Embryo development and sex determination in the Cassava whitefly (*Bemisia tabaci*)

Rebecca Louise Corkill

University of East Anglia

A thesis submitted for the degree of Doctor of Philosophy
September 2019

Abstract

The cassava whitefly, *Bemisia tabaci* (Hemiptera: Aleyrodidae), is a highly invasive and destructive agricultural pest, with a global distribution. *B. tabaci* insect has evolved resistance to multiple insecticides, and therefore, new control methods will have to be developed to control this insect. The self-limiting system, which produces a female-specific lethal heritable element and causes population suppression, works well for control of mosquitoes. The goal of this thesis is to assess the feasibility of creating a self-limiting system in *B. tabaci*.

The self-limiting system requires generating transgenic insects, and this is done via injection of constructs into eggs at the pre-germ-line cell stage (around blastoderm formation), this enables stable germline transformation. However, there has been limited research into *B. tabaci* early embryogenesis stages, and therefore it is unclear when transformation constructs will have to be introduced. In this thesis, I conducted confocal microscopy studies to determine the timing of early embryogenesis stages in *B. tabaci* MED. Unexpectedly, I revealed that early embryogenesis in *B. tabaci* starts before the eggs are oviposited, unlike in mosquitoes and fruitflies. Therefore, injecting laid *B. tabaci* eggs is too late for obtaining stable transgenic whitefly lines, though somatic transgenics may be obtained.

Past self-limiting transgenic insects used genes found in the sex determination pathway, for female-specific lethality, as these were expressed early and had sex-specific splicing. Genes of the sex determination pathways in *B. tabaci* and other hemipteran are largely unknown. Hence, I conducted a genome-wide search of 11 publicly available hemipteran genomes to identify their sex determination genes. Also, I conducted single-embryonic RNA-seq experiments in *B. tabaci* to assess if these genes undergo sex-specific splicing at the early embryogenesis stages in males and females. These results identified sex determination genes in *B. tabaci* and other hemipterans and revealed potential targets for future genetic control methods.

“If we knew what it was we were doing, it would not be called research, would it?” - *Albert Einstein*

Table of contents

Acknowledgements

Firstly, I must thank my incredible supervisory team, for the suggestions and encouragement over the last four years. Especially Saskia Hogenhout, for allowing me to conduct a PhD within her laboratory and encouraging me to be a strong independent woman in the field of science. To Sam Mugford, for tolerating my constant questions. Without his guidance, I would not have gotten anywhere close to where I am now. To Neil Morrison, who was my industry supervisor who helped my placement be seamless.

I must acknowledge the enormous support I have received from the members of the Hogenhout laboratory, that I have interacted with both professionally and personally. It is one of the main reasons why I have enjoyed my time at John Innes Centre; the friendliness and the willingness to help. I want to thank the JIC insectary staff. Without your help, my time in the insectary would have been a lot more painful and less productive.

Thank you to my fellow PhD friends, in which I have shared the last four years (especially Jimmy and Ben). You guys helped maintain my sanity with lunch dates and emergency tea and cake sessions.

I want to acknowledge my little unit in Norwich; Kris, Queenie and Bellatrix. Without you three, my life would have had a lot more stress and tears (surprisingly). Lastly, I would like to acknowledge my parents and family. Nothing would have been possible without your help. Thank you for understanding the distance I had to travel to accomplish my dreams and the support you provided throughout.

List of Figures

List of tables

List of abbreviations

BLAST	Basic local alignment search tool
CFP	Cyan Fluorescent Protein
CI	Cytoplasmic incompatibility
CSD/*Csd*	Complementary sex determiner
CZI	Carl Zaiss Format
DA/ *Da*	Daughterless
DAPI	4',6'-Diamidine-2'-phenylindole dihydrochloride
DM	Doublesex and mab-3 DNA binding motif
DMRT	Doublesex and mab-3 related transcription factor
DPN/ *Dpn*	Deadpan
DSF/ *Dsf*	Dissatisfaction
DSX/ *Dsx*	Doublesex
Em	Emission
EMC/ *Emc*	Extra-macrochaetae
Ex	Excitation
FEM/ *Fem*	Feminizer
FL(2)d/ *fl(2)d*	Female lethal d
FPKM	fragments per kilobase of exon model per million reads mapped
FRU / *Fru*	Fruitless
GFP	Green Fluorescent Protein
GRO/ *Gro*	Grouchu
Gtf	gene transfer format
HER/ *Her*	Hermaphrodite
HLH	Helix-loop-helix
IMP/ *Imp*	IGF-II mRNA binding protein
LTPG	Longest Transcript Per Gene
Mab-3	Male abnormal-3
MASC/ *Masc*	Masculinizer
MEAM1	Middle- East Asia minor 1
MED	Mediterranean
NCBI	National Centre for Biotechnology information
Nt	Nucleotide
PBS	Phosphate Buffered Saline
PHO	Putative Hemipteran Orthologue
PI	Propidium Iodide
piRNA	Piwi-interacting RNA
PSI/ *PSI*	P- element somatic inhibitor
RBBH	Recipricol Best Blast Hit
RFP	Red Fluorescent Protein
RIDL	Release of an insect with dominant lethal
RRM	RNA-recognition motif
SDG	Sex Determination Gene
SDL	Sex determination locus
SDP	Sex Determination Protein
SIT	Sterile insect technique

SNF/ *Snf*	Sans fille
SNP	Single nuceleotide polymorphism
SXL/ *Sxl*	Sex-lethal
TRA/ *Tra*	Transformer
TRA2/ *Tra2*	Transformer 2
tTAV	tetracycline- repressible transactivator
TYLCV	Tomato yellow leaf curl virus
UTR	Untranslated region
Vcf	variant call format
VIR/ *Vir*	Virilizer
YFP	Yellow Fluorescent Protein

Chapter 1: Introduction

1.1 Hemipteran

Hemiptera are an order of insects which covers 50,000 to 80,000 different species. Hemiptera all have the same common arrangements of the sucking and piercing mouthparts. Most hemipterans are herbivorous; however, some Hemiptera are not herbivorous and feed on other insects, small invertebrates and in some cases humans. Hemiptera diverged 320 million years ago from Holometabola (Hogenhout and Bos, 2011a; Wu and Baldwin, 2010).

For 350 million years, plants and insects have coexisted (Gatehouse, 2002). During this 350-million-year timeframe insects have evolved to locate, feed and oviposit on their host plant. Generalist herbivorous insects feed on many plant species, whereas specialist insects can only feed and oviposit on one of few plants within the same family (Wu and Baldwin, 2010).

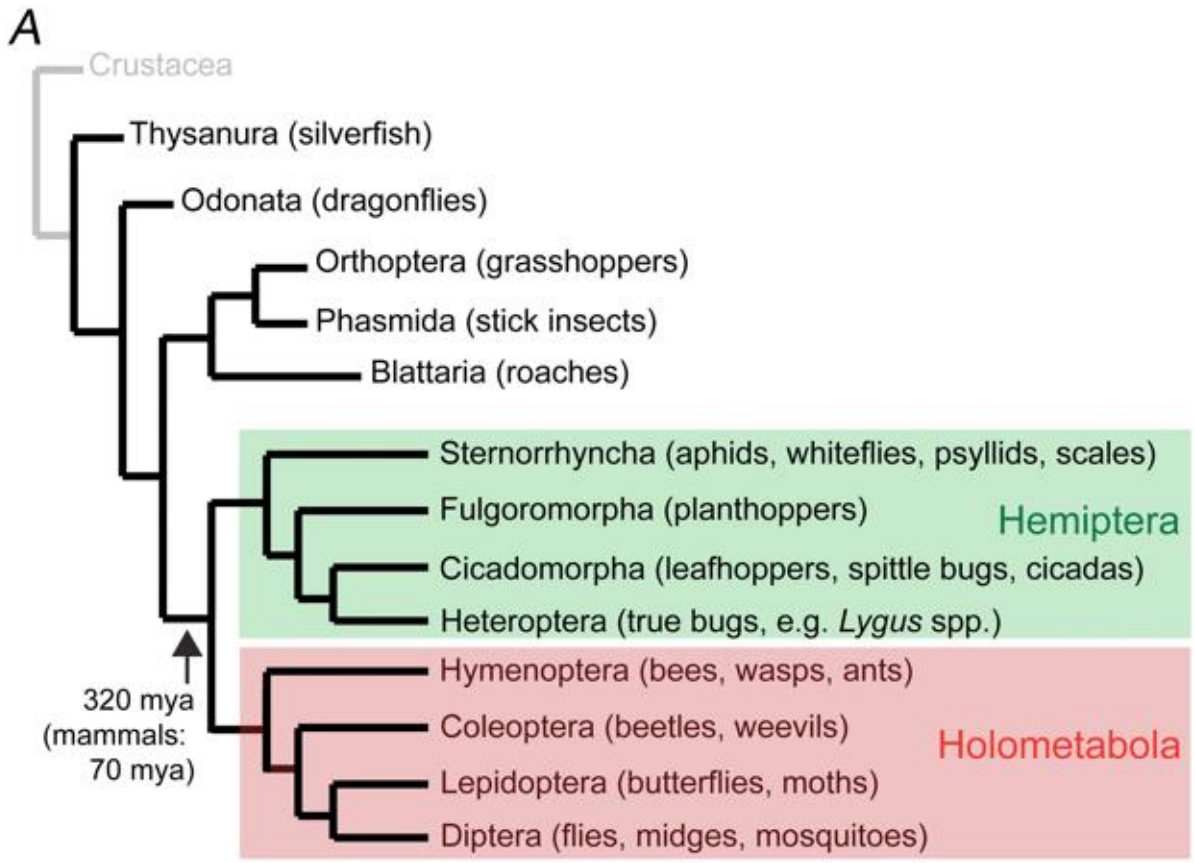

Figure 1. 1 Schematic overview of insect phylogeny
Figure produced from Hogenhout and Bos (2011a)

1.2. *Bemisia tabaci*

Bemisia tabaci (*B. tabaci*) (Cassava whitefly) is a sap-feeding herbivore of the hemipteran Order, Suborder Sternorrhyncha. The Sternorrhyncha suborder contains aphids, scale insects, psylloids and whiteflies and is a sister group to Auchenenorrynha (Campbell et al., 1994). Sternorrhyncha are characterised by the type of gut filter chamber, a stylet which rests between the bases of front legs and if the insect is winged, lack of vannus and vannal fold in the hind wing (Cranston and Gullan, 2009). Whiteflies are similar to aphids, they produce honeydew (Johnson, 1982), transmit plant viruses (Jones, 2003; Blackman and Cahill, 1998) and harbour endosymbionts (Baumann, 2005).

B. tabaci is a generalist feeder with an extensive host range, which is unlike most herbivorous insects. *B. tabaci* inflicts damage to crops, through transmitting viruses and causing physiological damage. Alongside this, *B. tabaci* have developed high insecticide resistance (Nauen and Denholm, 2005), causing the traditional control methods to be less effective. Therefore, it is necessary to identify novel methods of controlling this pest.

B. tabaci has a wide geographical range (Salvucci, 2000), and has always been considered a dangerous invasive species because it is prevalent throughout the majority of the world. The UK, Finland, Ireland, Sweden are protected zone (PZ) countries for *B. tabaci*; these countries have a policy to eradicate any plants containing whiteflies. The majority of *B. tabaci* introduced to PZ countries are through imported plant material. In the UK, 56% of interceptions have mainly been found on *Euphorbia pulcherima* (poinsettia) (Cuthbertson et al., 2011). Q/ MED biotype (discussed more later) is the predominant whitefly entering the UK and being intercepted at nurseries (Cuthbertson et al., 2012).

1.2.1 Life cycle and morphology of *Bemisia tabaci*

B. tabaci is a haplodiploid species; males hatch from unfertilised eggs and have 10 chromosomes compared to the 20 chromosomes of diploid females. Therefore males only inherit the maternal genome (Blackman and Cahill, 1998).

An ovariole mass forms B. tabaci ovaries. Ovary pairs can range from 12-22 and be at different developmental stages (Guo et al., 2010). Figure 1.2 shows the ovariole development.

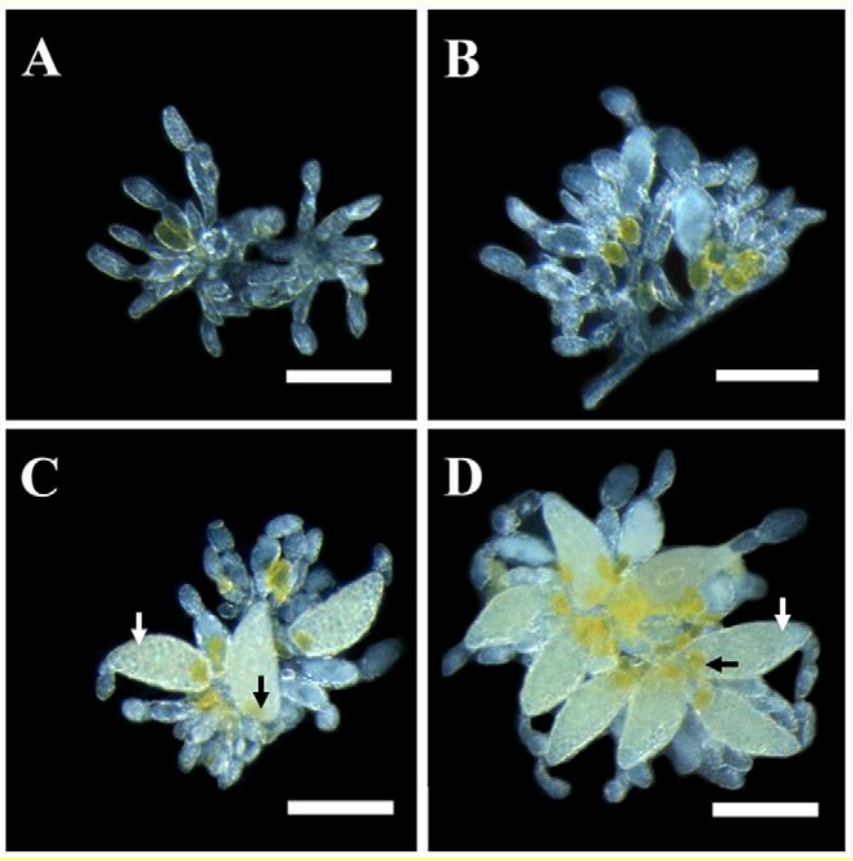

Figure 1. 2 The ovary of a B. tabaci Middle East Asia minor biotype (MEAM1) at different developmental stages after eclosion

The ovary of a *B. tabaci* (MEAM1) at different developmental stages after eclosion. A: The ovary of a female after emerging. B: 1-2day after emerging. C: 3-10 days after emerging and D: 11-14 days after emerging. White arrows indicate the mature oocytes and the black arrows indicate the bacteriocyte. The scale bar is 0.10mm. Figure produced from Guo *et al* (2010).

Each ovariole can contain different maturity stages of oocytes. There are four morphological stages of embryos; A, B, C and D (Figure 1.3). A type egg has no yolk protein. The yolk content identifies B and C stages, B has less than 50% yolk, and C contained more than 50% yolk. The D stage are mature eggs and contain the bacteriocyte, indicated by black arrow in Figure 1.3. The bacteriocyte enter the C eggs by the pedicle (Guo et al., 2010; Ghanim et al., 2001).

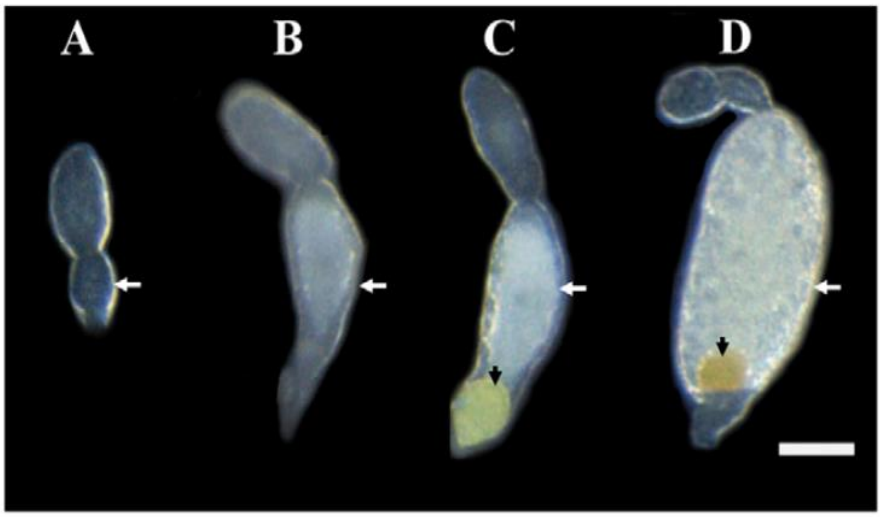

Figure 1. 3 Eggs of the different developmental stages in the ovarioles of B. tabaci MEAM1.
The white arrows indicate the different morphological stages. The black arrows indicate the bacteriocyte. Scale bar= 0.05 µm. Figure produced from Guo *et al* (2010).

Female whiteflies lay eggs, from 0.8-1mm on the underside of leaves. The eggs are attached to the leaf by the female adults creating a slit with their ovipositor and positioning the eggs within (Buckner et al., 2002). The egg has a pedicle to ensure plant attachment. The pedicle is a hook/peg-like extension structure (Byrne and Bellows, 1991). In other hemipterans, the chorion is the outer egg layer that comprises of polyphenol and lipoproteins (Beament, 1946). The chorion layer extends from the base of the egg to halfway down the pedicle(Buckner et al., 2002). The pedicle transports water from the leaf to the egg. Egg removal from the leaf causes mortality by dehydration. Adult females deposit a glue-like substance called the cement, during oviposition. The cement creates a seal that helps water/solutes move efficiently into the egg (Buckner et al., 2002).

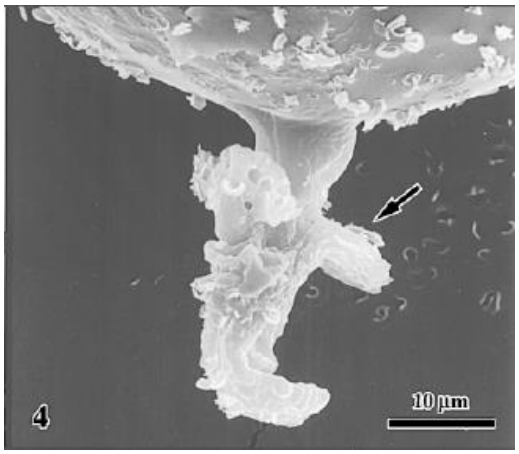

Figure 1. 4 SEM micrograph of an oviposited Silverleaf whitefly egg removed from a cotton leaf
Eggs extricated from leaves show portions of leaf epidermis (arrow) cemented to the pedicel. The pedicel tips within epidermal cells were curved or hooked. Picture produced from Buchner *et al* (2002).

Along with the cement, the mother also transfers symbiotic bacteria into the egg. Large cells contain the bacteria and form the bacteriocyte. The bacteriocytes are visible as large yellow spheres inside the egg and are located on both sites of the intestinal tract in the various nymphal stages and in adults (Costa et al., 1996).

There are two types of endosymbionts in the bacteriocyte in *B. tabaci*; primary (P-endosymbionts) and secondary. P-endosymbionts are defined as being essential for the survival of whiteflies, these are vertically transmitted and present in all individuals, whilst the knowledge on secondary symbionts function is poorly defined (Baumann, 2005). These P-endosymbionts convert the non-essential amino acids present in the plant phloem into amino acids essential to whiteflies and that cannot be obtained from the phloem diet alone (Costa et al., 1997). The P-endosymbiont in *B. tabaci* is *Candidatus Portiera aleyrodidarum* (Thao and Baumann, 2004). There are several common secondary endosymbionts in *B. tabaci,* these are from the *Arsenophonus, Cardinium, Hamiltonella, Wolbachia, Fritschea,* and *Rickettsia* species (Zchori-Fein and Brown, 2002; Gottlieb et al., 2006; Chiel et al., 2007; Gueguen et al., 2010; Karut and Tok, 2014; Tajebe et al., 2015). In one study, 95% of the *B. tabaci* samples had secondary endosymbionts and some symbionts are associated with specific cryptic species, such as *Hamiltonella* in MEAM1 and MED (Q1) (Gueguen et al., 2010).

Wolbachia is an interesting endosymbiont, as they can be involved in reproductive abnormalities such as cytoplasmic incompatibility (CI), male killing, feminization, male killing and pathogenesis induction (Zeh et al., 2005; Werren et al., 2008; Stouthamer et al., 1999). CI is the most common phenotype and has been used as a biological control in many different insects (Bourtzis, 2008;

Christodoulou, 2011; Hancock et al., 2011). CI occurs when male and female insects mate, but have different *Wolbachia* strains or infection status, this causes embryonic mortality in diploid species, but in haplodiploid organisms the sex ratio of the offspring can lean heavily towards the sex that is produced from unfertilised eggs (Feldhaar, 2011; Zhong and Li, 2014). CI can be induced in *B. tabaci* by; treating the insects with antibiotics and allowing them to mate with *Wolbachia* infected *B. tabaci* and infecting females with a *Wolbachia* strain from *D. melanogaster* and allowing them to mate with males with a native *Wolbachia* (Zhong and Li, 2013; Zhou and Li, 2016).

An oviposited egg takes approximately seven days for it to hatch under laboratory conditions. The nymph emerging from the egg is called the first instar nymph or 'crawler' stage. The crawler is mobile and will settle down when it finds a suitable place to feed (Walker et al., 2010). The subsequent second, third and fourth instars settle permanently in one spot. These latter stages have an oval and dorso-ventrally flattened morphology; within each nymphal instar, the height increases while the width and length remain constant (Gelman et al., 2002). Different *B. tabaci* biotypes (further discussed below), have minor differences in morphology which helps the identification process at the pupal stages (Walker et al., 2010).

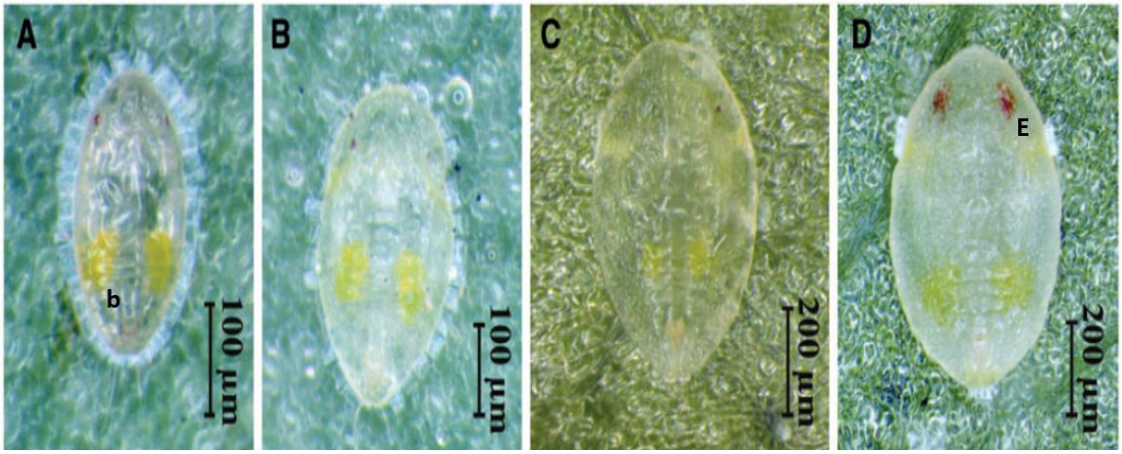

Figure 1. 5 Different developmental stages of the instar
A- first instar (crawler), B- second instar, C- third instar and D- fourth instar. E is the red-eye spot, which in the fourth instar is the sophisticated bipartite adult compound eye. 'b' is the bacteriome. Figure produced from Chaubey *et al* (2015).

The latter part of the 4th nymphal instar is the 'pupal stage', which is not a separate instar stage, but signifies the developmental change to adult. This stage typically lasts approximately eight days. The first signal, of this stage, is the enlargement of the red pinpoint eyes. Usually, around 24 hours after eye spots begin to enlarge, the eye has transformed into a sophisticated bipartite

adult compound eye. During the pupal stage, the wings begin to develop (Gelman et al., 2002) (Walker et al., 2010).

The insect abdomen contains the genitals. The genitals are small, and detail can only be studied with microscopy. The male anatomy has a 'clasper' structure and a single penis (Figure 1.6). The claspers are used to hold the female during copulation (Figure 1.7) (Walker et al., 2010).

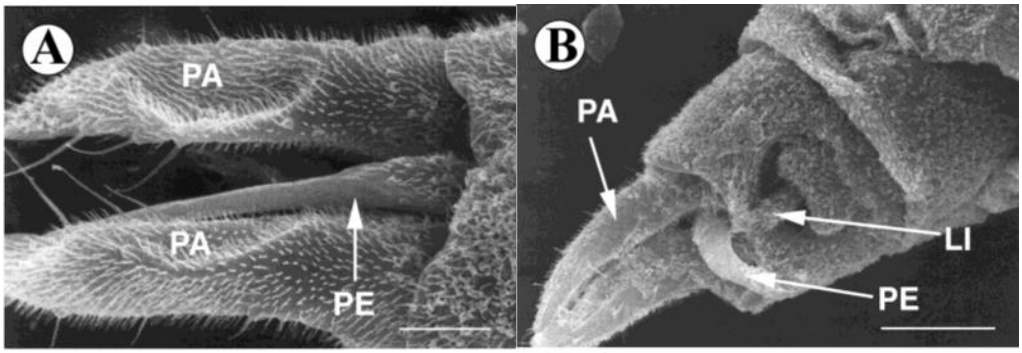

Figure 1. 6 SEM of the male B. tabaci genitalia
A- the Sagittal view of the paramers (PA) and the penis (PE), the scale bar represents= 20 µm. B- the dorsal view of the parameres (PA) and the penis (PE). The anus is located beneath the lingual (LI) Scale bar= 47.5 µm. Figure produced from Ghanim *et al* (2001).

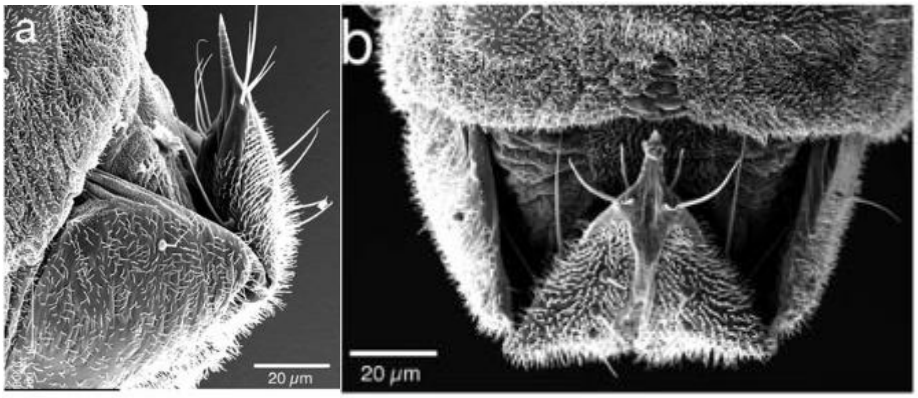

Figure 1. 7 SEM of the female B. tabaci genitalia
a) a laterally positioned retracted ovipositor b) a dorsally positioned retracted ovipositor. Figure produced from Walker *et al* (2009).

Courting is very specific in *B. tabaci*. There is a courtship cascade, which has 4 stages: male positioning; antennal drumming; abdominal undulation; and copulation. All these stages need to be performed to ensure successful copulation. Males actively search and position themselves parallel to the females. During antennal drumming both sexes raise their antennae and keep them at 45^{o} angles horizontally from their heads. The male drums the flagellum of female antennae at the medial section. If successful, the abdominal undulation phase begins in which the male raises his abdomen and positions it in a vertical plane and the antennal drumming is continued. Some biotypes show body pushing behaviour in which the males push the females to prepare them for

copulation. In the copulation phase, the male layers his wings over those of the female and his claspers hold the terminalia of the female (Li et al., 1989). The *B. tabaci* courtship cascade is complex and involves specific behaviours within each species. Therefore, pest control methods that use transgenic or sterile males should be species specific and are unlikely to be at risk for off target species involvement.

All stages of *B. tabaci* feed from the phloem in a leaf- a part of the vascular system in higher-level plants which provides a source of sugars, amino acids and other organic metabolites (Will et al., 2013). The adults feed by stylet bundle penetration- left and right mandibular and maxillary stylets (Walker et al., 2009). The bundle is very flexible and weaves around the cells towards the phloem (Walker and Perring, 1994). The adults primarily feed on phloem and minor veins that are typically accessed from the abaxial surface (Cohen et al., 1996). Nymphs employ a different mechanism than adults, as the stylet bundle forms a loop-like structure called the crumena (Walker et al., 2009).

Saliva plays a major role in the feeding behaviour of the *B. tabaci*. Normally there are two different types of saliva; watery and sheath. Sheath saliva is secreted during the stylets penetration phase and possesses gelling qualities. The exact functions of sheath saliva are unclear, but hypotheses include: providing a frictionless passage through the leaf; sealing the puncture wound caused by the stylets to ensure the phloem pressure stays like the pressure before the penetration; or to provide an inert barrier to ensure the wounding responses are low. Watery saliva does not have a gelling quality and normally contain salivary enzymes and metabolites (Walker et al., 2010).

Saliva in other hemipterans, such as aphids, have been known to contain effectors. These are small proteins that help modulate plant defence responses and increases the fecundity of the insects (Hogenhout and Bos, 2011c). Transcriptomic studies have been previously conducted on the salivary glands of *B. tabaci*, and comparative analysis of known aphid effectors (C002, Mp10 and Mp42) did not have homologues in *B. tabaci* (Su et al., 2012). However unpublished work by the Hogenhout lab indicates that *B. tabaci* does have an orthologue of Mp10.

1.2.2 Cryptic species

There are over 40 different morphologically identical 'biotypes' in *B. tabaci*. Previously, *B. tabaci* was classed into biotypes depending on different biological characteristics, such as the host range, silverleafing in squash, yellow vein in honeysuckle and begomovirus transmission (Bedford et al., 1994; De Barro et al., 2011b). The term biotype is now outdated and *B. tabaci* is recognised as a morphologically indistinguishable cryptic species complex. Molecular analysis of the mitochondrial cytochrome oxidase I (mtCOI) can differentiate between the different cryptic complex species (De Barro et al., 2011a). Few crossing experiment and observations have been conducted on the cryptic species so far. Some experiments show that the cryptic species are isolated by reproduction (Sun et al., 2011). However, laboratory-based hybridisation studies have produced no definite answers on whether the different cryptic species are truly different species. There have been no hybrids of different cryptic species found in the field (Saleh et al., 2012).

