **2000,** *97* (15), 8229-32.

89. Gossen, M.; Bujard, H., Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. *Proc Natl Acad Sci U S A* **1992,** *89* (12), 5547-51.

90. Gong, P.; Epton, M. J.; Fu, G.; Scaife, S.; Hiscox, A.; Condon, K. C.; Condon, G. C.; Morrison, N. I.; Kelly, D. W.; Dafa'alla, T.; Coleman, P. G.; Alphey, L., A dominant lethal genetic system for autocidal control of the Mediterranean fruitfly. *Nat Biotechnol* **2005,** *23* (4), 453-6.

91. Berger, S. L.; Cress, W. D.; Cress, A.; Triezenberg, S. J.; Guarente, L., Selective inhibition of activated but not basal transcription by the acidic activation domain of VP16: evidence for transcriptional adaptors. *Cell* **1990**, *61* (7), 1199-208.

92. Gill, G.; Ptashne, M., Negative effect of the transcriptional activator GAL4. *Nature* **1988,** *334* (6184), 721-4.

93. Harris, A. F.; Nimmo, D.; McKemey, A. R.; Kelly, N.; Scaife, S.; Donnelly, C. A.; Beech, C.; Petrie, W. D.; Alphey, L., Field performance of engineered male mosquitoes. *Nat Biotechnol* **2011,** *29* (11), 1034-7.

94. Harris, A. F.; McKemey, A. R.; Nimmo, D.; Curtis, Z.; Black, I.; Morgan, S. A.; Oviedo, M. N.; Lacroix, R.; Naish, N.; Morrison, N. I.; Collado, A.; Stevenson, J.; Scaife, S.; Dafa'alla, T.; Fu, G.; Phillips, C.; Miles, A.; Raduan, N.; Kelly, N.; Beech, C.; Donnelly, C. A.; Petrie, W. D.; Alphey, L., Successful suppression of a field mosquito population by sustained release of engineered male mosquitoes. *Nat Biotechnol* **2012,** *30* (9), 828-30.

95. Lacroix, R.; McKemey, A. R.; Raduan, N.; Kwee Wee, L.; Hong Ming, W.; Guat Ney, T.; Rahidah A A, S.; Salman, S.; Subramaniam, S.; Nordin, O.; Hanum A T, N.; Angamuthu, C.; Marlina Mansor, S.; Lees, R. S.; Naish, N.; Scaife, S.; Gray, P.; Labbé, G.; Beech, C.; Nimmo, D.; Alphey, L.; Vasan, S. S.; Han Lim, L.; Wasi A, N.; Murad, S., Open field release of genetically engineered sterile male *Aedes aegypti* in Malaysia. *PLoS One* **2012,** *7* (8), e42771.

96. Carvalho, D. O.; McKemey, A. R.; Garziera, L.; Lacroix, R.; Donnelly, C. A.; Alphey, L.; Malavasi, A.; Capurro, M. L., Suppression of a Field Population of *Aedes aegypti* in Brazil by Sustained Release of Transgenic Male Mosquitoes. *PLoS Negl Trop Dis* **2015,** *9* (7), e0003864.

97. Fu, G.; Lees, R. S.; Nimmo, D.; Aw, D.; Jin, L.; Gray, P.; Berendonk, T. U.; White-Cooper, H.; Scaife, S.; Kim Phuc, H.; Marinotti, O.; Jasinskiene, N.; James, A. A.; Alphey, L.,

Female-specific flightless phenotype for mosquito control. *Proc Natl Acad Sci U S A* **2010,** *107* (10), 4550-4.

98. Labbé, G. M.; Scaife, S.; Morgan, S. A.; Curtis, Z. H.; Alphey, L., Female-Specific Flightless (fsRIDL) Phenotype for Control of *Aedes albopictus*. *PLoS Negl Trop Dis* **2012,** *6* (7), e1724.

99. Fu, G.; Condon, K. C.; Epton, M. J.; Gong, P.; Jin, L.; Condon, G. C.; Morrison, N. I.; Dafa'alla, T. H.; Alphey, L., Female-specific insect lethality engineered using alternative splicing. *Nat Biotechnol* **2007,** *25* (3), 353-7.

100. Ant, T.; Koukidou, M.; Rempoulakis, P.; Gong, H. F.; Economopoulos, A.; Vontas, J.; Alphey, L., Control of the olive fruit fly using genetics-enhanced sterile insect technique. *BMC Biol* **2012,** *10*, 51.

101. Jin, L.; Walker, A. S.; Fu, G.; Harvey-Samuel, T.; Dafa'alla, T.; Miles, A.; Marubbi, T.; Granville, D.; Humphrey-Jones, N.; O'Connell, S.; Morrison, N. I.; Alphey, L., Engineered female-specific lethality for control of pest *Lepidoptera*. *ACS Synth Biol* **2013,** *2* (3), 160-6.

102. Cator, L. J.; Arthur, B. J.; Harrington, L. C.; Hoy, R. R., Harmonic convergence in the love songs of the dengue vector mosquito. *Science* **2009,** *323* (5917), 1077-9.

103. Galizi, R.; Doyle, L. A.; Menichelli, M.; Bernardini, F.; Deredec, A.; Burt, A.; Stoddard, B. L.; Windbichler, N.; Crisanti, A., A synthetic sex ratio distortion system for the control of the human malaria mosquito. *Nat Commun* **2014,** *5*, 3977.

104. Hall, A. B.; Basu, S.; Jiang, X.; Qi, Y.; Timoshevskiy, V. A.; Biedler, J. K.; Sharakhova, M. V.; Elahi, R.; Anderson, M. A.; Chen, X. G.; Sharakhov, I. V.; Adelman, Z. N.; Tu, Z., A male-determining factor in the mosquito *Aedes aegypti*. *Science* **2015,** *348* (6240), 1268-70.

105. Hertig, M.; Wolbach, S. B., Studies on *Rickettsia*-Like Micro-Organisms in Insects. *J Med Res* **1924,** *44* (3), 329-374.7.

106. Hilgenboecker, K.; Hammerstein, P.; Schlattmann, P.; Telschow, A.; Werren, J. H., How many species are infected with *Wolbachia*?--A statistical analysis of current data. *FEMS Microbiol Lett* **2008,** *281* (2), 215-20.

107. Sinkins, S. P., *Wolbachia* and cytoplasmic incompatibility in mosquitoes. *Insect Biochem Mol Biol* **2004,** *34* (7), 723-9.

108. Kittayapong, P.; Baisley, K. J.; Baimai, V.; O'Neill, S. L., Distribution and diversity of *Wolbachia* infections in Southeast Asian mosquitoes (*Diptera*: *Culicidae*). *J Med Entomol* **2000,** *37* (3), 340-5.

109. Ricci, I.; Cancrini, G.; Gabrielli, S.; D'Amelio, S.; Favi, G., Searching for *Wolbachia* (*Rickettsiales*: *Rickettsiaceae*) in mosquitoes (*Diptera*: *Culicidae*): large polymerase chain reaction survey and new identifications. *J Med Entomol* **2002,** *39* (4), 562-7.

110. Rasgon, J. L.; Scott, T. W., An initial survey for *Wolbachia* (*Rickettsiales*: *Rickettsiaceae*) infections in selected California mosquitoes (*Diptera*: *Culicidae*). *J Med Entomol* **2004,** *41* (2), 255-7.

111. Baldini, F.; Segata, N.; Pompon, J.; Marcenac, P.; Shaw, W. R.; Dabiré, R. K.; Diabaté, A.; Levashina, E. A.; Catteruccia, F., Evidence of natural *Wolbachia* infections in field populations of *Anopheles gambiae*. *Nat Commun* **2014,** *5*, 3985.

112. Bian, G.; Joshi, D.; Dong, Y.; Lu, P.; Zhou, G.; Pan, X.; Xu, Y.; Dimopoulos, G.; Xi, Z., *Wolbachia* invades *Anopheles stephensi* populations and induces refractoriness to *Plasmodium* infection. *Science* **2013,** *340* (6133), 748-51.

113. Xi, Z.; Khoo, C. C.; Dobson, S. L., *Wolbachia* establishment and invasion in an *Aedes aegypti* laboratory population. *Science* **2005,** *310* (5746), 326-8.

114. Ruang-Areerate, T.; Kittayapong, P., *Wolbachia* transinfection in *Aedes aegypti*: a potential gene driver of dengue vectors. *Proc Natl Acad Sci U S A* **2006,** *103* (33), 12534-9.

115. McMeniman, C. J.; Lane, R. V.; Cass, B. N.; Fong, A. W.; Sidhu, M.; Wang, Y. F.; O'Neill, S. L., Stable introduction of a life-shortening *Wolbachia* infection into the mosquito *Aedes aegypti*. *Science* **2009,** *323* (5910), 141-4.

116. Walker, T.; Johnson, P. H.; Moreira, L. A.; Iturbe-Ormaetxe, I.; Frentiu, F. D.; McMeniman, C. J.; Leong, Y. S.; Dong, Y.; Axford, J.; Kriesner, P.; Lloyd, A. L.; Ritchie, S. A.; O'Neill, S. L.; Hoffmann, A. A., The *w*Mel *Wolbachia* strain blocks dengue and invades caged *Aedes aegypti* populations. *Nature* **2011,** *476* (7361), 450-3.

117. Xi, Z.; Dean, J. L.; Khoo, C.; Dobson, S. L., Generation of a novel *Wolbachia* infection in *Aedes albopictus* (Asian tiger mosquito) via embryonic microinjection. *Insect Biochem Mol Biol* **2005,** *35* (8), 903-10.

118. Xi, Z.; Khoo, C. C.; Dobson, S. L., Interspecific transfer of *Wolbachia* into the mosquito disease vector *Aedes albopictus*. *Proc Biol Sci* **2006,** *273* (1592), 1317-22.

119. Suh, E.; Mercer, D. R.; Fu, Y.; Dobson, S. L., Pathogenicity of life-shortening *Wolbachia* in *Aedes albopictus* after transfer from *Drosophila melanogaster*. *Appl Environ Microbiol* **2009,** *75* (24), 7783-8.

120. Fu, Y.; Gavotte, L.; Mercer, D. R.; Dobson, S. L., Artificial triple *Wolbachia* infection in *Aedes albopictus* yields a new pattern of unidirectional cytoplasmic incompatibility. *Appl Environ Microbiol* **2010,** *76* (17), 5887-91.

121. Calvitti, M.; Moretti, R.; Lampazzi, E.; Bellini, R.; Dobson, S. L., Characterization of a new *Aedes albopictus* (*Diptera*: *Culicidae*)-*Wolbachia pipientis* (*Rickettsiales*: *Rickettsiaceae*) symbiotic association generated by artificial transfer of the *w*Pip strain from *Culex pipiens* (*Diptera*: *Culicidae*). *J Med Entomol* **2010,** *47* (2), 179-87.

122. Blagrove, M. S.; Arias-Goeta, C.; Failloux, A. B.; Sinkins, S. P., *Wolbachia* strain *w*Mel induces cytoplasmic incompatibility and blocks dengue transmission in *Aedes albopictus*. *Proc Natl Acad Sci U S A* **2012,** *109* (1), 255-60.

123. Andrews, E. S.; Fu, Y.; Calvitti, M.; Dobson, S. L., Interspecific Transfer of a *Wolbachia* Infection Into *Aedes albopictus* (*Diptera*: *Culicidae*) Yields a Novel Phenotype Capable of Rescuing a Superinfection. *J Med Entomol* **2014,** *51* (6), 1192-8.

124. Kittayapong, P.; Baimai, V.; O'Neill, S. L., Field prevalence of *Wolbachia* in the mosquito vector *Aedes albopictus*. *Am J Trop Med Hyg* **2002,** *66* (1), 108-11.

125. Sinkins, S. P.; Braig, H. R.; O'Neill, S. L., *Wolbachia* superinfections and the expression of cytoplasmic incompatibility. *Proc Biol Sci* **1995,** *261* (1362), 325-30.

126. Werren, J. H.; Baldo, L.; Clark, M. E., *Wolbachia*: master manipulators of invertebrate biology. *Nat Rev Microbiol* **2008,** *6* (10), 741-51.

127. Stouthamer, R.; Breeuwer, J. A.; Hurst, G. D., *Wolbachia pipientis*: microbial manipulator of arthropod reproduction. *Annu Rev Microbiol* **1999,** *53*, 71-102.

128. Laven, H., Eradication of *Culex pipiens fatigans* through cytoplasmic incompatibility. *Nature* **1967,** *216* (5113), 383-4.

129. Dobson, S. L.; Fox, C. W.; Jiggins, F. M., The effect of *Wolbachia*-induced cytoplasmic incompatibility on host population size in natural and manipulated systems. *Proc Biol Sci* **2002,** *269* (1490), 437-45.

130. Turelli, M.; Hoffmann, A. A., Cytoplasmic incompatibility in *Drosophila simulans*: dynamics and parameter estimates from natural populations. *Genetics* **1995,** *140* (4), 1319-38.

131. Bourtzis, K.; Nirgianaki, A.; Markakis, G.; Savakis, C., *Wolbachia* infection and cytoplasmic incompatibility in *Drosophila* species. *Genetics* **1996,** *144* (3), 1063-73.

132. Reynolds, K. T.; Hoffmann, A. A., Male age, host effects and the weak expression or non-expression of cytoplasmic incompatibility in *Drosophila* strains infected by maternally transmitted *Wolbachia*. *Genet Res* **2002,** *80* (2), 79-87.

133. Laven, H., [Crossing experiments with European and American strains of the *Culex pipiens* complex]. *Z Tropenmed Parasitol* **1954,** *5* (3), 317-23.

134. Sinkins, S. P.; Braig, H. R.; O'Neill, S. L., *Wolbachia pipientis*: bacterial density and unidirectional cytoplasmic incompatibility between infected populations of *Aedes albopictus*. *Exp Parasitol* **1995,** *81* (3), 284-91.

135. Werren, J. H., Biology of *Wolbachia*. *Annu Rev Entomol* **1997,** *42*, 587-609.

136. Serbus, L. R.; Casper-Lindley, C.; Landmann, F.; Sullivan, W., The genetics and cell biology of *Wolbachia*-host interactions. *Annu Rev Genet* **2008,** *42*, 683-707.

137. McGraw, E. A.; O'Neill, S. L., Evolution of *Wolbachia pipientis* transmission dynamics in insects. *Trends Microbiol* **1999,** *7* (7), 297-302.

138. Charlat, S.; Calmet, C.; Merçot, H., On the mod resc model and the evolution of *Wolbachia* compatibility types. *Genetics* **2001,** *159* (4), 1415-22.

139. Zabalou, S.; Apostolaki, A.; Pattas, S.; Veneti, Z.; Paraskevopoulos, C.; Livadaras, I.; Markakis, G.; Brissac, T.; Merçot, H.; Bourtzis, K., Multiple rescue factors within a *Wolbachia* strain. *Genetics* **2008,** *178* (4), 2145-60.

140. Bordenstein, S. R.; Werren, J. H., Effects of A and B *Wolbachia* and host genotype on interspecies cytoplasmic incompatibility in *Nasonia*. *Genetics* **1998,** *148* (4), 1833-44.

141. Duron, O.; Weill, M., *Wolbachia* infection influences the development of *Culex pipiens* embryo in incompatible crosses. *Heredity* **2006,** *96* (6), 493-500.

142. Ryan, S. L.; Saul, G. B., Post-fertilization effect of incompatibility factors in *Mormoniella*. *Mol Gen Genet* **1968,** *103* (1), 29-36.

143. Callaini, G.; Dallai, R.; Riparbelli, M. G., *Wolbachia*-induced delay of paternal chromatin condensation does not prevent maternal chromosomes from entering anaphase in incompatible crosses of *Drosophila simulans*. *J Cell Sci* **1997,** *110 (Pt 2)*, 271-80.

144. Tram, U.; Sullivan, W., Role of delayed nuclear envelope breakdown and mitosis in *Wolbachia*-induced cytoplasmic incompatibility. *Science* **2002,** *296* (5570), 1124-6.

145. Yu, K. R.; Saint, R. B.; Sullivan, W., The Grapes checkpoint coordinates nuclear envelope breakdown and chromosome condensation. *Nat Cell Biol* **2000,** *2* (9), 609-15.

146. Chen, Y.; Poon, R. Y., The multiple checkpoint functions of CHK1 and CHK2 in maintenance of genome stability. *Front Biosci* **2008,** *13*, 5016-29.