Due to the newness of the cryptic species classification, limited research has been conducted on the comparison between different species. Therefore, it is difficult to determine true differences between some of the cryptic species. The most invasive cryptic species from the complex are Middle East-Asia Minor 1 (MEAM1, known previously as B Biotype) and Mediterranean (MED, known previously as Q biotype) (Gueguen et al., 2010). Due to the invasive nature of MED and MEAM1 more research has been conducted on these then any other cryptic species. Both MED and MEAM1 are widely distributed in the tropics and subtropics (Tsai and Wang, 1996). Currently there are protected zones across the globe which have fought to maintain status against the invasive pests, this includes the UK (Cuthbertson and Vanninen, 2015). Despite this, there are countries that have been invaded across the years. For example, in Brazil MEAM1 was reported in the early 1990s, since then it is now widespread throughout the country (Marubayashi et al., 2013). Two decades later, MED was reported in Brazil (Barbosa et al., 2015).

MED and MEAM1 can co-exist in the same geographical location and on the same hosts. However significant switches of dominance between the species are observed (Bertin et al., 2018). Without an insecticide factor, MED dominated over MEAM1 on sweet pepper, however MEAM1 dominates over MED on cabbage cotton and tomato (Sun et al., 2013). When insecticide is a factor MED is more resistant than MEAM1, allowing the MED population to rise in these situations (Horowitz et al., 2005; Horowitz and Ishaaya, 2014).

1.1.3 Why is controlling *Bemisia tabaci* so important?

B. tabaci has caused much devastation in the past. However, it has been challenging to determine how much economic loss worldwide has occurred, because of the extensive host range and the wide range of viruses that *B. tabaci* transmits. *B. tabaci* has caused monetary losses of $1 billion annually in the USA (Czosnek and Brown, 2009).

B. tabaci transmits more than 100 different viruses (genus *Begomovirus* (Geminiviridae), *Crinivirus* (Closteroviridae), *Carlavirus* or *Ipomovirus* (Potyviridae)), with Begomoviruses being the most numerous causing 20-100% yield loss (Jones, 2003). The tomato yellow leaf curl (TYLCV; *Begomovirus*) is one of the most prominent; this virus was first reported in Israel in 1931 (Chouchane et al., 2007). Since then, epidemics have been widespread over the globe, in Africa, the Americas, Europe, Middle-East, and South-eastern Asia (Moriones and Navas-Castillo, 2000). This virus can cause up to 100% loss of tomatoes, and it also affects other crops. Symptoms include stunting, prominent upward curling of leaflet margins, size reduction of leaflets, yellowing of leaves and flower abortion (Moriones and Navas-Castillo, 2009). The virus can be detected in the ovaries and mature eggs of viruliferous insects (Goldman and Czosnek, 2002).

The physiological damage that *B. tabaci* causes is also considerable. The whitefly secretes a honeydew, (Johnson, 1982) which can be washed off, but acts as a substrate to sooty mould that blackens the leaves, thereby reducing the amount of photosynthesis that can take place and reducing yield (Stansly and Natwick, 2009). The honeydew is also detrimental to the factory workers, which process the cotton crop. The honeydew on the plant has adhesive properties, so when the cotton is collected dirt and trash attaches to the leaf surface. The dirt and trash causes dust particles when the crop is being processed. This dust can cause respiratory damage to the textile factory workers during the processing stage (Oliveira et al., 2001; Ayars et al., 1986)

1.2.3 Genomics of *Bemisia tabaci*

The *B. tabaci* genome is vast compared to most model organisms. The mean DNA content in males is approximately 1.04 pg per cell (five times greater than *Drosophila melanogaster*) and 2.08 pg for females (Czosnek and Brown, 2009). A recent K-mer analysis revealed that the male haploid genome size is 690Mb (Chen et al., 2015). Even with the large genome size it is thought that the gene number would be closer to 15,000, similar to *D. melanogaster*, with a gene density on average of 1 gene per 60,000bp (Czosnek and Brown, 2009).

Bemisia tabaci MEAM1 draft genome has been published within the last 3 years. This genome came from a single female established colony (Chen et al., 2016). *B. tabaci* MED draft genome is also published (Xie et al., 2017a.).

1.3 Sex-determination

Sex is the mixing of genomes by meiosis and fusion of gametes (Partridge, 1983). Sex produces genetic variance that is beneficial to individuals and populations (Otto, 2009). The sex determination pathway is fundamental biological knowledge. Understanding the sex determination pathways can bring an exceptional understanding of the genome evolution, function and organisation (Craves, 2008).

There is a vast amount of knowledge of sex determination mechanisms in model organisms; *D. melanogaster* (Cline and Meyer, 1996)*, Caenorhabditis elegans* (Hansen and Pilgrim, 1999) and the mouse *Mus musculus* (Mclaren, 1991). This vast research has highlighted the diversity, but also in some aspect's conservation of sex determination pathways.

1.3.1 Overall structure of the sex determination cascade

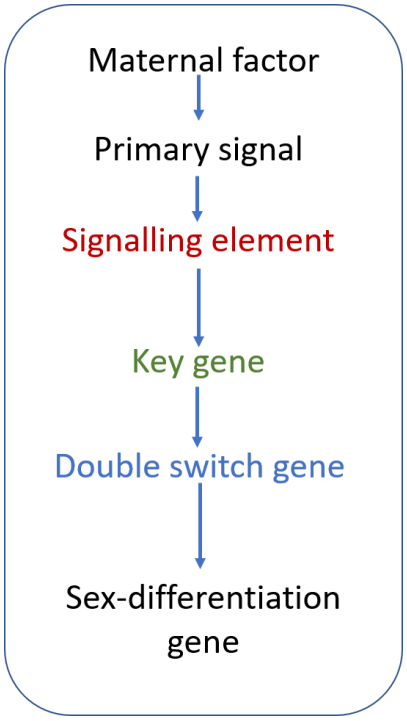

Figure 1. 8: A general diagram about how the sex determination genes usually are organised in* D. melanogaster, B. mori *and* Apis mellifera*. Adapted from Sawanth et al (2016)

Across divergent insects, the sex determination pathway shares many general features in common, although the molecular identities of all components may not be the same. Figure 1.8 shows a generic sex determination pathway describing these features. Signalling elements are usually a genetic determinant that distinguishes males and female, depending on the mating system of the insect, for example the Y chromosome. Next is the 'key' gene group; this group

interacts directly with the double switch gene (Sawanth et al., 2016). Double switch genes are found at the bottom of cascade and are the most conserved of the sex determination elements, and trigger sex-specific development and behaviour. The term double switch is used, because research has shown the gene at the bottom of the cascade has dual functions. In *Bombyx mori*, the protein encoded by the gene helps form sex-specific genitalia (Suzuki et al., 2001), but the female isoform also is a positive regulator for genes like vitellogenin (Suzuki et al., 2003).

1.3.2 Sex determination in other insects

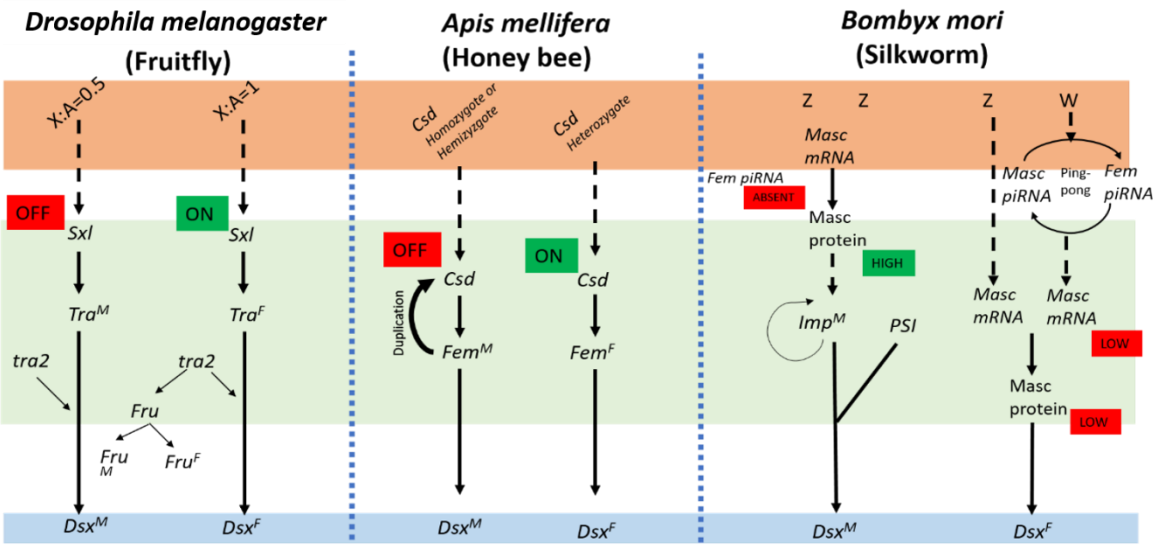

***Figure 1. 9 Sex determination pathways in* D. melanogaster, A. mellifera and B. mori**
The red section indicates the signalling elements. The green section is the 'key' genes, and the blue section is the double-switch genes. In *D. melanogaster*, when the X chromosome equals or exceeds one, the *Sxl* is transcribed. *Sxl* binds to *Tra* and *Tra* undergoes differential splicing depending on the male or female *Sxl*. *Tra* promotes sex-specific splicing in *Dsx*. In *A. mellifera* the signalling element depend on the CSD being heterozygote or homozygote. Homozygote CSD produces a female-specific splice variant of fem. The FEM protein creates a female specific *Dsx* gene which leads to female development. In *B. mori* the signalling element is chromosome-based. ZW individuals downregulate *Masc*, which fails to induce *ImpM* expression, and female-specific *Dsx* occurs leading to female development. In ZZ individuals there is no *fem*, so there is a high expression of *Masc*. *Masc* induces *Imp*. *Imp* and *PSI* form a male-specific *Dsx*, which leads to male development (Sakai et al., 2015; Yamamoto and Koganezawa, 2013; Herpin and Schartl, 2015; Hoff, 2009).

1.3.3 Signalling element and 'key' genes

Different signalling elements occur in different species. This signalling element can be sex-determining sex chromosomes, dose-dependent sex chromosome or haplodiploid species. A sex-determining chromosome is a common signalling element, especially in mammals. Sexual reproduction has led to female and males having different sex chromosome sizes (Charlesworth, 1996). In *Homo sapiens* females have homogametic sex chromosomes XX, whereas males are heterogametic with XY. The X chromosome has approximately 1000 genes, whereas the Y

chromosome may have a few dozen. In the early evolution of mammals, the X and Y chromosome originated from autosomes. Recombination restriction and gene loss has resulted in sex chromosome morphological differences (Bachtrog, 2013). Some dipteran insects have a male-determining Y chromosome; where females are XX, and the males are XY, this occurs in *Musca domestics* (Dubendorfer et al., 2002) and *Ceratitis capitata* (Pane et al., 2002). Butterflies differ; there is a female-determining W chromosome, males are ZZ and females are ZW (Maeki, 1981).

Dose-dependent sex chromosomes occurs in species like the chicken (*Gallus gallus*), tiger pufferfish (*Takifugu rubripes*), smooth tongue sole (*Cynoglossus semilaevis*) and fruitfly (*D. melanogaster*) (Bachtrog et al., 2014). The ratio of X-chromosome versus autosomes determine the *Caenorhabditis elegans* sex. (Farboud et al., 2013). Sex chromosomes undergo selective forces and therefore are the most changeable part of any animal or insect genome (Saunders et al., 2018).

Haplodiploidy occurs in many different insects, like thrips, whiteflies and a lot of Hymenoptera species. One form of haplodiploidy is called arrhenotokous haploidy. Arrhenotoky is where males develop from unfertilised eggs. In species like aphids, virgin females give birth to diploid males, and this is also arrhenotoky (Normark, 2003).

1.3.3.1 Signalling elements and 'Key' genes in *D. melanogaster*

D. melanogaster has a dose-dependent sex chromosome signalling element. A double dose of X chromosome determines sex femaleness, whereas males need only a single dose (Erickson and Quintero, 2007). *D. melanogaster* embryo's first action is to count the X chromosome (Salz et al., 1989). *D. melanogaster* has four different key genes; *Sex-lethal (Sxl), transformer (Tra)* and *transformer 2 (Tra2)* (see Figure 1.9).

Sxl is turned off in haplo- X male individuals and turned on in diplo-X female individuals; this causes sexual differentiation and dosage compensation (Cline, 1993). X-chromosome numerator proteins stimulate the female-specific activation of *Sxl* (Cline, 1988). Sex-specific regulation of *Sxl* is already determined by the time the embryos reaches the blastoderm stage. In haplo-X individuals, *Sxl* remains inactive during development (Salz et al., 1989). For sex determination pathway maintenance in diplo-X individuals, *Sxl* activity needs to be continuous. A positive feedback loop maintains the female-specific *Sxl*; the SXL protein binds and activates its promoter(Cline, 1993).

Sxl controls the processing of the *Tra* gene transcript, another 'key' gene. SXL directs the splicing of *Tra*, the splice site in the female-specific variant is in exon 3, and the non-specific is at exon 2.

This splicing creates female-specific and non-sex specific splicing variants. The non-sex-specific variants form a non-functional protein by having an early termination codon (Handa et al., 1999). The functional TRA protein in females interacts with a non-sex specific TRA2 (Tian and Maniatis, 1993; Verhulst et al., 2010).

TRA/TRA2 in *D. melanogaster* regulates the splicing of *fruitless (Fru)* in a female-specific manner; this leads to a non-functional FRU protein. *Fru* is a support gene and will be discussed later in this chapter. Males do not produce a functional TRA; a male-specific splicing occurs and produces a functional FRU (Ryner et al., 1996; Gailey et al., 2006; Verhulst et al., 2010).

1.3.3.2 Signalling elements and 'key' genes in *B. mori*

The sex-determining W chromosome is the signalling element. Transposable elements and other repeat elements are highly prevalent in the W chromosome (Abe et al., 2005; Katsuma et al., 2015). A piRNA comparative sequencing approach using piRNA derived from ovary and testis revealed female enriched piRNAs produced at the sex-determining region of the W chromosome (Kawaoka et al., 2011). piRNAs are 24-31nt small RNAs that can act as PIWI protein guides to silence transposon activity in animal gonads (Iwasaki et al., 2015).

The 'key' genes in *B. mori* are *feminizer (Fem), masculinizer (Masc), IGF-II mRNA binding protein (Imp)* and *P-element somatic inhibitor* (*PSI*), Figure 1.9. The first 'key' gene in *B. mori* is *Feminizer (Fem). Fem* is a female-specific PIWI-interacting RNA (piRNAs) precursor found in the sex-determining region of the W chromosome. When *Fem* is inhibited the male-specific variants of *BmDsx* (double switch gene) are produced (Kiuchi et al., 2014).

Masculinizer (Masc) is the target gene for *fem* piRNA (Suzuki et al., 2001), and is located on the Z chromosome. *Masc* encodes a CCCH-type zinc finger protein, which is conserved in Lepidoptera, such as *Trilochaa varians* and *Ostrinia furnacalis* (Lee et al., 2015; Fukui et al., 2015). *Masc* mRNA is needed for female-specific isoforms of the *BmDsx* in female embryos (Kiuchi et al., 2014; Lee et al., 2015).

Imp is localised on the Z chromosome (Suzuki et al., 2014). IMPs are part of a highly conserved family called VICKZ. VICKZ family have a KH RNA-binding; these recognise the *cis*-acting elements in RNAs. VICKZ family are related to cancer, cell proliferation, cell polarity and migration. The predicted domains present in IMP are 1 RNA-recognition motif (RRM) and 4 KH (hnRNP H-homology) domains. IMP interacts with the PSI via the KH domains (Suzuki et al., 2010). There are eight exons in *IMP,* and the terminal exon males-specific. Autoregulatory actions control the

male-specific splicing of *IMP* and require the A-rich elements in the terminal intron (Suzuki et al., 2014).

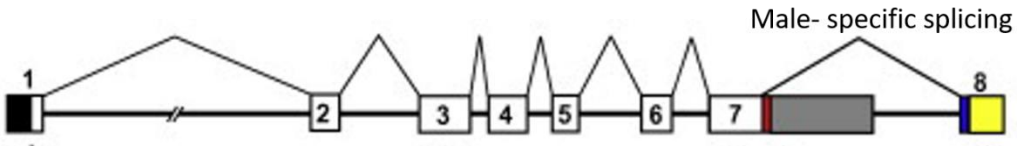

Figure 1. 10 The schematic diagram of the alternative splicing of the Imp pre-mRNA
Exons are the boxes, and the introns are lines. The exon number are numbers in a box. There is male-specific splicing between exons 7 and 8. The black box represents the 5' UTR. Figure from Suzuki *et al* (2014).

PSI is essential in the sexual development cascade. *Imp* and *PSI* interact with each other to regulates the males specific splicing in the double switch gene. Downregulation experiments of *PSI* leads to an increase in female-specific splicing of *Dsx*. Males exclude exons 3 and 4, in females, these exons are present, and they produce a female type *BmDsx* mRNA (Suzuki et al., 2008).

1.3.3.3 Signalling elements and 'Key' genes in *A. mellifera*

Apis mellifera is a haplodiploid organism. The unfertilised eggs have 16 chromosomes, whereas the fertilised has 32. The fertilised eggs that are homozygous at the sex determination locus (SDL) create diploid males, heterozygous eggs at the SDL develop into females. The worker bees kill diploid males once they hatch; however, inbreeding increases the probability of diploid males. Haploid eggs are fertile males that are hemizygous at SDL (Gempe et al., 2009).

The 'key' genes in *A. mellifera* are *complementary sex determiner (Csd)* and *Feminizer (Fem). Csd* is always heterozygous in the females and homo- or hemizygous in males; this state determines whether the *Csd* protein is active or not (Cook, 1993). *Csd* produces an SR-type protein which is a potential splicing factor (Gempe et al., 2009). The *Csd* has no transcription differences between sexes and is expressed after the blastoderm stage and throughout development (Beye et al., 2003).

CSD directly interacts with FEM. FEM is located on the sex determination locus (*SDL*), a region that is always heterozygous in females. *Fem* is located 12kb upstream of *Csd* and encodes an SR-type protein. *Fem* is an ancestrally conserved gene, in which *Csd* originates. *Fem* is an orthologue to the *D. melanogaster Tra* gene; it encodes a protein which has an Arg/Ser-rich domain and a proline- rich-domain. However, FEM lacks all conserved motifs found in TRA across different

species, apart from one; a 30-amino-acid motif in *C. capitata*. *Fem* has a male-specific splice variant, which produces a non-functioning protein as it contains has a premature stop codon. The female-specific splice variant encodes for a functional protein. Knockdown of female-specific *fem* splice variant, in *A. mellifera,* results in male bees (Hasselmann et al., 2008).

1.3.4 Double switch genes

Double switch genes are the last genes in the sex determination pathway that lead to male or female sex development (Marin and Baker, 1998; Oliveira et al., 2009). The double switch gene *doublesex* (*Dsx*) in many insects, *male abnormal-3 (Mab-3)* in *Caenorhabditis elegans* and *Dmrt1* in humans. Originally *Doublesex /Mab3* DNA-binding motif (DM) domain was found in *D. melanogaster* and *C. elegans*. The DM domain has been used to identify other sex-determination genes (Raymond et al., 1998), because the DM domain is conserved in DSX (Raymond et al., 1998; Yi et al., 2000). Other homologues of DSX genes occur encoding DM-domain proteins. Many of these also have roles in sexual development. Genes encoding proteins with DM domain are called *Dmrt* genes, and these genes are involved in sexual dimorphism (Kim et al., 2003).

Two functional oligomerisation domains characterise the DSX protein; the DM and the DSX-dimer. The DM and DSX dimer oligomerize through coiled-coil interactions, and they act as a DNA binding unit for sex-specific transcriptional activation or repression (An et al., 1996; Zhou and Li, 2016; Hodgkin, 2002). The zinc finger creates DNA binding sites (Erdman and Burtis, 1993). The DNA binding site regulated transcription and helps coordinate sex and tissue signals (Zhu et al., 2000).

The double switch genes can produce sex-specific proteins, by exon retention, intron retention or different transcription finishing site. In *D. melanogaster, Dsx* produces female and male-specific proteins- DSXF and DSXM. DSXM and DSXF are responsible for the dimorphic sexual characteristics of the fly (Erdman and Burtis, 1993). These proteins help regulate the differential gene expression, such as regulation for yolk protein (Burtis et al., 1991). The DSXM activates male differentiation while represses female differentiation (Jursnich and Burtis, 1993). The conserved residues are in the amino-terminal region in DSXF and DSXM (Ohbayashi et al., 2001). Figure 1.11 represents the different double switch proteins in insects and other phyla, highlighting the sex-specificity. The double switch genes in Insecta and Arachnida is *Dsx*, in Mammals and Zebra Fish it is *Dmrt1*, Sea urchin is *DmrtA2* and worm is *Mab-3*. All Insecta male and female isoform have DM and DSX dimer protein domains, and sex-specific exon retention (Xu et al., 2017).

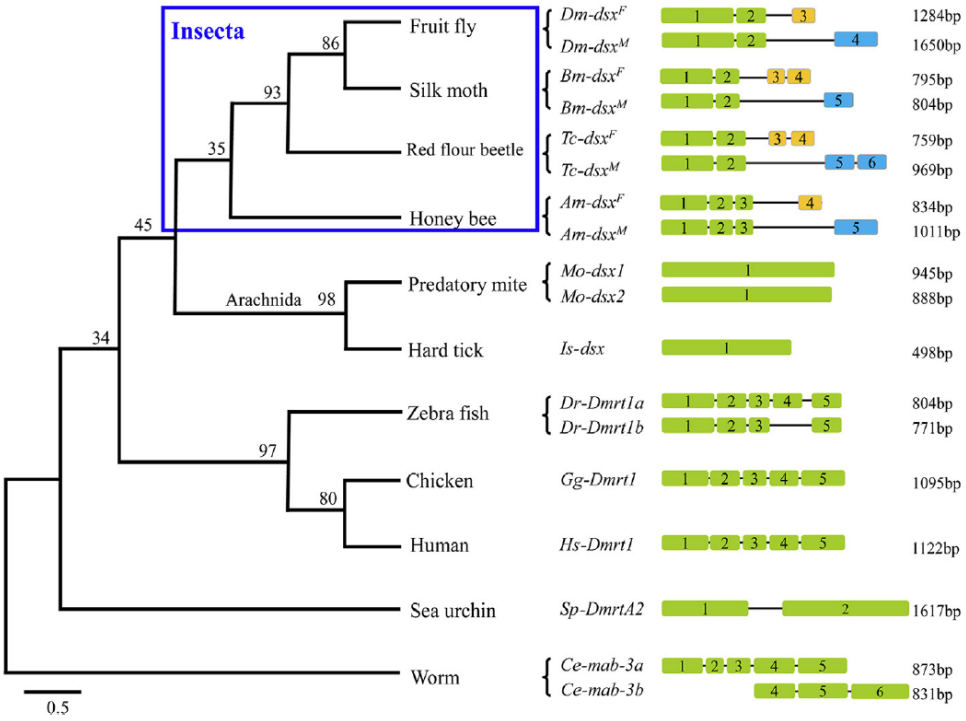

Figure 1. 11 Evolution of the C-terminal of DM domain genes in insects and other phyla
The right of the figure are the schematic representations of the gene structures. The green blocks represent the non-sex-specific spliced exons (1-6). Female and male-specific exons are represented in yellow and blue, respectively. Figure from Xu et al (2017).

1.3.5 Sex support genes

'Support' group genes are genes that help regulate some 'key' genes (Ellis et al., 1990) and produce specific courtship behaviours (Demir and Dickson, 2005; Finley et al., 1998). Genes assigned to the 'support' group and those that will be investigated further in Chapter 3 include, *daughterless (Da), deadpan (Dpn), extra-macrochaetae* (*Emc*)*, Female lethal d (Fl(2)d), grouchu (Gro), hermaphrodite (Her), sans fille (Snf), virilizer (Vir), dissatisfaction (Dsf)* and *fruitless (Fru)* of *D. melanogaster.* Some sex 'support' genes are required for regulation of *Sxl*; *Da, Emc, Fl(2)d, Dpn, Gro, Her, Snf* and *Vir.*

Da is a positive regulator of *Sxl,* while *Emc*, *Dpn* and *Gro* are negative regulators (Karandikar et al., 2005). The proteins encoded by these genes contain the Helix-loop-helix (HLH) protein domain. HLH protein domain present in proteins that contain sequence-specific DNA-binding proteins that can act as transcription factors. HLH contain a basic region needed for DNA-binding activity, and if this is not present then the HLH will function as negative regulators (*Emc* and *Dpn*) (Letunic and Bork, 2018; Ellis et al., 1990). GRO is a co-repressor which binds to hairy and hairy-related bHLH proteins encoded by the *Dpn* complex (Paroush et al., 1994). The Hairy/E (SPL) family is involved in DNA-binding transcription repressors which help regulate embryonic patterning, cell differentiation and other biological processes in vertebrate and invertebrates. The domain functions are transcription repressors involved in the regulation of differentiation, anteroposterior segmentation and sex determination in flies (Letunic and Bork, 2018; Davis and Turner, 2001). Regulation of *Sxl* needs *Fl(2)d, Snf* and *Vir* genes, specifically the sex-specific splicing of female *Sxl* RNA. There is a lack of information about the mechanism for these autoregulation *Sxl* genes. Conclusions have been based on *D. melanogaster* mutation work (Penalva et al., 2000; Granadino et al., 1990).

Three 'support' genes are essential for sex determination but are not involved in regulation of *Sxl*; *Her, Dsf* and *Fru*. Sexual differentiation in *D. melanogaster* requires *Her. Sxl* is at the top of the female-specific pathway for production yolk proteins whereas *Her* is at the top of the non-specific pathway (Li and Baker, 1998). *Dissatisfaction (Dsf)* is a gene that affects the sex-specific courtship behaviours and neural differentiation in both sexes. Expression of the female transformer cDNA under a *Dsf* enhancer leads to bisexual behaviour in males (Finley et al., 1998).

1.3.6 Current knowledge of the sex determination pathway in Hemiptera

What is known about the signalling elements is in Table 1.2.

Species	Signalling element	Reference
Silverleaf whitefly (*Bemisia tabaci*)	Haplodiploid	(Blackman and Cahill, 1998)
Asian citrus psyllid (*Diaphornia citri*)	XX/XY	(Yu and Killiny, 2018)
The soybean aphid (*Aphis glycine*)	XX/XO	(Morgan, 1909)
The pea aphid (*Acyrthosiphon pisum*)	XX/XO	(Morgan, 1909)
Green peach aphid (*Myzus persicae*)	XX/XO	(Morgan, 1909)
The brown planthopper (*Nilaparvata lugens*)	XX/XY	(Noda, 1990)
White-backed planthopper (*Sogatella furcifera*)	XX/XO	(Noda, 1990)
Bed bug (*Cimex lectularius*)	XX/XY *	(Sadilek et al., 2013)
Kissing bug (*Rhodnius prolixus*)	XX/XY	(Panzera et al., 2012)

***Table 1. 1 Sex determination in the Hemiptera;* B. tabaci, D. citri, A glycine, A. pisum, M. persicae, N. lugens, S. furcifera, C. lectularius *and* R. prolixus**

***The Bed bug has a strange sex-determining mechanism. There is often variation in the number of chromosomes, with the sex chromosomes showing extensive variation. The standard karyotype is 26 autosomes and varying number of supernumerary chromosomes which originate after the X chromosome fragmentation. Some populations can vary from 2 up to 14X chromosome (Sadilek et al., 2013).**

Some 'key' genes have been identified in some Hemiptera. *Tra2* has been found in *D. citri, N. lugens* and *B. tabaci. D. citri Tra2* is expressed throughout all developmental stages, and knockdown of the *Tra2* leads to lower progeny in *D. citri.* However, the sex ratio does not change (Yu and Killiny, 2018). When *N. lugens Tra2* is knocked down, the females become infertile pseudo males with undeveloped ovaries (Zhuo et al., 2017). In *B. tabaci Tra2* has no sex-specific isoforms (Liu et al., 2016); however, silencing causes malformation of male genitalia (Guo et al., 2018a). There is a *PSI* orthologue in *B. tabaci*, with 92 female-specific isoforms and 14 male-specific isoforms (Liu et al., 2016). There is a *B. tabaci Dsx,* which have 28 non-sex-specific isoforms silencing of *Dsx* caused malformation of male genitalia (Guo et al., 2018a). Full-length female

restricted *Tra* and male and female specific *Dsx* have been found in *R. prolixus*. However, the function of *R. prolixus Tra* and *Dsx* is currently unknown (Wexler et al., 2019).

Liu *et al* (2019) conducted a *B. tabaci* MED genome wide study looking for the sex determination genes. The putative sex determination genes (Key genes, Support genes and Double switch) where identified by TBLASTN along with a domain and phylogenetic analysis. The gene ID's and the locations are shown in Table 1.3. Sex-specific splicing analysis was conducted on transcriptomes from oviposited eggs, larvae stages and adult female and male stages. What was lacking in the paper was sex-specific splicing of sex determination genes in pre-oviposited eggs, which I hope to rectify later in this thesis.

Gene	Gene ID	Gene location
PSI*	BTA005137.1	Scaffold_15: 624811:653084 − 735 13 77.90 6.27
snf	BTA003437.1	Scaffold_1312# + 245 27.87 9.87
Sxl	BTA022642.1	Scaffold_506: 310383:334162 − 281 6 31.29 9.60
tra	BTA007388.2	Scaffold_1762: 3969:21280 − 455 11 52.82 10.06
tra2 *	BTA027641.1	Scaffold_800: 271218:282025 − 258 7 29.45 9.99
	BTA014916.3	Scaffold_301: 26103:39134 − 266 8 31.20 11.39
dpn	BTA027689.1	Scaffold_808:136323:145969 − partial
dsf	BTA004681.1	Scaffold_145# + 405 45.58 9.07
dsx *	BTA004042.1	Scaffold_1383# − 245 27.43 9.45
emc	BTA026084.1	Scaffold_684: 96020:103465 − 147 2 16.08 7.76
fl(2)d	BTA009661.1	Scaffold_209: 198137:205593 −
	BTA029394.1	Scaffold_92: 783237:790696 − 355 5 39.98 4.74
fru	BTA006255.2	Scaffold_1615: 100263:87130 − 420 7 46.67 6.47
gro	BTA020461.1	Scaffold_43: 104941:127175 − 705 18 76.20 7.16
Imp	BTA017124.1	Scaffold_349# + 592 63.77 9.09
vir	BTA014178.1	Scaffold_2874# + 1803 203.84 5.25

Table 1. 2 The putative sex-determination genes of* B. tabaci *MED. Data collected and adapted from Liu et al (2019).

****A gene that has been studied in a previous report (Liu et al., 2016; Guo et al., 2018a; Xie et al., 2014).***

#The gene sequence does not match well with the genome.

1.4 Control methods for *B. tabaci*

B. tabaci is an agricultural pest that needs to be controlled to minimise damage. Current methods involve biological control and insecticides. Current biological control methods for *B. tabaci* are *Amblyseius swirskii* (Cavalcante et al., 2015)*; Macrolophus caliginosus* (Lucas and Alomar, 2001); and *Nesdiocoris tenuis* (Arnó et al., 2009). *A. swirskii* is a predatory mite used for whitefly control. The mite consumes eggs and larvae of the whitefly (Calvo et al., 2015). *M. caliginosus* and *N. tenius* are Heteroptera, which feed on the pupae stage; both have multiple prey targets (Bonato et al., 2006; Arnó et al., 2009). Negatively, the two Heteroptera can damage the plant by weakening the apex and limiting plant growth, creates yield loss, and flower abortion (Sanchez, 2008; Arnó et al., 2009). Parasitic wasps are also a biological control agent against *B. tabaci*. Parasitic wasps can lay eggs next to (*Eretmocerus spp.*) or internally (*Encarsia spp.*) into early life stages of *B. tabaci* (Urbaneja et al., 2007; Arnó et al., 2009). Parasitoids have never been able to completely control the *B. tabaci* (Arnó et al., 2009). Biological control is hampered because of *B. tabaci* migration from surrounding areas.

Insecticides are the primary control method for whiteflies. However, aggressive applications leads to the evolution of insecticide resistance (Dittrich et al., 1985). *B. tabaci* MED and MEAM1 are resistant to many insecticides, MED more so than MEAM1 (Horowitz et al., 2005; Castle et al., 2009). Insecticide application can cause problematic side effects; such as, such as toxicity to livestock and humans and toxic residues in the environment. New control methods that do not pose resistance and off-target side effects need to be developed for *B. tabaci*. One such method is genetic control.

1.4.1 Genetic control

Genetic insect control is the introduction of genetic traits into a pest population that eliminates or reduces harmful insects (Curtis, 1985). The level of success of genetic insect control methods depends on the mating behaviour and the ways by which transferrable genetic traits pass to the next generation. Genetic control methods are species-specific; because, individuals of different species cannot or do not mate (due to i.e. genetic compatibility, spatial co-occurrence, niche, and mating behaviours), and this is a big advantage over chemical control. Genetic control is an area-wide method, so it is not dependent on the individual purchasing the control method, which may be a limiting factor in more impoverished areas. Genetic insect control reduces the competent vectors in an area (suppression technique) or reduces the vector capacities in a target population (replacement technique) (Alphey, 2014a). This method of control could be self-limiting or self-sustaining. Self-limiting approach introduces a genetic trait into the native population. In the long

term, the overall fitness is lower in transgenic insects, and the genetic trait disappears from the population. Therefore, to sustain the population control there needs to be frequent releases of the transgenic insects. In a self-sustaining approach the genetic trait may have neutral effects for the overall fitness of the insect pest and remains in the population requiring less frequent release of insect populations with the genetic trait (Alphey, 2014a).

Genetic control has many advantages compared to pesticide-based control methods. Many insect species have evolved resistance to insecticides, for example *B. tabaci* (Nauen and Denholm, 2005), *Anopheles funestus* (Hargreaves et al., 2000) and *Myzus persicae* (Bass et al., 2014). New insecticides need to be discovered. However, the number of companies that develop pesticides have declined due to the long production time of new active compounds (3.7-4.2 years) (Sparks, 2013).

Genetic control methods for many insect pest species have been developed. One of the oldest technique is the sterile insect technique (SIT), which is a self-limiting system (Alphey, 2014a). SIT is a species-specific technique that involves mass-reared males, that are irradiated by ionising radiation and released into the wild to compete with the wild males for the attention of females (Knipling, 1959). A proportion of the irradiated males will produce damaged sperm and are unable to fertilise the eggs, resulting in overall decline of the pest population growth (Bourtzis et al., 2016). SIT has been successful in the control of a number of insect pests, including for example *Glossina austeni* in Unguja, Zanzibar, (Vreysen et al., 2000), and the Mediterranean fruit fly (Medfly*) C. capitata* (Alphey et al., 2010). SIT has been tried to control *B. tabaci* as well (Calvitti et al., 2000), but no follow up studies were published since this paper in 2000, suggesting that SIT may not be the best method for controlling the whiteflies.

There are many problems with traditional SIT. Firstly, it has proven difficult to remove all females from an irradiated population. Moreover, separating the males from the females can be time-consuming, even with automated systems (Marois et al., 2012). Secondly, irradiated males are less competitive, in Medfly and Mexican fruit fly. Irradiation causes damage of somatic cells, and the gut microbiota (Ben Ami et al., 2010; Lauzon and Potter, 2012). However, irradiation gives sufficient sterility to males of *Anopheles arabiensis* and *Aedes albopictus* without impacting the competitiveness compared to the wild males of these species (Bellini et al., 2013) (Ageep et al., 2014).

1.4.2 The self-limiting gene technique

The self-limiting gene technique, previously known as release of an insect with a dominant lethal (RIDL), is a variation of the traditional SIT in the sense that the progeny does not survive. The self-

limiting gene technique requires the generation of transgenic insects with a sex-specific lethal construct (Figure 1.12), which is activated in the absence of tetracycline (Thomas et al., 2000). There are several essential components present in the sex-specific lethal construct: (1) the (DsRed) fluorescence marker for the detection of transgenic insects; (2) transposable elements for integration into the insect genome; (3) the tetracycline-repressible transactivator tTAV, which is lethal at high concentrations; and (4) the female-specific expression of tTAV (Figure 1.12).

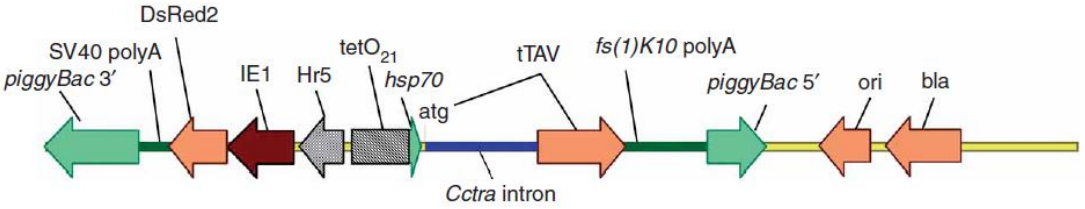

Figure 1. 12 Diagrammatic representation of the linearised plasmid pLA3097 with the sex-specific lethal construct flanked by the piggyBac transposon elements used in the* Ceratitis capitata *(Medfly).
The line represents the DNA sequence. The block arrows represent genes or promoters. Abbreviations: *piggyBac3'* and *piggyBac5'* = the left and rights borders, respectively, of the *piggyBac* transposable element; DsRed2 = red fluorescent marker with downstream SV40 polyA tail to enhance transcript stability; IE1 = baculovirus IE-1 promoter driving the expression of DsRed2; Hr5 = the enhancer of the IE-1 promoter; $tetO_{21}$ = 21 copies of the operator sequence that binds the (non-tetracycline binding) active tTAV transactivator; *hsp70* = promoter sequence that drives the expression of tTAV; *Cctra* intron = intron that is spliced out in females, but not in males; tTAV = transactivator composed of a fusion between the promoter-activating portion of the herpes simplex virus protein 16 (VP16) and Escherichia coli TetR, which is inactive when bound to tetracycline; *fs(1)K10 polyA* = poly-A tail enhancing the stability of the tTAV transcript; ori = origin of replication required for replication of the plasmid; *bla* = gene encoding the enzyme beta-lactamase giving antibiotic resistance to transformed bacteria. Upon introduction of the plasmid into the eggs, the plasmid may go to the germline cells, and the DNA between the *piggyBac* transposable elements is inserted into the insect genome. Insect cells that carry the *piggyBac*-flanked DNA will express the DsRed fluorescent gene and can be detected because of their red fluorescence. In males, the *Cctra* intron will remain unspliced, and the tTAV transactivator is not translated. Hence, the construct is not lethal for males. In females, the *Cctra* intron will be spliced out leading to the expression of tTAV. The tTAV protein will bind to tetO when there is no tetracycline present and will produce more tTAV giving a positive feedback loop producing high concentrations of tTAV. The VP16 portion of tTAV binds many promoter sequences in the insect leading to lethality. When females are reared in the presence of tetracycline the intron will still splice out, but tTAV is not expressed leading to survival of the females. Diagram was obtained from Fu et al. (2007).

1.4.2.1 The female-specificity in genetic control methods

In SIT trials, male-only releases are more effective compared to female mixed releases because females 'distract' the SIT males (Rendon et al., 2004). The transgenic genetic control systems rely on sex specificity driving the expression of a lethal gene in females only. In this system only females die, whereas males survive and can spread the self-limiting gene within the population; the male offspring of transgenic males (which do not die) are heterozygous for the self-limiting gene construct and the male and female progeny of the heterozygous males will have a 50% chance to inherit the self-limiting gene construct leading to 50% death of females (Thomas et al., 2000; Labbe et al., 2012). Thus, a system to control sex-specific expression is required.

Previous studies have reported the generation of self-limiting gene constructs that are female-specific and late acting. The female-specific gene *tra* is present in *D. melanogaster* and *Ae. aegypti* (Fu et al., 2007)(see Section 1.3 for more information). This gene is part of the sex determination cascade, which generates different splice variants for the *tra* genes in females and males (Figure 1.12). The intron specific for the female *tra* splice variant is inserted between the *tetO* sequence and the tTAV gene generating a one-component self-liming construct that is only lethal for females (Figure 1.12). Another female-specific splice variant is *Ae. aegypti* Actin-4 (AeAct-4), which is spliced only in the developing wing muscles of female L4 instar (Munoz et al., 2004). When the AeAct-4 intron is used in the self-limiting construct, then flightless female adults emerge (Fu et al., 2010). The aims of this thesis are to understand the SDG pathway and sex-specific splicing in order to identify elements in *B. tabaci* suitable for the self-limiting system.

1.4.2.2 Insect transformation systems

A self-limiting system needs germline transformation; therefore, the self-limiting construct needs to be introduced into the egg by microinjection before germ cells have developed (pre-blastoderm stage). Pre-blastoderm stage varies depending on the organism; for *Anopheles gambiae* and *Aedes aegypti* this is 1-2.5 hours after oviposition. The construct is deposited in the periplasmic space or near the yolk to promote DNA transfer into germline cells (Eggleston, 2014). One of the aims of this thesis is to better understand *B. tabaci* embryogenesis to guide this approach.

1.4.2.3 Advantages and disadvantages of the self-limiting system

There are many advantages to the self-limiting system. Firstly, insecticide-susceptible alleles can be introduced into a wild population (Alphey, 2014b; Alphey et al., 2009; Alphey et al., 2007). Secondly, there is an excellent level of flexibility using various versions of the same constructs, specifically the use of different promoters that drive the expression of the self-limiting genes. The

promoters can be active in females only or during specific developmental stages of the insect, for example during development of the flight muscles causing flightless females (Fu et al., 2010). Late acting fatalities may be beneficial as the developing eggs, larvae and pupae, and transgenic insects compete for resources with the native ones of the same species (Dye, 1984). Thirdly, insect populations can be reduced for an entire region and does not involve individual circumstances (for example, decisions of who should use the bed net within a larger family).

The disadvantages of the self-limiting system are the public perception of genetically modified organisms. The self-limiting system is invasive because it is not possible to limit the flight and spread of the released insects and the genes throughout the insect population (Alphey, 2014a). Genetic control can lead to political disagreements about the release of genetically modified organisms. The self-limiting system cannot eradicate the pests, because insects from other areas may invade the release area. Public perception has generally been positive (Alphey, 2014a), however recently Oxitec has been under attack by a paper suggesting that background genetics lead to increase hybrid vigor in the mosquito (Evans et al., 2019). Suggesting that the transgenic mosquitoes carry on breading and transmitting the transgenic element longer than it previously thought. Since this paper, Oxitec has complained to Nature Research as the studies does not identify anything that was not unanticipated, and the paper was very speculative and dramatizing (Oxitec, 2019).

Aims of the project

Bemisia tabaci is an important agricultural and economical pest. Resistance to pesticides is increasing in *B. tabaci*, therefore new control methods need to be developed. The industrial partner connected to this PhD is Oxitec, a company who have created a successful genetic control method; the self-limiting system. The aim of this research was to assess the feasibility of using the self-limiting system on *B. tabaci*.

For a self-limiting system to succeed, researchers must be aware of when they can insert the transgenic element and what they can insert. Traditionally, a transgenic element is inserted into the egg before the blastoderm has formed. Both the process of embryogenesis and sex determination are poorly understood in whitefly. The focus of this thesis is to build knowledge of these processes using a combination of approaches.

Overview of thesis contents

Chapter 3: Sex determination genes appear conserved across Hemiptera with diverse sexual lifecycles

This chapter asses the question of which genes are likely to be involved in the sex determination pathway in whiteflies and broadens this to investigate the pathway in hemipteran insects more generally. Comparative genomic analysis can be used on hemipteran genomes to discover the orthologues. 11 hemipteran genomes are publicly available. The aphids; *Aphis glycines, Acyrthosiphon pisum, Myzus persicae Clone 0* and *M. persicae G006*. The whiteflies; *B. tabaci MEAM1* and *B. tabaci* MED. The pysllid; *Diaphorina citri*. The planthoppers; *Nilaparvata lugens* and *Sogatella furcifera*. The public health pests; *Cimex lectularius* (bed bug) and *Rhodnius prolixus* (kissing bug).

Sex determination Genes (SDGs) were taken from *D. melanogaster*, *B. mori* and *A. mellifera*, and a reciprocal best blast hit technique (RBBH) was used to identify hemipteran SDGs orthologues. It was discovered that hemipteran insect species have potential orthologues of several sex determinations genes of *D. melanogaster* and *B. mori*, but not of *A. mellifera.*

Further investigation concluded that DSX hemipteran orthologues only contained one of the two domains usually found in functional DSX protein, the DM protein domain. Other known proteins that contain only the DM domain are also involved in sex determination/ differentiation. Hemipteran DM proteins were investigated and discovered they fell into certain clades; a

Hemiptera-specific, a *dmrt99b* clade, a *dmrt11e* clade and a *dmrt93b* clade. There is no true hemipteran DSX orthologue found in this study.

Chapter 4: Confocal microscopic investigation reveals embryogenesis stages of *B. tabaci* pre-oviposition eggs

This chapter aims to understand the stage at which the blastulation occurs during *B. tabaci* egg development. Germline transformation requires insertion of a transgenic element before the blastoderm stage. Before this thesis there was no published knowledge of embryogenesis stages in *B. tabaci*. *D. melanogaster* is a model organism for insect embryogenesis. Comparative analysis of the cell division numbers by staining nuclei in *D. melanogaster,* against the nuclei of novel embryos can help determine embryogenesis stages. Nuclei staining protocols were developed and optimised for *B. tabaci* in this chapter.

This chapter shows that unlike *Drosophila* and other dipterans, embryogenesis has already begun in pre-oviposition eggs. Morphological analysis of the pre-oviposition eggs shows migration of nuclei to the peripheral edges, indicating these eggs are starting to form the blastoderm. The number of nuclei in *B. tabaci* pre-oviposition eggs concurs with the morphological analysis.

Chapter 5: Single egg RNA-seq narrows down the developmental stages of *B. tabaci* eggs

This chapter uses a transcriptomic approach to further address the timing development of events and sex determination gene expression. Female lethality in self-limiting systems require genes to be expressed early and female-specific isoforms. SDGs are used in self-limiting system as they are conserved, expressed during embryogenesis and can have sex-specific isoforms. *B. tabaci* is a haplo-diploid organism; males are haploid. In this chapter, single pre-oviposition egg RNA-seq samples were successfully split into male and female, depending on SNPs percentage.

An embryonic development marker gene was used to test the RNA-seq bioinformatic pipeline and confirm the embryonic stages in more detail. The conserved embryonic gene, *Vasa* expressed at all embryonic and adult stages. *Vasa* is a gene that is expressed when germline cells are created and at or after blastoderm formation. Therefore, *Vasa* being expressed at all stages indicates the earliest eggs are after blastoderm formation. This result concurs with Chapter 4. SDGs were identified in Chapter 3. Sex-specific isoform and expression of these genes were investigated further. DM proteins and the 'key' genes had isoform data in the embryo stage and adults.

In this chapter, I discovered sex-specific isoforms in *Btdmrt3* (orthologue of *dmrt99b*) and other key genes. Most of the genes had expression in early embryogenesis (apart from *PSI*) and the expression varied with the sex of the samples.

Contributions to thesis

All experiments in this thesis were conducted by me, unless acknowledged.

Contributions of those who shared data, expertise or knowledge are listed in Table 1.3.

Name	Affiliation	Chapter	Contribution
Thomas Mathers	John Innes Centre	3	Provided the protein databases for hemipteran genomes Provided the early release of *A. glycine* genome Provided the differential expression data for Table 3.36
Sam Mugford	John Innes Centre	3	1:1 orthologue list of genes from *M. persicae* G006 and Clone 0 for Table 3.36
Adi Kliot	John Innes Centre	4	Provided equipment for the dissection of the *B. tabaci* adults
Michael Giolai	Earlham Institute	5	Sequenced the egg samples
Archana Singh	Earlham Institute	5	Provided the count data analysis for the gene expression analysis

Table 1. 3 Contribution of everyone (apart from myself) who shared data, expertise or knowledge in this thesis

Chapter 2: Materials and methods

2.1 Sex determination gene discovery

2.1.1 Pipeline for the bioinformatics

The bioinformatic pipeline for SDGs discovery is Figure 2.1.

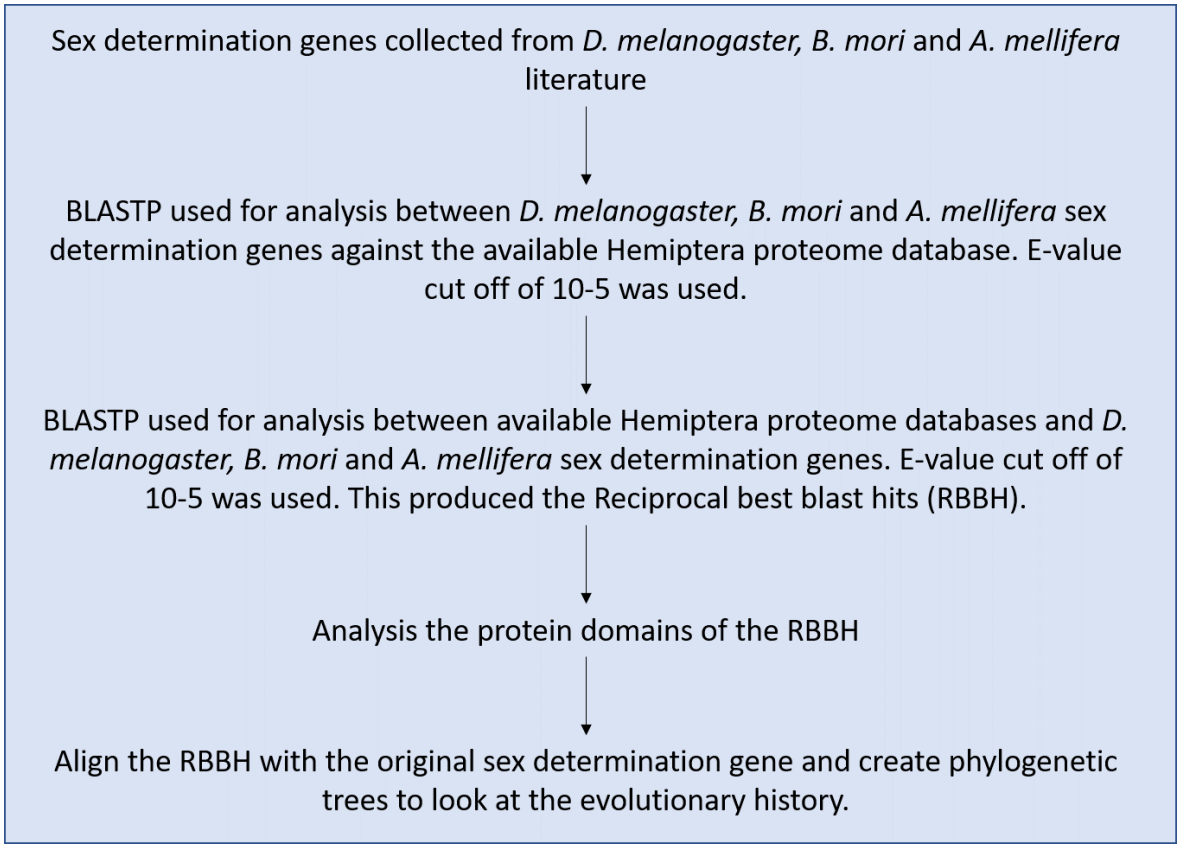

Figure 2. 1 Bioinformatic pipeline for sex determination gene discovery in Hemiptera
In *R. prolixus, C. lectularius, S. furcifera, N. lugens, M. persicae G006 and Clone 0, A. pisum, A. glycines, D. citri, B. tabaci MEAM1* and *MED.*

Sex determination proteins used for this analysis were described in the introduction. The Genebank accession numbers for each gene is found in Table 2.2. 11 Hemiptera genomes were used in this bioinformatic pipeline (*R. prolixus, C. lectularius, S. furcifera, N. lugens, M. persicae G006 and Clone 0, A. pisum, A. glycines, D. citri, B. tabaci MEAM1* and *MED)*. Table 2.1 shows the metadata of the genomes used in the analysis.

Thomas Mathers from the Hogenhout laboratory constructed a database of hemipteran protein sequences using gene predictions from 11 publicly available genome sequences (Table 2.1). For each species, the longest transcript per gene (LTPG) was selected as the representative transcript to avoid hits to multiple isoforms of the same gene when searching for RBBH. I used the *D. melanogaster, A. mellifera* and *B. mori* SDPs (Table 2.2) to create a full-length SDP database.

Specifications to the BLASTP results were conducted according to (He et al., 2015). The E-value threshold was set to 10^{-5}, this value has been used in previous papers for the cut-off for RBBH (Rachamim et al., 2015; Thorpe et al., 2016). E value is the expected value, the lower the E-value the more 'significant' the match is.

The sex determination proteins were 'blast searched' (using BLASTP function, version 2.2.30) against the hemipteran genomes. The genomes were Blasted back against the sex determination protein database. The top hits from both results gave the reciprocal best blast hit (RBBH). RBBH have a higher probability of being an orthologue to the sex determination genes.

2.1.2 Genomes from the sex determination pipeline

The genomes that were used are shown in Table 2.1.

Species	Date gathered	Release date	Notes	Link	Reference
Drosophila melanogaster	24/10/2017	23/08/2017	FB2017_04	ftp://ftp.flybase.net/genomes/Drosophila_melanogaster/current/	(Gramates et al., 2017)
Bombyx mori	24/10/2017	4/09/2013	GenesetA	http://sgp.dna.affrc.go.jp/ComprehensiveGeneSet/	(Suetsugu et al., 2013)
Apis mellifera	24//10/2017	Feb 2011	Amel 4.5	https://metazoa.ensembl.org/Apis_mellifera/Info/Index	(Weinstock et al., 2006)
Aphis glycines	20/02/2019	25/09/2019 (publicly), however had access earlier		https://zenodo.org/record/3453468#.XYyXxG5FzQY	(Mathers, 2019)
Acyrthosiphon pisum	24/10/2017	23/02/2010	ACYPI proteins v2.1b	http://bipaa.genouest.org/is/aphidbase/acyrthosiphon_pisum/downloads/	(Richards et al., 2010)
Myzus persicae Clone 0	24/10/2017	13/02/2017		http://bipaa.genouest.org/is/aphid	(Mathers et al., 2017)

				base/myzus_persicae/downloads/	
Myzus persicae G006	24/10/2017	13/02/2017		http://bipaa.genouest.org/is/aphidbase/myzus_persicae/downloads/	(Mathers et al., 2017)
Diaphorina citri	24/10/2017	04/07/2014	Annotation Release 100	https://i5k.nal.usda.gov/data/Arthropoda/diacit-%28Diaphorina_citri%29/Current%20Genome%20Assembly/	Genome Sequencing Project-RefSeq: Accession: PRJNA29447 ID: 29447 (http://www.ncbi.nlm.nih.gov/bioproject/PRJNA29447)
Bemisia tabaci MEAM1	24/10/2017	14/12/2016		ftp://www.whiteflygenomics.org/pub/MEAM1/MEAM1/	(Chen et al., 2016)
Bemisia tabaci MED	24/10/2017	19/04/2017		http://gigadb.org/dataset/100286	(Xie et al., 2017a)
Nilaparvata lugens	24/10/2017	20/04/2015		http://gigadb.org/dataset/100139	(Xue et al., 2014)
Sogatella furcifera	24/10/2017	15/11/2016		http://gigadb.org/dataset/100255	(Wang et al., 2017)
Cimex lectularius	24/10/2017	02/02/2016		https://www.ncbi.nlm.nih.gov/Traces/wgs/?val=JRLE01#contigs	(Rosenfeld et al., 2016)
Rhodnius prolixus	24/10/2017	August 2015	Rhodnius-prolixus-CDC_CONTIGS_RproC3.fa.gz	https://www.vectorbase.org/downloadinfo/rhodnius-prolixus-cdccontigsrproc3fagz	(Giraldo-Calderon et al., 2015)

Table 2. 1: All the genomes gathered for the sex determination analysis.

2.1.3 Sex determination protein NCBI

The sex determination proteins described in bioinformatic pipeline in Section 2.2.1

Gene	Species	Protein
Da	*Drosophila melanogaster*	NP_001260340.1
Dpn	*Drosophila melanogaster*	NP_476923.1
Dsf	*Drosophila melanogaster*	NP_001260109.1
Emc	*Drosophila melanogaster*	sp|P18491.2
fl(2)d	*Drosophila melanogaster*	NP_001246306.1
Fru	*Drosophila melanogaster*	NP_001262712.1
Gro	*Drosophila melanogaster*	NP_001287539.1
Her	*Drosophila melanogaster*	NP_001260506.1
Vir	*Drosophila melanogaster*	sp|Q9W1R5.1
Snf	*Drosophila melanogaster*	NP_511045.1
Sxl	*Drosophila melanogaster*	NP_001303551.1
Fem	*Bombyx mori*	-
Dsx	*Drosophila melanogaster*	NP_001262353.1
Masc	*Bombyx mori*	NP_001296506.1
Tra	*Drosophila melanogaster*	NP_524114.1
Tra2	*Drosophila melanogaster*	NP_995835.1
Imp	*Bombyx mori*	XP_004929907.1
PSI	*Bombyx mori*	NP_001103813.1
CSD	*Apis mellifera*	ABU68670.1
Fem	*Apis mellifera*	NP_001128300.1

Table 2. 2: The proteins used for the blastp analysis and the genbank ID on ncbi

2.1.4 Multiple sequence alignments

The protein sequences were analysed using the program CLC Main Workbench 8 (Qiagen Bioinformatics, Aarhus, Denmark), from now on the programme will be referred to as CLC. The RBBH protein sequences were placed into a list with the original query. The alignments were created using the CLC alignment function, with the default alignment algorithm parameters.

2.1.5 Full-length protein pairwise comparison

The full proteins were aligned as Section 2.1.4. The pairwise comparison was conducted using the program CLC Main Workbench 8 (Qiagen Bioinformatics, Aarhus, Denmark). Typically, only the pairwise identity was compared, and only between the query and the orthologue. The percentage identity is the percentage of identical residues in alignment positions to overlapping alignment positions between the two sequences.

2.1.6 Phylogenetic trees

The phylogenetic trees were created using MEGA 7.0.26 (Kumar et al., 2016). The protein sequences were aligned by “MUSCLE” alignment algorithm. The best-fitting model for any of the alignments was determined using the model selection (ML) function in MEGA 7.0.26, using the maximum likelihood statistical method. The best model was used to conduct a maximum likelihood tree with a bootstrap of 1000. Edits of the tree were in the Interactive tree of life (ITOL) (Letunic and Bork, 2016).

2.2 Whole mount *in-situ* protocol for *B.tabaci* MED embryos

2.2.1 *B. tabaci* MED maintenance

A stock colony of ***B. tabaci*** MED (Acquired from Alison Clarke, Rothemstead, collected from Mexico) which was assigned to a cryptic species by molecular analysis, was reared continuously on Cotton (Delta pine) in cages (52 cmx 52cm x50 cm) with a 16 h day and 8 h night photoperiod at 22 °C.

2.2.2 *B. tabaci* MED virgin collection

A mature cotton leaf was collected from the stock cage. Last stage pupae were identified on the cotton leaf. The pupae and the surrounding cotton leaf were separated from the main leaf by a single edge razor blade (Agar scientific LTD). Melted 1% water agar (Formedium limited) was placed into the wells of a 12-Well CytoOne flat bottom Plate (Starlab (UK) LTD). Leaf discs containing individual pupae were placed into the wells, with the pupae facing up. 100 mm x 38 m parafilm (Slaughter LTD, R&L) was stretched across the wells, with surgical 12.5 mm x 10 m micropore tape (Slaughter LTD, R&L) secured the parafilm to the plate.

The plates were left for 3 days, until adults emerged, in a controlled environment room with a 10 h day and 14 h night photoperiod at 22 °C. Adult females were identified by the genitalia and then dissected as per Section 2.2.4.

2.2.3 *B. tabaci* MED non-virgin collection

A mature cotton leaf was collected from the stock cage. The adults were removed from the leaf. The leaf was placed into a 140 mm triple vent Petri dish (Slaughter LTD, R&L) in 1% water agar (Formedium limited), abaxial side facing up. The lids were placed on the Petri dish, and surgical 12.5 mm x 10 m micropore tape (Slaughter LTD, R&L) secured the plate.

The plates were left for 3 days, until adults emerged, in a controlled environment room with a 10 h day and 14 h night photoperiod at 22 °C. Adult females were identified by the genitalia and then dissected as per Section 2.2.4.