147. Reed, K. M.; Werren, J. H., Induction of paternal genome loss by the paternal-sex-ratio chromosome and cytoplasmic incompatibility bacteria (*Wolbachia*): a comparative study of early embryonic events. *Mol Reprod Dev* **1995,** *40* (4), 408-18.

148. Tram, U.; Fredrick, K.; Werren, J. H.; Sullivan, W., Paternal chromosome segregation during the first mitotic division determines *Wolbachia*-induced cytoplasmic incompatibility phenotype. *J Cell Sci* **2006,** *119* (Pt 17), 3655-63.

149. Lassy, C. W.; Karr, T. L., Cytological analysis of fertilization and early embryonic development in incompatible crosses of *Drosophila simulans*. *Mech Dev* **1996,** *57* (1), 47-58.

150. Brennan, L. J.; Haukedal, J. A.; Earle, J. C.; Keddie, B.; Harris, H. L., Disruption of redox homeostasis leads to oxidative DNA damage in spermatocytes of *Wolbachia*-infected *Drosophila simulans*. *Insect Mol Biol* **2012,** *21* (5), 510-20.

151. Landmann, F.; Orsi, G. A.; Loppin, B.; Sullivan, W., *Wolbachia*-mediated cytoplasmic incompatibility is associated with impaired histone deposition in the male pronucleus. *PLoS Pathog* **2009,** *5* (3), e1000343.

152. Renault, A. D.; Zhang, X. H.; Alphey, L. S.; Frenz, L. M.; Glover, D. M.; Saunders, R. D.; Axton, J. M., giant nuclei is essential in the cell cycle transition from meiosis to mitosis. *Development* **2003,** *130* (13), 2997-3005.

153. Bressac, C.; Rousset, F., The reproductive incompatibility system in *Drosophila simulans*: DAPI-staining analysis of the *Wolbachia* symbionts in sperm cysts. *J Invertebr Pathol* **1993,** *61* (3), 226-30.

154. Clark, M. E.; Veneti, Z.; Bourtzis, K.; Karr, T. L., The distribution and proliferation of the intracellular bacteria *Wolbachia* during spermatogenesis in *Drosophila*. *Mech Dev* **2002,** *111* (1-2), 3-15.

155. Presgraves, D. C., A genetic test of the mechanism of *Wolbachia*-induced cytoplasmic incompatibility in *Drosophila*. *Genetics* **2000,** *154* (2), 771-6.

156. Breeuwer, J. A.; Werren, J. H., Microorganisms associated with chromosome destruction and reproductive isolation between two insect species. *Nature* **1990,** *346* (6284), 558-60.

157. Kose, H.; Karr, T. L., Organization of *Wolbachia pipientis* in the *Drosophila* fertilized egg and embryo revealed by an anti-*Wolbachia* monoclonal antibody. *Mech Dev* **1995,** *51* (2-3), 275-88.

158. Poinsot, D.; Charlat, S.; Merçot, H., On the mechanism of *Wolbachia*-induced cytoplasmic incompatibility: confronting the models with the facts. *Bioessays* **2003,** *25* (3), 259-65.

159. Tram, U.; Ferree, P. M.; Sullivan, W., Identification of *Wolbachia*--host interacting factors through cytological analysis. *Microbes Infect* **2003,** *5* (11), 999-1011.

160. Ferree, P. M.; Sullivan, W., A genetic test of the role of the maternal pronucleus in *Wolbachia*-induced cytoplasmic incompatibility in *Drosophila melanogaster*. *Genetics* **2006,** *173* (2), 839-47.

161. Wu, M.; Sun, L. V.; Vamathevan, J.; Riegler, M.; Deboy, R.; Brownlie, J. C.; McGraw, E. A.; Martin, W.; Esser, C.; Ahmadinejad, N.; Wiegand, C.; Madupu, R.; Beanan, M. J.; Brinkac, L. M.; Daugherty, S. C.; Durkin, A. S.; Kolonay, J. F.; Nelson, W. C.; Mohamoud, Y.; Lee, P.; Berry, K.; Young, M. B.; Utterback, T.; Weidman, J.; Nierman, W. C.; Paulsen, I. T.; Nelson, K. E.; Tettelin, H.; O'Neill, S. L.; Eisen, J. A., Phylogenomics of the reproductive parasite *Wolbachia pipientis w*Mel: a streamlined genome overrun by mobile genetic elements. *PLoS Biol* **2004,** *2* (3), E69.

162. Klasson, L.; Westberg, J.; Sapountzis, P.; Naslund, K.; Lutnaes, Y.; Darby, A. C.; Veneti, Z.; Chen, L.; Braig, H. R.; Garrett, R.; Bourtzis, K.; Andersson, S. G., The mosaic genome structure of the *Wolbachia w*Ri strain infecting *Drosophila simulans*. *Proc Natl Acad Sci U S A* **2009,** *106* (14), 5725-30.

163. Ellegaard, K. M.; Klasson, L.; Näslund, K.; Bourtzis, K.; Andersson, S. G., Comparative genomics of *Wolbachia* and the bacterial species concept. *PLoS Genet* **2013,** *9* (4), e1003381.

164. Metcalf, J. A.; Jo, M.; Bordenstein, S. R.; Jaenike, J., Recent genome reduction of *Wolbachia* in *Drosophila recens* targets phage WO and narrows candidates for reproductive parasitism. *PeerJ* **2014,** *2*, e529.

165. Klasson, L.; Walker, T.; Sebaihia, M.; Sanders, M. J.; Quail, M. A.; Lord, A.; Sanders, S.; Earl, J.; O'Neill, S. L.; Thomson, N.; Sinkins, S. P.; Parkhill, J., Genome evolution of *Wolbachia* strain *w*Pip from the *Culex pipiens* group. *Mol Biol Evol* **2008,** *25* (9), 1877-87.

166. Pinto, S. B.; Stainton, K.; Harris, S.; Kambris, Z.; Sutton, E. R.; Bonsall, M. B.; Parkhill, J.; Sinkins, S. P., Transcriptional Regulation of *Culex pipiens* Mosquitoes by *Wolbachia* Influences Cytoplasmic Incompatibility. *PLoS Pathog* **2013,** *9* (10), e1003647.

167. Mavingui, P.; Valiente Moro, C.; Tran-Van, V.; Wisniewski-Dye, F.; Raquin, V.; Minard, G.; Tran, F. H.; Voronin, D.; Rouy, Z.; Bustos, P.; Lozano, L.; Barbe, V.; Gonzalez, V., Whole-Genome Sequence of *Wolbachia* Strain *w*AlbB, an Endosymbiont of Tiger Mosquito Vector *Aedes albopictus*. *J Bacteriol* **2012,** *194* (7), 1840.

168. Foster, J.; Ganatra, M.; Kamal, I.; Ware, J.; Makarova, K.; Ivanova, N.; Bhattacharyya, A.; Kapatral, V.; Kumar, S.; Posfai, J.; Vincze, T.; Ingram, J.; Moran, L.; Lapidus, A.; Omelchenko, M.; Kyrpides, N.; Ghedin, E.; Wang, S.; Goltsman, E.; Joukov, V.; Ostrovskaya, O.; Tsukerman, K.; Mazur, M.; Comb, D.; Koonin, E.; Slatko, B., The *Wolbachia* genome of *Brugia malayi*: endosymbiont evolution within a human pathogenic nematode. *PLoS Biol* **2005,** *3* (4), e121.

169. Saha, S.; Hunter, W. B.; Reese, J.; Morgan, J. K.; Marutani-Hert, M.; Huang, H.; Lindeberg, M., Survey of endosymbionts in the *Diaphorina citri* metagenome and assembly of a *Wolbachia w*Di draft genome. *PLoS One* **2012,** *7* (11), e50067.

170. Ramírez-Puebla, S. T.; Servín-Garcidueñas, L. E.; Ormeño-Orrillo, E.; Vera-Ponce de León, A.; Rosenblueth, M.; Delaye, L.; Martínez, J.; Martínez-Romero, E., Species in *Wolbachia*? Proposal for the designation of '*Candidatus Wolbachia bourtzisii*', '*Candidatus Wolbachia onchocercicola*', '*Candidatus Wolbachia blaxteri*', '*Candidatus Wolbachia brugii*', '*Candidatus Wolbachia taylori*', '*Candidatus Wolbachia collembolicola*' and '*Candidatus Wolbachia multihospitum*' for the different species within *Wolbachia* supergroups. *Syst Appl Microbiol* **2015,** *38* (6), 390-9.

171. Mosavi, L. K.; Cammett, T. J.; Desrosiers, D. C.; Peng, Z. Y., The ankyrin repeat as molecular architecture for protein recognition. *Protein Sci* **2004,** *13* (6), 1435-48.

172. Fenn, K.; Blaxter, M., Are filarial nematode *Wolbachia* obligate mutualist symbionts? *Trends Ecol Evol* **2004,** *19* (4), 163-6.

173. Guillemaud, T.; Pasteur, N.; Rousset, F., Contrasting levels of variability between cytoplasmic genomes and incompatibility types in the mosquito *§*. *Proc Biol Sci* **1997,** *264* (1379), 245-51.

174. Duron, O.; Boureux, A.; Echaubard, P.; Berthomieu, A.; Berticat, C.; Fort, P.; Weill, M., Variability and expression of ankyrin domain genes in *Wolbachia* variants infecting the mosquito Culex pipiens. *J Bacteriol* **2007,** *189* (12), 4442-8.

175. Sinkins, S. P.; Walker, T.; Lynd, A. R.; Steven, A. R.; Makepeace, B. L.; Godfray, H. C.; Parkhill, J., *Wolbachia* variability and host effects on crossing type in *Culex* mosquitoes. *Nature* **2005,** *436* (7048), 257-60.

176. Walker, T.; Klasson, L.; Sebaihia, M.; Sanders, M. J.; Thomson, N. R.; Parkhill, J.; Sinkins, S. P., Ankyrin repeat domain-encoding genes in the wPip strain of *Wolbachia* from the *Culex pipiens* group. *BMC Biol* **2007,** *5*, 39.

177. Iturbe-Ormaetxe, I.; Burke, G. R.; Riegler, M.; O'Neill, S. L., Distribution, expression, and motif variability of ankyrin domain genes in *Wolbachia* pipientis. *J Bacteriol* **2005,** *187* (15), 5136-45.

178. Yamada, R.; Iturbe-Ormaetxe, I.; Brownlie, J. C.; O'Neill, S. L., Functional test of the influence of *Wolbachia* genes on cytoplasmic incompatibility expression in *Drosophila melanogaster*. *Insect Mol Biol* **2011,** *20* (1), 75-85.

179. Williams, E. H.; Fields, S.; Saul, G. B., Transfer of incompatibility factors between stocks of *Nasonia* (= *Mormoniella*) *vitripennis*. *J Invertebr Pathol* **1993,** *61* (2), 206-10.

180. Sanogo, Y. O.; Eitam, A.; Dobson, S. L., No evidence for bacteriophage WO orf7 correlation with *Wolbachia*-induced cytoplasmic incompatibility in the *Culex pipiens* complex (*Culicidae*: *Diptera*). *J Med Entomol* **2005,** *42* (5), 789-94.

181. Duron, O.; Fort, P.; Weill, M., Hypervariable prophage WO sequences describe an unexpected high number of *Wolbachia* variants in the mosquito *Culex pipiens*. *Proc Biol Sci* **2006,** *273* (1585), 495-502.

182. Duron, O.; Bernard, C.; Unal, S.; Berthomieu, A.; Berticat, C.; Weill, M., Tracking factors modulating cytoplasmic incompatibilities in the mosquito *Culex pipiens*. *Mol Ecol* **2006,** *15* (10), 3061-71.

183. Gavotte, L.; Henri, H.; Stouthamer, R.; Charif, D.; Charlat, S.; Boulétreau, M.; Vavre, F., A Survey of the bacteriophage WO in the endosymbiotic bacteria *Wolbachia*. *Mol Biol Evol* **2007,** *24* (2), 427-35.

184. Saridaki, A.; Sapountzis, P.; Harris, H. L.; Batista, P. D.; Biliske, J. A.; Pavlikaki, H.; Oehler, S.; Savakis, C.; Braig, H. R.; Bourtzis, K., *Wolbachia* prophage DNA adenine methyltransferase genes in different *Drosophila*-*Wolbachia* associations. *PLoS One* **2011,** *6* (5), e19708.

185. Sasaki, T.; Massaki, N.; Kubo, T., *Wolbachia* variant that induces two distinct reproductive phenotypes in different hosts. *Heredity (Edinb)* **2005,** *95* (5), 389-93.

186. McGraw, E. A.; Merritt, D. J.; Droller, J. N.; O'Neill, S. L., *Wolbachia*-mediated sperm modification is dependent on the host genotype in *Drosophila*. *Proc Biol Sci* **2001,** *268* (1485), 2565-70.

187. Xi, Z.; Gavotte, L.; Xie, Y.; Dobson, S. L., Genome-wide analysis of the interaction between the endosymbiotic bacterium *Wolbachia* and its *Drosophila* host. *BMC Genomics* **2008,** *9*, 1.

188. Hurst, D.; Rylett, C. M.; Isaac, R. E.; Shirras, A. D., The *Drosophila* angiotensin-converting enzyme homologue Ance is required for spermiogenesis. *Dev Biol* **2003,** *254* (2), 238-47.

189. Clark, M. E.; Heath, B. D.; Anderson, C. L.; Karr, T. L., Induced paternal effects mimic cytoplasmic incompatibility in *Drosophila*. *Genetics* **2006,** *173* (2), 727-34.

190. Zheng, Y.; Ren, P. P.; Wang, J. L.; Wang, Y. F., *Wolbachia*-induced cytoplasmic incompatibility is associated with decreased *Hira* expression in male *Drosophila*. *PLoS One* **2011,** *6* (4), e19512.

191. Zheng, Y.; Wang, J. L.; Liu, C.; Wang, C. P.; Walker, T.; Wang, Y. F., Differentially expressed profiles in the larval testes of *Wolbachia* infected and uninfected *Drosophila*. *BMC Genomics* **2011,** *12*, 595.

192. Liu, C.; Wang, J. L.; Zheng, Y.; Xiong, E. J.; Li, J. J.; Yuan, L. L.; Yu, X. Q.; Wang, Y. F., *Wolbachia*-Induced Paternal Defect in *Drosophila* is likely by Interaction with the Juvenile Hormone Pathway. *Insect Biochem Mol Biol* **2014,** *49*, 49-58.

193. Beckmann, J. F.; Fallon, A. M., Detection of the *Wolbachia* Protein WPIP0282 in Mosquito Spermathecae: Implications for Cytoplasmic Incompatibility. *Insect Biochem Mol Biol* **2013,** *43* (9), 867-878.

194. Beckmann, J. F.; Markowski, T. W.; Witthuhn, B. A.; Fallon, A. M., Detection of the *Wolbachia*-encoded DNA binding protein, HU beta, in mosquito gonads. *Insect Biochem Mol Biol* **2013,** *43* (3), 272-9.

195. Yuan, L. L.; Chen, X.; Zong, Q.; Zhao, T.; Wang, J. L.; Zheng, Y.; Zhang, M.; Wang, Z.; Brownlie, J. C.; Yang, F.; Wang, Y. F., Quantitative Proteomic Analyses of Molecular Mechanisms Associated with Cytoplasmic Incompatibility in *Drosophila melanogaster* Induced by *Wolbachia*. *J Proteome Res* **2015**.

196. Hoffmann, A. A.; Montgomery, B. L.; Popovici, J.; Iturbe-Ormaetxe, I.; Johnson, P. H.; Muzzi, F.; Greenfield, M.; Durkan, M.; Leong, Y. S.; Dong, Y.; Cook, H.; Axford, J.; Callahan, A. G.; Kenny, N.; Omodei, C.; McGraw, E. A.; Ryan, P. A.; Ritchie, S. A.; Turelli, M.; O'Neill, S. L., Successful establishment of *Wolbachia* in *Aedes* populations to suppress dengue transmission. *Nature* **2011,** *476* (7361), 454-7.

197. Chauvatcharin, N.; Ahantarig, A.; Baimai, V.; Kittayapong, P., Bacteriophage WO-B and *Wolbachia* in natural mosquito hosts: infection incidence, transmission mode and relative density. *Mol Ecol* **2006,** *15* (9), 2451-61.

198. Iturbe-Ormaetxe, I.; Howie, J.; O'Neill, S. L., Development of *Wolbachia* transformation by homologous recombination. In *Progress report meeting for the grand challenges in human health grant, Heron Island, Queensland, Australia* **2007**.