2.2.4 *B. tabaci* MED dissection

0-3-day old females were collected from a Petri dish or a 12 well plate depending on if they were virgin or non-virgin. A short Pasteur pipette (Slaughter LTD, R&L) with sponge acting as a filter was used to move the adult females. The adult whiteflies were knocked out by CO_2 and placed on a 78 x 26 mm single cavity slides (Sigma-Aldrich CO LTD) with 1X Phosphate buffered saline (PBS) (Formedium). The dissection tools are made by melting an entomological pin into a 5ml

Eppendorf tube (this equipment was kindly given by A. Kliot). The female abdomens were cut with the dissection tool and the eggs dissected out.

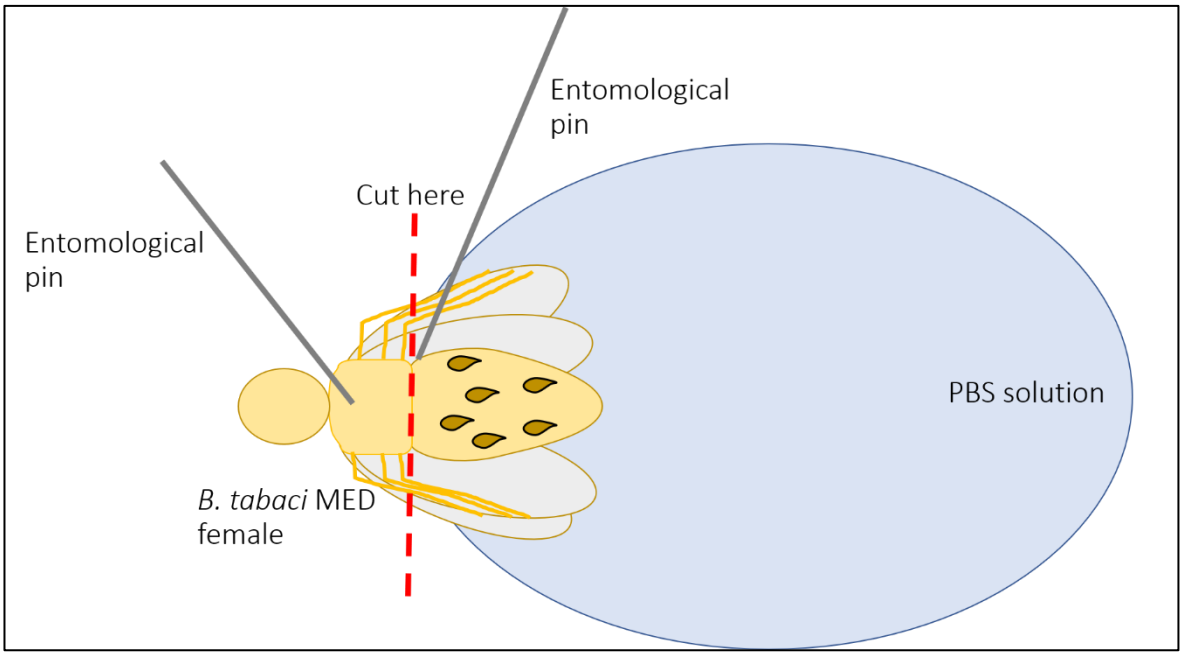

Figure 2. 2: A diagram of* B. tabaci *MED dissection, for collection of pre-oviposited eggs

2.2.5 Original DAPI protocol with oviposited eggs

The eggs were collected from the plant using the adapted entomological tool (described above) and the eggs were placed in a Carnoy fixative (Chloroform: Ethanol: Acetic acid in a 6:3:1 ratio) overnight. Post-oviposited eggs underwent a 2-hour bleaching (100 µl of 6% hydrogen peroxide in 400 µl of 99% ethanol). The eggs were placed in the hybridisation buffer (HB) (1 ml of 1M pH8 Tris, 9 ml of 5 M NaCl, 200 ml of 5% SDS, 25 ml of dH_20 and 15 ml of 30% formamide) with 0.5 ug/ml of DAPI (Sigma-Aldrich CO LTD). The eggs were left in the HB with DAPI for 10 minutes. The eggs were washed with HB and placed onto a slide with Vectashield (Vector Laboratories) was applied. A number 11 22 x 50 mm coverslip (Slaughter LTD, R&L) was placed onto the slide. Generic nail polish was used to seal the slide. The slides were stored in the dark at 4 °C until imaged on the Zeiss 780 Confocal microscope.

2.2.6 Propidium iodide staining

Propidium iodide has a different wavelength than DAPI and DAPI has the same wavelength as *B. tabaci* autofluorescence. Therefore, the propidium iodide staining was used to see if the nuclear structures are better visualised using this dye. The protocol was adapted from chapter 9 of Fluorescent analysis of Drosophila Embryos (Rothwell and Sullivan, 2000). There were 5 stages of the protocol; dechorionation, embryo fixation, rehydration of embryos, Propidium iodide (PI) staining and mounting.

Dechorination

The eggs were collected as described in Section 2.3.3. The eggs were dechorinated by a 30% bleach wash for 5 minutes at room temperature.

Embryo fixation

Embryo wash solution contained 70 g of 7% NaCl, 5 ml of 0.5% Triton X-100 and H_20 added to 1 litre. The eggs were washed with the embryo wash solution and 1 ml heptane. PEM buffer (0.1 M PIPES, 1 mM $MgCl_2$, 1 mM EGTA and the pH was adjusted with KOH) was added to 3.7% formaldehyde in a 50:50 ratio. The bottom formaldehyde layer was removed from the eggs. 1 ml of methanol was added to the eggs for 1 minute before the removal of the top heptane layer was then removed.

Embryo rehydration

250 µl of PBTA solution (50 ml 10X PBS, 5 g 1% BSA, 250 µl 0.05% Triton X-100, 0.1 g of 0.02% sodium azide, buffer was adjusted to 500 ml of water) was added to the eggs for 15 minutes at room temperature; this rehydrated the eggs.

Propidium iodide (PI) staining and mounting

PBTA buffer was removed. The eggs were submerged in 10mg/ml of RNase and incubated at 37 °C for 2 hours. RNase was removed. The eggs were washed with PBTA solution and then PBS-azide (1X PBS and 0.02% sodium azide). The PBS-azide was removed and 40 µl of PI mounting medium was added to the eggs (10 mg/ml 1,4-phenylenediamine in 10x PBS- 10ml of this mixture was added to 90 ml of glycerol, 1 µg/ml of PI added). The embryos were sealed under a cover slip and imaged immediately on the Zeiss 780 Confocal microscope.

2.2.7 Final optimised DAPI-staining protocol

The female bodies were removed from the slide and carnoy fixative (Chloroform: Ethanol: Acetic acid at 6:3:1) was placed onto the slide for 10 minutes, covering the eggs. The fixative was removed and left in a hybridisation buffer (1 ml of 1 M pH8 Tris, 9 ml of 5 M NaCl, 100 ml of 5% SDS, 25 ml of dH_2O and 15 ml of 30% formamide) overnight at 4°C. 1.5 µl of 10 µg/ml of DAPI slides (Sigma-Aldrich CO LTD) was placed into 100 µl hybridisation buffer creating a DAPI infused buffer. The DAPI infused hybridisation buffer was placed onto the eggs for 10 minutes at room temperature. The excess DAPI infused buffer was removed and antifade mounting medium Vectashield (Vector laboratories) was applied. A number 11 22x50 mm coverslip (Slaughter LTD, R&L) was placed onto the slide. Generic nail polish was used to seal the slide. The slides were stored in the dark at 4 °C until used on the Zeiss 780 Confocal microscope.

2.2.8 Confocal microscopy

Nuclei counting of embryos was conducted on a Zeiss LSM780 confocal microscope (located in John Innes Centre). Eggs were imaged under air medium and a magnification of 20x, in both brightfield and under excitation. DAPI was excited at 405 nm with a Diode 405-30 laser. The DAPI fluorescent emission was detected between 436-475 nm. The exposure was kept constant within experiments. The confocal microscope experiment was conducted per the standard operation protocol set out by the Bioimaging facility in JIC (Calder, 2017). A brightfield image and a Z-stack fluorescence image was taken for each egg.

2.2.9 Embryo analysis in FIJI

The confocal microscope produced Carl Zeiss format (CZI) files that were analysed using FIJI software (National Institutes of health, USA; v1.52c) (Schindelin et al., 2012). A scale of 10 µm was placed on each image. The files were converted into 16-bit images for DAPI signal analysis and then into graphics interchange format (GIF). Slices of individual Z-stack could be analysed and saved as JPEG. A maximum intensity Z projection was used to count the nuclei.

2.3 Single embryo RNA-seq

2.3.1 Sample preparation

Virgin and potentially non-virgin females were collected, as detailed in Section 2.2.2 and 2.2.3. The females were dissected in ice-cold 1X Phosphate buffered saline (PBS) (Formedium). The morphological developmental stage of the eggs was catalogued into A, B/C and D (described in Section 1.2.1) and placed into a 1.5 ml Eppendorf tube with a buffer; 0.2% (vol/vol) Triton X-100 and 2 µl RNase inhibitor. RNA was extracted using the protocol described in Picelli *et al* (2014) and performed by Michael Giolai from the Hogenhout laboratory (Picelli et al., 2014).

2.3.2 Library preparation method and quality control

M. Giolai sequenced with NextSeq500 and produced 150bp pair-end reads. The reads were trimmed using cutadapt. End-trimming was done by a quality threshold of 20, all read pairings with at least one read shorter than 100 bases post adapter/ quality trimming was also removed.

2.3.3 Determining the sex of the samples

STAR 2-pass method (version star-2.6.0c) was used to align the reference genome with the individual RNA-seq files (Dobin et al., 2013). Picard was used to read groups, sort, mark duplicates and create an index (Institute, 2019).

The GATK tool (version 3.7.0), Split'N'Trim split reads in the exon segments and hard clip any sequences overhanging into the intronic regions. GATK's ReassignOneMappingQuality read filter was used to reassign good alignments to the value of 60 (McKenna et al., 2010).

The GATK tool, HaplotypeCaller, calls the variants producing a .vcf file. The GT values were collected, and the homozygosity and heterozygosity percentage calculated.

2.3.4 Creating a transcriptome assembly

Reads were aligned to the *B. tabaci* MED/Q reference genome (Xie et al., 2017b) by HISat2 (v2.1.0) (Pertea et al., 2016). Sequence alignment map (SAM) files were converted to Binary AlignmetMap (BAM) file using Samtools (v1.7) (Li et al., 2009). Transcriptome compilation of all fastq files completed by stringtie (v1.3.5) (Pertea et al., 2015) to produce a gene transfer format (gtf) file. Individual embryo gtf files were converted back to cDNA by cufflinks (v2.2.1) Gff read function (Trapnell et al., 2010).

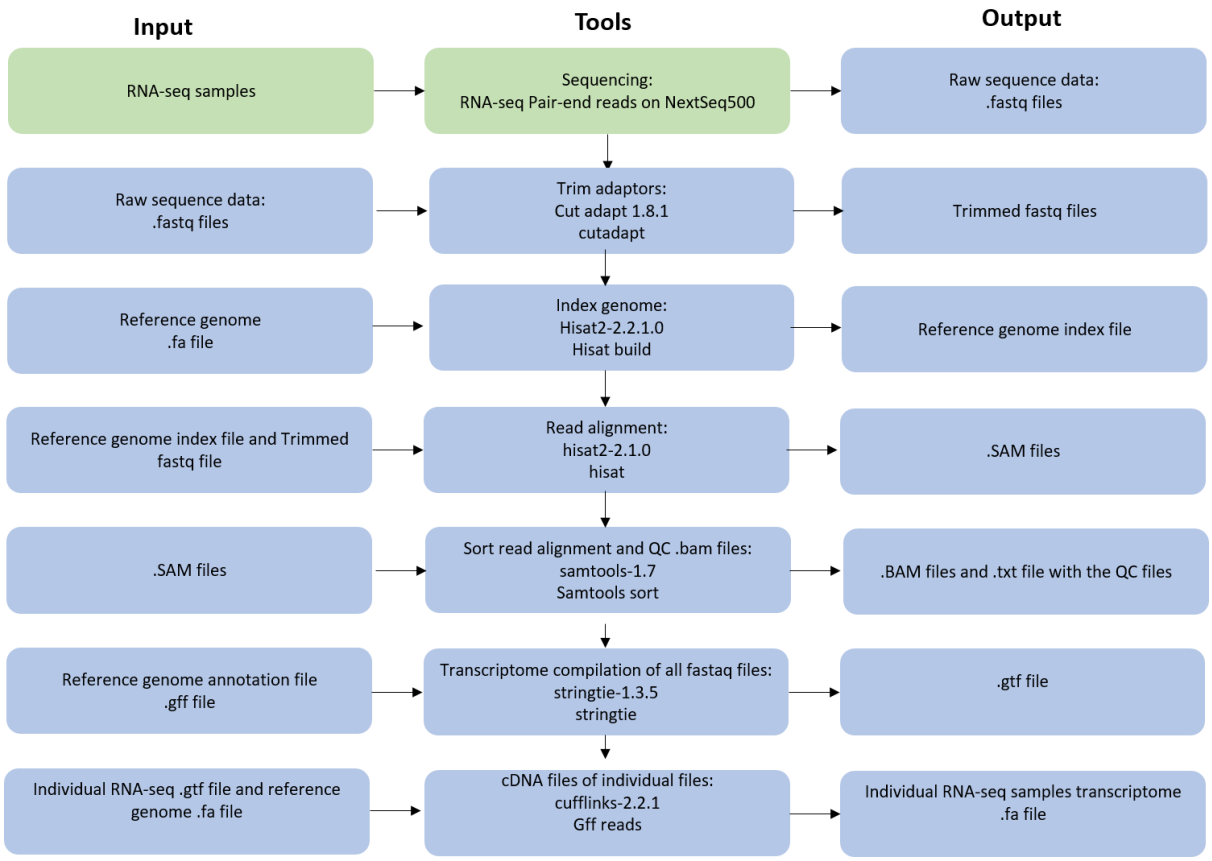

Figure 2. 3 Bioinformatic pipeline for transcriptome creation

2.3.5 Creating files for gene expression analysis

A transcriptome file was created with Samtools (v1.7) Samtool merge function (Li et al., 2009). The merged transcriptome files were converted back to cDNA by cufflinks (version 2.2.1) Gff read function (Trapnell et al., 2010). Transcriptome file was indexed by Kallisto (version 0.42.3) usinf kallisto index function (Bray et al., 2016). The single embryo RNA-seq fastq files were aligned to the new transcriptome by Kallisto (v0.42.4) using the Kallisto quant function (Bray et al., 2016). The count matrix was completed by Archana Singh from the Hogenhout laboratory.

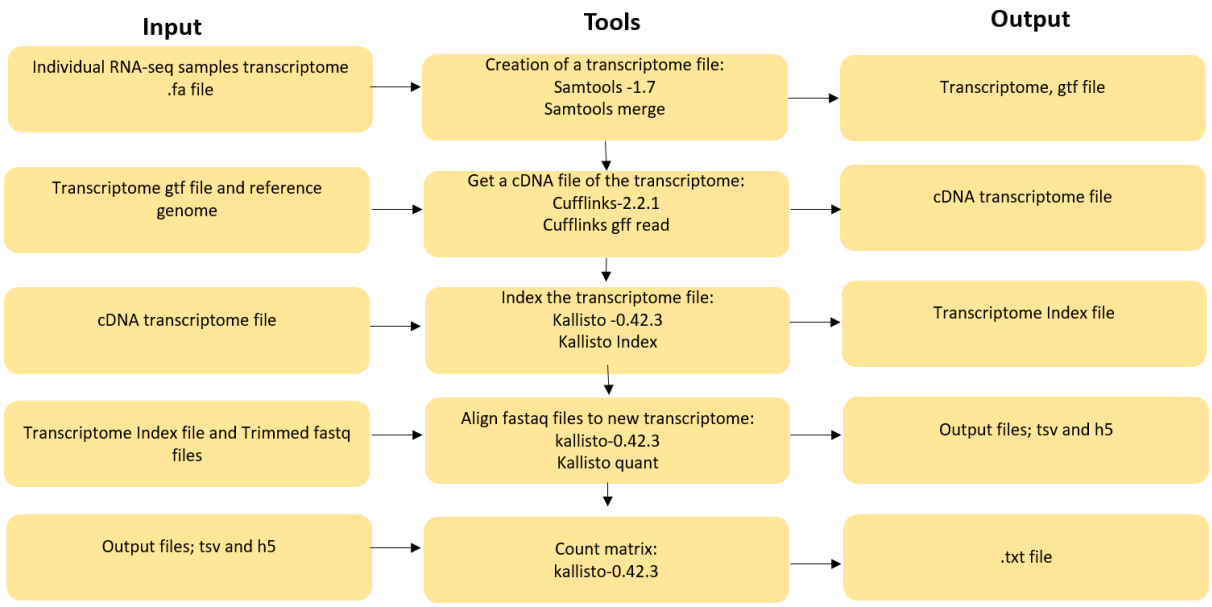

Figure 2. 4 Bioinformatic pipeline for gene expression analysis

Chapter 3: Sex determination genes appear conserved across Hemiptera with diverse sexual lifecycles

3.1 Introduction

Insects of the order Hemiptera are known as true bugs; they are a large order of insects. Hemiptera split from holometabolous insects 320 million years ago (see Figure 1.1) (Hogenhout and Bos, 2011b; Hogenhout et al., 2009). Hemiptera have common mouthparts, and many are herbivores and economic pests.

Insecticide resistance is on the rise in many hemipteran species (Nauen and Denholm, 2005). New control methods need to be developed to tackle agricultural and public health pests. One potential control method is genetic control. Oxitec is a company that has created the self-limiting system; this system causes a female-specific lethality. Traditionally, sex determination genes (SDGs) are the target for female-specific lethality due to expression in early developmental stages. Some SDGs have female-specific isoforms (Xu et al., 2017). The development of genetic control tools in Hemiptera will require in-depth knowledge of sex-determination genes. Currently, there is a lack of knowledge of sex-determination genes in Hemiptera; this chapter will rectify this.

The term orthologue, in this thesis, refers to the genes in different species that have evolved from a common ancestral speciation. Thereby, I will be using the term as I will be trying to find sex determination genes in hemipteran species. Orthologues can retain the same function; any candidate orthologues found in this chapter will potentially have the same function.

Sex determination pathways in the holometabolous insects, *Bombyx mori* (Lepidoptera), *Drosophila melanogaster* (Dipteran) and *Apis mellifera* (Hymenoptera), have been well described (Cline and Meyer, 1996). SDGs were identified and functionally characterised in these three species. The sex determination pathways of these three insect species are different, though their cascade structures are similar ((Sakai et al., 2015; Yamamoto and Koganezawa, 2013; Herpin and Schartl, 2015; Hoff, 2009); see Fig. 1.9 of main introduction). Other insect species have sex determination gene orthologues. Complete sex determination pathways are under-researched in species other than the three holometabolous insects (mentioned above).

A detailed description of the sex-determination genes in *Bombyx mori* (Lepidoptera), *Drosophila melanogaster* (Dipteran) and *Apis mellifera* (Hymenoptera) are provided in the general introduction of this thesis (Sawanth et al., 2016). Briefly, genes of the sex determination pathway form three main groups. The first group contains one gene named *doublesex* (*Dsx*). This gene is conserved in most organisms for which sex determination genes are identified, including invertebrates and humans. The second group contains the 'key genes', so-called because it regulates processes upstream of *Dsx* directly or indirectly. Genes assigned to the 'key gene' group include *D. melanogaster Sex-lethal (Sxl), transformer (Tra), transformer 2 (Tra2)*, *B. mori*

masculinizer (Masc), IGF-II mRNA binding protein (Imp), feminizer (Fem) and *P-element somatic inhibitor* (*PSI*) and *A. mellifera CSD and Feminizer (Fem)*. The latter is the orthologue of *D. melanogaster Tra,* and *Fem* and *Tra* directly regulate *Dsx* in *A. mellifera* and *D. melanogaster*, respectively (Hasselmann et al., 2008). Finally, the third group is named the 'support group' and contains genes that help regulate some 'key genes' and produce specific courtship behaviours. Genes assigned to the 'support group' include, for example, *daughterless (Da), deadpan (Dpn), extra-macrochaetae* (*Emc*)*, Female lethal d (Fl(2)d), grouchu (Gro), hermaphrodite (Her), sans fille (Snf), virilizer (Vir), dissatisfaction (Dsf)* and *fruitless (Fru)* of *D. melanogaster.* Members of the 'support genes' group are necessary for the function of the 'key gene' group, specifically for their regulation (Ellis et al., 1990). There are support genes that take a sex differentiation role and regulate, for instance, sex-specific courtship behaviours (Demir and Dickson, 2005; Finley et al., 1998).

Specific protein domains characterise sex determination proteins (SDPs); orthologues are found in organisms using the protein domains. For example, DSX has a doublesex/Mab3 DNA binding motif (DM) which is conserved among all DSX proteins identified so far (Marin and Baker, 1998). The DM domain is essential for the sex determination function of DSX because it binds to the promoter of proteins regulated by DSX (Erdman and Burtis, 1993; Zhu et al., 2000). The *D. melanogaster* DSX has a second domain; DSX dimerization domain. DSX dimerization domain is less conserved among orthologues of organisms other than the drosophilids. Whereas the DM domain is conserved in DSX, it is not the only protein that has this domain. DMRT proteins contain DM domains and are involved in sex determination or differentiation. Therefore, the DM domain is useful to find genes involved in sex determination in other organisms (Kato et al., 2008).

The sex determination pathway in Hemiptera is poorly defined. Some SDGs/ SDPs in Cassava whitefly (*Bemisia tabaci* MED*)* already been identified (see Section 1.3.6) (Liu et al., 2019), including one *Dsx* orthologue and some SDGs in the 'key gene' and 'support gene' groups. In this chapter, reciprocal best blast hit (RBBH), and protein domain analysis have identified SDGs/SDPs in *B. tabaci* and other hemipteran species. The hemipteran species investigated are those with public genomes available; *Aphis glycines, Acyrthosiphon pisum, Myzus persicae Clone 0, M. persicae G006*, *B. tabaci MEAM1* and MED, *Diaphorina citri*, *Nilaparvata lugens*, *Sogatella furcifera*, *Cimex lectularius* and *Rhodnius prolixus* (see Table 2.1, materials and methods, for more information).

There is variation in signalling elements among the 11 hemipteran populations tested. The signalling element differences are; different life cycles (sexual versus asexual reproduction of

aphids versus only sexual reproduction in other hemipteran species), and variations in genome structures between the sexes (diploid versus haploid of whiteflies versus XX/XO or XX/XY systems of other insect species). The hypothesis is that the variation in sex determination pathways will be due to the signalling elements (more detail in Section 1.2.4).

In this chapter, the results show that the hemipteran insect species have potential orthologues of several sex determinations genes of *D. melanogaster* and *B. mori*, but not of *A. mellifera*. Surprisingly, whereas proteins with DM domains are present and conserved among hemipterans, a direct homolog of DSX was not identified. Most SDGs/SDPs appear conserved among the hemipterans, except for *D. melanogaster* HER that was not present in aphids. This research chapter has elucidated some surprising findings that did not confirm some published data. SDGs/SDPs evolution in hemipterans will be investigated and discussed later in the chapter.

3.2 Results

3.2.1 First assessment of the presence of sex determination genes in Hemiptera insects

This chapter describes the investigation of hemipteran sex determination proteins. Firstly, by reciprocal best blast hit (RBBH) analysis, and secondly by protein domain investigation. Section 3.2.1 will cover the RBBH analysis.

Basic Local Alignment Search Tool (BLAST) finds regions of similarity between biological sequences, i.e. the query sequence against the subject sequence. BLASTP is a programme which compares protein sequences to a protein database and calculates the statistical significance. In this Chapter, BLASTP analysis was conducted on full- length SDPs (found in *D. melanogaster, A. mellifera* and *B. mori,* Section 1.3) and hemipteran protein databases (from the 11 public genomes, Table 2.1). In this Chapter, all BLASTP analysis is done to specification of Figure 2.1, unless otherwise stated. Reciprocal Best Blast Hit (RBBH) contains an extra BLASTP analysis of the hemipteran proteins against the full-length SDP database.

T. Mathers constructed the hemipteran protein databases (Table 2.3) from all 11 public hemipteran genomes (Table 2.1). Construction of hemipteran protein databases had two stages. Firstly, the identification of the longest transcript per gene (LTPG), and secondly, the translation of LTPG into protein sequence. Hemipteran protein database, which only contains the LTPG, prevents hits to multiple isoforms in the same gene in the RBBH analysis (Thorpe et al., 2016). I used the *D. melanogaster, A. mellifera* and *B. mori* SDPs (Table 2.2) to create a full-length SDP database.

One of the statistics that BLASTP conducts is the percentage similarity. The numbers (0-100) in Figure 3.1 are percentage similarity. RBBH analysis revealed that hemipteran have putative SDP orthologues (Figure 3.1). All 11 Hemiptera had putative orthologues for *D. melanogaster* DSX (Figure 3.1), at 60% protein identity or higher. All 11 Hemiptera had putative 'key' SDP orthologues; *B. mori* IMP, MASC and PSI, and *D. melanogaster* SXL and TRA2. However, there were no results for four 'key' SDP orthologues; *A. mellifera* CSD and FEM, *B. mori* FEM and *D. melanogaster* TRA. RBBH analysis has stricter requirements than BLASTP analysis; BLASTP analysis of the four 'Key' SDP also revealed no candidate orthologues.

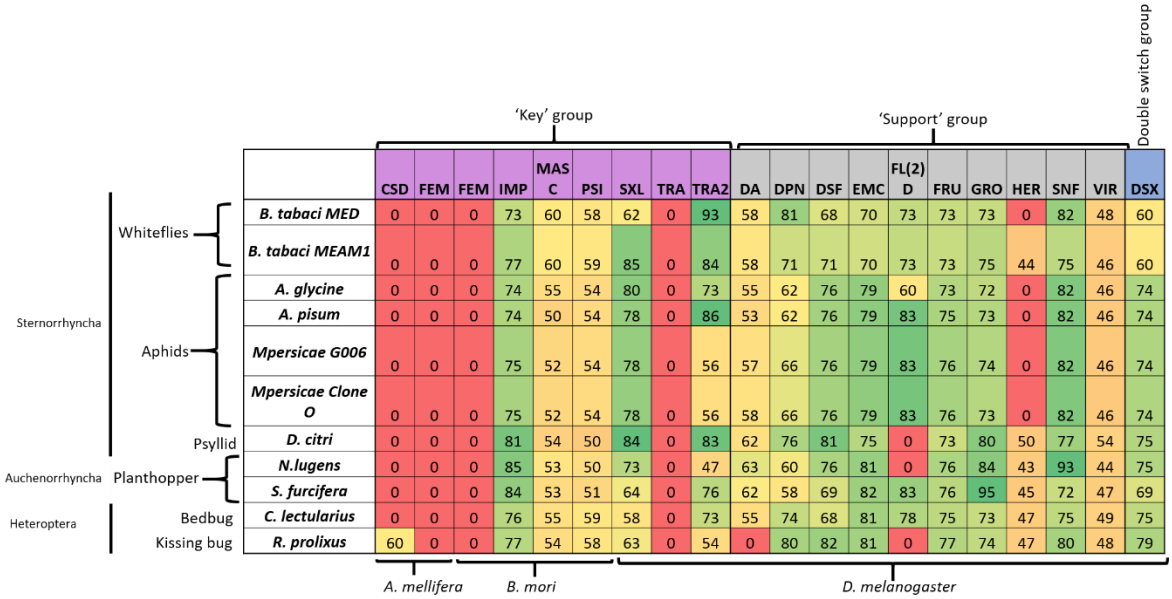

Figure 3. 1: A summary table of the positive-scoring matches of the RBBH results.
Hemipteran proteins (rows) with the highest similarity scores to sex determination proteins of *D. melanogaster, B. mori* and *A. mellifera* (columns, see Fig. 1.6 Introduction). Similarity scores in % are shown for each protein-protein comparison with the lowest scores highlighted in red and highest scores in the darkest green. Please note there are two columns for FEM, however one is from *A. mellifera* whilst the other is from *B. mori*. The two FEMs have the same name but are not orthologues of each other.

Whereas these findings in this section are promising, more in-depth analyses whether the identified hemipteran proteins are sex-determination proteins are needed, and this includes assessments of the presence of conserved domains. The next sections of this chapter contain further analysis for each SDP.

3.2.2 Assessment of the presence of DSX homologs in Hemiptera

RBBH analyses above (methodology; fig. 2.1) identified DSX putative hemipteran orthologues (PHO) in 11 hemipteran genomes. BLASTP analysis produces a report of the alignment statistics; e-value of the alignment between query and subject (more information in Section 2.1.1), length of BLASTP alignments in amino acids, query coverage, percentage of identity and the percentage of similarity. Figure 3.2 contains the BLASTP alignment statistics between the known SDPs and the hemipteran protein databases. Percentage similarity (Table 3.1) is of the alignment and not the whole protein. Table 3.3 contains the percentage identity of the pairwise comparison between the SDP and putative hemipteran SDP full- length protein.

BLASTP alignments showed high protein identities and similarities of the putative hemipteran DSX orthologues to *D. melanogaster* DSX (Table 3.1). However, CLC alignments of whole proteins showed the identity percentages were low (Table 3.2).

Hemipteran species	Subject (hemipteran protein) ID	E-value	Lengths of blastp alignments in amino acids	Query Coverage	Identity (%)	Similarity (%)
B. tabaci MEAM1	Bta13246	2.00E-21	107	40	49.53	59.81
B. tabaci MED	BTA013024.1	2.00E-21	107	40	49.53	59.81
D. citri	D_citri_rna13694	5.00E-22	73	27	58.9	75.34
A. glycines	A_glycine_01584	4.00E-21	62	11	64.52	74.19
A. pisum	ACYPI004122-RA	1.00E-19	66	25	62.12	74.24
M. persicae G006	MYZPE13164_G006_v1.0_000112290.1	7.00E-20	66	25	62.12	74.24
M. persicae Clone O	MYZPE13164_0_v1.0_000078560.2	7.00E-20	66	25	62.12	74.24
N. lugens	NLU005873.1	9.00E-20	64	24	62.5	75
S. furcifera	Sfur-8.45	9.00E-14	67	25	49.25	68.66
C. lectularius	CLEC005903	1.00E-20	67	25	61.19	74.63
R. prolixus	RPRC008869-RA	9.00E-26	70	26	70	78.57

Table 3. 1 Results of BLASTP analysis of the full-length* D. melanogaster *DSX protein (Query) against all hemipteran PHO (Subject).

Species	Subject ID	Identity (%) of full-length aligned protein against the query
B. tabaci MEAM1	Bta13246	8.78
B. tabaci MED	BTA013024.1	8.78
D. citri	D_citri_rna13694	10.32
A. glycines	A_glycine_01584	13.43
A. pisum	ACYPI004122-RA	13.43
M. persicae G006	MYZPE13164_G006_v1.0_000112290.1	13.27
M. persicae Clone O	MYZPE13164_0_v1.0_000078560.2	13.27
N. lugens	NLU005873.1	13.09
S. furcifera	Sfur-8.45	10.27
C. lectularius	CLEC005903	8.70
R. prolixus	RPRC008869-RA	8.97

Table 3. 2: Percentage identity of full-length* D. melanogaster *protein against the full-length proteins of the putative DSX from hemipterans identified in RBBH.

DSX in *D. melanogaster* contains a DM domain and DSX dimer domain (Fig. 3.2). DM is present in most DSX proteins, but also in some other sex differentiation genes. I analysed all of the DSX-PHO for the presence of protein domains using the SMART protein analysis (Letunic and Bork, 2018). The DSX-PHO of 11 hemipteran genomes only have the DM protein domain, but no DSX dimer domain (Fig. 3.2).

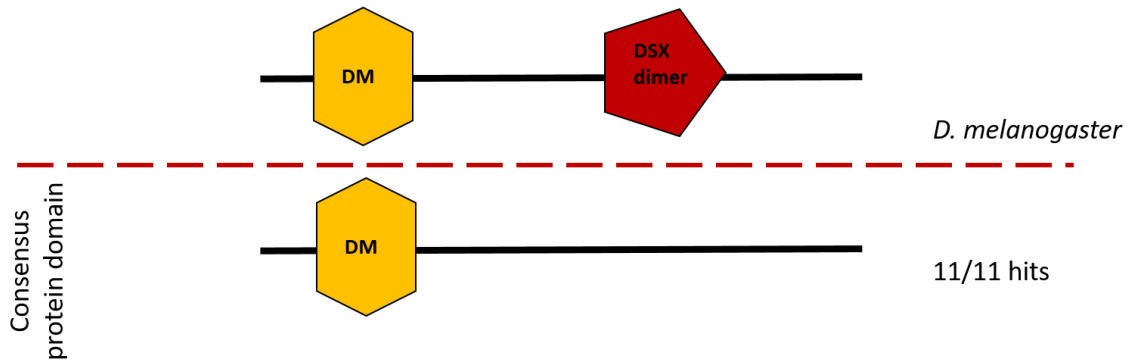

Figure 3. 2 A simplified schematic drawing of domains found in* D. melanogaster *DSX and putative DSX proteins of 11 hemipteran genomes.

There are no DSX dimer domains present in the RBBH hemipteran DSX proteins. However, BLASTP analysis may find the DSX dimer protein domain in Hemiptera. Firstly, I conducted BLASTP analysis to Figure 2.1 specifications. There was no blast hit results with Figure 2.1 specifications, indicating the DSX domain at this threshold does not show any similarities to hemipteran protein. Secondly, I conducted the BLASTP analysis with higher E-value threshold (10^{-2}); there was one blast hit result from *S. furcifera* (Sfur-199.19).

The DSX dimer in hemipteran could be highly divergent from *D. melanogaster*; Sfur-199.19 could be used to reveal hemipteran specific DSX dimer. BLASTP search between Sfur-199.19 DSX dimer domain and the hemipteran proteins revealed one result; *N. lugens* (NLU018480.1). Table 3.3 contains the BLASTP results. NLU018480.1 and Sfur-199.19 both have a DM protein domain and a DSX dimer domain; so are probably DSX proteins.

Species	Subject Seq-id	E-value	Raw score	Alignment length	Query Coverage Per Subject	Percentage of identical matches
S. furcifera	Sfur-199.19	0	218	40	100	100
N.lugens	NLU018480.1	1.00E-15	168	36	90	80.56

Table 3. 3 Potential DSX dimer domain found in Hemiptera.
The only BLASTP hit results of DSX dimer (*S. furcifera*) against the rest of the hemipteran genomes.

D. melanogaster has 4 DM domain proteins; DSX, DMRT93B, DMRT99B, and DMRT11E. The DMRT proteins are found in other organisms beyond *D. melanogaster* and have a sex differentiation function. The RBBH analyses I have conducted has identified 1 DM protein. There may be more than 1 DM protein in the 11 hemipteran protein databases. I used BLASTP to search the hemipteran proteins using the DM protein domain as the query, results in Table 3.4.

Species	Subject Seq-id	RBBH	e-value	Raw score	Alignment length	Query Coverage Per Subject	Percentage of identical matches	Percentage of positive-scoring matches
A. pisum	ACYPI004122-RA	Yes	5.00 E-18	192	52	96	69.23	76.92
A. pisum	ACYPI007850-RA	No	8.00 E-16	173	52	96	61.54	73.08
A. pisum	ACYPI43968-RA	No	2.00 E-08	122	49	91	55.1	65.31
A. glycines	A_glycine_015300	No	2.00 E-09	127	49	91	55.1	65.31
A. glycines	A_glycine_01584	Yes	6.00 E-18	192	52	96	69.23	76.92
B. tabaci MED	BTA004042.1	No	7.00 E-09	122	48	89	52.08	60.42
B. tabaci MED	BTA011988.1	No	9.00 E-14	157	52	96	57.69	69.23
B. tabaci MED	BTA013024.1	Yes	4.00 E-19	192	52	96	69.23	76.92
B. tabaci MED	BTA021616.1	No	1.00 E-18	200	52	96	67.31	75
B. tabaci MEAM1	Bta02428	No	2.00 E-10	135	51	94	50.98	58.82
B. tabaci MEAM1	Bta11938	No	8.00 E-18	191	52	96	67.31	75
B. tabaci MEAM1	Bta13246	Yes	4.00 E-19	192	52	96	69.23	76.92
B. tabaci MEAM1	Bta14520	No	2.00 E-14	163	52	96	57.69	69.23

C. lectularius	CLEC003032	No	1.00 E-15	171	52	96	63.46	69.23
C. lectularius	CLEC005903	Yes	7.00 E-19	190	52	96	69.23	76.92
C. lectularius	CLEC025047	No	4.00 E-05	94	50	93	40	56
D. citri	D_citri_rna10946	No	1.00 E-16	182	52	96	65.38	75
D. citri	D_citri_rna10947	No	6.00 E-17	185	52	96	65.38	75
D. citri	D_citri_rna13694	Yes	1.00 E-17	186	54	100	68.52	79.63
D. citri	D_citri_rna19818	No	9.00 E-18	193	54	100	68.52	79.63
D. citri	D_citri_rna19899	No	3.00 E-18	191	52	96	69.23	76.92
M. persicae Clone O	MYZPE13164_0_v1.0_00006455 0.1	No	1.00 E-09	128	49	91	55.1	65.31
M. persicae Clone O	MYZPE13164_0_v1.0_00007856 0.2	Yes	5.00 E-18	192	52	96	69.23	76.92
M. persicae Clone O	MYZPE13164_0_v1.0_00011091 0.1	No	6.00 E-16	174	52	96	61.54	73.08
M. persicae Clone G006	MYZPE13164_G006_v1.0_0001 12290.1	Yes	5.00 E-18	192	52	96	69.23	76.92
M. persicae Clone	MYZPE13164_G006_v1.0_0001 16760.1	No	2.00 E-09	128	49	91	55.1	65.31

G006								
M. persicae Clone G006	MYZPE13164_G006_v1.0_00125820.1	No	6.00 E-16	174	52	96	61.54	73.08
N. lugens	NLU005395.1	No	5.00 E-11	135	55	96	50.91	61.82
N. lugens	NLU005396.1	No	1.00 E-13	156	52	96	53.85	65.38
N. lugens	NLU005873.1	Yes	1.00 E-17	191	52	96	69.23	76.92
N. lugens	NLU018480.1	No	1.00 E-16	177	50	93	70	72
R. prolixus	RPRC008869-RA	Yes	4.00 E-20	198	52	96	73.08	80.77
S. furcifera	Sfur-199.19	No	6.00 E-14	165	52	96	59.62	67.31
S. furcifera	Sfur-695.1	No	4.00 E-11	139	50	93	52	64
S. furcifera	Sfur-8.45	Yes	2.00 E-16	184	52	96	61.54	76.92

Table 3. 4 The results of blastp analyses of all* D. melanogaster *proteins with DM against protein databases of 11 hemipteran genomes.

DM domain proteins of *D. melanogaster* and all the hemipterans (Table 3.4) were aligned using MUSCLE, and the section of the alignment of the DM domains were used to generate a maximum likelihood phylogenetic tree with a bootstrap of 1000 (Figure 3.3).

The phylogenetic tree (Figure 3.3) shows five distinct clades; a Hemiptera-only clade, DSX, DMRT99B, DMRT93B and DMRT11E. The majority of the RBBH DM-containing proteins of the hemipteran species (identified in Table 3.4 and marked with a purple dot in Figure 3.3) group in the DMRT99B clade; these may be true orthologues of DMRT99B. The high bootstrap number (0.818) supports the DMRT99B clade.

There are two RBBH hemipteran SDPs in the DSX clade; *D. citri* (D_citri_rna13694), and *R. prolixus* (RPRC008869-RA) however, these lack the DSX dimer so are probably not true orthologues of DSX.

The two DM and DSX dimer-containing proteins (Table 3.3) are present in the DSX clade; therefore, *N. lugens* (NLU018480.1) and *S. furcifera* (Sfur-199.19) are probably true orthologues of DSX. In the DMRT93B clade, there is one *S. furcifera* RBBH DM-containing protein (Sfur8.45), so this is probably a true DMRT93B orthologue. Figure 3.3 contains a Hemiptera-specific clade that is labelled 'unknown'. Interestingly, this included the *B. tabaci* protein BTA004042.1 that was DSX by Guo et al. (2018a). Based on analyses shown herein, there is no evidence that *B. tabaci* BTA004042.1 is a true homolog of DSX. Firstly, the protein was not RBBH (Table 3.4). Secondly, it does not fall within the *D. melanogaster* DSX clade. Finally, it does not have a DSX dimer domain. SDG expression and isoform analysis would determine whether the 'unknown' clade may have sex determination qualities (more information in Chapter 5).

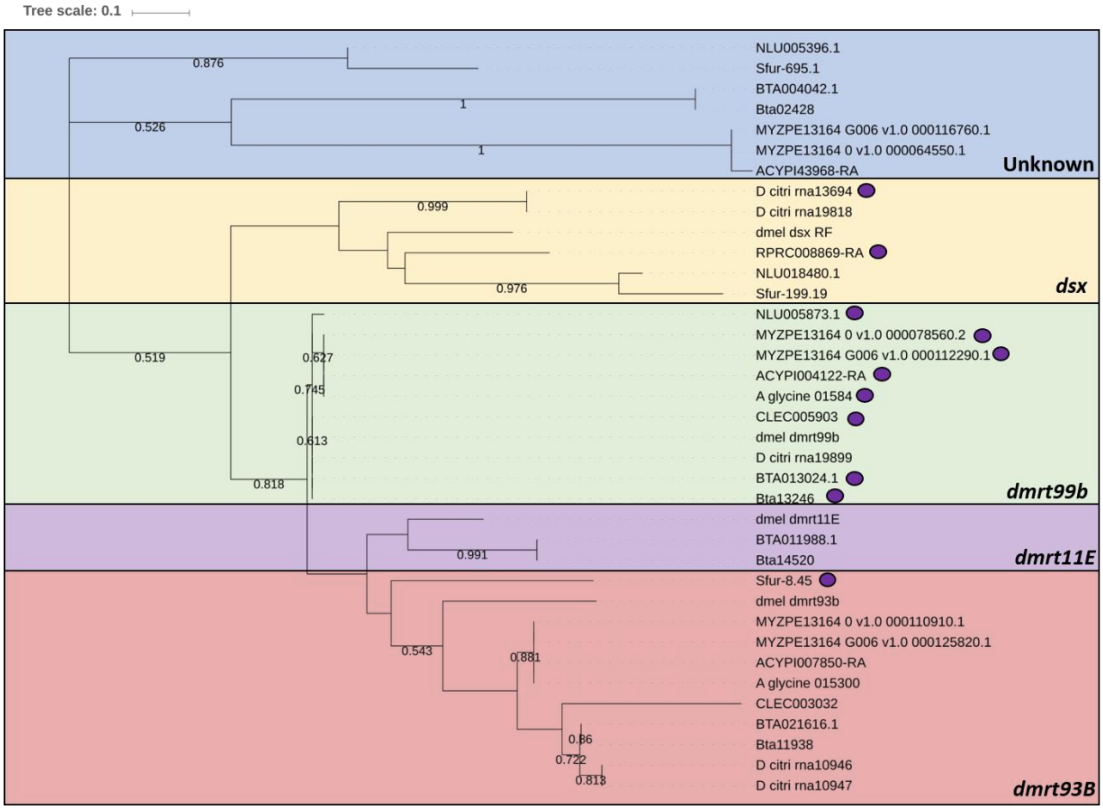

Figure 3. 3 Molecular Phylogenetic analysis by Maximum Likelihood method of the DM- containing proteins in* D. melanogaster *and Hemiptera

DMRT93B, DMRT99B and DMRT11E and hemipteran DM-containing proteins identified by RBBH in Table 3.4. The evolutionary history was inferred by using the Maximum Likelihood method based on the Whelan And Goldman model (Whelan and Goldman, 2001). The best model was predicted by MEGA7, from the aligned DM protein domains (Kumar et al., 2016; Felsenstein, 1985; Letunic and Bork, 2016). Evolutionary analyses were conducted in MEGA7. The purple dots are the RBBH from Table 3.2.1.1. *N. lugens* (NLU005873.1) and *C. lectularius* (CLEC025047) from the tree, as they had longer branch lengths than the other proteins and therefore dramatically affected the tree. *N. lugens* (NLU005873.1) and *C. lectularius* (CLEC025047) were removed from the tree, due to the branch lengths.

3.2.3 *B. mori* 'key' SDPs are found in Hemiptera

Hemiptera contain putative *B. mori* SDP orthologues; MASC, IMP and PSI but not FEM (Figure 3.1). Section 3.2.3 describes the more in-depth investigation of hemipteran SDPs knowledge gained from Section 3.2.1. Firstly, BLASTP analysis, and secondly by protein domain investigation.

3.2.3.1 Investigating Masc orthologues

Masc is a 'key gene' that is found in the *Bombyx mori* sex determination cascade. RBBH analysis (methodology; fig. 2.1) identified MASC-PHO in all 11 hemipteran genomes. Comparison of the BLASTP alignment statistics (Table 3.5) and whole-protein percentage identity (Table 3.6) shows a big difference; percentage identity is low.

Hemipteran Species	**Subject ID**	**E-value**	**Query length**	**Query Coverage**	**Identity (%)**	**Similarity (%)**
B. tabaci MEAM1	Bta04973	9.00E-11	70	12	38.57	60
B. tabaci MED	BTA009179.1	4.00E-10	70	12	38.57	60
D. citri	D_citri_rna2394	1.00E-08	71	12	35.21	53.52
A. glycines	A_glycine_07362	1.00E-08	67	11	34.33	55.22
A. pisum	ACYPI28312-RA	8.00E-08	80	14	33.75	50
M. Persicae G006	MYZPE13164_G006_v1.0_000000900.12	2.00E-08	81	14	28.4	51.85
M. persicae Clone 0	MYZPE13164_0_v1.0_000036270.2	1.00E-08	81	14	28.4	51.85
N. lugens	NLU010975.2	5.00E-06	86	14	31.4	53.49
S. furcifera	Sfur-473.13	2.00E-07	86	14	30.23	53.49
C. lectularius	CLEC008482	6.00E-09	74	13	35.14	55.41
R. prolixus	RPRC004468-RA	7.00E-09	81	14	35.8	54.32

Table 3. 5 Results of BLASTP analysis of the full-length* B. mori *MASC protein (Query) against all hemipteran MASC-PHO (Subject).
The subject ID corresponds with the proteome unique numbers.

Species	**Subject ID**	**Percentage identity of full protein against the query**
B. tabaci MEAM1	Bta04973	10.67
B. tabaci MED	BTA009179.1	8.57
D. citri	D_citri_rna2394	8.38
A. glycines	A_glycine_07362	10.50
A. pisum	ACYPI28312-RA	10.02
M. Persicae G006	MYZPE13164_G006_v1.0_000000900.12	10.50
M. persicae Clone 0	MYZPE13164_0_v1.0_000036270.2	10.50
N. lugens	NLU010975.2	10.51
S. furcifera	Sfur-473.13	11.59
C. lectularius	CLEC008482	9.70
R. prolixus	RPRC004468-RA	9.02

Table 3. 6 The percentage identity of the hemipteran putative Masc against the full-length* B. mori *Masc. **The pairwise comparison was conducted via CLC.**

B. mori MASC contains two zinc fingers (Figure 3.4). Zinc fingers are small stable protein motifs. The motifs can make tandem contact with the target molecule by the multiple finger-like protrusions that are present. Despite the name, the motifs can bind with other metals and non-metals. The non-metals can be used to form salt bridges to stabilise the finger-like folds. The Zinc finger that is present in the *B. mori* MASC has a CCCH domain. The CCCH-domain Zinc finger proteins are typically involved with cell cycle or growth phase regulation, regulatory proteins involved in regulating responses to growth factors (Letunic and Bork, 2018).

I analysed all of the putative hemipteran MASC orthologues (identified in Table 3.5) protein domains using the SMART protein analysis (Letunic and Bork, 2018). Figure 3.4 is a schematic of the protein's domains in *B. mori* MASC and putative hemipteran MASC orthologues (not to scale). 8/11 of the MASC-PHO have three zinc fingers, Whereas, 3/11 MASC-PHO hits have two zinc fingers (same as *B. mori* MASC).

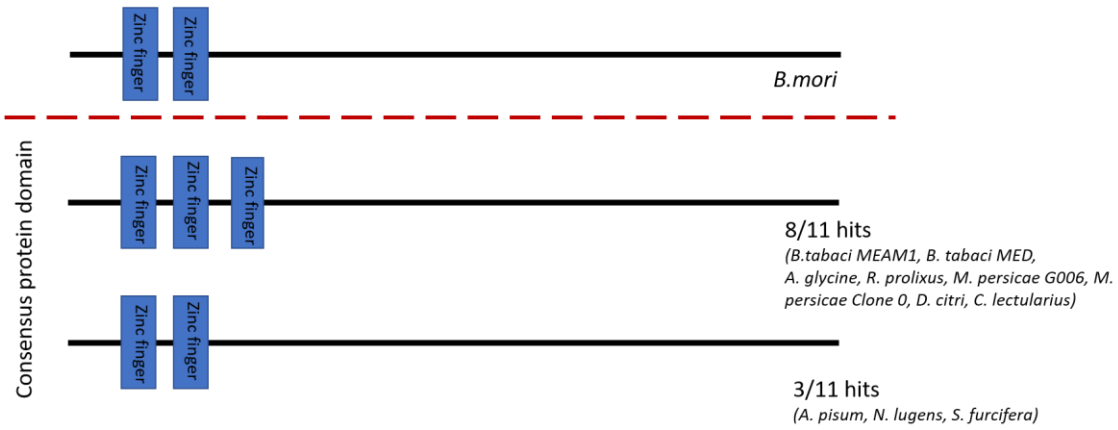

Figure 3. 4: A simplified schematic drawing of* B.mori *Masc protein structure.
All the RBBH proteins were analysed by SMART protein domain (Letunic and Bork, 2018). The 3/11 hits are *A. pisum*, *N. lugens* and *S. furcifera.*

The Zinc fingers may be conserved regions. Full-length MASC-PHO was aligned to *B. mori* MASC, using CLC (Section 2.14 and 2.1.6), and the Zinc finger domains identified (Figure 3.5b). The zinc finger motifs are highly conserved. All the aphids are very similar to each other within the first two zinc finger domains. The middle regions of Znf1 and Znf2 domains are more conserved than the outer regions. A maximum-likelihood phylogeny tree analysis was conducted to see the evolutionary relationship (Figure 3.5a). The branches range from being highly supported (bootstrap value of 0.97, between *S. furcifera* and *N. lugens*) to being very low supported (bootstrap value of 0.25 between the aphid clade and the whitefly).

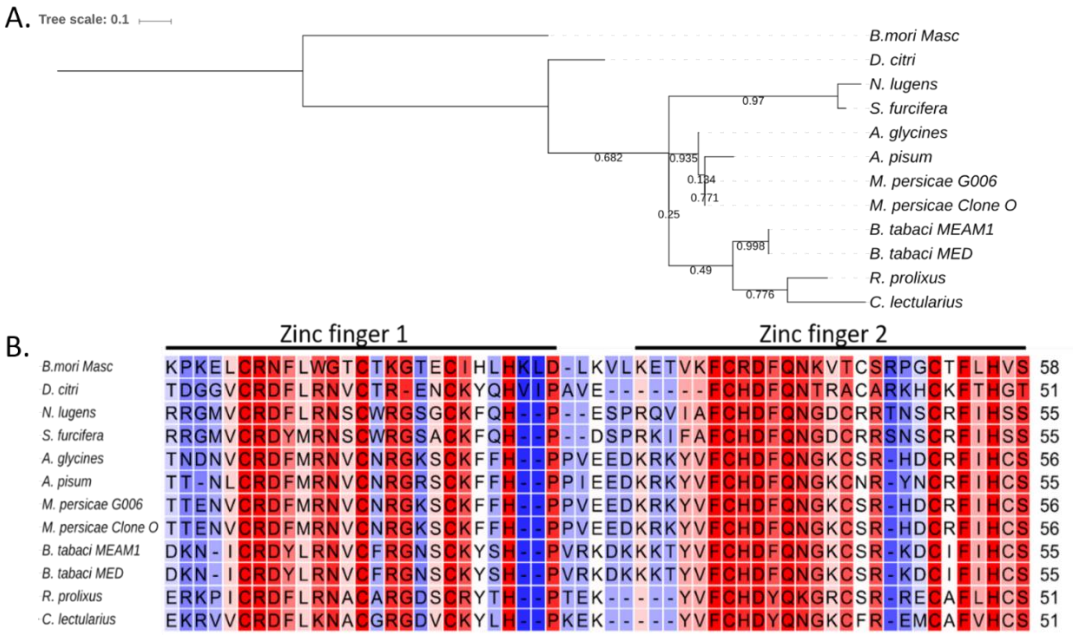

Figure 3. 5 Molecular Phylogenetic analysis of the Masc RBBH orthologues in Hemiptera by Maximum Likelihood method.

A.The phylogenetic tree was based on the Dayhoff matrix based model- which was decided by the MEGA7 best model analysis. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. Evolutionary analyses were conducted in MEGA7 with a bootstrap of 1000. It was edited in the Interactive tree of life. This tree is from the RBBH of the Masc protein- specifically at the Zinc finger domains (alignment) for the *B. mori* and the rest of the hemipteran species. The tree is rooted at *B. mori* MASC. B. Is the alignment of the Zinc fingers (Le and Gascuel, 2008; Kumar et al., 2016; Felsenstein, 1985; Letunic and Bork, 2016).

Masculinization of *B. mori* MASC does not require the zinc fingers, but requires the presence of Cys-301 and Cys-304 (Kiuchi et al., 2019). A protein alignment, with Ramsul colouring, was conducted to investigate whether Cys-301 and Cys-304 was present in Hemiptera (Figure 3.6). The *B. mori* Cys-301 and Cys-304 in Figure 3.6 is present and highlighted; however, no hemipteran MASC-PHO contains these residues.

The MASC-PHOs may not have the same functions as *B. mori* MASC; MASC-PHOs lack vital masculinizing residues, have a low similarity between the BLASTP alignment, and low identity between the full- length *B. mori* MASC and the full-length MASC-PHOs. All the reasons described indicate a low probability that the MASC-PHO proteins have the same function as *B. mori* MASC.

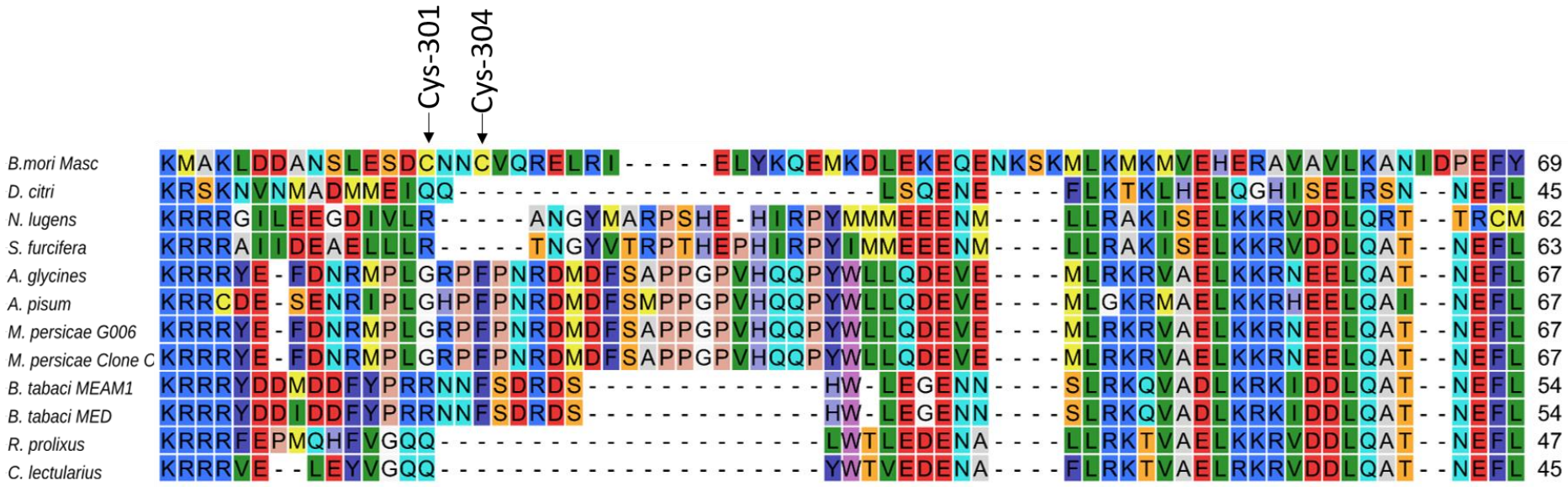

Figure 3. 6 Protein alignment of the* B. mori *Masc against the putative hemipteran MASC.

Highlighted on the top is the position of the Cys-301 and Cys-304. The amino acid properties are coloured by Ramsol colouring from CLC.

3.2.3.2 Investigating Imp orthologues

B. mori IMP is part of the 'key' group. RBBH analysis (methodology; fig. 2.1) identified IMP-PHO in all 11 hemipteran genomes. E-value are very low in BLASTP alignment statistics (Table 3.7), and the protein similarity is high which indicates conserved matches. Protein identity scores from the pairwise comparison analysis is high (Table 3.8).

Hemipteran species	**Subject ID**	**E-value**	**Query length**	**Query Coverage**	**Identity (%)**	**Similarity (%)**
B. tabaci MEAM1	Bta13278	0	500	95	67.2	77.4
B. tabaci MED	BTA017124.1	0	525	95	64	73.33
D. citri	D_citri_rna20833	3.00E-151	308	69	70.78	81.49
A. glycines	A_glycine_05034	0	516	94	61.82	74.22
A. pisum	ACYPI004277-RA	0	499	95	62.12	74.15
M. persicae G006	MYZPE13164_G006_v1.0_000030270.2	0	492	95	63.21	75.41
M. persicae Clone O	MYZPE13164_0_v1.0_000039230.2	0	492	95	63.21	75.41
N. lugens	NLU026974.1	0	365	74	78.08	85.48
S. furcifera	Sfur-409.13	0	395	76	76.96	84.3
C. lectularius	CLEC011792	0	496	95	65.32	76.41
R. prolixus	RPRC000571-RA	0	513	98	68.42	76.61

Table 3. 7 Results of BLASTP analysis of the full-length* B. mori *IMP protein (Query) against all hemipteran IMP-PHO (Subject).

Hemipteran Species	Subject ID	Percentage identity of full protein against the query
B. tabaci MEAM1	Bta13278	48.54
B. tabaci MED	BTA017124.1	51.06
D. citri	D_citri_rna20833	40.22
A. glycines	A_glycine_05034	49.92
A. pisum	ACYPI004277-RA	53.34
M. persicae G006	MYZPE13164_G006_v1.0_000030270.2	45.94
M. persicae Clone O	MYZPE13164_0_v1.0_000039230.2	45.94
N. lugens	NLU026974.1	48.52
S. furcifera	Sfur-409.13	20.35
C. lectularius	CLEC011792	55.13
R. prolixus	RPRC000571-RA	58.48

Table 3. 8 Percentage identity of the full-length IMP protein against the full-length putative hemipteran IMP orthologues.

B. mori IMP contains four protein domains (Figure 3.7); one RNA-recognition motif domain (RRM) and four KH-1 domains. RRM domains bind to single-stranded RNA, and proteins with RRM usually have alternative splicing. KH domains bind to RNA and has RNA recognition function. Proteins with KH-domains are typically nucleic acid-binding proteins (Letunic and Bork, 2018).

All the IMP-PHO (identified in Table 3.7) underwent SMART protein analysis for protein domain identification (Letunic and Bork, 2018). Figure 3.7 is a schematic of the protein's domains in *B. mori* IMP and IMP-PHO (not to scale). Four PHO do not have an RRM domain (*D. citri, C. lectularius, N. lugens* and *S. furcifera*). The other seven PHO have both the RRM domain and all four KH domains.