199. Rasgon, J. L.; Ren, X.; Petridis, M., Can *Anopheles gambiae* be infected with *Wolbachia* pipientis? Insights from an in vitro system. *Appl Environ Microbiol* **2006,** *72* (12), 7718-22.

200. Hughes, G. L.; Pike, A. D.; Xue, P.; Rasgon, J. L., Invasion of *Wolbachia* into *Anopheles* and other insect germlines in an ex vivo organ culture system. *PLoS One* **2012**, *7*, e36277.

201. Jin, C.; Ren, X.; Rasgon, J. L., The virulent *Wolbachia* strain *w*MelPop efficiently establishes somatic infections in the malaria vector *Anopheles gambiae*. *Appl Environ Microbiol* **2009,** *75* (10), 3373-6.

202. Kambris, Z.; Blagborough, A. M.; Pinto, S. B.; Blagrove, M. S.; Godfray, H. C.; Sinden, R. E.; Sinkins, S. P., *Wolbachia* stimulates immune gene expression and inhibits plasmodium development in *Anopheles gambiae*. *PLoS Pathog* **2010,** *6* (10).

203. Beard, C. B.; O'Neill, S. L.; Mason, P.; Mandelco, L.; Woese, C. R.; Tesh, R. B.; Richards, F. F.; Aksoy, S., Genetic transformation and phylogeny of bacterial symbionts from tsetse. *Insect Mol Biol* **1993,** *1* (3), 123-31.

204. Turelli, M.; Hoffmann, A. A.; McKechnie, S. W., Dynamics of cytoplasmic incompatibility and mtDNA variation in natural *Drosophila simulans* populations. *Genetics* **1992,** *132* (3), 713-23.

205. Rasgon, J. L., Using predictive models to optimize *Wolbachia*-based strategies for vector-borne disease control. *Adv Exp Med Biol* **2008,** *627*, 114-25.

206. Allen, M. L.; O'Brochta, D. A.; Atkinson, P. W.; Levesque, C. S., Stable, germ-line transformation of *Culex quinquefasciatus* (*Diptera*: *Culicidae*). *J Med Entomol* **2001,** *38* (5), 701-10.

207. Catteruccia, F.; Nolan, T.; Loukeris, T. G.; Blass, C.; Savakis, C.; Kafatos, F. C.; Crisanti, A., Stable germline transformation of the malaria mosquito *Anopheles stephensi*. *Nature* **2000,** *405* (6789), 959-62.

208. Coates, C. J.; Jasinskiene, N.; Miyashiro, L.; James, A. A., Mariner transposition and transformation of the yellow fever mosquito, *Aedes aegypti*. *Proc Natl Acad Sci U S A* **1998,** *95* (7), 3748-51.

209. Atyame, C. M.; Cattel, J.; Lebon, C.; Flores, O.; Dehecq, J. S.; Weill, M.; Gouagna, L. C.; Tortosa, P., *Wolbachia*-Based Population Control Strategy Targeting *Culex quinquefasciatus* Mosquitoes Proves Efficient under Semi-Field Conditions. *PLoS One* **2015,** *10* (3), e0119288.

210. Chambers, E. W.; Hapairai, L.; Peel, B. A.; Bossin, H.; Dobson, S. L., Male Mating Competitiveness of a *Wolbachia*-Introgressed *Aedes polynesiensis* Strain under Semi-Field Conditions. *PLoS Negl Trop Dis* **2011,** *5* (8), e1271.

211. O'Connor, L.; Plichart, C.; Sang, A. C.; Brelsfoard, C. L.; Bossin, H. C.; Dobson, S. L., Open release of male mosquitoes infected with a *Wolbachia* biopesticide: field performance and infection containment. *PLoS Negl Trop Dis* **2012,** *6* (11), e1797.

212. Brelsfoard, C. L.; Dobson, S. L., *Wolbachia*-based straegies to control insect pests and disease vectors. *AsPac J Mol Biol Biotechnol* **2009,** *17* (3), 55-63.

213. Sharma, V. P.; Subarrao, S. K.; Adak, T.; Razdan, R. K., Integration of gamma irradiation and cytoplasmic incompatibility in *Culex pipiens fatigans* (*Diptera*: *Culicidae*). *J Med Entomol* **1979,** *15* (2), 155-6.

214. Arunachalam, N.; Curtis, C. F., Integration of radiation with cytoplasmic incompatibility for genetic control in the *Culex pipiens* complex (*Diptera*: Culicidae). *J Med Entomol* **1985,** *22* (6), 648-53.

215. Zhang, D.; Lees, R. S.; Xi, Z.; Gilles, J. R.; Bourtzis, K., Combining the Sterile Insect Technique with *Wolbachia*-Based Approaches: II- A Safer Approach to *Aedes albopictus* Population Suppression Programmes, Designed to Minimize the Consequences of Inadvertent Female Release. *PLoS One* **2015,** *10* (8), e0135194.

216. Hedges, L. M.; Brownlie, J. C.; O'Neill, S. L.; Johnson, K. N., *Wolbachia* and virus protection in insects. *Science* **2008,** *322* (5902), 702.

217. Teixeira, L.; Ferreira, A.; Ashburner, M., The bacterial symbiont *Wolbachia* induces resistance to RNA viral infections in *Drosophila melanogaster*. *PLoS Biol* **2008,** *6* (12), e2.

218. Osborne, S. E.; Leong, Y. S.; O'Neill, S. L.; Johnson, K. N., Variation in antiviral protection mediated by different *Wolbachia* strains in *Drosophila simulans*. *PLoS Pathog* **2009,** *5* (11), e1000656.

219. Panteleev, D. I. u.; Goriacheva, I. I.; Andrianov, B. V.; Reznik, N. L.; Lazebnyĭ, O. E.; Kulikov, A. M., [The endosymbiotic bacterium *Wolbachia* enhances the nonspecific resistance to insect pathogens and alters behavior of *Drosophila melanogaster*]. *Genetika* **2007,** *43* (9), 1277-80.

220. Glaser, R. L.; Meola, M. A., The native *Wolbachia* endosymbionts of *Drosophila melanogaster* and *Culex quinquefasciatus* increase host resistance to West Nile virus infection. *PLoS One* **2010,** *5* (8), e11977.

221. Mousson, L.; Zouache, K.; Arias-Goeta, C.; Raquin, V.; Mavingui, P.; Failloux, A. B., The Native *Wolbachia* Symbionts Limit Transmission of Dengue Virus in *Aedes albopictus*. *PLoS Negl Trop Dis* **2012,** *6*, e1989.

222. Bian, G.; Xu, Y.; Lu, P.; Xie, Y.; Xi, Z., The endosymbiotic bacterium *Wolbachia* induces resistance to dengue virus in *Aedes aegypti*. *PLoS Pathog* **2010,** *6* (4), e1000833.

223. Moreira, L. A.; Iturbe-Ormaetxe, I.; Jeffery, J. A.; Lu, G.; Pyke, A. T.; Hedges, L. M.; Rocha, B. C.; Hall-Mendelin, S.; Day, A.; Riegler, M.; Hugo, L. E.; Johnson, K. N.; Kay, B. H.; McGraw, E. A.; van den Hurk, A. F.; Ryan, P. A.; O'Neill, S. L., A *Wolbachia* symbiont in *Aedes aegypti* limits infection with dengue, Chikungunya, and *Plasmodium*. *Cell* **2009,** *139* (7), 1268-78.

224. Hussain, M.; Lu, G.; Torres, S.; Edmonds, J. H.; Kay, B. H.; Khromykh, A. A.; Asgari, S., Effect of *Wolbachia* on replication of West Nile Virus in mosquito cell line and adult mosquitoes. *J Virol* **2013,** *87* (2), 851-8.

225. Kambris, Z.; Cook, P. E.; Phuc, H. K.; Sinkins, S. P., Immune activation by life-shortening *Wolbachia* and reduced filarial competence in mosquitoes. *Science* **2009,** *326* (5949), 134-6.

226. Hughes, G. L.; Koga, R.; Xue, P.; Fukatsu, T.; Rasgon, J. L., *Wolbachia* infections are virulent and inhibit the human malaria parasite *Plasmodium falciparum* in *Anopheles gambiae*. *PLoS Pathog* **2011,** *7* (5), e1002043.

227. van den Hurk, A. F.; Hall-Mendelin, S.; Pyke, A. T.; Frentiu, F. D.; McElroy, K.; Day, A.; Higgs, S.; O'Neill, S. L., Impact of *Wolbachia* on Infection with Chikungunya and Yellow Fever Viruses in the Mosquito Vector *Aedes aegypti*. *PLoS Negl Trop Dis* **2012,** 6, e1892.

228. Ye, Y. H.; Carrasco, A. M.; Frentiu, F. D.; Chenoweth, S. F.; Beebe, N. W.; van den Hurk, A. F.; Simmons, C. P.; O'Neill, S. L.; McGraw, E. A., *Wolbachia* Reduces the Transmission Potential of Dengue-Infected *Aedes aegypti*. *PLoS Negl Trop Dis* **2015,** *9* (6), e0003894.

229. Frentiu, F. D.; Zakir, T.; Walker, T.; Popovici, J.; Pyke, A. T.; van den Hurk, A.; McGraw, E. A.; O'Neill, S. L., Limited Dengue Virus Replication in Field-Collected *Aedes aegypti* Mosquitoes Infected with *Wolbachia*. *PLoS Negl Trop Dis* **2014,** *8* (2), e2688.

230. Hoffmann, A. A.; Iturbe-Ormaetxe, I.; Callahan, A. G.; Phillips, B. L.; Billington, K.; Axford, J. K.; Montgomery, B.; Turley, A. P.; O'Neill, S. L., Stability of the wMel *Wolbachia* Infection following Invasion into *Aedes aegypti* Populations. *PLoS Negl Trop Dis* **2014,** *8* (9), e3115.

231. Lu, P.; Bian, G.; Pan, X.; Xi, Z., *Wolbachia* induces density-dependent inhibition to dengue virus in mosquito cells. *PLoS Negl Trop Dis* **2012,** *6* (7), e1754.

232. Zele, F.; Nicot, A.; Berthomieu, A.; Weill, M.; Duron, O.; Rivero, A., *Wolbachia* increases susceptibility to *Plasmodium* infection in a natural system. *Proc Biol Sci* **2014,** *281* (1779), 20132837.

233. Dodson, B. L.; Hughes, G. L.; Paul, O.; Matacchiero, A. C.; Kramer, L. D.; Rasgon, J. L., *Wolbachia* Enhances West Nile Virus (WNV) Infection in the Mosquito *Culex tarsalis*. *PLoS Negl Trop Dis* **2014,** *8* (7), e2965.

234. Hughes, G. L.; Vega-Rodriguez, J.; Xue, P.; Rasgon, J. L., *Wolbachia* strain *w*AlbB enhances infection by the rodent malaria parasite *Plasmodium berghei* in *Anopheles gambiae* mosquitoes. *Appl Environ Microbiol* **2012,** *78* (5), 1491-5.

235. Osborne, S. E.; Iturbe-Ormaetxe, I.; Brownlie, J. C.; O'Neill, S. L.; Johnson, K. N., Antiviral protection and the importance of *Wolbachia* density and tissue tropism in *Drosophila simulans*. *Appl Environ Microbiol* **2012,** *78* (19), 6922-9.

236. Chrostek, E.; Marialva, M. S.; Esteves, S. S.; Weinert, L. A.; Martinez, J.; Jiggins, F. M.; Teixeira, L., *Wolbachia* Variants Induce Differential Protection to Viruses in *Drosophila melanogaster*: A Phenotypic and Phylogenomic Analysis. *PLoS Genet* **2013,** *9* (12), e1003896.

237. Micieli, M. V.; Glaser, R. L., Somatic *Wolbachia* (*Rickettsiales*: *Rickettsiaceae*) levels in *Culex quinquefasciatus* and *Culex pipiens* (*Diptera*: *Culicidae*) and resistance to West Nile virus infection. *J Med Entomol* **2014,** *51* (1), 189-99.

238. Iturbe-Ormaetxe, I.; Walker, T.; SL, O. N., *Wolbachia* and the biological control of mosquito-borne disease. *EMBO Rep* **2011**.

239. Rolff, J.; Reynolds, S., In *Insect infection and immunity: evolution, ecology, and mechanisms* **2009**.

240. Zug, R.; Hammerstein, P., *Wolbachia* and the insect immune system: what reactive oxygen species can tell us about the mechanisms of *Wolbachia*-host interactions. *Front Microbiol* **2015,** *6*, 1201.

241. Wang, Y.; Oberley, L. W.; Murhammer, D. W., Evidence of oxidative stress following the viral infection of two lepidopteran insect cell lines. *Free Radic Biol Med* **2001,** *31* (11), 1448-55.

242. Skulachev, V. P., Possible role of reactive oxygen species in antiviral defense. *Biochemistry (Mosc)* **1998,** *63* (12), 1438-40.

243. Brennan, L. J.; Keddie, B. A.; Braig, H. R.; Harris, H. L., The endosymbiont *Wolbachia* pipientis induces the expression of host antioxidant proteins in an *Aedes albopictus* cell line. *PLoS One* **2008,** *3* (5), e2083.

244. Pan, X.; Zhou, G.; Wu, J.; Bian, G.; Lu, P.; Raikhel, A. S.; Xi, Z., *Wolbachia* induces reactive oxygen species (ROS)-dependent activation of the Toll pathway to control dengue virus in the mosquito *Aedes aegypti*. *Proc Natl Acad Sci U S A* **2012,** *109* (1), E23-31.

245. Wong, Z. S.; Hedges, L. M.; Brownlie, J. C.; Johnson, K. N., *Wolbachia*-mediated antibacterial protection and immune gene regulation in *Drosophila*. *PLoS One* **2011,** *6* (9), e25430.

246. Thomas, P.; Kenny, N.; Eyles, D.; Moreira, L. A.; O'Neill, S. L.; Asgari, S., Infection with the wMel and wMelPop strains of *Wolbachia* leads to higher levels of melanization in the hemolymph of *Drosophila melanogaster*, *Drosophila simulans* and *Aedes aegypti*. *Dev Comp Immunol* **201**1, *35* (3), 360-5.

247. Wong, Z. S.; Brownlie, J. C.; Johnson, K. N., Oxidative stress correlates with *Wolbachia*-mediated antiviral protection in *Wolbachia-Drosophila* associations. *Appl Environ Microbiol* **2015,** *81* (9), 3001-5.

248. Chrostek, E.; Marialva, M. S.; Yamada, R.; O'Neill, S. L.; Teixeira, L., High anti-viral protection without immune upregulation after interspecies *Wolbachia* transfer. *PLoS One* **2014,** *9* (6), e99025.

249. Ye, Y. H.; Woolfit, M.; Rances, E.; O'Neill, S. L.; McGraw, E. A., *Wolbachia*-Associated Bacterial Protection in the Mosquito *Aedes aegypti*. *PLoS Negl Trop Dis* **2013,** *7* (8), e2362.

250. Rances, E.; Ye, Y. H.; Woolfit, M.; McGraw, E. A.; O'Neill, S. L., The relative importance of innate immune priming in *Wolbachia*-mediated dengue interference. *PLoS Pathog* **2012,** *8* (2), e1002548.

251. Cho, K. O.; Kim, G. W.; Lee, O. K., *Wolbachia* bacteria reside in host Golgi-related vesicles whose position is regulated by polarity proteins. *PLoS One* **2011,** *6* (7), e22703.

252. Lin, M.; Rikihisa, Y., *Ehrlichia chaffeensis* and *Anaplasma phagocytophilum* lack genes for lipid A biosynthesis and incorporate cholesterol for their survival. *Infect Immun* **2003,** *71* (9), 5324-31.

253. Eaton, S., Multiple roles for lipids in the Hedgehog signalling pathway. *Nat Rev Mol Cell Biol* **2008,** *9* (6), 437-45.

254. Kabouridis, P. S.; Janzen, J.; Magee, A. L.; Ley, S. C., Cholesterol depletion disrupts lipid rafts and modulates the activity of multiple signaling pathways in T lymphocytes. *Eur J Immunol* **2000,** *30* (3), 954-63.

255. Feldlaufer, M. F.; Weirich, G. F.; Imberski, R. B.; Svoboda, J. A., Ecdysteroid production in *Drosophila melanogaster* reared on defined diets. *Insect Biochem Mol Biol* **1995,** *25* (6), 709-12.

256. Mackenzie, J. M.; Khromykh, A. A.; Parton, R. G., Cholesterol manipulation by West Nile virus perturbs the cellular immune response. *Cell Host Microbe* **2007,** *2* (4), 229-39.