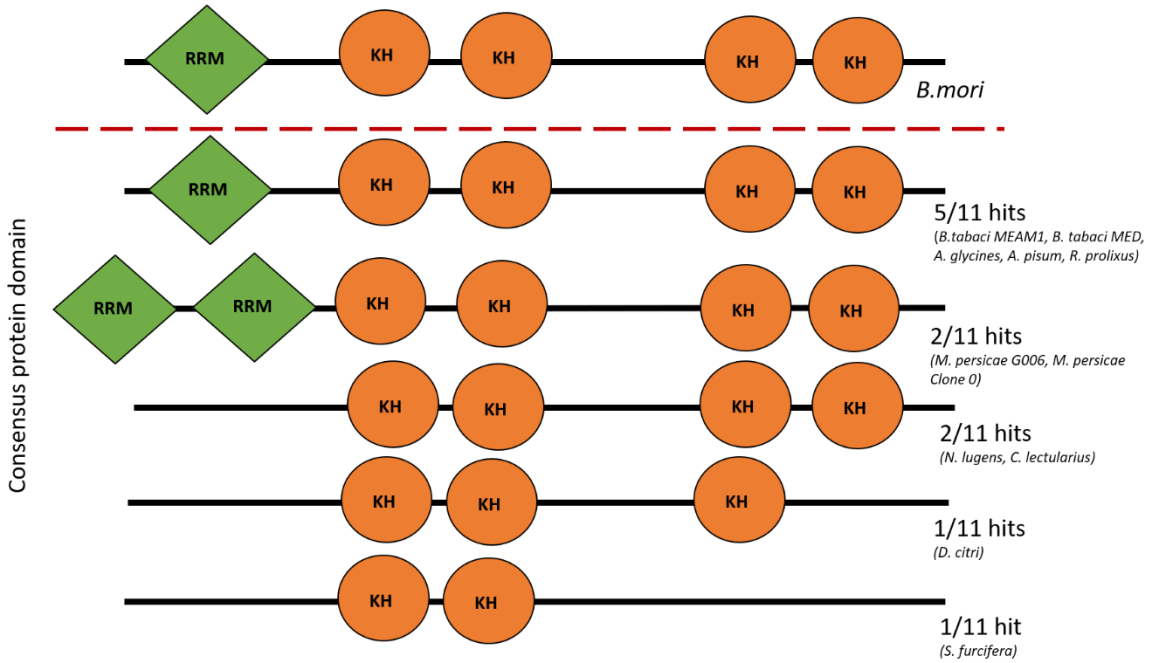

Figure 3. 7 A simplified schematic drawing of* B. mori *IMP protein structure and IMP-PHO structures.
The orange circles are the KH domains whilst the green diamonds represent the RRM domains.

There is a high percentage similarity from the BLASTP data analyses and most of the proteins found had at high similarity in protein domains as *B. mori* IMP. There is a high probability that Hemiptera have a functional IMP protein.

3.2.3.3 Investigating PSI orthologues

B. mori PSI is part of the 'key' group. RBBH analysis (methodology; fig. 2.1) identified PSI-PHO in all 11 hemipteran genomes. E-value are very low in BLASTP alignment statistics (Table 3.9), and the protein similarity is over 49%, which indicates possible conserved matches. Protein identity scores from the pairwise comparison analysis is lower than BLASTP protein similarity scores (Table 3.10).

Hemipteran species	**Subject ID**	**E-value**	**Query length**	**Query Coverage**	**Identity (%)**	**Similarity (%)**
B. tabaci MEAM1	Bta01264	8.00E-132	569	76	46.92	59.23
B. tabaci MED	BTA005137.1	1.00E-145	685	91	44.82	57.66
D. citri	D_citri_rna6836	1.00E-66	445	71	38.2	49.89
A. glycines	A_glycine_04622	6.00E-121	702	92	40.88	53.56
A. pisum	ACYPI006827-RA	3.00E-123	697	92	41.32	53.95
M. persicae G006	MYZPE13164_G006_v1.0_000020840.1	4.00E-124	699	92	41.2	53.93
M. persicae Clone O	MYZPE13164_0_v1.0_000093180.4	4.00E-124	699	92	41.2	53.93
N. lugens	NLU014116.1	4.00E-116	726	87	39.39	50.41
S. furcifera	Sfur-15.267	1.00E-125	752	91	41.49	51.46
C. lectularius	CLEC002109	1.00E-84	406	68	44.33	58.62
R. prolixus	RPRC006910-RA	2.00E-81	396	74	45.2	58.33

Table 3. 9 Results of BLASTP analysis of the full-length* B. mori *PSI protein (Query) against all hemipteran PSI-PHO (Subject).

Species	**Subject ID**	**Percentage identity of full protein against the query**
B. tabaci MEAM1	Bta01264	35.02
B. tabaci MED	BTA005137.1	39.65
D. citri	D_citri_rna6836	30.04
A. glycines	A_glycine_04622	34.19
A. pisum	ACYPI006827-RA	34.23
M. persicae G006	MYZPE13164_G006_v1.0_000020840.1	34.60
M. persicae Clone O	MYZPE13164_0_v1.0_000093180.4	34.23
N. lugens	NLU014116.1	28.78
S. furcifera	Sfur-15.267	36.49
C. lectularius	CLEC002109	32.82
R. prolixus	RPRC006910-RA	26.70

Table 3. 10: The percentage identity of the pairwise comparison of the PSI-PHO against the full-length* B. mori *PSI protein.

B. mori PSI contains four KH protein domains (Figure 3.8). A description of the KH function is in Section 3.2.3.2. All the PSI-PHO (identified in Table 3.9) underwent SMART protein analysis for protein domain identification (Letunic and Bork, 2018). Figure 3.8 is a schematic of the protein domains in *B. mori* PSI and PSI-PHO (not to scale). All PHO have at least three KH domains. *A. pisum, B. tabaci MED* and *MEAM1, M. persicae Clone 0* and *G006, A. glycines, R. prolixus, D. citri and N. lugens* all have domains that are the same as *B. mori* PSI.

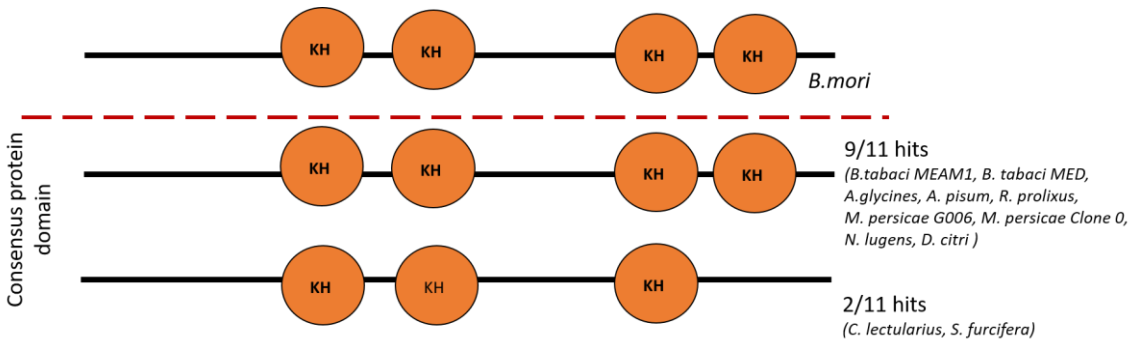

Figure 3. 8: A simplified schematic drawing of* B. mori *PSI and PSI-PHO protein structure.

There are orthologues of the MASC, IMP and PSI in Hemiptera. They are very similar in the protein domains so may indicate the functionality of these proteins. MASC determining masculizing factor is the Cys regions, when aligned the hemipteran RBBH do not contain this. So, unless the genes have a masculizing factor elsewhere, this is not like masc. We would need to do a functionality test to determine this. Overall, MASC PHO has a low probability of having the same function as *B. mori* MASC. Whereas IMP and PSI proteins have a high probability of the same function.

3.2.4 *A. mellifera* unique SDP are not conserved in Hemiptera

In Section 3.2.1, there are no unique *A. mellifera* SDP present in hemipterans, when using this pipeline. *A. mellifera* has a haplodiploid life cycle, like *B. tabaci*. Using the hypothesis suggested in the introduction, *B. tabaci* should have matching sex determination pathway as *A. mellifera*. The opposite is true; *B. tabaci* shares no similar 'key' SDPs and therefore the hypothesis, in this case, is false.

3.2.5 How conserved are the SDGs 'key' genes from the *D. melanogaster* sex determination cascade?

There are three 'key' SDP in *D. melanogaster* (SXL, TRA and TRA2). There are SXL and TRA2 in Hemiptera (Section 3.2.1). Despite using RBBH and BLASTP analysis, TRA is not detected in any Hemiptera studied.

3.2.5.1 Investigating Sxl orthologues

D. melanogaster SXL is part of the 'key' group. RBBH analysis (methodology; fig. 2.1) identified SXL-PHO in all 11 hemipteran genomes. E-values are very low in BLASTP alignment statistics (Table 3.11), and the protein similarity is over 55%, which indicates possible conserved matches. Protein identity scores from the pairwise comparison analysis is lower than BLASTP protein similarity scores (Table 3.12).

Hemipteran species	**Subject ID**	**E-value**	**Length**	**Query Coverage**	**Identity (%)**	**Similarity (%)**
B. tabaci MEAM1	Bta03294	2.00E-84	176	50	64.2	85.23
B. tabaci MED	BTA022642.1	1.00E-51	211	60	45.5	61.61
D. citri	D_citri_rna1282	9.00E-74	165	47	61.82	83.64
A. glycines	A_glycine_01063	2.00E-79	179	54	63.69	79.89
A. pisum	ACYPI000005-RA	7.00E-80	184	52	61.96	78.26
M. persicae G006	MYZPE13164_G006_v1.0_000089020.3	1.00E-79	184	52	62.5	78.26
M. persicae Clone O	MYZPE13164_0_v1.0_000097000.3	1.00E-79	184	52	62.5	78.26
N. lugens	NLU018908.1	4.00E-68	184	52	55.43	73.37
S. furcifera	Sfur-24.76	1.00E-47	185	49	42.7	63.78
C. lectularius	CLEC004462	1.00E-44	224	61	37.95	57.59
R. prolixus	RPRC001543-RA	6.00E-48	185	49	42.16	63.24

Table 3. 11: Results of BLASTP analysis of the full-length* D. melanogaster *SXL protein (Query) against all hemipteran SXL- PHO (Subject).

Hemipteran Species	Subject ID	Percentage identity of full protein against the query
B. tabaci MEAM1	Bta03294	33.62
B. tabaci MED	BTA022642.1	16.64
D. citri	D_citri_rna1282	27.96
A. glycines	A_glycine_01063	29.37
A. pisum	ACYPI000005-RA	34.31
M. persicae G006	MYZPE13164_G006_v1.0_000089020.3	29.20
M. persicae Clone O	MYZPE13164_0_v1.0_000097000.3	29.20
N. lugens	NLU018908.1	30.96
S. furcifera	Sfur-24.76	19.07
C. lectularius	CLEC004462	19.67
R. prolixus	RPRC001543-RA	21.74

Table 3. 12: The percentage identity of the pairwise comparison of the Hemipteran proteins against the full-length* D. melanogaster *SXL protein.

D. melanogaster SXL contains two RRM protein domains (Figure 3.9). A description of the RRM function is in Section 3.2.3.2. All the SXL-PHO (identified in Table 3.11) underwent SMART protein analysis for protein domain identification (Letunic and Bork, 2018). Figure 3.9 is a schematic of the protein's domains in *D. melanogaster* SXL and SXL-PHO (not to scale). Most hemipterans had two RRM protein domains. The exceptions were; *B. tabaci* MED with one RRM domain and *C. lectularius* and *S. furcifera* with three RRM domains.

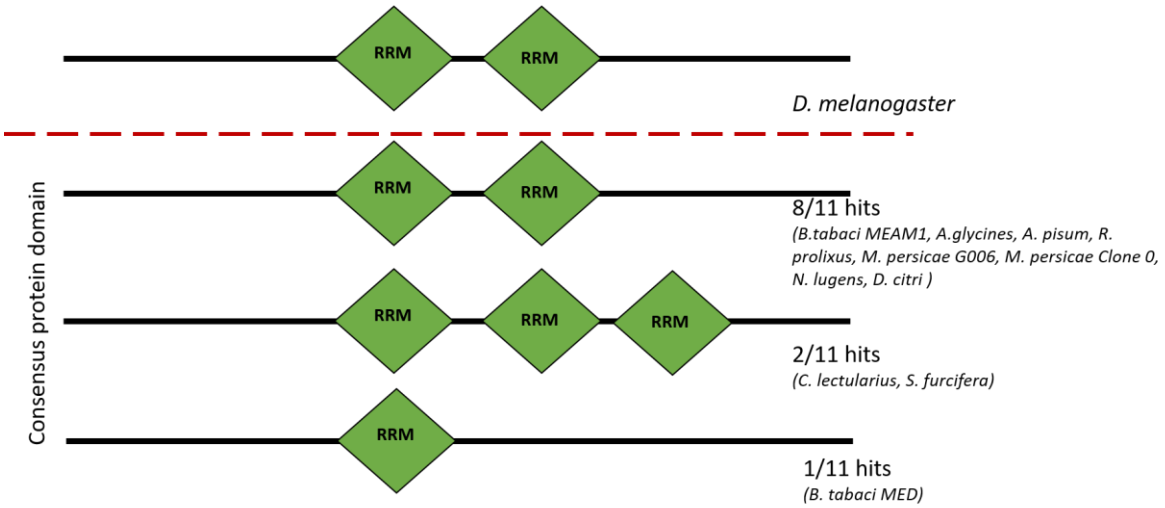

Figure 3. 9 A simplified schematic drawing of* D. melanogaster *SXL and SXL-PHO

RRM protein domains may be conserved regions. Full-length SXL-PHO was aligned to *D. melanogaster* SXL using CLC, and the two RRM domains identified (Figure 3.10b and Figure 3.10c). Both RRM domains are highly conserved. However, the first RRM domain (Figure 3.10b) is more conserved than the second RRM domain (Figure 3.10c). All the aphids are very similar to each other within the first RRM domains. The second RRM domain is more variable than the first, as there is a gap in the sequence at *C. lectularius, S. furcifera* and *R. prolixus.* A maximum-likelihood phylogeny tree analysis was conducted to evaluate the evolutionary relationship (Figure 3.10a). The branches range from being highly supported (bootstrap value of 1, between *S. furcifera* and *R. prolixus*) to being very low supported (bootstrap value of 0.08 in the aphid clade).

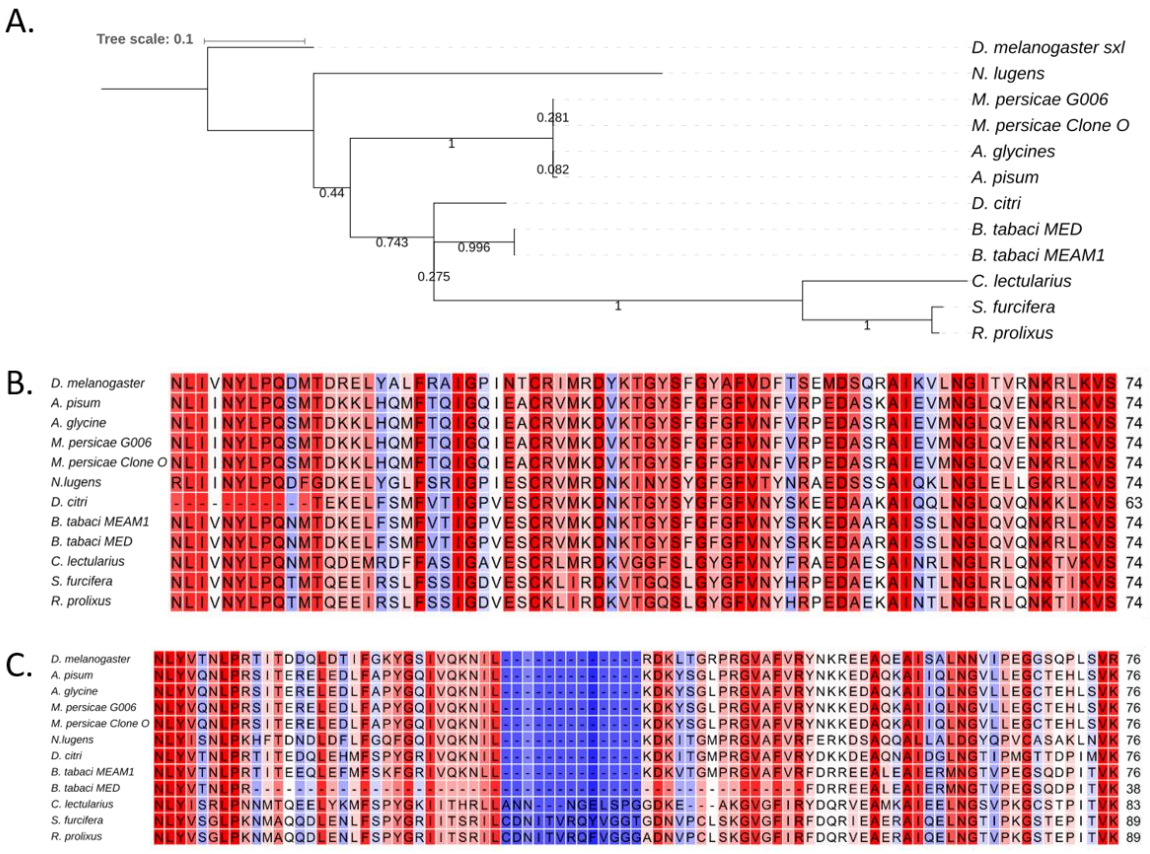

Figure 3. 10: Molecular Phylogenetic analysis of the SXL RBBH orthologue proteins at the RRM1 and RRM2 domain in Hemiptera by Maximum Likelihood method.

A.The phylogenetic tree was based on the Le_Gascuel_2008 model (Le and Gascuel, 2008)- which was decided by the MEGA7 best model analysis. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. Evolutionary analyses were conducted in MEGA7 with a bootstrap of 1000. It was edited in the Interactive tree of life. This tree is from the RBBH of the SXL protein-specifically at the RRM1 and RMM2 (alignment) for the *D. melanogaster* and the rest of the hemipteran species. The tree is rooted at *D. melanogaster* SXL. B. The top alignment is RRM1 and the C. bottom is RRM (Kumar et al., 2016; Letunic and Bork, 2016; Felsenstein, 1985).

SXL RRM protein domains may be more conserved than full-length proteins. A pairwise comparison was conducted on Figure 3.10b and 3.10c alignments. The RRM domains are more conserved than the rest of the protein (Table 3.13).

Hemipteran Species	**Subject ID**	**Percentage identity of full protein against the query**	**Percentage identity of RRM1 domain against the *D. melanogaster* RRM1 domain**	**Percentage identity of RRM2 domain against the *D. melanogaster* RRM2 domain**
B. tabaci MEAM1	Bta03294	33.62	62.16	63.16
B. tabaci MED	BTA022642.1	16.64	62.16	28.95
D. citri	D_citri_rna1282	27.96	44.59	67.11
A. glycines	A_glycine_01063	29.37	60.81	63.16
A. pisum	ACYPI000005-RA	34.31	59.46	63.16
M. persicae G006	MYZPE13164_G006_v1.0_000089020.3	29.20	60.81	63.16
M. persicae Clone O	MYZPE13164_0_v1.0_000097000.3	29.20	60.81	63.16
N. lugens	NLU018908.1	30.96	56.76	53.95
S. furcifera	Sfur-24.76	19.07	50	33.71
C. lectularius	CLEC004462	19.67	50	36.05
R. prolixus	RPRC001543-RA	21.74	50	32.58

Table 3. 13 percentage identity scores of the* D. melanogaster *SXL full-length protein against the SXL-PHO, and the RRM domains.

3.2.5.2 Investigating Tra2 orthologues

D. melanogaster TRA2 is part of the 'key' group. RBBH analysis (methodology; fig. 2.1) identified TRA2-PHO in all 11 hemipteran genomes. The BLASP analysis between *D. melanogaster* and TRA2-PHO in Table 3.14. E-value are very low in BLASTP alignment statistics (Table 3.14), and the protein similarity is over 50%, which indicates possible conserved matches. Protein identity scores from the pairwise comparison analysis is lower than BLASTP protein similarity scores (Table 3.15).

Hemipteran species	**Subject ID**	**E-value**	**Length**	**Query Coverage**	**Identity (%)**	**Similarity (%)**
B. tabaci MEAM1	Bta02209	1.00E-50	118	42	66.1	83.9
B. tabaci MED	BTA014916.3	1.00E-52	101	36	74.26	93.07
D. citri	D_citri_rna3604	8.00E-55	126	44	73.81	82.54
A. glycine	A_glycine_04284	1.00E-42	142	54	53.52	73.24
A. pisum	ACYPI007316-RA	8.00E-47	102	36	66.67	86.27
M. persicae G006	MYZPE13164_G006_v1.0_000099530.2	7.00E-11	95	33	34.74	55.79
M. persicae Clone O	MYZPE13164_0_v1.0_000075270.1	6.00E-11	95	33	34.74	55.79
N. lugens	NLU021361.1	2.00E-15	172	56	34.88	46.51
S. furcifera	Sfur-518.6	2.00E-57	152	49	65.13	75.66
C. lectularius	CLEC007595	7.00E-59	156	55	61.54	73.08
R. prolixus	RPRC007424-RA	7.00E-11	108	35	30.56	53.7

Table 3. 14 Results of BLASTP analysis of the full-length* D. melanogaster *TRA2 protein (Query) against all hemipteran TRA2-PHO (Subject).

Hemipteran species	Subject ID	Percentage identity of full protein against the query
B. tabaci MEAM1	Bta02209	36.39
B. tabaci MED	BTA014916.3	31.79
D. citri	D_citri_rna3604	26.11
A. glycine	A_glycine_04284	10.79
A. pisum	ACYPI007316-RA	33.55
M. persicae G006	MYZPE13164_G006_v1.0_000099530.2	5.91
M. persicae Clone O	MYZPE13164_0_v1.0_000075270.1	5.91
N. lugens	NLU021361.1	12.53
S. furcifera	Sfur-518.6	10.60
C. lectularius	CLEC007595	29.05
R. prolixus	RPRC007424-RA	29.75

Table 3. 15 Percentage identity of the pairwise comparison of the full-length proteins of* D. melanogaster *TRA2 and TRA2-PHO.

D. melanogaster TRA2 contains one RRM protein domain (Figure 3.11). A description of the RRM domain function is in Section 3.2.3.2. All the TRA2-PHO (identified in Table 3.14) underwent SMART protein analysis for protein domain identification (Letunic and Bork, 2018). Figure 3.11 is a schematic of the protein's domains in *D. melanogaster* TRA2 and SXL-TRA2 (not to scale). All 11 hemipterans have one RRM protein domain.

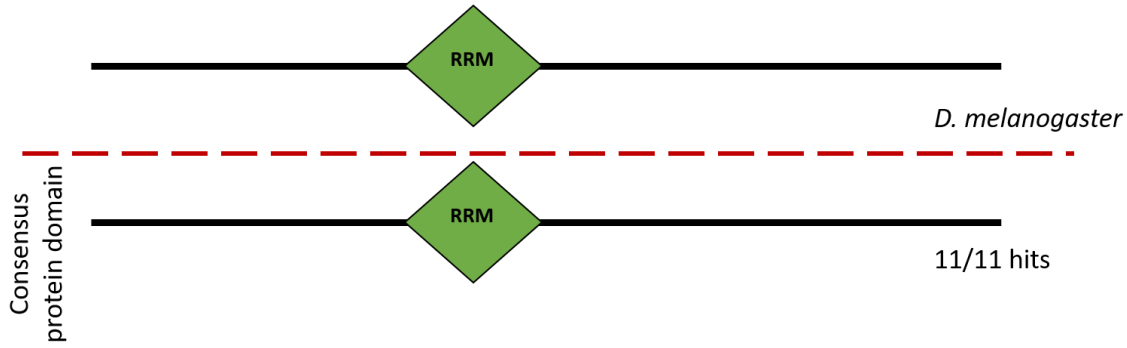

Figure 3. 11: A simplified schematic drawing of* D. melanogaster *TRA2 and TRA2-PHO protein structure.

RRM protein domain may be conserved regions. Full-length TRA2-PHOs were aligned to *D. melanogaster* TRA2 using CLC, and the RRM domain identified (Figure 3.12b). The RRM domain is

highly conserved. A maximum-likelihood phylogeny tree analysis was conducted to investigate the evolutionary relationship (Figure 3.12a). The branches range from being highly supported (bootstrap value of 1, between *M. persicae G006* and *Clone 0*) to being very low supported (bootstrap value of 0.08 between *B. tabaci* MED and *S. furcifera*).

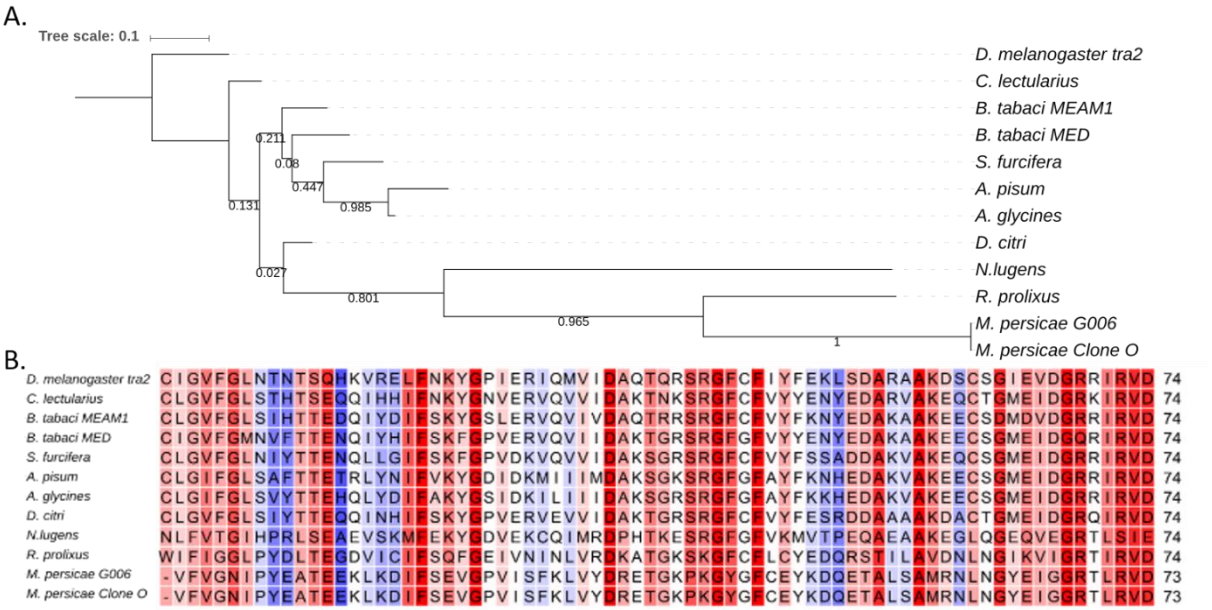

Figure 3. 12 Molecular Phylogenetic analysis RRM domain in* D. melanogaster *TRA2 and TRA2-PHOs, aligned by the RRM domain by Maximum Likelihood method

A. Evolutionary analyses were conducted in MEGA7 with a bootstrap of 1000. The tree was based on the Maximum Likelihood method based on the Le_Gascuel_2008 model (which was determined as the best model by MEGA7). The analysis involved 12 amino acid sequences. Edited in ITOL. Rooted at *D. melanogaster* TRA2. B. The alignment is of the RRM domain for all the hemipteran (Le and Gascuel, 2008; Felsenstein, 1985; Letunic and Bork, 2016; Kumar et al., 2016).

3.2.6 Sex determination support genes are found in Hemiptera

Support genes are from the *D. melanogaster* sex determination pathway. The support genes include; *Da, Dpn, Dsf, Emc, Fl(2)d, Fru, Gro, Her, Snf* and *Vir* (more information located in Section 1.3.6). *Da, Dsf, Emc, Fl2d, Fru, Gro, Her, Snf* are all part of the regulation of SXL. *Dpn* and *Vir* are involved in courtship behaviours. SXL is in all Hemipteran (Dection 3.2.5).

3.2.6.1 Investigating Da orthologues

Daughterless (Da) is a 'support' gene; found in *D. melanogaster* sex determination cascade, (more information in Section 1.3.6). *Da* is a positive regulator of Sxl. RBBH analysis (methodology; fig. 2.1) identified DA-PHO in 10 hemipteran species; missing is *R. prolixus* DA. E-value are very low in BLASTP alignment statistics (Table 3.13), and the protein similarity is over 50% which indicates possible conserved matches. Protein identity scores from the pairwise comparison analysis is lower than BLASTP protein similarity scores (Table 3.14).

Hemipteran species	**Subject Seq-id**	**e-value**	**Length**	**Query Coverage**	**Identity (%)**	**Similarity (%)**
B. tabaci MEAM1	Bta11931	4E-63	309	40	48.87	58.25
B. tabaci MED	BTA018608.1	4E-63	309	40	48.87	58.25
D. citri	D_citri_rna13465	1E-64	296	41	50.68	61.82
A. glycines	A_glycine_011973	4E-64	375	63	47.47	55.2
A. pisum	ACYPI003796-RA	3E-62	401	63	45.64	53.12
M. persicae G006	MYZPE13164_G006_v1.0_000018360.1	2E-65	375	63	48.53	56.8
M. persicae Clone O	MYZPE13164_0_v1.0_000036540.1	2E-64	299	53	50.5	57.53
N. lugens	NLU002710.1	1E-62	273	38	50.92	63
S. furcifera	Sfur-24.227	6E-61	271	38	51.29	61.62
C. lectularius	CLEC011986	2E-45	329	45	46.81	54.71

Figure 3. 13: Results of BLASTP analysis of the full-length* D. melanogaster *DA protein (Query) against all hemipteran DA-PHO (Subject).

Hemipteran Species	Subject Seq-id	Percentage identity of full protein against the query
B. tabaci MEAM1	Bta11931	28.72
B. tabaci MED	BTA018608.1	28.68
D. citri	D_citri_rna13465	25.38
A. glycines	A_glycine_011973	32.57
A. pisum	ACYPI003796-RA	31.50
M. persicae G006	MYZPE13164_G006_v1.0_000018360.1	32.57
M. persicae Clone O	MYZPE13164_0_v1.0_000036540.1	31.63
N. lugens	NLU002710.1	20.0
S. furcifera	Sfur-24.227	27.18
C. lectularius	CLEC011986	24.60

***Figure 3. 14: Percentage identity of the pairwise comparison of the DA-PHO and* D. melanogaster DA.**

D. melanogaster DA does not have any specific protein domains but a low complexity region. The other hemipteran DA orthologues show similar lack of protein domains.

3.2.6.2 Investigating Dpn orthologues

Deadpan (*Dpn*) is a 'support gene; found in *D. melanogaster* sex determination cascade (more information in Section 1.3.6). DPN is a negative regulator of SXL. RBBH analysis (methodology; fig. 2.1) identified DPN-PHO in all 11 hemipteran genomes. E-value are low in the BLASTP analysis (Table 3.16), and the protein similarity score is high; ranging from 58.39% (*S. furcifera*) to 81.48% (*B. tabaci* MEAM1). Protein identity scores from the pairwise comparison analysis is lower than BLASTP protein similarity scores (Table 3.17).

Hemipteran Species	Subject Seq-id	E-value	Length	Query Coverage	Identity (%)	Similarity (%)
B. tabaci MED	BTA027689.1	5E-52	135	31	64.44	81.48
B. tabaci MEAM1	Bta06040	2E-72	225	51	55.11	71.11
D. citri	D_citri_rna243	4E-61	183	42	60.11	75.96
A. glycines	A_glycine_05420	2E-48	207	46	44.44	61.84
A. pisum	ACYPI004499-RA	3E-45	195	43	45.64	61.54
M. persicae G006	MYZPE13164_G006_v1.0_000127320.1	1E-45	176	36	51.14	65.91
M. persicae Clone O	MYZPE13164_0_v1.0_000192730.1	1E-45	176	36	51.14	65.91
N. lugens	NLU017783.1	5E-40	205	42	43.9	59.51
S. furcifera	Sfur-159.30	2E-24	149	30	38.93	58.39
C. lectularius	CLEC009505	4E-71	200	46	57.5	73.5
R. prolixus	RPRC000496-RA	1E-43	118	27	61.86	79.66

Table 3. 16: Results of BLASTP analysis of the full-length* D. melanogaster *DPN protein (Query) against all hemipteran DPN-PHO (Subject).

Hemipteran Species	Subject Seq-id	Percentage identity of full protein against the query
B. tabaci MED	BTA027689.1	32.37
B. tabaci MEAM1	Bta06040	22.17
D. citri	D_citri_rna243	21.74
A. glycines	A_glycine_05420	20.67
A. pisum	ACYPI004499-RA	20.83
M. persicae G006	MYZPE13164_G006_v1.0_000127320.1	24.49
M. persicae Clone O	MYZPE13164_0_v1.0_000192730.1	24.49
N. lugens	NLU017783.1	11.74
S. furcifera	Sfur-159.30	13.55
C. lectularius	CLEC009505	29.61
R. prolixus	RPRC000496-RA	19.87

Table 3. 17: Percentage identity of the pairwise comparison of the* D. melanogaster *DPN orthologue against the full-length DPN-PHO.

D. melanogaster DPN contains one HLH and one Orange protein domain (Figure 3.15). More information about HLH domain is available in Section 1.3.6. All DPN-PHO (identified in Table 3.16) underwent SMART protein analysis for protein domain identification (Letunic and Bork, 2018). Figure 3.15 is a schematic of the protein's domains in *D. melanogaster* DPN and DPN-PHO (not to scale). All 11 hemipterans have one HLH protein domain. However, only 8/11 PHO contain orange domain; *A. pisum, B. tabaci MED* and *R. prolixus* do not have the orange domain.

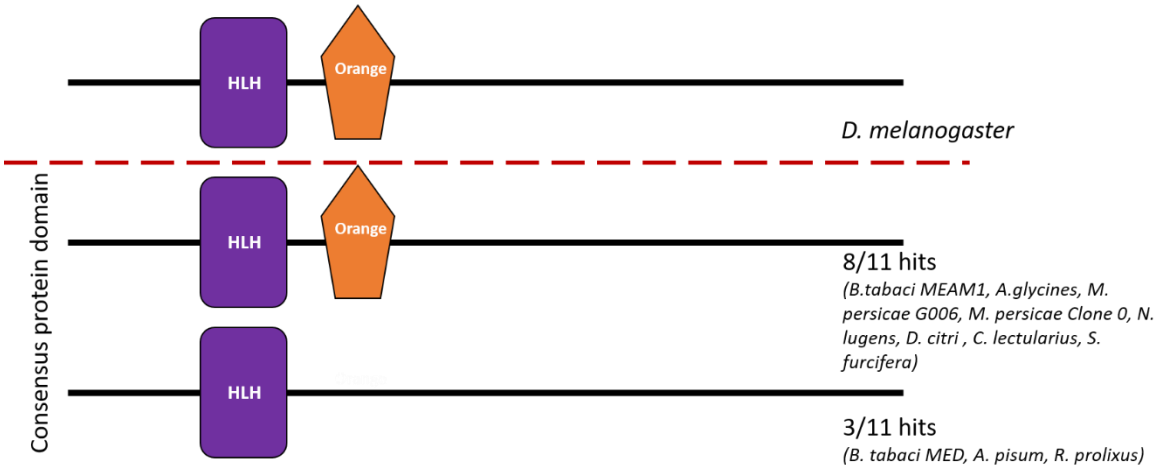

Figure 3. 15: A simplified schematic drawing of* D. melanogaster *DPN structure.

HLH protein domain may be conserved region. Full-length DPN-PHOs were aligned to *D. melanogaster* DPN using CLC, and the HLH domain identified (Figure 3.16b). The HLH domain is highly conserved. A maximum-likelihood phylogeny tree analysis was conducted to see the evolutionary relationship (Figure 3.16a). The branches range from being highly supported (bootstrap value of 1 between *A. glycine* and *A. pisum*).

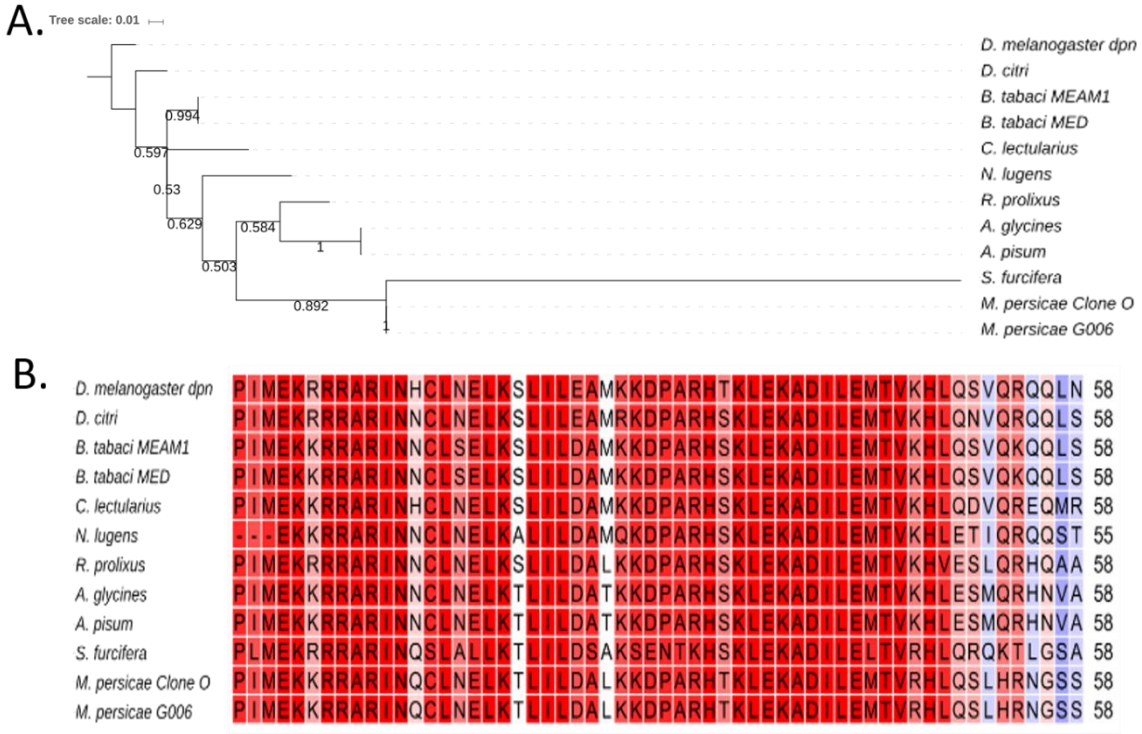

Figure 3. 16 Molecular Phylogenetic analysis of hemipteran orthologues of DPN aligned at the HLH domain.

A. Evolutionary analyses were conducted in MEGA7 with a bootstrap of 1000. The tree was based on the Maximum Likelihood method based on the Le_Gascuel_2008 model (which was determined as the best model by MEGA7). The analysis involved 12 amino acid sequences. Edited in ITOL. It is rooted at *D. melanogaster* DPN. B. The alignment is of the HLH for all the Hemiptera. Red indicates a conserved region (Le and Gascuel, 2008; Kumar et al., 2016; Letunic and Bork, 2016; Felsenstein, 1985).

Orange protein domain may be a conserved region. Full-length DPN-PHOs were aligned to *D. melanogaster* DPN using CLC, and the Orange domain identified (Figure 3.17b). The Orange domain is slightly conserved. A maximum-likelihood phylogeny tree analysis was conducted to see the evolutionary relationship (Figure 3.17a). The branches range from being highly supported (bootstrap value of 0.996, between *A. glycine* and *A. pisum*) to lowly supported (any branch lengths without a value is under 0.5).

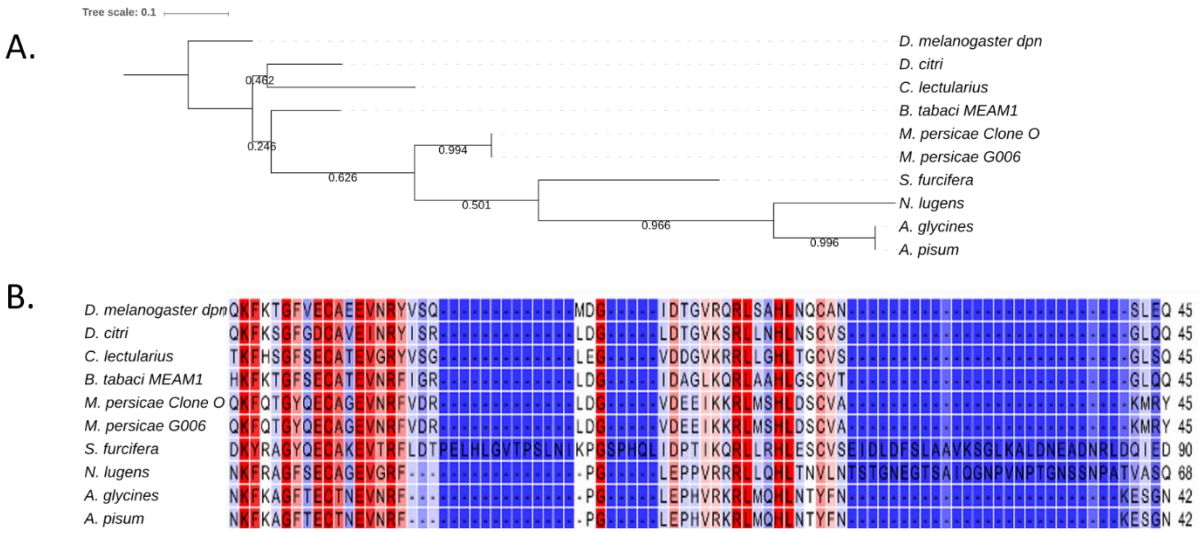

Figure 3. 17 Molecular Phylogenetic analysis of hemipteran orthologues of dpn aligned at the orange domain.

A. Evolutionary analyses were conducted in MEGA7 with a bootstrap of 1000. The tree was based on the Maximum Likelihood method based on the General Reverse Transcriptase model (which was determined as the best model by MEGA7). The analysis involved 10 amino acid sequences. Edited in ITOL. It is rooted at *D. melanogaster* DPN. B. The alignment is of the orange domain, from Figure 3.15, for all the Hemiptera. Red indicates conserved region (Le and Gascuel, 2008; Kumar et al., 2016; Letunic and Bork, 2016; Felsenstein, 1985; Dimmic et al., 2002).

DPN orange and HLH protein domains may be more conserved than full-length proteins. A pairwise comparison with Figure 3.16b and 3.17b alignments is shown in Table 3.18. The protein domains are more conserved than the rest of the protein (Table 3.18). However, HLH domain is more conserved than the orange domain.

Hemipteran species	Subject Seq-id	Percentage identity of full protein against the query	Percentage identity of HLH domain against the *D. melanogaster* HLH domain	Percentage identity of Orange domain against the *D. melanogaster* Orange domain
B. tabaci MED	BTA027689.1	32.37	87.93	54.35
B. tabaci MEAM1	Bta06040	22.17	87.93	40.91
D. citri	D_citri_rna243	21.74	91.38	52.17
A. glycines	A_glycine_05420	20.67	79.31	45.65
A. pisum	ACYPI004499-RA	20.83	79.31	45.65
M. persicae G006	MYZPE13164_G006_v1.0_000127320.1	24.49	79.31	43.48
M. persicae Clone O	MYZPE13164_0_v1.0_000192730.1	24.49	79.31	43.48
N. lugens	NLU017783.1	11.74	75.86	25.00
S. furcifera	Sfur-159.30	13.55	56.90	18.89
C. lectularius	CLEC009505	29.61	87.93	47.83
R. prolixus	RPRC000496-RA	19.87	82.76	23.61

Table 3. 18: percentage identity scores of the* D. melanogaster *DPN full-length protein against the DPN-PHO between the HLH and Orange protein domain.

The HLH containing DPN depends on the 'WRPW' (Trp-Arg-Pro-Trp) motif at the C-terminal for activity (Wainwright and Ishhorowicz, 1992). Figure 3.18 is of the DPN-PHO alignment at the WRPW region. All DPN-PHO have the WRPW motif, except *B.tabaci MED*. Presence of WRPW indicates a high probability of the DPN-PHO having the same function as *D. melanogaster* DPN.

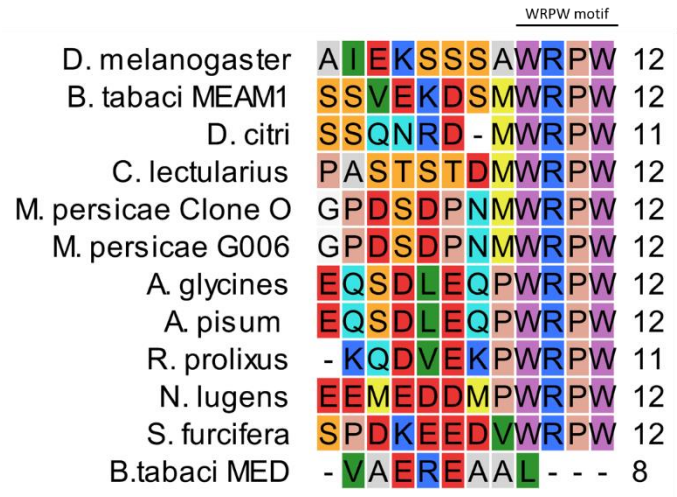

Figure 3. 18:* D. melanogaster *DPN protein aligned at the Hemipteran hits at the WRPW motif at the C-terminus.

3.2.6.3 Investigating *Emc* orthologues

Extra-macrochaetae (*Emc*) is a 'support gene; found in *D. melanogaster* sex determination cascade (more information in Section 1.3.6). EMC is a negative regulator of SXL. RBBH analysis (methodology; fig. 2.1) identified EMC-PHO in all 11 hemipteran genomes. E-value are low in the BLASTP analysis (Table 3.19), and similarity scores range from 70.1% (*B. tabaci* MEAM1) to 82.14% (*S. furcifera*). Protein identity scores from the pairwise comparison analysis is lower than BLASTP protein similarity scores (Table 3.20).

Hemipteran species	**Subject Seq-id**	**e-value**	**Length**	**Query Coverage**	**Identitiy (%)**	**Similarity (%)**
B. tabaci MED	BTA026084.1	4E-23	97	47	53.61	70.1
B. tabaci MEAM1	Bta13293	5E-24	101	49	52.48	70.3
D. citri	D_citri_rna16519	2E-22	76	37	57.89	75
A. glycines	A_glycine_05577	9E-18	61	31	55.74	78.69
A. pisum	ACYPI002529-RA	1E-18	63	32	57.14	79.37
M. persicae G006	MYZPE13164_G006_v1.0_000153470.1	9E-19	67	34	56.72	79.1
M. persicae Clone O	MYZPE13164_0_v1.0_000158600.1	9E-19	67	34	56.72	79.1
N. lugens	NLU011228.1	1E-27	84	41	63.1	80.95
S. furcifera	Sfur-74.201	1E-25	84	41	63.1	82.14
C. lectularius	CLEC007449	6E-20	62	31	64.52	80.65
R. prolixus	RPRC005973-RA	8E-22	62	31	66.13	80.65

Table 3. 19: Results of BLASTP analysis of the full-length* D. melanogaster *EMC protein (Query) against all hemipteran EMC-PHO (Subject).

Species	Subject Seq-id	Percentage identity of full protein against the query
B. tabaci MED	BTA026084.1	27.49
B. tabaci MEAM1	Bta13293	25.62
D. citri	D_citri_rna16519	26.17
A. glycines	A_glycine_05577	21.30
A. pisum	ACYPI002529-RA	22.13
M. persicae G006	MYZPE13164_G006_v1.0_000153470.1	22.03
M. persicae Clone O	MYZPE13164_0_v1.0_000158600.1	22.03
N. lugens	NLU011228.1	28.00
S. furcifera	Sfur-74.201	10.65
C. lectularius	CLEC007449	21.91
R. prolixus	RPRC005973-RA	24.09

Table 3. 20: Percentage identity of the pairwise comparison of the D. melanogaster EMC orthologue against the full-length EMC-PHO.

D. melanogaster EMC contains one HLH protein domain (Figure 3.19). More information about HLH domain is available in Section 1.3.6. All EMC-PHO (identified in Table 3.19) underwent SMART protein analysis for protein domain identification (Letunic and Bork, 2018). Figure 3.19 is a schematic of the protein's domains in *D. melanogaster* EMC and EMC-PHO (not to scale). All 11 hemipteran orthologues have one HLH protein domain.

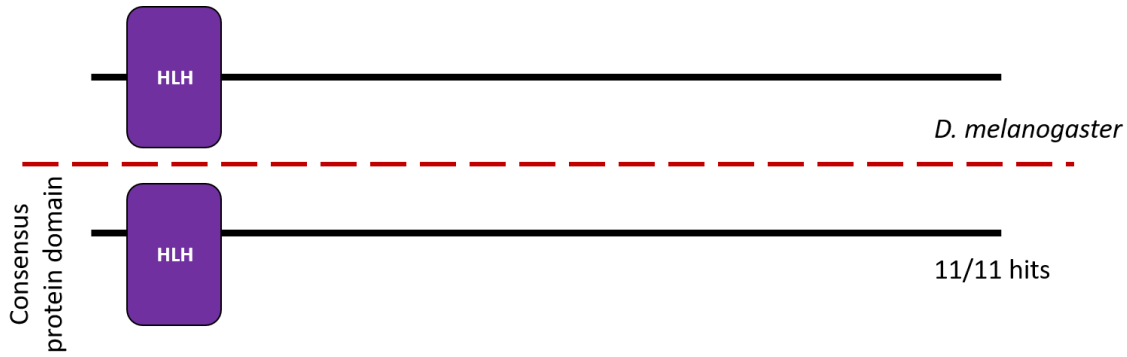

Figure 3. 19: A simplified schematic drawing of* D. melanogaster *EMC protein structure.

EMC HLH protein domains may be more conserved than full-length proteins. Full-length EMC-PHOs were aligned to *D. melanogaster* EMC using CLC, and the HLH domain identified (Figure 3.19b). The HLH domain is conserved. A maximum-likelihood phylogeny tree analysis was conducted to explore the evolutionary relationship (Figure 3.19a). The branches range from being highly supported (bootstrap value of 0.996, between *B. tabaci* MED and MEAM1).

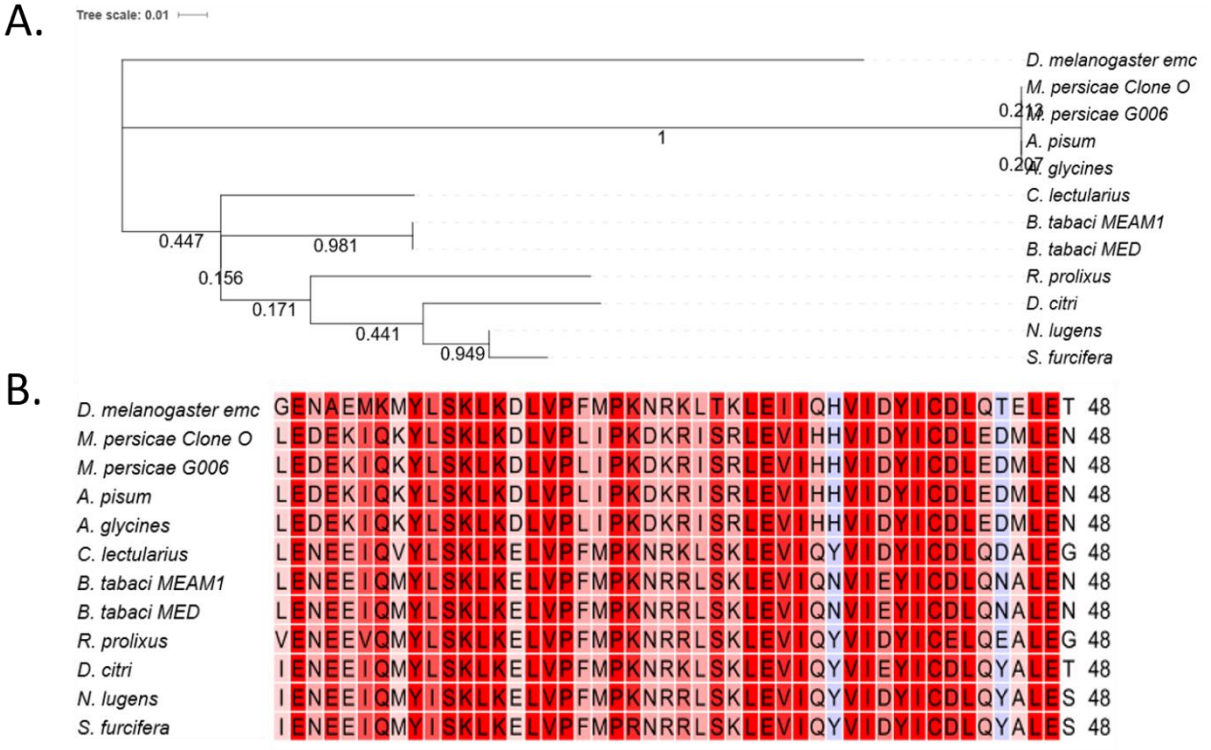

Figure 3. 20: Molecular Phylogenetic analysis of hemipteran orthologues of EMC aligned at the HLH domain.

A. Evolutionary analyses were conducted in MEGA7 with a bootstrap of 1000. The tree was based on the Maximum Likelihood method based on the Le_Gascuel_2008 model (which was determined as the best model by MEGA7). The analysis involved 12 amino acid sequences. Edited in ITOL. It is rooted at *D. melanogaster* EMC. B. The alignment is of the HLH for all the Hemiptera on the EMC protein. Red indicates conserved region (Le and Gascuel, 2008; Kumar et al., 2016; Letunic and Bork, 2016; Felsenstein, 1985).

Both EMC and DPN have HLH domains. HLH protein domains lack a basic domain repress transcription, repressing the DNA binding. *D. melanogaster* EMC lacks the basic domain. CLC aligned the EMC and DPN PHO (Figure 3.21). There is a difference between the DPN and EMC HLH protein domain. Specifically, the 'RRAR' and 'PARSH' motif is missing in EMC, which may indicate a difference in functionality.

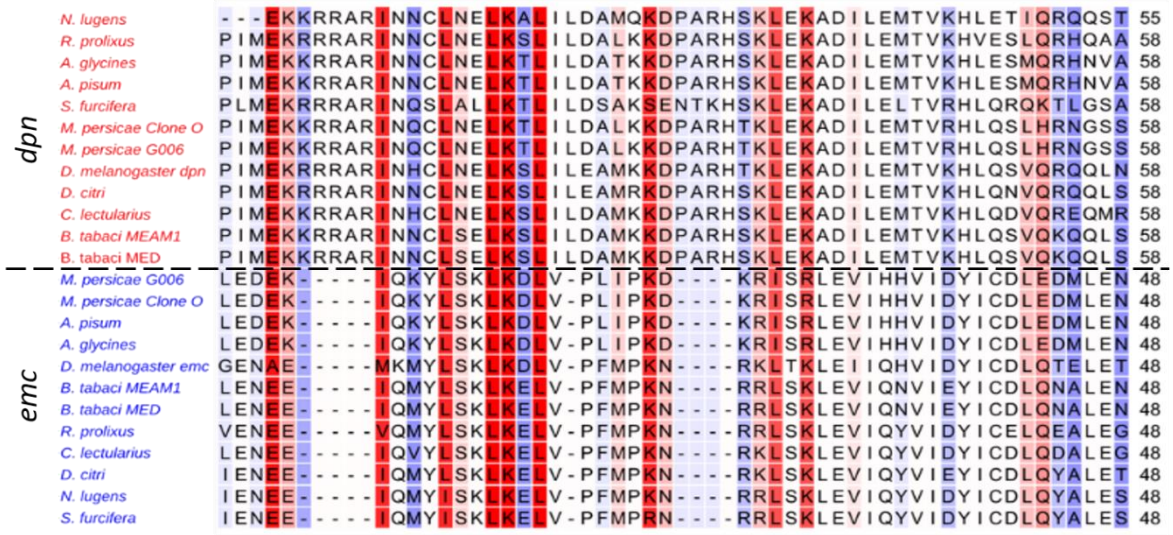

Figure 3. 21 Alignment of the EMC and DPN HLH domain from the full protein hemipteran orthologues.
This alignment is from the RBBH of the DPN (coloured red) and EMC (coloured blue) HLH protein domain.

3.2.6.4 Investigating fl(2)d orthologues

Female lethal d (Fl(2)d) is a 'support' gene; found in *D. melanogaster* sex determination cascade, (more information in Section 1.3.6). SXL requires FL(2)D for autoregulation. RBBH analysis (methodology; fig. 2.1) identified FL(2)D-PHO in 9 hemipteran species. E-value are low in the BLASTP analysis (Table 3.21), and similarity scores range from 73.25% (*B. tabaci* MEAM1 and MED) to 83.93% (*A. glycines*). Protein identity scores from the pairwise comparison analysis is lower than BLASTP protein similarity scores (Table 3.22).

Hemipteran species	Subject Seq-id	E-value	Length	Query Coverage	Identity (%)	Similarity (%)
B. tabaci MED	BTA029394.1	2.00E-80	228	43	60.53	73.25
B. tabaci MEAM1	Bta14734	7.00E-80	228	43	60.53	73.25
A. glycines	A_glycine_06187	5.00E-68	168	31	67.86	83.93
A. pisum	ACYPI005891-RA	4.00E-67	169	31	65.09	83.43
M. persicae G006	MYZPE13164_G006_v1.0_000135450.2	9.00E-67	167	31	65.87	83.23
M. persicae Clone O	MYZPE13164_0_v1.0_000103830.4	9.00E-67	167	31	65.87	83.23
N. lugens	NLU025797.1	3.00E-85	197	37	67.51	82.74
S. furcifera	Sfur-504.7	4.00E-75	203	38	67	81.28
C. lectularius	CLEC004180	1.00E-69	226	42	57.52	73.89

Table 3. 21: Results of BLASTP analysis of the full-length* D. melanogaster *FL2)D protein (Query) against all hemipteran FL(2)D-PHO (Subject).

Species	Subject Seq-id	Percentage identity of full protein against the query
B. tabaci MEAD	BTA029394.1	29.37
B. tabaci MEAM1	Bta14734	28.62
A. glycines	A_glycine_06187	24.82
A. pisum	ACYPI005891-RA	24.82
M. persicae G006	MYZPE13164_G006_v1.0_000135450.2	25.55
M. persicae Clone O	MYZPE13164_0_v1.0_000103830.4	25.37
N. lugens	NLU025797.1	27.37
S. furcifera	Sfur-504.7	16.78
C. lectularius	CLEC004180	27.21

Table 3. 22: Percentage identity of the pairwise comparison of the FL(2)D PHO against the full-length* D. melanogaster FL*(2)D protein

D. melanogaster FL(2)D and FL(2)D-PHO contains no protein domains, so no protein domain consensus diagram. Despite a lack of protein domains, there are two features of interest in the *D. melanogaster* FL(2)D; a histidine-rich area and a glutamine-rich area. The histidine rich area is present at residues 56-69, and the glutamine rich area is present at residues 72-95 (Penalva et al., 2000). When glutamine-rich domains are near histamine stretches, these are typically associated with transcription factors. The domains help promote protein-protein interactions and an alignment of the FL(2)D-PHO and *D. melanogaster* FL(2)D at the Histidine-rich and Glutamine-rich domain. No FL(2)D-PHO have histidine-rich and glutamine-rich domains (Figure 3.22). The FL(2)D - PHO may not have the same function as *D. melanogaster* FL(2)D, due to the lack of domains.

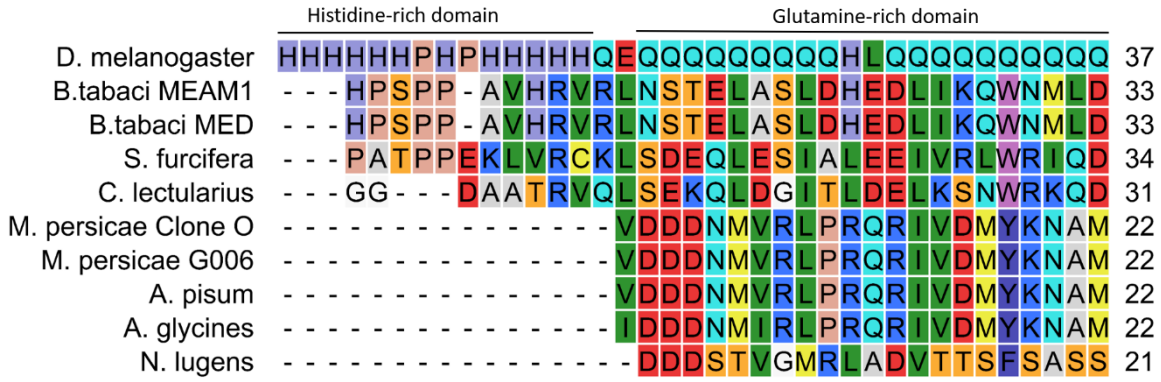

Figure 3. 22: The* D. melanogaster *FL(2)D aligned with FL(2)D-PHO at the Histidine and Glutamine rich areas.