257. Acosta, E. G.; Castilla, V.; Damonte, E. B., Functional entry of dengue virus into *Aedes albopictus* mosquito cells is dependent on clathrin-mediated endocytosis. *J Gen Virol* **2008,** *89* (Pt 2), 474-84.

258. Caragata, E. P.; Rances, E.; Hedges, L. M.; Gofton, A. W.; Johnson, K. N.; O'Neill, S. L.; McGraw, E. A., Dietary cholesterol modulates pathogen blocking by *Wolbachia*. *PLoS Pathog* **2013,** *9* (6), e1003459.

259. Heaton, N. S.; Perera, R.; Berger, K. L.; Khadka, S.; Lacount, D. J.; Kuhn, R. J.; Randall, G., Dengue virus nonstructural protein 3 redistributes fatty acid synthase to sites of viral replication and increases cellular fatty acid synthesis. *Proc Natl Acad Sci U S A* **2010,** *107* (40), 17345-50.

260. Heaton, N. S.; Randall, G., Dengue virus-induced autophagy regulates lipid metabolism. *Cell Host Microbe* **2010,** *8* (5), 422-32.

261. Perera, R.; Riley, C.; Isaac, G.; Hopf-Jannasch, A. S.; Moore, R. J.; Weitz, K. W.; Pasa-Tolic, L.; Metz, T. O.; Adamec, J.; Kuhn, R. J., Dengue virus infection perturbs lipid homeostasis in infected mosquito cells. *PLoS Pathog* **2012,** *8* (3), e1002584.

262. Samsa, M. M.; Mondotte, J. A.; Iglesias, N. G.; Assunção-Miranda, I.; Barbosa-Lima, G.; Da Poian, A. T.; Bozza, P. T.; Gamarnik, A. V., Dengue virus capsid protein usurps lipid droplets for viral particle formation. *PLoS Pathog* **2009,** *5* (10), e1000632.

263. Krejbich-Trotot, P.; Gay, B.; Li-Pat-Yuen, G.; Hoarau, J. J.; Jaffar-Bandjee, M. C.; Briant, L.; Gasque, P.; Denizot, M., Chikungunya triggers an autophagic process which promotes viral replication. *Virol J* **2011,** *8*, 432.

264. Voronin, D.; Cook, D. A.; Steven, A.; Taylor, M. J., Autophagy regulates *Wolbachia* populations across diverse symbiotic associations. *Proc Natl Acad Sci U S A* **2012,** *109* (25), E1638-46.

265. Bull, J. J.; Turelli, M., *Wolbachia* versus dengue: Evolutionary forecasts. *Evol Med Public Health* **2013,** *2013* (1), 197-207.

266. Sim, S.; Cirimotich, C.; Ramirez, J.; Souza-Neto, J.; Dimopoulos, G., Dengue Virus–Mosquito Interactions and Molecular Methods of Vector Control. In *Dengue and Dengue Hemorrhagic Fever*, CABI: **2014**.

267. Zhang, G.; Hussain, M.; O'Neill, S. L.; Asgari, S., *Wolbachia* uses a host microRNA to regulate transcripts of a methyltransferase, contributing to dengue viral inhibition in *Aedes aegypti*. *Proc Natl Acad Sci U S A* **2013,** *110* (25), 10276-81.

268. Hoffmann, A. A.; Clancy, D.; Duncan, J., Naturally-occurring *Wolbachia* infection in *Drosophila simulans* that does not cause cytoplasmic incompatibility. *Heredity (Edinb)* **1996,** *76 (Pt 1)*, 1-8.

269. James, A. C.; Ballard, J. W., Expression of cytoplasmic incompatibility in *Drosophila simulans* and its impact on infection frequencies and distribution of *Wolbachia pipientis*. *Evolution* **2000,** *54* (5), 1661-72.

270. Charlat, S.; Le Chat, L.; Merçot, H., Characterization of non-cytoplasmic incompatibility inducing *Wolbachia* in two continental African populations of *Drosophila simulans*. *Heredity (Edinb)* **2003,** *90* (1), 49-55.

271. Kriesner, P.; Hoffmann, A. A.; Lee, S. F.; Turelli, M.; Weeks, A. R., Rapid sequential spread of two *Wolbachia* variants in *Drosophila simulans*. *PLoS Pathog* **2013,** *9* (9), e1003607.

272. Sutton, E. R.; Harris, S. R.; Parkhill, J.; Sinkins, S. P., Comparative genome analysis of *Wolbachia* strain *w*Au. *BMC Genomics* **2014,** *15* (1), 928.

273. Choi, J. Y.; Bubnell, J. E.; Aquadro, C. F., Population genomics of infectious and integrated *Wolbachia* pipientis genomes in Drosophila ananassae. *Genome Biol Evol* **2015,** *7* (8), 2362-82.

274. Duplouy, A.; Iturbe-Ormaetxe, I.; Beatson, S.A.; Szubert, J.M.; Brownlie, J.C.; McMeniman C.J.; McGraw, E.A.; Hurst G.D.D.; Charlat, S.; O'Neill S.L.; Woolfit, M., Draft genome sequence of the male-killing *Wolbachia* strain *w*Bol1 reveals recent horizontal gene transfers from diverse sources. *BMC Genomics* **2013,** *14*, 20.

275. Iturbe-Ormaetxe, I.; Woolfit, M.; Rancès, E.; Duplouy, A.; O'Neill, S. L., A simple protocol to obtain highly pure *Wolbachia* endosymbiont DNA for genome sequencing. *J Microbiol Methods* **2011,** *84* (1), 134-6.

276. Livak, K. J., Organization and mapping of a sequence on the *Drosophila melanogaster* X and Y chromosomes that is transcribed during spermatogenesis. *Genetics* **1984,** *107* (4), 611-34.

277. Ye, J.; Coulouris, G.; Zaretskaya, I.; Cutcutache, I.; Rozen, S.; Madden, T. L., Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics* **2012,** *13*, 134.

278. Quail, M. A.; Kozarewa, I.; Smith, F.; Scally, A.; Stephens, P. J.; Durbin, R.; Swerdlow, H.; Turner, D. J., A large genome center's improvements to the Illumina sequencing system. *Nat Methods* **2008,** *5* (12), 1005-10.

279. Quail, M. A.; Swerdlow, H.; Turner, D. J., Improved protocols for the illumina genome analyzer sequencing system. *Curr Protoc Hum Genet* **2009,** *Chapter 18*, Unit 18.2.

280. Quail, M. A.; Otto, T. D.; Gu, Y.; Harris, S. R.; Skelly, T. F.; McQuillan, J. A.; Swerdlow, H. P.; Oyola, S. O., Optimal enzymes for amplifying sequencing libraries. *Nat Methods* **2012,** *9* (1), 10-1.

281. Zerbino, D. R.; Birney, E., Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* **2008,** *18* (5), 821-9.

282. Chin, C. S.; Alexander, D. H.; Marks, P.; Klammer, A. A.; Drake, J.; Heiner, C.; Clum, A.; Copeland, A.; Huddleston, J.; Eichler, E. E.; Turner, S. W.; Korlach, J., Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat Methods* **2013,** *10* (6), 563-9.

283. SMALT. http://www.sanger.ac.uk/resources/software/smalt.

284. Otto, T. D.; Sanders, M.; Berriman, M.; Newbold, C., Iterative Correction of Reference Nucleotides (iCORN) using second generation sequencing technology. *Bioinformatics* **2010,** *26* (14), 1704-7.

285. Seemann, T., Prokka: Prokaryotic Genome Annotation System. http://vicbioinformatics.com.

286. Nawrocki, E. P.; Eddy, S. R., Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* **2013,** *29* (22), 2933-5.

287. Laslett, D.; Canback, B., ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res* **2004,** *32* (1), 11-6.

288. Lagesen, K.; Hallin, P.; Rødland, E.; Stærfeldt, H.; Rognes, T.; Ussery, D., RNammer: consistent annotation of rRNA genes in genomic sequences. *Nucleic Acids Res* **2007,** *35* (9), 3100-8.

289. Hyatt, D.; Chen, G. L.; Locascio, P. F.; Land, M. L.; Larimer, F. W.; Hauser, L. J., Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **2010,** *11*, 119.

290. Pruitt, K. D.; Tatusova, T.; Klimke, W.; Maglott, D. R., NCBI Reference Sequences: current status, policy and new initiatives. *Nucleic Acids Res* **2009,** *37* (Database issue), D32-6.

291. Fu, L.; Niu, B.; Zhu, Z.; Wu, S.; Li, W., CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **2012,** *28* (23), 3150-2.

292. UniProt Consortium, Update on activities at the Universal Protein Resource (UniProt) in 2013. *Nucleic Acids Res* **2013,** *41* (Database issue), D43-7.

293. Carver, T. J.; Rutherford, K. M.; Berriman, M.; Rajandream, M. A.; Barrell, B. G.; Parkhill, J., ACT: the Artemis Comparison Tool. *Bioinformatics* **2005,** *21* (16), 3422-3.

294. Kearse, M.; Moir, R.; Wilson, A.; Stones-Havas, S.; Cheung, M.; Sturrock, S.; Buxton, S.; Cooper, A.; Markowitz, S.; Duran, C.; Thierer, T.; Ashton, B.; Meintjes, P.; Drummond, A., Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **2012,** *28* (12), 1647-9.

295. Hunter, S.; Jones, P.; Mitchell, A.; Apweiler, R.; Attwood, T. K.; Bateman, A.; Bernard, T.; Binns, D.; Bork, P.; Burge, S.; de Castro, E.; Coggill, P.; Corbett, M.; Das, U.; Daugherty, L.; Duquenne, L.; Finn, R. D.; Fraser, M.; Gough, J.; Haft, D.; Hulo, N.; Kahn, D.; Kelly, E.; Letunic, I.; Lonsdale, D.; Lopez, R.; Madera, M.; Maslen, J.; McAnulla, C.; McDowall, J.; McMenamin, C.; Mi, H.; Mutowo-Muellenet, P.; Mulder, N.; Natale, D.; Orengo, C.; Pesseat, S.; Punta, M.; Quinn, A. F.; Rivoire, C.; Sangrador-Vegas, A.; Selengut, J. D.; Sigrist, C. J.; Scheremetjew, M.; Tate, J.; Thimmajanarthanan, M.; Thomas, P. D.; Wu, C. H.; Yeats, C.; Yong, S. Y., InterPro in 2011: new developments in the family and domain prediction database. *Nucleic Acids Res* **2012,** *40* (Database issue), D306-12.

296. SNAP. http://www.hiv.lanl.gov.

297. Carter, G. C.; Bernstone, L.; Sangani, D.; Bee, J. W.; Harder, T.; James, W., HIV entry in macrophages is dependent on intact lipid rafts. *Virology* **2009,** *386* (1), 192-202.

298. Ashburner, M.; Ball, C. A.; Blake, J. A.; Botstein, D.; Butler, H.; Cherry, J. M.; Davis, A. P.; Dolinski, K.; Dwight, S. S.; Eppig, J. T.; Harris, M. A.; Hill, D. P.; Issel-Tarver, L.; Kasarskis, A.; Lewis, S.; Matese, J. C.; Richardson, J. E.; Ringwald, M.; Rubin, G. M.; Sherlock, G., Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **2000,** *25* (1), 25-9.

299. Sonnhammer, E. L.; Durbin, R., A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis. *Gene* **1995,** *167* (1-2), GC1-10.

300. Brand, A. H.; Perrimon, N., Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* **1993,** *118* (2), 401-15.

301. O'Malley, R. C.; Alonso, J. M.; Kim, C. J.; Leisse, T. J.; Ecker, J. R., An adapter ligation-mediated PCR method for high-throughput mapping of T-DNA inserts in the *Arabidopsis* genome. *Nat Protoc* **2007,** *2* (11), 2910-7.

302. Wang, C.; Dickinson, L. K.; Lehmann, R., Genetics of *nanos* localization in *Drosophila*. *Dev Dyn* **1994,** *199* (2), 103-15.

303. Forrest, K. M.; Gavis, E. R., Live imaging of endogenous RNA reveals a diffusion and entrapment mechanism for nanos mRNA localization in *Drosophila*. *Curr Biol* **2003**, *13*, 1159-68.

304. McKearin, D. M.; Spradling, A. C., *bag-of-marbles*: a *Drosophila* gene required to initiate both male and female gametogenesis. *Genes Dev* **1990,** *4* (12B), 2242-51.

305. Yamada, R.; Floate, K. D.; Riegler, M.; O'Neill, S. L., Male development time influences the strength of *Wolbachia*-induced cytoplasmic incompatibility expression in *Drosophila melanogaster*. *Genetics* **2007,** *177* (2), 801-8.

306. Hoffmann, A. A.; Turelli, M.; Harshman, L. G., Factors affecting the distribution of cytoplasmic incompatibility in *Drosophila simulans*. *Genetics* **1990,** *126* (4), 933-48.

307. Wilcoxon-Mann-Whitney Test Calculator. http://www.ccb.uni-saarland.de/software/wtest.

308. Kruskal-Wallis Test Spreadsheet. udel.edu/~mcdonald/statkruskalwallis.xls.

309. Masui, S.; Kuroiwa, H.; Sasaki, T.; Inui, M.; Kuroiwa, T.; Ishikawa, H., Bacteriophage WO and virus-like particles in *Wolbachia*, an endosymbiont of arthropods. *Biochem Biophys Res Commun* **2001,** *283* (5), 1099-104.

310. Ross, W.; Gosink, K. K.; Salomon, J.; Igarashi, K.; Zou, C.; Ishihama, A.; Severinov, K.; Gourse, R. L., A third recognition element in bacterial promoters: DNA binding by the alpha subunit of RNA polymerase. *Science* **1993,** *262* (5138), 1407-13.

311. Brownlie, J. C.; Adamski, M.; Slatko, B.; McGraw, E. A., Diversifying selection and host adaptation in two endosymbiont genomes. *BMC Evol Biol* **2007,** *7*, 68.

312. Chrostek E.; Teixeira L., Mutualism Breakdown by Amplification of *Wolbachia* Genes. *PLoS Biology* **2015,** *13* (2), e1002065.

313. Quail, M. A.; Smith, M.; Coupland, P.; Otto, T. D.; Harris, S. R.; Connor, T. R.; Bertoni, A.; Swerdlow, H. P.; Gu, Y., A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics* **2012,** *13*, 341.

314. Au, K. F.; Underwood, J. G.; Lee, L.; Wong, W. H., Improving PacBio long read accuracy by short read alignment. *PLoS One* **2012,** *7* (10), e46679.

315. Drost, J. B.; Lee, W. R., Biological basis of germline mutation: comparisons of spontaneous germline mutation rates among *Drosophila*, mouse, and human. *Environ Mol Mutagen* **1995,** *25 Suppl 26*, 48-64.

316. Hoffmann, A. A.; Turelli, M., Unidirectional incompatibility in *Drosophila simulans*: inheritance, geographic variation and fitness effects. *Genetics* **1988,** *119* (2), 435-44.

317. Darby, A. C.; Gill, A. C.; Armstrong, S. D.; Hartley, C. S.; Xia, D.; Wastling, J. M.; Makepeace, B. L., Integrated transcriptomic and proteomic analysis of the global response of *Wolbachia* to doxycycline-induced stress. *ISME J* **2014,** *8* (4), 925-37.

318. Mayoral, J. G.; Hussain, M.; Joubert, D. A.; Iturbe-Ormaetxe, I.; O'Neill, S. L.; Asgari, S., *Wolbachia* small noncoding RNAs and their role in cross-kingdom communications. *Proc Natl Acad Sci U S A* **2014,** *111* (52), 18721-6.

319. Mayoral, J. G.; Etebari, K.; Hussain, M.; Khromykh, A. A.; Asgari, S., *Wolbachia* Infection Modifies the Profile, Shuttling and Structure of MicroRNAs in a Mosquito Cell Line. *PLoS One* **2014,** *9* (4), e96107.

320. Woolfit, M.; Algama, M.; Keith, J. M.; McGraw, E. A.; Popovici, J., Discovery of Putative Small Non-Coding RNAs from the Obligate Intracellular Bacterium *Wolbachia pipientis*. *PLoS One* **2015,** *10* (3), e0118595.

321. Darby, A. C.; Armstrong, S. D.; Bah, G. S.; Kaur, G.; Hughes, M. A.; Kay, S. M.; Koldkjær, P.; Rainbow, L.; Radford, A. D.; Blaxter, M. L.; Tanya, V. N.; Trees, A. J.; Cordaux, R.; Wastling, J. M.; Makepeace, B. L., Analysis of gene expression from the *Wolbachia* genome of a filarial nematode supports both metabolic and defensive roles within the symbiosis. *Genome Res* **2012,** *22* (12), 2467-77.

322. Gutzwiller, F.; Carmo, C. R.; Miller, D. E.; Rice, D. W.; Newton, I. L.; Hawley, R. S.; Teixeira, L.; Bergman, C. M., Dynamics of *Wolbachia* pipientis Gene Expression Across the *Drosophila melanogaster* Life Cycle. *G3 (Bethesda)* **2015,** *5* (12), 2843-56.

323. Graveley, B. R.; Brooks, A. N.; Carlson, J. W.; Duff, M. O.; Landolin, J. M.; Yang, L.; Artieri, C. G.; van Baren, M. J.; Boley, N.; Booth, B. W.; Brown, J. B.; Cherbas, L.; Davis, C. A.; Dobin, A.; Li, R.; Lin, W.; Malone, J. H.; Mattiuzzo, N. R.; Miller, D.; Sturgill, D.; Tuch, B. B.; Zaleski, C.; Zhang, D.; Blanchette, M.; Dudoit, S.; Eads, B.; Green, R. E.; Hammonds, A.; Jiang, L.; Kapranov, P.; Langton, L.; Perrimon, N.; Sandler, J. E.; Wan, K. H.; Willingham, A.; Zhang, Y.; Zou, Y.; Andrews, J.; Bickel, P. J.; Brenner, S. E.; Brent, M. R.; Cherbas, P.; Gingeras, T. R.; Hoskins, R. A.; Kaufman, T. C.; Oliver, B.; Celniker, S. E., The developmental transcriptome of *Drosophila melanogaster*. *Nature* **2011,** *471* (7339), 473-9.

324. Brown, J. B.; Boley, N.; Eisman, R.; May, G. E.; Stoiber, M. H.; Duff, M. O.; Booth, B. W.; Wen, J.; Park, S.; Suzuki, A. M.; Wan, K. H.; Yu, C.; Zhang, D.; Carlson, J. W.; Cherbas, L.; Eads, B. D.; Miller, D.; Mockaitis, K.; Roberts, J.; Davis, C. A.; Frise, E.; Hammonds, A. S.; Olson, S.; Shenker, S.; Sturgill, D.; Samsonova, A. A.; Weiszmann, R.; Robinson, G.; Hernandez, J.; Andrews, J.; Bickel, P. J.; Carninci, P.; Cherbas, P.; Gingeras, T. R.; Hoskins, R. A.; Kaufman, T. C.; Lai, E. C.; Oliver, B.; Perrimon, N.; Graveley, B. R.; Celniker, S. E., Diversity and dynamics of the *Drosophila transcriptome*. *Nature* **2014,** *512* (7515), 393-9.

325. Duff, M. O.; Olson, S.; Wei, X.; Garrett, S. C.; Osman, A.; Bolisetty, M.; Plocik, A.; Celniker, S. E.; Graveley, B. R., Genome-wide identification of zero nucleotide recursive splicing in *Drosophila*. *Nature* **2015,** *521* (7552), 376-9.

326. Luck, A. N.; Anderson, K. G.; McClung, C. M.; VerBerkmoes, N. C.; Foster, J. M.; Michalski, M. L.; Slatko, B. E., Tissue-specific transcriptomics and proteomics of a filarial nematode and its *Wolbachia* endosymbiont. *BMC Genomics* **2015,** *16* (1), 920.

327. Sim, S.; Jupatanakul, N.; Ramirez, J. L.; Kang, S.; Romero-Vivas, C. M.; Mohammed, H.; Dimopoulos, G., Transcriptomic profiling of diverse *Aedes aegypti* strains reveals increased basal-level immune activation in dengue virus-refractory populations and identifies novel virus-vector molecular interactions. *PLoS Negl Trop Dis* **2013,** *7* (7), e2295.

328. Voronin, D.; Tran-Van, V.; Potier, P.; Mavingui, P., Transinfection and growth discrepancy of *Drosophila Wolbachia* strain *w*Mel in cell lines of the mosquito *Aedes albopictus*. *J Appl Microbiol* **2010,** *108* (6), 2133-41.

329. Niu, H.; Xiong, Q.; Yamamoto, A.; Hayashi-Nishino, M.; Rikihisa, Y., Autophagosomes induced by a bacterial Beclin 1 binding protein facilitate obligatory intracellular infection. *Proc Natl Acad Sci U S A* **2012,** *109* (51), 20800-7.

330. Hawley, W. A., The biology of *Aedes albopictus*. *J Am Mosq Control Assoc Suppl* **1988,** *1*, 1-39.

331. Y, O.; H, T., Elimination of *Wolbachia pipientis* from *Aedes albopictus*. *Medical entomology and zoology* **1997,** *48* (3), 257-60.

332. Blagrove, M. S.; Arias-Goeta, C.; Di Genua, C.; Failloux, A. B.; Sinkins, S. P., A *Wolbachia w*Mel transinfection in *Aedes albopictus* is not detrimental to host fitness and inhibits Chikungunya virus. *PLoS Negl Trop Dis* **2013,** *7* (3), e2152.

333. FlyBase. flybase.org.

334. GenBank. http://www.ncbi.nlm.nih.gov/genbank.

335. Langmead, B.; Salzberg, S. L., Fast gapped-read alignment with Bowtie 2. *Nat Methods* **2012,** *9* (4), 357-9.

336. Kim, D.; Pertea, G.; Trapnell, C.; Pimentel, H.; Kelley, R.; Salzberg, S. L., TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* **2013,** *14* (4), R36.

337. Broad Institute, Picard Tools. http://picard.sourceforge.net.

338. Andrews, S., FastQC A Quality Control tool for High Throughput Sequence Data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

339. Martin, M., Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **2011,** *17* (1), 10-2.

340. Trapnell, C.; Williams, B. A.; Pertea, G.; Mortazavi, A.; Kwan, G.; van Baren, M. J.; Salzberg, S. L.; Wold, B. J.; Pachter, L., Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* **2010,** *28* (5), 511-5.

341. Roberts, A.; Pimentel, H.; Trapnell, C.; Pachter, L., Identification of novel transcripts in annotated genomes using RNA-Seq. *Bioinformatics* **2011,** *27* (17), 2325-9.

342. Trapnell, C.; Roberts, A.; Goff, L.; Pertea, G.; Kim, D.; Kelley, D. R.; Pimentel, H.; Salzberg, S. L.; Rinn, J. L.; Pachter, L., Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* **2012,** *7* (3), 562-78.

343. Haas, B. J.; Papanicolaou, A.; Yassour, M.; Grabherr, M.; Blood, P. D.; Bowden, J.; Couger, M. B.; Eccles, D.; Li, B.; Lieber, M.; Macmanes, M. D.; Ott, M.; Orvis, J.; Pochet, N.; Strozzi, F.; Weeks, N.; Westerman, R.; William, T.; Dewey, C. N.; Henschel, R.; Leduc, R. D.; Friedman, N.; Regev, A., *De novo* transcript sequence reconstruction from RNA-seq using

the Trinity platform for reference generation and analysis. *Nat Protoc* **2013,** *8* (8), 1494-512.

344. Kumar, S., contig_stats.pl. https://github.com/joshuaburkhart/bio/blob/master/contig_stats.pl.

345. Li, B.; Fillmore, N.; Bai, Y.; Collins, M.; Thomson, J. A.; Stewart, R.; Dewey, C. N., Evaluation of de novo transcriptome assemblies from RNA-Seq data. *Genome Biol* **2014,** *15* (12), 553.

346. Trapnell, C.; Hendrickson, D. G.; Sauvageau, M.; Goff, L.; Rinn, J. L.; Pachter, L., Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat Biotechnol* **2013,** *31* (1), 46-53.

347. Goff, L. A.; Trapnell, C.; Kelley, D. CummeRbund: Visualization and Exploration of Cufflinks High-throughput Sequencing Data, **2013**.

348. Conesa, A.; Götz, S.; García-Gómez, J. M.; Terol, J.; Talón, M.; Robles, M., Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **2005,** *21* (18), 3674-6.

349. Carlson, M. GO.db: A set of annotation maps describing the entire Gene Ontology. R package version 3.0.0.

350. Sessions, O. M.; Barrows, N. J.; Souza-Neto, J. A.; Robinson, T. J.; Hershey, C. L.; Rodgers, M. A.; Ramirez, J. L.; Dimopoulos, G.; Yang, P. L.; Pearson, J. L.; Garcia-Blanco, M. A., Discovery of insect and human dengue virus host factors. *Nature* **2009,** *458* (7241), 1047-50.

351. Young, M. D.; Wakefield, M. J.; Smyth, G. K.; Oshlack, A., Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol* **2010,** *11* (2), R14.

352. Kanehisa, M.; Goto, S., KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* **2000,** *28* (1), 27-30.

353. Kanehisa, M.; Goto, S.; Sato, Y.; Kawashima, M.; Furumichi, M.; Tanabe, M., Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res* **2014,** *42* (Database issue), D199-205.

354. Zhu, Y.; Li, M.; Sousa, A. M.; Sestan, N., XSAnno: a framework for building ortholog models in cross-species transcriptome comparisons. *BMC Genomics* **2014,** *15*, 343.

355. Derrien, T.; Estellé, J.; Marco Sola, S.; Knowles, D. G.; Raineri, E.; Guigó, R.; Ribeca, P., Fast computation and applications of genome mappability. *PLoS One* **2012,** *7* (1), e30377.

356. Chen, X. G.; Jiang, X.; Gu, J.; Xu, M.; Wu, Y.; Deng, Y.; Zhang, C.; Bonizzoni, M.; Dermauw, W.; Vontas, J.; Armbruster, P.; Huang, X.; Yang, Y.; Zhang, H.; He, W.; Peng, H.; Liu, Y.; Wu, K.; Chen, J.; Lirakis, M.; Topalis, P.; Van Leeuwen, T.; Hall, A. B.; Thorpe, C.; Mueller, R. L.; Sun, C.; Waterhouse, R. M.; Yan, G.; Tu, Z. J.; Fang, X.; James, A. A., Genome sequence of the Asian Tiger mosquito, *Aedes albopictus*, reveals insights into its biology, genetics, and evolution. *Proc Natl Acad Sci U S A* **2015,** *112* (44), e5907-15.

357. Luck, A. N.; Evans, C. C.; Riggs, M. D.; Foster, J. M.; Moorhead, A. R.; Slatko, B. E.; Michalski, M. L., Concurrent transcriptional profiling of *Dirofilaria immitis* and its *Wolbachia* endosymbiont throughout the nematode life cycle reveals coordinated gene expression. *BMC Genomics* **2014,** *15* (1), 1041.

358. Miller, W. J.; Riegler, M., Evolutionary dynamics of wAu-like *Wolbachia* variants in neotropical *Drosophila* spp. *Appl Environ Microbiol* **2006,** *72* (1), 826-35.

359. Hedstrom, L., Serine protease mechanism and specificity. *Chem Rev* **2002,** *102* (12), 4501-24.

360. Madala, P. K.; Tyndall, J. D.; Nall, T.; Fairlie, D. P., Update 1 of: Proteases universally recognize beta strands in their active sites. *Chem Rev* **2010,** *110* (6), PR1-31.

361. Barrett, A. J., Classification of peptidases. *Methods Enzymol* **1994,** *244*, 1-15.

362. Rawlings, N. D.; Barrett, A. J., Families of serine peptidases. *Methods Enzymol* **1994,** *244*, 19-61.

363. Zdobnov, E. M.; von Mering, C.; Letunic, I.; Torrents, D.; Suyama, M.; Copley, R. R.; Christophides, G. K.; Thomasova, D.; Holt, R. A.; Subramanian, G. M.; Mueller, H. M.; Dimopoulos, G.; Law, J. H.; Wells, M. A.; Birney, E.; Charlab, R.; Halpern, A. L.; Kokoza, E.; Kraft, C. L.; Lai, Z.; Lewis, S.; Louis, C.; Barillas-Mury, C.; Nusskern, D.; Rubin, G. M.; Salzberg, S. L.; Sutton, G. G.; Topalis, P.; Wides, R.; Wincker, P.; Yandell, M.; Collins, F. H.; Ribeiro, J.; Gelbart, W. M.; Kafatos, F. C.; Bork, P., Comparative genome and proteome analysis of *Anopheles gambiae* and *Drosophila melanogaster*. *Science* **2002,** *298* (5591), 149-59.

364. Shah, P. K.; Tripathi, L. P.; Jensen, L. J.; Gahnim, M.; Mason, C.; Furlong, E. E.; Rodrigues, V.; White, K. P.; Bork, P.; Sowdhamini, R., Enhanced function annotations for *Drosophila* serine proteases: a case study for systematic annotation of multi-member gene families. *Gene* **2008,** *407* (1-2), 199-215.

365. Findlay, G. D.; Yi, X.; Maccoss, M. J.; Swanson, W. J., Proteomics reveals novel *Drosophila* seminal fluid proteins transferred at mating. *PLoS Biol* **2008,** *6* (7), e178.

366. Mueller, J. L.; Ravi Ram, K.; McGraw, L. A.; Bloch Qazi, M. C.; Siggia, E. D.; Clark, A. G.; Aquadro, C. F.; Wolfner, M. F., Cross-species comparison of *Drosophila* male accessory gland protein genes. *Genetics* **2005,** *171* (1), 131-43.

367. Gubb, D.; Sanz-Parra, A.; Barcena, L.; Troxler, L.; Fullaondo, A., Protease inhibitors and proteolytic signalling cascades in insects. *Biochimie* **2010,** *92* (12), 1749-59.

368. Monsma, S. A.; Harada, H. A.; Wolfner, M. F., Synthesis of two *Drosophila* male accessory gland proteins and their fate after transfer to the female during mating. *Dev Biol* **1990,** *142* (2), 465-75.

369. Park, M.; Wolfner, M. F., Male and female cooperate in the prohormone-like processing of a *Drosophila melanogaster* seminal fluid protein. *Dev Biol* **1995,** *171* (2), 694-702.

370. Peng, J.; Chen, S.; Büsser, S.; Liu, H.; Honegger, T.; Kubli, E., Gradual release of sperm bound sex-peptide controls female postmating behavior in *Drosophila*. *Curr Biol* **2005,** *15* (3), 207-13.

371. Ravi Ram, K.; Sirot, L. K.; Wolfner, M. F., Predicted seminal astacin-like protease is required for processing of reproductive proteins in *Drosophila melanogaster*. *Proc Natl Acad Sci U S A* **2006,** *103* (49), 18674-9.

372. Pilpel, N.; Nezer, I.; Applebaum, S. W.; Heifetz, Y., Mating-increases trypsin in female *Drosophila* hemolymph. *Insect Biochem Mol Biol* **2008,** *38* (3), 320-30.

373. Lawniczak, M. K.; Begun, D. J., A genome-wide analysis of courting and mating responses in *Drosophila melanogaster* females. *Genome* **2004,** *47* (5), 900-10.

374. Braswell, W. E.; Andrés, J. A.; Maroja, L. S.; Harrison, R. G.; Howard, D. J.; Swanson, W. J., Identification and comparative analysis of accessory gland proteins in *Orthoptera*. *Genome* **2006,** *49* (9), 1069-80.

375. Sirot, L. K.; Poulson, R. L.; McKenna, M. C.; Girnary, H.; Wolfner, M. F.; Harrington, L. C., Identity and transfer of male reproductive gland proteins of the dengue vector mosquito, *Aedes aegypti*: potential tools for control of female feeding and reproduction. *Insect Biochem Mol Biol* **2008,** *38* (2), 176-89.

376. Dacheux, J. L.; Gatti, J. L.; Dacheux, F., Contribution of epididymal secretory proteins for spermatozoa maturation. *Microsc Res Tech* **2003,** *61* (1), 7-17.

377. Aarts, M. G.; Hodge, R.; Kalantidis, K.; Florack, D.; Wilson, Z. A.; Mulligan, B. J.; Stiekema, W. J.; Scott, R.; Pereira, A., The *Arabidopsis* MALE STERILITY 2 protein shares similarity with reductases in elongation/condensation complexes. *Plant J* **1997,** *12* (3), 615-23.

378. Molloy, J. C.; Sinkins, S. P., *Wolbachia* Do Not Induce Reactive Oxygen Species-Dependent Immune Pathway Activation in *Aedes albopictus*. *Viruses* **2015,** *7* (8), 4624-39.

379. Beutler, B., Innate immunity: an overview. *Mol Immunol* **2004,** *40* (12), 845-59.

380. Iwanaga, S.; Lee, B. L., Recent advances in the innate immunity of invertebrate animals. *J Biochem Mol Biol* **2005,** *38* (2), 128-50.

381. Chauhan, C.; Behura, S. K.; Debruyn, B.; Lovin, D. D.; Harker, B. W.; Gomez-Machorro, C.; Mori, A.; Romero-Severson, J.; Severson, D. W., Comparative expression profiles of midgut genes in dengue virus refractory and susceptible *Aedes aegypti* across critical period for virus infection. *PLoS One* **2012,** *7* (10), e47350.

382. Oliveira, J. H.; Gonçalves, R. L.; Lara, F. A.; Dias, F. A.; Gandara, A. C.; Menna-Barreto, R. F.; Edwards, M. C.; Laurindo, F. R.; Silva-Neto, M. A.; Sorgine, M. H.; Oliveira, P. L., Blood meal-derived heme decreases ROS levels in the midgut of *Aedes aegypti* and allows proliferation of intestinal microbiota. *PLoS Pathog* **2011,** *7* (3), e1001320.

383. Kumar, S.; Molina-Cruz, A.; Gupta, L.; Rodrigues, J.; Barillas-Mury, C., A peroxidase/dual oxidase system modulates midgut epithelial immunity in *Anopheles gambiae*. *Science* **2010,** *327* (5973), 1644-8.

384. Kawahara, T.; Quinn, M. T.; Lambeth, J. D., Molecular evolution of the reactive oxygen-generating NADPH oxidase (Nox/Duox) family of enzymes. *BMC Evol Biol* **2007,** *7*, 109.

385. Kremer, N.; Voronin, D.; Charif, D.; Mavingui, P.; Mollereau, B.; Vavre, F., *Wolbachia* interferes with ferritin expression and iron metabolism in insects. *PLoS Pathog* **2009,** *5* (10), e1000630.

386. Brownlie, J. C.; Cass, B. N.; Riegler, M.; Witsenburg, J. J.; Iturbe-Ormaetxe, I.; McGraw, E. A.; O'Neill, S. L., Evidence for metabolic provisioning by a common invertebrate endosymbiont, *Wolbachia pipientis*, during periods of nutritional stress. *PLoS Pathog* **2009,** *5* (4), e1000368.

387. Gill, A. C.; Darby, A. C.; Makepeace, B. L., Iron Necessity: The Secret of *Wolbachia*'s Success? *PLoS Negl Trop Dis* **2014,** *8* (10), e3224.

388. Molloy, J. *Wolbachia*-host metabolic and immune interactions and genetically engineered female-specific lethality for *Aedes* mosquito control. University of Oxford, **2015**.

389. Mayack, C.; Naug, D., Energetic stress in the honeybee *Apis mellifera* from *Nosema ceranae* infection. *J Invertebr Pathol* **2009,** *100* (3), 185-8.

390. Cloutier, C., Amino acid utilization in the aphid *Acyrthosiphon pisum* infected by the parasitoid *Aphidius smithi*. *J Insect Physiol* **1986,** *32*, 263-7.

391. Turley, A. P.; Moreira, L. A.; O'Neill, S. L.; McGraw, E. A., *Wolbachia* infection reduces blood-feeding success in the dengue fever mosquito, *Aedes aegypti*. *PLoS Negl Trop Dis* **2009,** *3* (9), e516.

392. Ye, Y. H.; Woolfit, M.; Huttley, G. A.; Rances, E.; Caragata, E. P.; Popovici, J.; O'Neill, S. L.; McGraw, E. A., Infection with a Virulent Strain of *Wolbachia* Disrupts Genome Wide-Patterns of Cytosine Methylation in the Mosquito. *PLoS One* **2013,** *8* (6), e66482.

393. Negri, I.; Franchini, A.; Gonella, E.; Daffonchio, D.; Mazzoglio, P. J.; Mandrioli, M.; Alma, A., Unravelling the *Wolbachia* evolutionary role: the reprogramming of the host genomic imprinting. *Proc Biol Sci* **2009,** *276* (1666), 2485-91.

394. Correa, C. C.; Ballard, J. W., *Wolbachia* gonadal density in female and male *Drosophila* vary with laboratory adaptation and respond differently to physiological and environmental challenges. *J Invertebr Pathol* **2012,** *111* (3), 197-204.

395. Correa, C. C.; Ballard, J. W., What can symbiont titres tell us about co-evolution of *Wolbachia* and their host? *J Invertebr Pathol* **2014,** *118*, 20-7.

396. Hussain, M.; Frentiu, F. D.; Moreira, L. A.; O'Neill, S. L.; Asgari, S., *Wolbachia* uses host microRNAs to manipulate host gene expression and facilitate colonization of the dengue vector *Aedes aegypti*. *Proc Natl Acad Sci U S A* **2011,** *108* (22), 9250-5.

397. Osei-Amo, S.; Hussain, M.; O'Neill, S. L.; Asgari, S., *Wolbachia*-induced aae-miR-12 miRNA negatively regulates the expression of MCT1 and MCM6 genes in *Wolbachia*-infected mosquito cell line. *PLoS One* **2012,** *7* (11), e50049.

398. Zhang, G.; Hussain, M.; Asgari, S., Regulation of arginine methyltransferase 3 by a *Wolbachia*-induced microRNA in *Aedes aegypti* and its effect on *Wolbachia* and dengue virus replication. *Insect Biochem Mol Biol* **2014,** *53*, 81-8.

399. Craig, G. B. J.; Hickey, W. A., Genetics of *Aedes aegypti*. In *Genetics of Insect Vectors of Disease* **1967,** pp67-131.

400. McClelland, G. A. H., Sex-linkage in *Aedes aegypti*. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **1962,** *56*, 4.

401. Craig, G. B.; Hickey, W. A., Current status of the formal genetics of *Aedes aegypti*. *Bull World Health Organ* **1967,** *36* (4), 559-62.

402. McDonald, P. T.; Rai, K. S., Correlation of linkage groups with chromosomes in the mosquito, *Aedes aegypti*. *Genetics* **1970,** *66* (3), 475-85.

403. Newton, M. E.; Southern, D. I.; Wood, R. J., X and Y chromosomes of *Aedes aegypti* (L.) distinguished by Giemsa C-banding. *Chromosoma* **1974,** *49* (1), 41-9.

404. Severson, D. W.; Meece, J. K.; Lovin, D. D.; Saha, G.; Morlais, I., Linkage map organization of expressed sequence tags and sequence tagged sites in the mosquito, *Aedes aegypti*. *Insect Mol Biol* **2002,** *11* (4), 371-8.

405. Timoshevskiy, V. A.; Severson, D. W.; Debruyn, B. S.; Black, W. C.; Sharakhov, I. V.; Sharakhova, M. V., An integrated linkage, chromosome, and genome map for the yellow fever mosquito *Aedes aegypti*. *PLoS Negl Trop Dis* **2013** *7*, e2052.

406. Carvalho, A. B.; Vibranovski, M. D.; Carlson, J. W.; Celniker, S. E.; Hoskins, R. A.; Rubin, G. M.; Sutton, G. G.; Adams, M. D.; Myers, E. W.; Clark, A. G., Y chromosome and other heterochromatic sequences of the *Drosophila melanogaster* genome: how far can we go? *Genetica* **2003,** *117* (2-3), 227-37.

407. Mysore, K.; Sun, L.; Tomchaney, M.; Sullivan, G.; Adams, H.; Piscoya, A. S.; Severson, D. W.; Syed, Z.; Duman-Scheel, M., siRNA-Mediated Silencing of doublesex during Female Development of the Dengue Vector Mosquito *Aedes aegypti*. *PLoS Negl Trop Dis* **2015,** *9* (11), e0004213.

408. Whyard, S.; Erdelyan, C. N.; Partridge, A. L.; Singh, A. D.; Beebe, N. W.; Capina, R., Silencing the buzz: a new approach to population suppression of mosquitoes by feeding larvae double-stranded RNAs. *Parasit Vectors* **2015,** *8*, 96.

409. Marinotti, O.; Jasinskiene, N.; Fazekas, A.; Scaife, S.; Fu, G.; Mattingly, S. T.; Chow, K.; Brown, D.; Alphey, L.; James, A. A., Development of a population suppression strain of the human malaria vector. *Malar J* **2013,** *12* (1), 142.

410. Kemphues, K. J.; Raff, R. A.; Kaufman, T. C.; Raff, E. C., Mutation in a structural gene for a beta-tubulin specific to testis in *Drosophila melanogaster*. *Proc Natl Acad Sci U S A* **1979,** *76* (8), 3991-5.

411. Smith, R. C.; Walter, M. F.; Hice, R. H.; O'Brochta, D. A.; Atkinson, P. W., Testis-specific expression of the beta2 tubulin promoter of *Aedes aegypti* and its application as a genetic sex-separation marker. *Insect Mol Biol* **2007,** *16* (1), 61-71.

412. Scolari, F.; Schetelig, M. F.; Bertin, S.; Malacrida, A. R.; Gasperi, G.; Wimmer, E. A., Fluorescent sperm marking to improve the fight against the pest insect *Ceratitis capitata* (Wiedemann; *Diptera*: Tephritidae). *Nat Biotechnol* **2008,** *25* (1), 76-84.

413. Olivieri, G.; Olivieri, A., Autoradiographic study of nucleic acid synthesis during spermatogenesis in *Drosophila melanogaster*. *Mutat Res* **1965,** *2* (4), 366-80.

414. Gould-Somero, M.; Holland, L., The timing of RNA synthesis for spermiogenesis in organ cultures of *Drosophila melanogaster* testes. *Wilhelm Rouxs Arch* **1974,** *174*, 133-148.

415. Barreau, C.; Benson, E.; White-Cooper, H., *Comet* and *cup* genes in *Drosophila* spermatogenesis: the first demonstration of post-meiotic transcription. *Biochem Soc Trans* **2008,** *36* (Pt 3), 540-2.

416. Barreau, C.; Benson, E.; Gudmannsdottir, E.; Newton, F.; White-Cooper, H., Post-meiotic transcription in *Drosophila* testes. *Development* **2008,** *135* (11), 1897-902.

417. Vibranovski, M. D.; Chalopin, D. S.; Lopes, H. F.; Long, M.; Karr, T. L., Direct evidence for postmeiotic transcription during *Drosophila melanogaster* spermatogenesis. *Genetics* **2010,** *186* (1), 431-3.

418. Schafer, M.; Nayernia, K.; Engel, W.; Schafer, U., Translational control in spermatogenesis. *Dev Biol* **1995,** *172* (2), 344-52.

419. Akbari, O. S.; Antoshechkin, I.; Amrhein, H.; Williams, B.; Diloreto, R.; Sandler, J.; Hay, B. A., The Developmental Transcriptome of the Mosquito *Aedes aegypti*, an Invasive Species and Major Arbovirus Vector. *G3 (Bethesda)* **2013,** *3* (9), 1493-509.

420. Baker, D. A.; Nolan, T.; Fischer, B.; Pinder, A.; Crisanti, A.; Russell, S., A comprehensive gene expression atlas of sex- and tissue-specificity in the malaria vector, *Anopheles gambiae*. *BMC Genomics* **2011,** *12*, 296.

421. Magnusson, K.; Mendes, A. M.; Windbichler, N.; Papathanos, P. A.; Nolan, T.; Dottorini, T.; Rizzi, E.; Christophides, G. K.; Crisanti, A., Transcription regulation of sex-biased genes during ontogeny in the malaria vector *Anopheles gambiae*. *PLoS One* **2011,** *6* (6), e21572.

422. Andrews, J.; Bouffard, G. G.; Cheadle, C.; Lü, J.; Becker, K. G.; Oliver, B., Gene discovery using computational and microarray analysis of transcription in the *Drosophila melanogaster* testis. *Genome Res* **2000,** *10* (12), 2030-43.

423. Parisi, M.; Nuttall, R.; Edwards, P.; Minor, J.; Naiman, D.; Lü, J.; Doctolero, M.; Vainer, M.; Chan, C.; Malley, J.; Eastman, S.; Oliver, B., A survey of ovary-, testis-, and soma-biased gene expression in *Drosophila melanogaster* adults. *Genome Biol* **2004,** *5* (6), R40.

424. Chintapalli, V. R.; Wang, J.; Dow, J. A., Using FlyAtlas to identify better *Drosophila melanogaster* models of human disease. *Nat Genet* **2007,** *39* (6), 715-20.

425. Ayyar, S.; Jiang, J.; Collu, A.; White-Cooper, H.; White, R. A., *Drosophila* TGIF is essential for developmentally regulated transcription in spermatogenesis. *Development* **2003,** *130* (13), 2841-52.

426. Doyle, J.; Doyle, J., A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Pytochem. Bull.* **1987,** (19), 11-5.

427. VectorBase. http://www.vectorbase.org.

428. Quinlan, A. R.; Hall, I. M., BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **2010,** *26* (6), 841-2.

429. Robinson, M. D.; McCarthy, D. J.; Smyth, G. K., edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **2010,** *26* (1), 139-40.

430. Camacho, C.; Coulouris, G.; Avagyan, V.; Ma, N.; Papadopoulos, J.; Bealer, K.; Madden, T. L., BLAST+: architecture and applications. *BMC Bioinformatics* **2009,** *10*, 421.

431. McKenna, A.; Hanna, M.; Banks, E.; Sivachenko, A.; Cibulskis, K.; Kernytsky, A.; Garimella, K.; Altshuler, D.; Gabriel, S.; Daly, M.; DePristo, M. A., The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **2010,** *20* (9), 1297-303.

432. Danecek, P.; Auton, A.; Abecasis, G.; Albers, C. A.; Banks, E.; DePristo, M. A.; Handsaker, R. E.; Lunter, G.; Marth, G. T.; Sherry, S. T.; McVean, G.; Durbin, R.; Group, G. P. A., The variant call format and VCFtools. *Bioinformatics* **2011,** *27* (15), 2156-8.

433. Angiuoli, S. V.; Salzberg, S. L., Mugsy: fast multiple alignment of closely related whole genomes. *Bioinformatics* **2011,** *27* (3), 334-42.

434. Chi-squared calculator. http://graphpad.com/quickcalcs/contingency1.cfm.

435. Sequence Read Archive. http://www.ncbi.nlm.nih.gov/sra.

436. Hannon Lab, FASTX Toolkit. http://hannonlab.cshl.edu/fastx_toolkit/index.html.

437. Joshi, N. A.; Fass, J. N., Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files. https://github.com/najoshi/sickle.

438. NCBI Genomes FTP site. ftp://ftp.ncbi.nlm.nih.gov/genomes.

439. Messeguer, X.; Escudero, R.; Farré, D.; Núñez, O.; Martínez, J.; Albà, M. M., PROMO: detection of known transcription regulatory elements using species-tailored searches. *Bioinformatics* **2002,** *18* (2), 333-4.

440. Loots, G. G.; Ovcharenko, I., rVISTA 2.0: evolutionary analysis of transcription factor binding sites. *Nucleic Acids Res* **2004,** *32* (Web Server issue), W217-21.

441. Okumura, T.; Makiguchi, H.; Makita, Y.; Yamashita, R.; Nakai, K., Melina II: a web tool for comparisons among several predictive algorithms to find potential motifs from promoter regions. *Nucleic Acids Res* **2007,** *35* (Web Server issue), W227-31.

442. Nene, V.; Wortman, J. R.; Lawson, D.; Haas, B.; Kodira, C.; Tu, Z. J.; Loftus, B.; Xi, Z.; Megy, K.; Grabherr, M.; Ren, Q.; Zdobnov, E. M.; Lobo, N. F.; Campbell, K. S.; Brown, S. E.; Bonaldo, M. F.; Zhu, J.; Sinkins, S. P.; Hogenkamp, D. G.; Amedeo, P.; Arensburger, P.; Atkinson, P. W.; Bidwell, S.; Biedler, J.; Birney, E.; Bruggner, R. V.; Costas, J.; Coy, M. R.; Crabtree, J.; Crawford, M.; Debruyn, B.; Decaprio, D.; Eiglmeier, K.; Eisenstadt, E.; El-Dorry, H.; Gelbart, W. M.; Gomes, S. L.; Hammond, M.; Hannick, L. I.; Hogan, J. R.; Holmes, M. H.; Jaffe, D.; Johnston, J. S.; Kennedy, R. C.; Koo, H.; Kravitz, S.; Kriventseva, E. V.; Kulp, D.; Labutti, K.; Lee, E.; Li, S.; Lovin, D. D.; Mao, C.; Mauceli, E.; Menck, C. F.; Miller, J. R.; Montgomery, P.; Mori, A.; Nascimento, A. L.; Naveira, H. F.; Nusbaum, C.; O'leary, S.; Orvis, J.; Pertea, M.; Quesneville, H.; Reidenbach, K. R.; Rogers, Y. H.; Roth, C. W.; Schneider, J. R.;

Schatz, M.; Shumway, M.; Stanke, M.; Stinson, E. O.; Tubio, J. M.; Vanzee, J. P.; Verjovski-Almeida, S.; Werner, D.; White, O.; Wyder, S.; Zeng, Q.; Zhao, Q.; Zhao, Y.; Hill, C. A.; Raikhel, A. S.; Soares, M. B.; Knudson, D. L.; Lee, N. H.; Galagan, J.; Salzberg, S. L.; Paulsen, I. T.; Dimopoulos, G.; Collins, F. H.; Birren, B.; Fraser-Liggett, C. M.; Severson, D. W., Genome sequence of *Aedes aegypti*, a major arbovirus vector. *Science* **2007,** *316* (5832), 1718-23.

443. Juneja, P.; Osei-Poku, J.; Ho, Y. S.; Ariani, C. V.; Palmer, W. J.; Pain, A.; Jiggins, F. M., Assembly of the Genome of the Disease Vector *Aedes aegypti* onto a Genetic Linkage Map Allows Mapping of Genes Affecting Disease Transmission. *PLoS Negl Trop Dis* **2014,** *8* (1), e2652.

444. Harker, B. W.; Behura, S. K.; deBruyn, B. S.; Lovin, D. D.; Mori, A.; Romero-Severson, J.; Severson, D. W., Stage-specific transcription during development of *Aedes aegypti*. *BMC Dev Biol* **2013,** *13*, 29.

445. Dissanayake, S. N.; Ribeiro, J. M.; Wang, M. H.; Dunn, W. A.; Yan, G.; James, A. A.; Marinotti, O., aeGEPUCI: a database of gene expression in the dengue vector mosquito, *Aedes aegypti*. *BMC Res Notes* **2010,** *3*, 248.

446. Robinson, J. T.; Thorvaldsdóttir, H.; Winckler, W.; Guttman, M.; Lander, E. S.; Getz, G.; Mesirov, J. P., Integrative genomics viewer. *Nat Biotechnol* **2011,** *29* (1), 24-6.

447. Li, K.; Xu, E. Y.; Cecil, J. K.; Turner, F. R.; Megraw, T. L.; Kaufman, T. C., *Drosophila* centrosomin protein is required for male meiosis and assembly of the flagellar axoneme. *J Cell Biol* **1998,** *141* (2), 455-67.

448. Hall, A. B.; Qi, Y.; Timoshevskiy, V.; Sharakhova, M. V.; Sharakhov, I. V.; Tu, Z., Six novel Y chromosome genes in *Anopheles* mosquitoes discovered by independently sequencing males and females. *BMC Genomics* **2013,** *14*, 273.

449. Hall, A. B.; Timoshevskiy, V. A.; Sharakhova, M. V.; Jiang, X.; Basu, S.; Anderson, M. A.; Hu, W.; Sharakhov, I. V.; Adelman, Z. N.; Tu, Z., 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* **2014,** *6* (1), 179-91.

450. Akbari, O. S.; Papathanos, P. A.; Sandler, J. E.; Kennedy, K.; Hay, B. A., Identification of germline transcriptional regulatory elements in *Aedes aegypti*. *Sci Rep* **2014,** *4*, 3954.

451. Akbari, O. S.; Chen, C. H.; Marshall, J. M.; Huang, H.; Antoshechkin, I.; Hay, B. A., Novel Synthetic Medea Selfish Genetic Elements Drive Population Replacement in *Drosophila*; a Theoretical Exploration of Medea-Dependent Population Suppression. *ACS Synth Biol* **2012,** *3* (12), 915-28.

452. Shen, S.; Park, J. W.; Huang, J.; Dittmar, K. A.; Lu, Z. X.; Zhou, Q.; Carstens, R. P.; Xing, Y., MATS: a Bayesian framework for flexible detection of differential alternative splicing from RNA-Seq data. *Nucleic Acids Res* **2012,** *40* (8), e61.

453. Nimmo, D. D.; Alphey, L.; Meredith, J. M.; Eggleston, P., High efficiency site-specific genetic engineering of the mosquito genome. *Insect Mol Biol* **2006,** *15* (2), 129-36.

454. Klowden, M. J., Sexual receptivity in *Anopheles gambiae* mosquitoes: absence of control by male accessory gland substances. *J Insect Physiol* **2001,** *47* (7), 661-6.

455. Miyatake, T.; Chapman, T.; Partridge, L., Mating-induced inhibition of remating in female Mediterranean fruit flies *Ceratitis capitata*. *J Insect Physiol* **1999,** *45* (11), 1021-1028.

456. Mossinson, S.; Yuval, B., Regulation of sexual receptivity of female Mediterranean fruit flies: old hypotheses revisited and a new synthesis proposed. *J Insect Physiol* **2003,** *49* (6), 561-7.

457. Kraaijeveld, K.; Chapman, T., Effects of male sterility on female remating in the mediterranean fruitfly, *Ceratitis capitata*. *Proc Biol Sci* **2004,** *271 Suppl 4*, S209-11.

458. Seo, S.; Vargas, R.; Gilmore, J.; Kurashima, R.; Fujimoto, M., Sperm transfer in normal and gamma-irradiated, laboratory-reared Mediterranean fruit flies (*Diptera*:*Tephritidae*). *J. Econ. Entomol.* **1990,** *83*, 1947-53.

459. Taylor, P.; Kaspi, R.; Mossinson, S.; Yuval, B., Age-dependent insemination success of sterile Mediterranean fruit flies. *Entomol. Exp. Appl.* **2001,** *98*, 27-33.

460. Kuba, H.; Itô, Y., Remating inhibition in the melon fly, *Bactrocera* (=*Dacus*) *cucurbitae* (*Diptera*: Tephritidae): copulation with spermless males inhibits female remating. *Journal of Ethology* **1993,** *11*, 23-8.

461. Harmer, A. M.; Radhakrishnan, P.; Taylor, P. W., Remating inhibition in female Queensland fruit flies: effects and correlates of sperm storage. *J Insect Physiol* **2006,** *52* (2), 179-86.

462. Radhakrishnan, P.; Pérez-Staples, D.; Weldon, C.; Taylor, P., Multiple mating and sperm depletion in male Queensland fruit flies: effects on female remating behaviour. *Animal Behaviour* **2009,** *78* (4), 839-846.

463. Oxford Nanopore. http://www.nanoporetech.com.

464. Segoli, M.; Hoffmann, A. A.; Lloyd, J.; Omodei, G. J.; Ritchie, S. A., The Effect of Virus-Blocking *Wolbachia* on Male Competitiveness of the Dengue Vector Mosquito, *Aedes aegypti*. *PLoS Negl Trop Dis* **2014,** *8* (12), e3294.

465. Lee, H. L.; Vasan, S.; Ahmad, N. W.; Idris, I.; Hanum, N.; Selvi, S.; Alphey, L.; Murad, S., Mating compatibility and competitiveness of transgenic and wild type *Aedes aegypti* (L.) under contained semi-field conditions. *Transgenic Res* **2013,** *22* (1), 47-57.

466. Murdock, C. C.; Blanford, S.; Hughes, G. L.; Rasgon, J. L.; Thomas, M. B., Temperature alters *Plasmodium* blocking by *Wolbachia*. *Sci Rep* **2014,** *4*, 3932.

467. McGraw, E. A.; Merritt, D. J.; Droller, J. N.; O'Neill, S. L., *Wolbachia* density and virulence attenuation after transfer into a novel host. *Proc Natl Acad Sci U S A* **2002,** *99* (5), 2918-23.

468. Labbe, G. M.; Nimmo, D. D.; Alphey, L., piggybac- and PhiC31-mediated genetic transformation of the Asian tiger mosquito, *Aedes albopictus* (Skuse). *PLoS Negl Trop Dis* **2010,** *4* (8), e788.

469. Gabrieli, P.; Smidler, A.; Catteruccia, F., Engineering the control of mosquito-borne infectious diseases. *Genome Biol* **2014,** *15* (11), 535.

Appendices

Appendix 1 – *Drosophila* food recipes

Recipe 1

For one batch of food:

component	amount
plant agar	8g
inactive yeast	16g
organic rolled oats	80g
Tate and Lyle black treacle	138g
acid mix	17ml
10 % Nippogene in ethanol	20ml
H_2O	1.33l

For 500ml acid mix:

component	amount
phosphoric acid	21ml
propionic acid	206ml
dH_2O	273ml

Recipe 2

For one batch of food:

component	amount
agar	16g
yeast powder	80g
white sugar	100g
propionic acid	5ml
10 % Nippogene in ethanol	30ml
H_2O	1l

Appendix 2 – Primer sequences

Primers used in Chapter 2

qPCR primers for genomic DNA purity assessment

WSP forward	ATCTTTTATAGCTGGTGGTGGT
WSP reverse	GGAGTGATAGGCATATCTTCAAT
HTH forward	TGGTCCTATATTGGCGAGCTA
HTH reverse	TCGTTTTTGCAAGAAGGTCA
RpL32 forward	CGGTTACGGATCGAACAAGC
RpL32 reverse	CTTGCGCTTCTTGGAGGAGA
D. simulans mitochondrial rRNA forward	TCGTCCAACCATTCATTCCAGCCT
D. simulans mitochondrial rRNA reverse	TGGCCGCAGTATTTTGACTGTGC
Ae. albopictus mitochondrial rRNA forward	ACTGATAGATATTAAAGGGCCGCA
Ae. albopictus mitochondrial rRNA reverse	ATGCTACCTTCGCACAGTCA

PCR primers for confirmation of transgene presence

WD0626 forward	CACAAGCAATTGGCCTCTCTGCTAA
WD0626 reverse	TGGAATGGCCACACGTCCTTGC
28S rRNA forward	TACCGTGAGGGAAAGTTGAAA
28S rRNA reverse	AGACTCCTTGGTCCGTGTTT

PCR primers for screening for *Wolbachia* infection

WSP forward	TGGTCCAATAAGTGATGAAGAAAC
WSP reverse	AAAAATTAAACGCTACTCCA
28S rRNA forward	TACCGTGAGGGAAAGTTGAAA
28S rRNA reverse	AGACTCCTTGGTCCGTGTTT

qPCR primers for expression analysis

WD0626 forward	CACAAGCAATTGGCCTCTCTGCTAA
WD0626 reverse	TGGAATGGCCACACGTCCTTGC
RpL32 forward	CGGTTACGGATCGAACAAGC
RpL32 reverse	CTTGCGCTTCTTGGAGGAGA

Primers used in Chapter 4

PCR primers for experimental testing of candidate testis-specifically expressed genes

Ae. aegypti

Novel gene 1 forward	ACCGAAGCTACAAGGACGC
Novel gene 1 reverse	CAGCAGCCGACCCGGAA
AAEL000463 forward	GGCCGTTTTGGCACATTCTT
AAEL000463 reverse	CACGCGTTTTTCACTTTGCG
AAEL001333 forward	CGAGACCGTTCTCCTCAACC
AAEL001333 reverse	TCCTCGCACATTTTTCTCCG
AAEL003021 forward	GGCCGTTTTGGCACATTCTT
AAEL003021 reverse	CACGCGTTTTTCACTTTGCG
AAEL004293 forward	GCGAGAAAAAGAAGCGTAAAAC
AAEL004293 reverse	TCCTCAGTTGACTCACGTCG
AAEL006665 forward	TCAAAGCCCTCAAGAATCGGA
AAEL006665 reverse	TCATTGCTTGTGCACTCACC
AAEL009267 forward	GTGAAGACGACCTTGCTCCA
AAEL009267 reverse	CCATGTGACGTTCCTTTGCG
AAEL010268 forward	CCTGCACCGATCGAAGTCAA
AAEL010268 reverse	TGGTGGTGGCAGATGAAACA
AAEL011838 forward	GGCAATCGGTTGTACCTGGA
AAEL011838 reverse	ATCACATCGCCAGGCTTGAA
AAEL012065 forward	CGCGAATACCATAAGCCCCT
AAEL012065 reverse	CGCAGCTACCACTACTACCG
AAEL012239 forward	GGTGCTAGTGGTGGAACGAA
AAEL012239 reverse	TCTTTGGCTTCACGCTTTGC
RpL22 forward	AAACAACTCCTGCGAGGCAA
RpL22 reverse	CACGCAGGCTGTTCTTCTTC

C. capitata

LOC101449084 forward	TACTGAGACAACAGCGCCAG
LOC101449084 reverse	TCGAACGCTCCGATACCAAG
LOC101449780 forward	TTGCCGAAACTTGCCCAAAG
LOC101449780 reverse	CACTCAACCGGATCACCCAA
LOC101451111 forward	CAATTTATGGGCCTCTTTCGG
LOC101451111 reverse	CGATTCGCCAACGTAAGCAG
LOC101451785 forward	AGCGTTGGGTTGTGTGATGA
LOC101451785 reverse	GACAATCCCGATGGAGCAGT
LOC101454909 forward	TGCTCATGTTGCTGTTGTGC
LOC101454909 reverse	GCCGATGATGTAGATGCCGA
LOC101455161 forward	GCGGAAGTACTCGGCTATTG
LOC101455161 reverse	CCCGTTGGTAGGACATGCT
LOC101457895 forward	CTACAGCGAGCTACGTCCAT
LOC101457895 reverse	CCGGACCCTCTGCTTTCATT
LOC101459316 forward	GAAAATCGCAAAGCTGCCCA
LOC101459316 reverse	TGATGCTTGCTGGCATCTCA
LOC101459689 forward	CCCACACTGATCGGCTACTC
LOC101459689 reverse	ATGTAGCCTGCAGCAAAGGT
LOC101462854 forward	GTAGTCGTACTCACCGAGCC
LOC101462854 reverse	CGCGCAGTGATTATCTCGTT
RpL22 forward	CAACACGGATCCAATCACGC
RpL22 reverse	TGCTGCTGCATCTGCTACTT

qPCR primers for experimental testing of candidate testis-specifically expressed genes

Ae. aegypti

AAEL001333 forward	CGTACTCCACGGAGGTCATC
AAEL001333 reverse	GTCCTCGCACATTTTTCTCCG
AAEL003021 forward	GGAGCTGCATGAGAGGTGTT
AAEL003021 reverse	CATGCTGACCAGGACCATGA
AAEL006665 forward	TCAAAGCCCTCAAGAATCGGA
AAEL006665 reverse	TCGTTTGATGGCGTACAGGG
AAEL009267 forward	TGAGATTAAGTTCCACGACGAC
AAEL009267 reverse	CCGTTTCAGTGCGACTTGTG
AAEL010268 forward	GCTTTGACAGTGAAACTGTGGC
AAEL010268 reverse	CGCAAATCAGTGAAGCTTGGA
AAEL012065 forward	CGCGAATACCATAAGCCCCT
AAEL012065 reverse	TGGTGGTACGTAACCCGTTG
AAEL012239 forward	GCCAACGGAGATCTAGAAGTT
AAEL012239 reverse	CGCTTTGCCTTGACATGAGC
α-tubulin forward	TCGGACAAAACCATCGGAGG
α-tubulin reverse	ACGGTTGGTTCCAGATCGAC

C. capitata

LOC101449084 forward	TGTTGAGAGCATCACGATCAC
LOC101449084 reverse	TGCCTCATCGTCGGAATGTT
LOC101449780 forward	TTGAAGCTCCCCGATGTTGC
LOC101449780 reverse	TGACCAATAGCACCACTGGC
LOC101451111 forward	CCCAAAGCAGCAGGTAGACA
LOC101451111 reverse	GCATTCACGTCCAATGTCGC
LOC101451785 forward	GACCGTTATGTGCGTCATGC
LOC101451785 reverse	GCCAACGGTGTTCTTCAACG
LOC101457895 forward	TCTGAAACGTACCCAGAAGGC
LOC101457895 reverse	CCCGGTCATATCTGTCAGCC
LOC101459316 forward	CGCTCCATAAGCGTCGTATCT
LOC101459316 reverse	CCCAAACCGATGGTTGAGAA
LOC101459689 forward	AAACGCAAAGCCACCTTGAC
LOC101459689 reverse	TGGGTTGGTGCCACGAATTA
LOC101462854 forward	CCATTGTTAAACGAGGCAACGA
LOC101462854 reverse	TGGCACCATTATACCCTGCG
α-tubulin forward	TGCCCTACCCACGTATTCAC
α-tubulin reverse	ACCATCTGGTTGGCTGGTTC

PCR primers for experimental testing of candidate testis-specifically spliced genes

Primers for amplifying both predicted testis-specific and other splice forms

Ae. aegypti

AAEL001916 forward	GGCCGGGCTCTTAATGTGAT
AAEL001916 reverse	GCAACGACAATGGCTGGCAA
AAEL005651 forward	CGTAGTTCGCTTCGTGGAGT
AAEL005651 reverse	ACTAAGCACCGGTTTCTCCG
AAEL005935 forward	CGGGGTCCTTTTACACCGAA
AAEL005935 reverse	GAGAATCAAGTGGTCGAGCG
AAEL007258 forward	CCCGTCGCCTATTTCACCAA
AAEL007258 reverse	GTTTGGGGATAACGTATCGC
AAEL008110 intron 2 forward	CAAATGCATGATGCGTTCCT
AAEL008110 intron 2 reverse	GCTGCAAGCGTTGGATATGG
AAEL008428 forward	GGGACTTTGACTCGCACAGA
AAEL008428 reverse	AGCAGCAATTCGGTTTTGCAT
AAEL010509 forward	TAAATCGCCGCCATCACCAT
AAEL010509 reverse	GATTTAGAGCGGCCGAGTGA
AAEL011105 forward	TCCGAGTGGTCGCCATTTTG
AAEL011105 reverse	TCACGTTGTCCTCATGGGTG
AAEL011968 forward	TCCAACTGTTCAGCTCGTCG
AAEL011968 reverse	TGGCACAGTAAGCCTGTTGA
AAEL012262 forward	CAGTGCCAGCAAAGCAGAAG
AAEL012262 reverse	TACAACCGAGCGTCATTGCT
AAEL013571 forward	GTCGAGAAGGTACTCGAGCG
AAEL013571 reverse	GACGTCGAAGCAACTGGGTT
AAEL014957 forward	TGCACGAAAAAGGAGCAACG
AAEL014957 reverse	AGTTTGTGTAGCGGTGCGTT
AAEL017395 forward	TGATTCGTCGGTTGGCTGTT
AAEL017395 reverse	CACCACTCGTCTGTGTCTCG
AAEL018039 forward	TCAGCAGCAACCGTTCTACC
AAEL018039 reverse	CCGCGTTGAGTCAGGAACAT
AAEL018211 intron 1 forward	CCATCTTTCGACGCGTGTTC
AAEL018211 intron 1 reverse	TCCAAGTCCGCAAGACAAGG
AAEL018350 forward	ATGGTGGCCCAAAACCAATG
AAEL018350 reverse	TTGTTTGTGATGCGGGACTG

C. capitata

LOC101449503 forward	CCGCCCTTTAGGTGTATGGG
LOC101449503 reverse	TGTTTTCGTATTGATTTTCCCTGG
LOC101450565 forward	AACGGGGTTTGGCAACTGTA
LOC101450565 reverse	GAAGGATCGCGTTGAGACCA
LOC101453609 forward	TGAGGAGGGACACAGGTCAA
LOC101453609 reverse	TGCAGCTTCCGATGTCTGTT
LOC101457260 forward	CTCACCACCGGGTGATATGG
LOC101457260 reverse	ACCGCGAGATCCGAAAATGT
LOC101459514 intron 1 forward	GGAGAAGCAACTGGGGTGTT
LOC101459514 intron 1 reverse	TAGGGCAAACTACTGCACCG
LOC101459514 intron 2 forward	TGGCGATCCTGCTCACCGAA
LOC101459514 intron 2 reverse	TAGGGCAAACTACTGCACCG

Primers for amplifying predicted testis-specific splice form

Ae. aegypti

AAEL000028 forward	TTTGGTTGTCCAGGTCCCAC
AAEL000028 reverse	CGTTTTAGTTGGAAAATTGTTTGCT
AAEL001898 forward	TGGGCTAAAAAGAGAGGCAA
AAEL001898 reverse	CCCAGTACATTTAAAAAGCATGAAG
AAEL002083 forward	CAGATTCGTCCTGATCGCCA
AAEL002083 reverse	ATGCTAACAAGCGCTGTCCA
AAEL008110 intron 1 forward	GGCAATCAAACCCTTGATTCTGT
AAEL008110 intron 1 reverse	AATGCAAGTTTCGTCCGTGC
AAEL008110 intron 2 forward	CTTGCTCATTTGCTCAGCATC
AAEL008110 intron 2 reverse	GCTGCAAGCGTTGGATATGG
AAEL008428 forward	GGGACTTTGACTCGCACAGA
AAEL008428 reverse	CGGTTTTGCATGTCTTCATAAGC
AAEL011153 forward	TCTGCCTTGCCAGTGATTGA
AAEL011153 reverse	TGCTCGGCTTTGAAGAACCC
AAEL012262 forward	AAACAGCCATCAAGCAATGTG
AAEL012262 reverse	CGCTACGCAATGAAGGGAGA
AAEL013571 forward	TCATCCTGGTAATTGTCACTCAA
AAEL013571 reverse	GGGTTTAATAGACGATGGAAACGG
AAEL018211 intron 1 forward	CCTTCACGCAATGTCTGGGA
AAEL018211 intron 1 reverse	CGGTCCAGTTCCCTAATTGAA
AAEL018211 intron 2 forward	CCTTCACGCAATGTCTGGGA
AAEL018211 intron 2 reverse	TTGTCAAAGACTGGCTAGAGTGA
AAEL018350 forward	ACTCAAAAGAAATTAGTAATAAGGA
AAEL018350 reverse	CACTTTCTTTGCTTGTTCCC

C. capitata

LOC101448609 forward	TACTATTGACTACGGGTGAAGC
LOC101448609 reverse	TCGAAAATGCGTCGCTGTTG
LOC101449153 intron 1 forward	GCTTCGCCTGCAGGTAGTC
LOC101449153 intron 1 reverse	CATGGGTTTGGCATGTGACG
LOC101449153 intron 2 forward	TCAATGTTAGCTCAATATTTGCCAT
LOC101449153 intron 2 reverse	GACTGAAACGAAGGCGCAAG
LOC101449153 intron 3 forward	AATTGTCGATTGCATTTTGCTCA
LOC101449153 intron 3 reverse	CTCGAGGGCATGGTTACACA
LOC101449503 forward	TGCCAATGTCCATGTAGCAC
LOC101449503 reverse	TTTATTCGAGTGGTACGTCAGC
LOC101450641 forward	TGGTAGGTAGTTAAGTTATTTGACA
LOC101450641 reverse	AGGCTTTTGAAGCTTTTCTCA
LOC101452861 forward	TTCTGGTGGCATACAGGTCG
LOC101452861 reverse	CGTACTGTGCAACTGGAGGT
LOC101452984 forward	ACTTAATAATCACACAATGGCCGA
LOC101452984 reverse	TCTCATCCAATTCGGCAGGC
LOC101457260 forward	TCAGTATCGGCGGCATCAGT
LOC101457260 reverse	TGGAAACCATCGTCTCTATGCT
LOC101459514 intron 1 forward	GCCAAGAGTCGTTGTGGTGT
LOC101459514 intron 1 reverse	TCGCTTTGTTTTCATCCGTCAATTA
LOC101459514 intron 2 forward	AATTGGCGATCCTGCTCAC
LOC101459514 intron 2 reverse	TTTAGAGGTAGACCAGTTATTTCAA
LOC101461218 forward	AACAACTTCGAAAACCTGTAAGCG
LOC101461218 reverse	CTAAGCGAATCCTCGGGTCC

Primers for amplifying other splice forms

Ae. aegypti

AAEL000028 forward	TTTGGTTGTCCAGGTCCCAC
AAEL000028 reverse A	TGTCATTATTTACAAGGTTTGCTGT
AAEL000028 reverse B	TACTTCCGAGAATGCGAGCC
AAEL001898 forward	GACTGGATGCCACTGTCCTG
AAEL001898 reverse	CTAGAGCGTATCCTGCCAGC
AAEL002083 forward A	CAGATTCGTCCTGATCGCCA
AAEL002083 forward B	GTTTTCCCACGTAGCTTCCG
AAEL002083 reverse	TGGCGTTGTCCATGTCTGTT
AAEL008110 forward	AAGGTCTTTCAGCCTCTGGC
AAEL008110 reverse	ACGAATGCGCCGAAATGAAG
AAEL008428 forward	TGCCTTGTCAAGAAAATCCCA
AAEL008428 reverse	TCGTTATCGCTGAGGCATCC
AAEL011153 forward	TCTGCCTTGCCAGTGATTGA
AAEL011153 reverse	CTTGGAAAAGAACCCTCACGC
AAEL018211 forward	CCATCTTTCGACGCGTGTTC
AAEL018211 reverse	TTCAAAGCGCACGTTGATCC
AAEL018350 forward	ACAAGCTGCTCCAACTCCTC
AAEL018350 reverse	TCGGACGGTGATGGTTTTCC

C. capitata

LOC101448609 forward	TAGGCCATGCACAAGTTGGT
LOC101448609 reverse	TCGAAAATGCGTCGCTGTTG
LOC101449153 forward	TTTGCTCACTCTTCTCGCGT
LOC101449153 reverse	CTCGAGGGCATGGTTACACA
LOC101450641 forward	CAAGCCTTTCGCCTTCGTTC
LOC101450641 reverse	TGCCAACTCGGTGTTGTCTG
LOC101452861 forward	CTGTTGCTGTTGGAGAAGGGT
LOC101452861 reverse	TCTGCCATCACATCAGCAGTT
LOC101452984 forward A	GGCAGTAAACAATCGACGAC
LOC101452984 forward B	CCGAACTCGTTGCTTTCCG
LOC101452984 reverse	GCGAAATCATTGGGACCCTC
LOC101457260 forward	CTGTAGACGCACGGTTTCCA
LOC101457260 reverse	AACGTGTCGTAGATGAGGACG
LOC101459514 forward	ATGATGTTACCAAATGGCCAAATA
LOC101459514 reverse	CATTGCACCGATTTATTGGTTTC
LOC101461218 forward	TCGCCTGTACATTCTTGCGT
LOC101461218 reverse	CATCCGATAAAACGCCAGCC

qPCR primers for experimental testing of candidate testis-specifically spliced genes

Primers for amplifying predicted testis-specific splice form

Ae. aegypti

AAEL000028 forward	CAGTTGGGAGGCTTCGGTAG
AAEL000028 reverse	CCGTTTTAGTTGGAAAATTGTTTGC
AAEL001898 forward	TTCAGTGTGCTTTGGAGGTGT
AAEL001898 reverse	CCCAGTACATTTAAAAAGCATGAAG
AAEL008110 intron 1 forward	GGCAATCAAACCCTTGATTCTGT
AAEL008110 intron 1 reverse	TGGCCGATGTCGAGTAAACC
AAEL011153 forward	TCTGCCTTGCCAGTGATTGA
AAEL011153 reverse	TGCTCGGCTTTGAAGAACCC
AAEL012262 forward	TCGCCACTACTCGTGCTATTC
AAEL012262 reverse	TCACATTGCTTGATGGCTGT
AAEL018211 intron 1 forward	ATCAAAGCCGCATAGCTTGG
AAEL018211 intron 1 reverse	CGGTCCAGTTCCCTAATTGAA
AAEL018211 intron 2 forward	GCTCACCTCGCTTCAGGATT
AAEL018211 intron 2 reverse	TTGTCAAAGACTGGCTAGAGTGA
AAEL018350 forward A	CCCAAAACCAATGAGTAATAAGGA
AAEL018350 forward B	ACTCAAAAGAAATTAGTAATAAGGA
AAEL018350 reverse	GTAGGTGTCTTAGCAGTGAC

C. capitata

LOC101449153 intron 1 forward	GCTTCGCCTGCAGGTAGTC
LOC101449153 intron 1 reverse	CGCCACGTTCCAATAGTTGT
LOC101449153 intron 3 forward	AATTGTCGATTGCATTTTGCTCA
LOC101449153 intron 3 reverse	AGAGAATGCGTGAACAACAAGAT
LOC101449503 forward	TCTTCGACGGTCCAGGACTA
LOC101449503 reverse	TTTATTCGAGTGGTACGTCAGC
LOC101450641 forward	TGGTAGGTAGTTAAGTTATTTGACA
LOC101450641 reverse	AGGCTTTTGAAGCTTTTCTCA
LOC101452861 forward	TTCTGGTGGCATACAGGTCG
LOC101452861 reverse	AGCTGTCGAATGAAGTGCGT
LOC101457260 forward	CCTTTCGGCTGGCTTGGATA
LOC101457260 reverse	TGGAAACCATCGTCTCTATGCT
LOC101459514 intron 1 forward	TCTTCTGGAGAAGCAACTGGG
LOC101459514 intron 1 reverse	TCGCTTTGTTTTCATCCGTCAATTA
LOC101459514 intron 2 forward	TGAGCAGCGATTTCTTCTGTT
LOC101459514 intron 2 reverse	TTTAGAGGTAGACCAGTTATTTCAA

Primers for amplifying other splice forms

Ae. aegypti

AAEL000028 forward	CAGTTGGGAGGCTTCGGTAG
AAEL000028 reverse A	TGTCATTATTTACAAGGTTTGCTGT
AAEL000028 reverse B	AAAGCCACTGGTTTGCTGTG
AAEL001898 forward	GCTGTCTATTGTTGGAGTCCTTTT
AAEL001898 reverse	ACCCAACAGCCATCGATACG
AAEL008110 intron 1 forward	TTCCGACTTTGCCTGGAACA
AAEL008110 intron 1 reverse A	ATGCAAGTTTCGTCCGTCCT
AAEL008110 intron 1 reverse B	CTCGGATGCTGAGGTCCTG
AAEL011153 forward	TCTGCCTTGCCAGTGATTGA
AAEL011153 reverse	CTTGGAAAAGAACCCTCACGC
AAEL01226 forward	TCGCCACTACTCGTGCTATTC
AAEL01226 reverse A	GCGTTTGCGACCTGTTTGT
AAEL01226 reverse B	GGGATGCAGGATATCTGAAACTG
AAEL018211 intron 1 forward A	TTTGAGGGCATTTCCTCGCT
AAEL018211 intron 1 reverse A	TCCAAGTCCGCAAGACAAGG
AAEL018211 intron 1 forward B	AGGGCATTTCATTTGGAGCA
AAEL018211 intron 1 reverse B	CGCACGTTGATCCCACAGAT
AAEL018211 intron 2 forward	GCAGCCTCTCACAAATGTCG
AAEL018211 intron 2 reverse A	CCTGAAGCGAGGAAATGCCC
AAEL018211 intron 2 reverse B	GCTCCAAATGAAATGCCCTCA
AAEL018350 forward A	CCCAAAACCAATGAGTAATAAGGA
AAEL018350 forward B	ACTAGCAGTATAGGTAATAAGGAGA
AAEL018350 reverse	GTAGGTGTCTTAGCAGTGAC

C. capitata

LOC101449153 forward	TTTGCTCACTCTTCTCGCGT
LOC101449153 reverse	TGGCAAATATTGAGTACCAACTGA
LOC101449503 forward	CGTAGGTCGAATAAATCTCAACATC
LOC101449503 reverse	TGTTTTCGTATTGATTTTCCCTGG
LOC101450641 forward	ATAACAAGAGCACCCGCATC
LOC101450641 reverse A	ATTTGGAAAAACTTAACTACCTACC
LOC101450641 reverse B	CGAGGACCCACACGAGAAAG
LOC101452861 forward	TTCTCTGCTGTGCCTCATCG
LOC101452861 reverse	TCTGCCATCACATCAGCAGTT
LOC101457260 forward	GTGTATCTGTAGACGCACGGT
LOC101457260 reverse	ACCTTTGAAGTGAAAATTTTGGATG
LOC101459514 forward A*	ACTGGTCTACCTCTAAATGAACA
LOC101459514 reverse A*	TGAGCAATAGCCAGTTGTCC
LOC101459514 forward B	TTGCAGAATAACTGGTCTACCTCT
LOC101459514 reverse B	TAGGGCAAACTACTGCACCG

*Also targets predicted testis-specific splice forms