3.2.6.5 Investigating gro orthologues

Groucho (Gro) is a 'support' gene; found in *D. melanogaster* sex determination cascade, (more information in Section 1.3.6). SXL requires GRO for negative regulation. RBBH analysis (methodology; fig. 2.1) identified GRO-PHO in all 11 hemipteran genomes. E-values are low in the BLASTP analysis (Table 3.23), and similarity scores range from 71.75% (*A. glycines*) to 94.57% (*S. furcifera*). Protein identity scores from the pairwise comparison analysis is lower than BLASTP protein similarity scores (Table 3.22). Protein identity scores from the pairwise comparison analysis is lower than BLASTP protein similarity scores (Table 3.24).

Hemipteran species	**Subject Seq-id**	**e-value**	**Length**	**Query Coverage**	**Identity (%)**	**Similarity (% of positive-scoring matches**
B. tabaci MED	BTA020461.1	0	724	98	61.05	72.51
B. tabaci MEAM1	Bta02857	0	738	97	66.94	75.2
D. citri	D_citri_rna1875	1.00E-141	299	48	72.58	79.6
A. glycines	A_glycine_017732	0	676	87	65.24	71.75
A. pisum	ACYPI005368-RA	0	765	99	65.88	73.46
M. persicae G006	MYZPE13164_G006_v1.0_000102370.1	0	750	97	66.27	73.6
M. persicae Clone O	MYZPE13164_0_v1.0_000105990.1	0	750	97	66.27	73.2
N. lugens	NLU003109.1	0	421	57	79.33	83.85
S. furcifera	Sfur-89.110	0	350	49	90	94.57
C. lectularius	CLEC000827	0	666	89	66.37	72.97
R. prolixus	RPRC006150-RA	0	737	97	67.3	74.49

Table 3. 23: Results of BLASTP analysis of the full-length* D. melanogaster *GRO protein (Query) against all hemipteran GRO (Subject).

Species	**Subject Seq-id**	**Percentage identity of full protein against the query**
B. tabaci MED	BTA020461.1	63.53
B. tabaci MEAM1	Bta02857	49.59
D. citri	D_citri_rna1875	38.59
A. glycines	A_glycine_017732	56.96
A. pisum	ACYPI005368-RA	55.36
M. persicae G006	MYZPE13164_G006_v1.0_000102370.1	64.34
M. persicae Clone O	MYZPE13164_0_v1.0_000105990.1	64.34
N. lugens	NLU003109.1	44.22
S. furcifera	Sfur-89.110	63.22
C. lectularius	CLEC000827	56.90
R. prolixus	RPRC006150-RA	63.19

Table 3. 24 Percentage identity of the pairwise comparison of the GRO PHO against the full-length* D. melanogaster *GRO protein

D. melanogaster GRO contains seven WD40 domains (Figure 3.23). WD40 are short amino acid motifs, which often terminate at a Trp-Asp (W-D) dipeptide. They have a variety of functions from cell cycle control and apoptosis to signal transduction and transcription regulation. The repeated motifs act as protein-protein interaction sites, and these proteins serve as the platform for the assembly of protein complexes. All GRO-PHO (identified in Table 3.23) underwent SMART protein analysis for protein domain identification (Letunic and Bork, 2018). Figure 3.23 is a schematic of the proteins domains in *D. melanogaster* GRO and GRO-PHO (not to scale). 10 hemipterans have seven WD40 protein domains; the exception is *D. citri* with five WD40 domains.

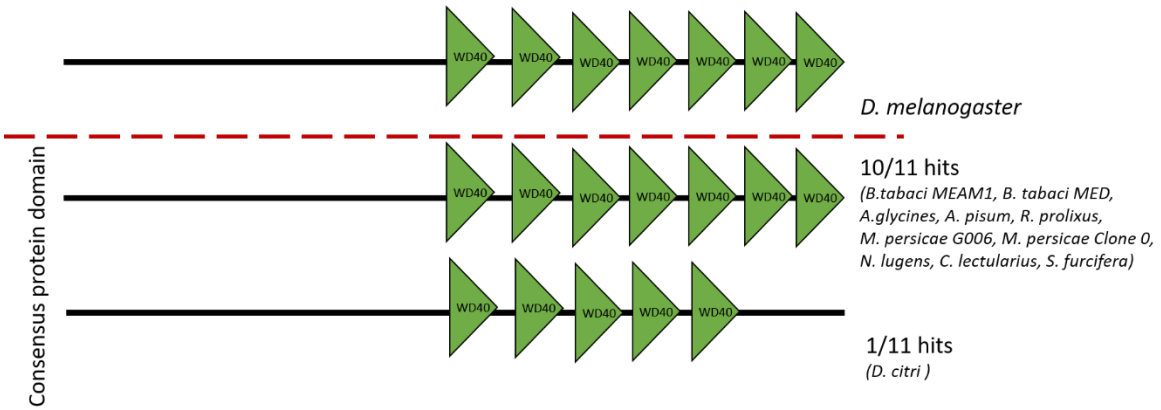

Figure 3. 23: A simplified schematic drawing of* D. melanogaster *GRO protein structure.

GRO WD40 protein domains may be more conserved than full-length proteins. Full-length GRO-PHOs were aligned to *D. melanogaster* GRO using CLC, and the WD40 domains identified (Figure 3.24). Overall, the WD40 domains seem conserved. *D. citri* has gaps in WD40 number 4, 5 and 6 which may the reason for the missing two domains. *N. lugens* is missing the WD40 at the WD40 number 7 position, but because the diagram does not consider the length another WD40 may be located in another position further in the protein.

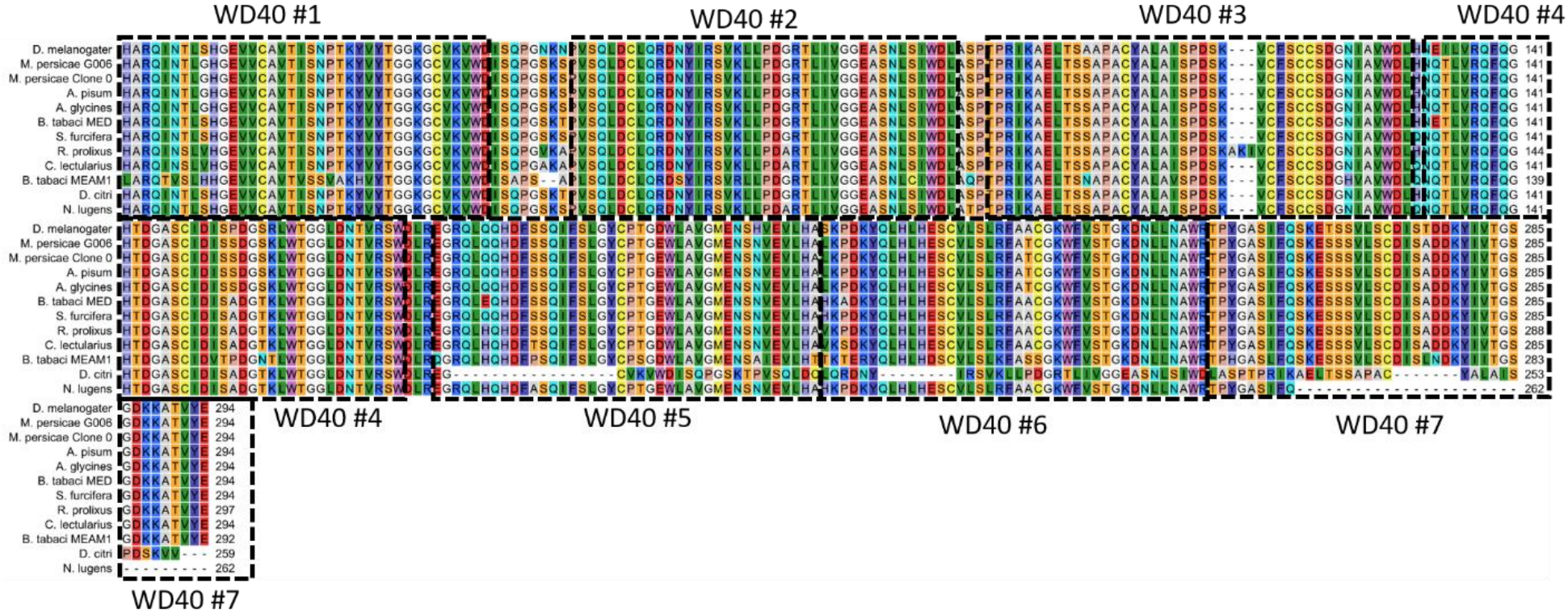

Figure 3. 24: The protein alignment of the GRO PHO and* D. melanogaster *protein at the WD40 sites.

The different WD40 protein domains are highlighted by a black lined box.

3.2.6.6 Investigating Her orthologues

Hermaphrodite (Her) is a 'support' gene; found in *D. melanogaster* sex determination cascade, (more information in Section 1.3.6). Sexual differentiation requires HER. RBBH analysis (methodology; fig. 2.1) identified HER-PHO in 6 hemipteran species (*B. tabaci MEAM1, D. citri, N. lugens, C. lectularius, R. prolixus*). E-value is not as low as some proteins studied in this chapter (Table 3.25), and similarity scores range from 42.75% (*N. lugens*) to 49.51% (*D. citri*). Protein identity scores from the pairwise comparison analysis is lower than BLASTP protein similarity scores (Table 3.26).

Hemipteran species	**Subject Seq-id**	**e-value**	**Length**	**Query Coverage**	**Identity (%)**	**Similarity (%)**
B. tabaci MEAM1	Bta05677	9.00E-06	173	32	28.32	43.93
D. citri	D_citri_rna4721	2.00E-09	103	21	34.95	49.51
N. lugens	NLU029082.3	4.00E-10	131	27	32.82	42.75
S. furcifera	Sfur-280.24	2.00E-11	159	28	30.19	45.28
C. lectularius	CLEC013447	2.00E-11	150	28	32	47.33
R. prolixus	RPRC009966-RA	3.00E-10	128	28	35.94	46.88

Table 3. 25: Results of BLASTP analysis of the full-length* D. melanogaster *HER protein (Query) against all hemipteran HER-PHO (Subject).

Species	**Subject Seq-id**	**Percentage identity of full protein against the query**
B. tabaci MEAM1	Bta05677	5.17
D. citri	D_citri_rna4721	6.73
N. lugens	NLU029082.3	6.66
S. furcifera	Sfur-280.24	9.62
C. lectularius	CLEC013447	6.24
R. prolixus	RPRC009966-RA	8.41

Table 3. 26: Percentage identity of the pairwise comparison of the HER PHO against the full-length* D. melanogaster *HER protein

D. melanogaster HER contains four zinc finger protein domains (Figure 3.2.3.1). More information about zinc fingers are in Section 3.2.3.1. All HER-PHO (identified in Table 3.25) underwent SMART protein analysis for protein domain identification (Letunic and Bork, 2018). Figure 3.25 is a schematic of the protein's domains in *D. melanogaster* HER and HER-PHO (not to scale). None of the hemipteran HER orthologues contains the same number of *D. melanogaster* zinc fingers. The hemipteran HER orthologues vary from three (*D. citri*) to ten (*C. lectularius*).

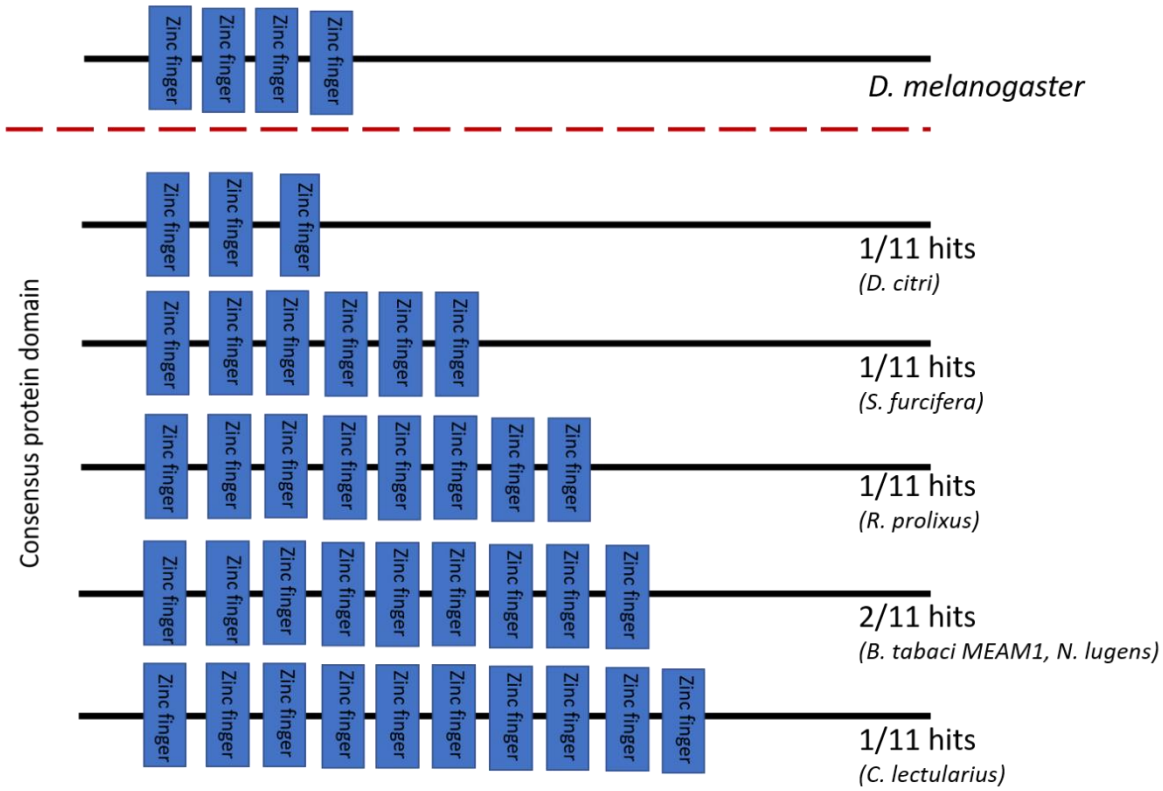

Figure 3. 25: A simplified schematic drawing of* D. melanogaster *Her protein structure.

3.2.6.7 Investigating Snf orthologues

Sans fille (Snf) is a 'support' gene; found in *D. melanogaster* sex determination cascade (more information in Section 1.3.6). SXL autoregulation requires SNF. RBBH analysis (methodology; fig. 2.1) identified SNF-PHO in all 11 hemipteran genomes. The E-value is low (Table 3.27), and similarity scores are high ranging from 72.44% (*S. furcifera*) to 93.27% (*N. lugens).* Protein identity scores from the pairwise comparison analysis is lower than BLASTP protein similarity scores (Table 3.28).

Hemipteran species	Subject Seq-id	e-value	Length	Query Coverage	Identity (%)	Similarity (%)
B. tabaci MEAM1	Bta15033	4.00E-102	245	99	64.49	74.69
B. tabaci MED	BTA003437.1	1.00E-43	100	46	75	82
D. citri	D_citri_rna10032	8.00E-108	246	100	67.07	77.24
A. glycines	A_glycine_09061	9.00E-103	223	100	70.4	81.61
A. pisum	ACYPI003668-RA	2.00E-102	223	100	70.4	81.61
M. persicae G006	MYZPE13164_G006_v1.0_000007620.1	2.00E-102	223	100	70.4	81.61
M. persicae Clone O	MYZPE13164_0_v1.0_000060450.1	2.00E-102	223	100	70.4	81.61
N. lugens	NLU004616.1	5.00E-49	104	48	84.62	93.27
S. furcifera	Sfur-572.4	9.00E-100	254	100	66.14	72.44
C. lectularius	CLEC000567	1.00E-81	223	100	65.92	74.89
R. prolixus	RPRC007750-RA	3.00E-104	234	100	70.94	80.34

Table 3. 27: Results of BLASTP analysis of the full-length* D. melanogaster *SNF protein (Query) against all hemipteran SNF-PHO (Subject).

Species	Subject Seq-id	Percentage identity of full protein against the query
B. tabaci MEAM1	Bta15033	63.10
B. tabaci MED	BTA003437.1	40.18
D. citri	D_citri_rna10032	66.40
A. glycines	A_glycine_09061	67.38
A. pisum	ACYPI003668-RA	67.38
M. persicae G006	MYZPE13164_G006_v1.0_000007620.1	67.38
M. persicae Clone O	MYZPE13164_0_v1.0_000060450.1	67.38
N. lugens	NLU004616.1	65.62
S. furcifera	Sfur-572.4	65.88
C. lectularius	CLEC000567	59.59
R. prolixus	RPRC007750-RA	69.29

Table 3. 28: Percentage identity of the pairwise comparison of the SNF PHO against the full-length* D. melanogaster *SNF protein

D. melanogaster SNF contains two RRM protein domains. More information about RRM are in Section 3.2.3.2. All SNF-PHO (identified in Table 3.27) underwent SMART protein analysis for protein domain identification (Letunic and Bork, 2018). Figure 3.26 is a schematic of the protein's domains in *D. melanogaster* SNF and SNF-PHO (not to scale). Nine PHO contained two RRM domains. However, two RBBH differ this is; *B. tabaci* MED (with 1 RRM) and *C. lectularius* (with no RRM).

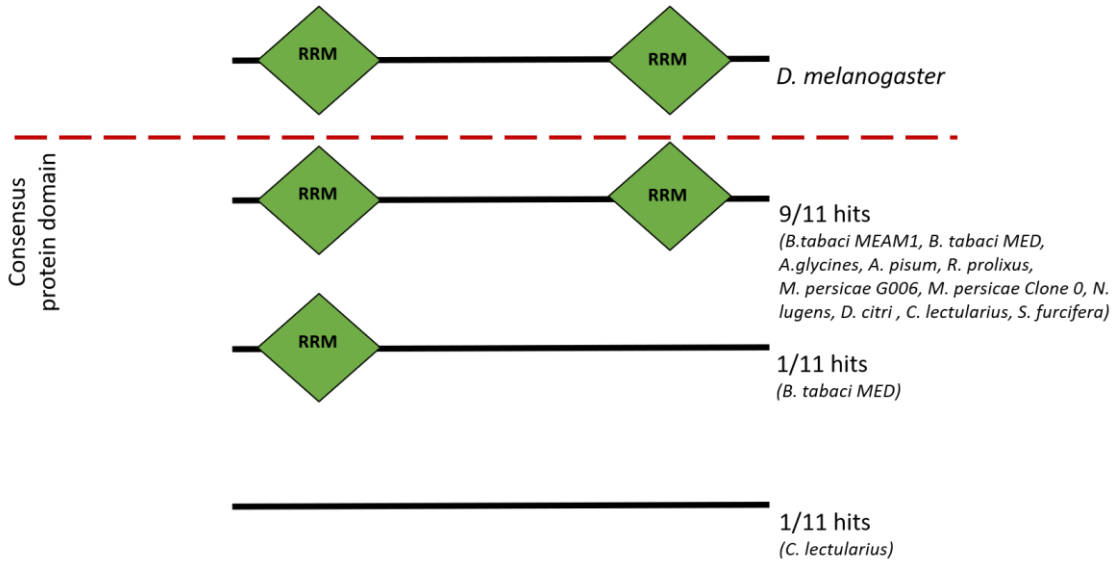

Figure 3. 26: A simplified schematic drawing of* D. melanogaster *SNF protein structure.

3.2.6.8 Investigating Vir orthologues

Virilizer (Vir) is a 'support' gene; found in *D. melanogaster* sex determination cascade (more information in Section 1.3.6). VIR is needed for maintenance of late SXL (Schutt et al., 1998). RBBH analysis (methodology; fig. 2.1) identified VIR-PHO in all 11 hemipteran genomes. The E-value is low (Table 3.29), and similarity scores range from 44% (*N. lugens*) to 54.35% (*D. citri*). Protein identity scores from the pairwise comparison analysis is lower than BLASTP protein similarity scores (Table 3.30). *D. melanogaster* VIR and VIR- PHO do not have any protein domains.

Hemipteran species	**Subject Seq-id**	**e-value**	**Length**	**Query Coverage**	**Identity (%)**	**Similarity (%)**
B. tabaci MEAM1	Bta14879	4.00E-115	1558	85	27.47	46.34
B. tabaci MED	BTA014178.1	8.00E-57	819	42	26.62	48.35
D. citri	D_citri_rna13774	7.00E-27	276	14	31.16	54.35
A. glycines	A_glycine_09166	6.00E-29	615	51	23.9	45.69
A. pisum	ACYPI002647-RA	1.00E-27	609	56	23.48	46.31
M. persicae G006	MYZPE13164_G006_v1.0_000129550.1	2.00E-29	619	33	22.13	45.72
M. persicae Clone O	MYZPE13164_0_v1.0_000187070.1	3.00E-30	619	35	22.13	46.04
N. lugens	NLU010196.1	1.00E-108	1441	76	26.3	44
S. furcifera	Sfur-131.52	2.00E-79	858	77	27.27	47.2
C. lectularius	CLEC013248	1.00E-92	957	90	28.84	49.22
R. prolixus	RPRC014736-RA	2.00E-135	1474	85	28.7	47.69

Table 3. 29 Results of BLASTP analysis of the full-length* D. melanogaster *VIR protein (Query) against all hemipteran VIR-PHO (Subject).

Species	Subject Seq-id	Percentage identity of full protein against the query
B. tabaci MEAM1	Bta14879	23.31
B. tabaci MED	BTA014178.1	10.96
D. citri	D_citri_rna13774	5.84
A. glycines	A_glycine_09166	13.17
A. pisum	ACYPI002647-RA	12.55
M. persicae G006	MYZPE13164_G006_v1.0_000129550.1	12.125
M. persicae Clone O	MYZPE13164_0_v1.0_000187070.1	12.69
N. lugens	NLU010196.1	17.93
S. furcifera	Sfur-131.52	18.87
C. lectularius	CLEC013248	21.79
R. prolixus	RPRC014736-RA	22.90

Table 3. 30: Percentage identity of the pairwise comparison of the VIR PHO against the full-length* D. melanogaster *VIR protein

3.2.6.9 Investigating Dsf orthologues

Dissatisfaction (Dsf) is a 'support' gene; found in *D. melanogaster* sex determination cascade, (more information in Section 1.3.6). DSF affects the courtship behaviours in *D. melanogaster*. TRA has three targets one is *Dsf*, the others are *Fru* and *Dsx.* Each of these genes are the top of different pathways contributing to the central nervous system in *D. melanogaster.* RBBH analysis (methodology; fig. 2.1) identified DSF-PHO in all 11 hemipteran genomes. The E-value is low (Table 3.31), and similarity scores range from 67.97% (*B.tabaci MEAM1*) to 82.2% (*R. prolixus*). Protein identity scores from the pairwise comparison analysis is lower than BLASTP protein similarity scores (Table 3.32).

Hemipteran species	Subject Seq-id	E-value	Length	Query Coverage	Identity (%)	Similarity (%)
B. tabaci MED	BTA012985.2	7E-84	256	37	57.03	67.97
B. tabaci MEAM1	Bta07918	5E-98	273	52	59.34	71.43
D. citri	D_citri_rna8036	5E-89	196	45	71.43	81.12
A. glycines	A_glycine_010412	5E-93	236	47	65.68	75.85
A. pisum	ACYPI56792-RA	3E-93	236	47	65.68	76.27
M. persicae G006	MYZPE13164_G006_v1.0_000188390.1	5E-93	236	47	65.68	75.85
M. persicae Clone O	MYZPE13164_0_v1.0_000032910.1	5E-93	236	47	65.68	75.85
N. lugens	NLU012929.1	5E-94	238	47	63.87	76.47
S. furcifera	Sfur-22.71	9E-91	267	46	59.18	69.29
C. lectularius	CLEC025112	3E-88	265	47	57.36	68.3
R. prolixus	RPRC010625-RA	3E-92	191	27	69.11	82.2

Table 3. 31 Results of BLASTP analysis of the full-length* D. melanogaster *DSF protein (Query) against all DSF-PHO (Subject).

Species	Subject Seq-id	Percentage identity of full protein against the query
B. tabaci MED	BTA012985.2	36.16
B. tabaci MEAM1	Bta07918	13.67
D. citri	D_citri_rna8036	34.15
A. glycines	A_glycine_010412	36.65
A. pisum	ACYPI56792-RA	39.83
M. persicae G006	MYZPE13164_G006_v1.0_000188390.1	39.69
M. persicae Clone O	MYZPE13164_0_v1.0_000032910.1	39.69
N. lugens	NLU012929.1	33.42
S. furcifera	Sfur-22.71	18.45
C. lectularius	CLEC025112	33.47
R. prolixus	RPRC010625-RA	18.99

Table 3. 32 Percentage identity of the pairwise comparison of the DSF PHO against the full-length* D. melanogaster DSF *protein

D. melanogaster DSF contains one Zinc finger and one HOLI domain. More information about the zinc finger domain is in Section 3.2.3.1. HOLI protein domain is a nuclear receptor and is one of the most abundant regulators in animals. The domain can bind to hydrophobic molecules. All DSF-PHO (identified in Table 3.31) underwent SMART protein analysis for protein domain identification (Letunic and Bork, 2018). Figure 3.27 is a schematic of the protein domains in *D. melanogaster* DSF and DSF-PHO (not to scale). Nine PHO contained both zinc finger and HOLI protein domains. However, two RBBH differ; *B. tabaci* MED and *R. prolixus* with only HOLI domain.

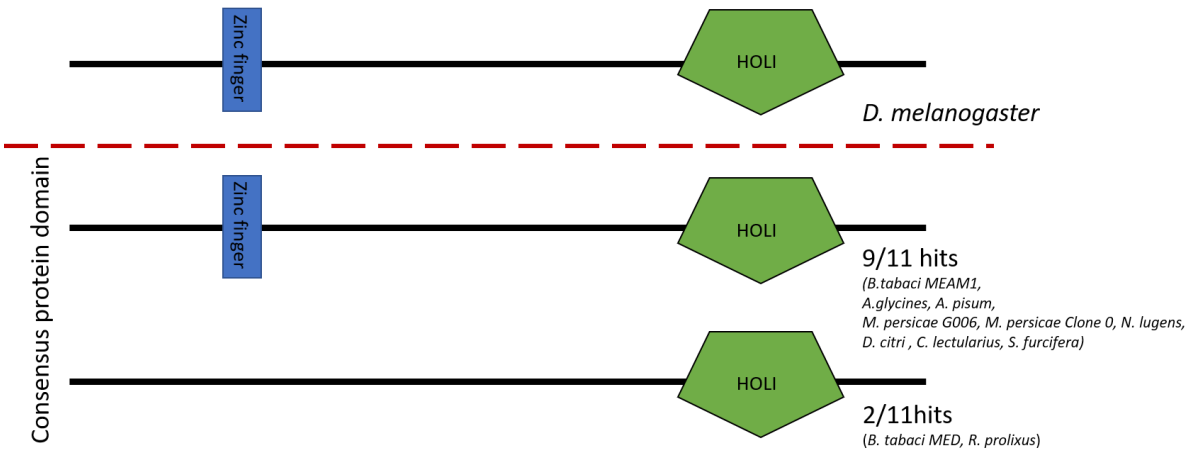

Figure 3. 27: A simplified schematic drawing of* D. melanogaster *DSF protein structure.

DSF HOLI protein domains may be more conserved than full-length proteins. Full-length DSF-PHOs were aligned to *D. melanogaster* DSF using CLC, and the HOLI domain identified (Figure 3.28b). The HOLI domain is conserved. A maximum-likelihood phylogeny tree analysis was conducted to explore the evolutionary relationship (Figure 3.2a). The branches range from being highly supported (bootstrap value of 0.943, between *C. lectularius* and *R. prolixus*).

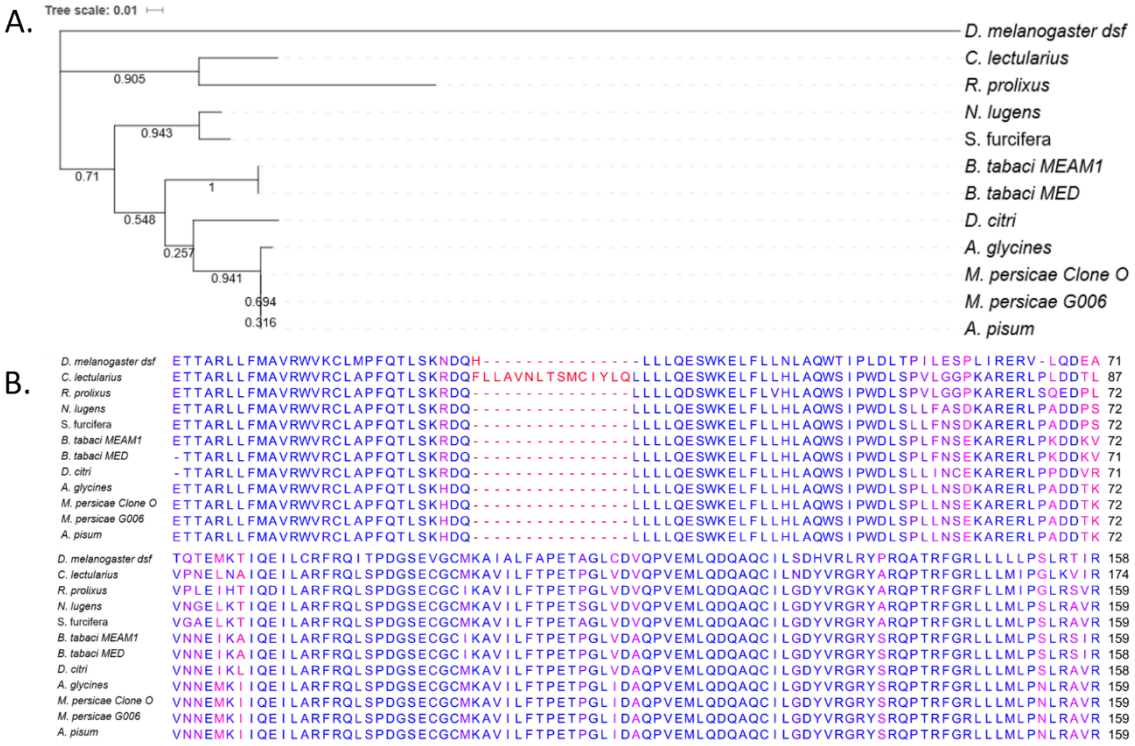

Figure 3. 28: Molecular Phylogenetic analysis of the DSF HOLI domain from the Hemiptera DSF orthologues by Maximum Likelihood method.

A. The evolutionary history was inferred by using the Maximum Likelihood method based on the Le_Gascuel_2008 model. The best model was calculated by Mega7. The analysis involved 12 amino acid sequences. It was edited in the Interactive tree of life. This tree is from the hemipteran orthologues of *D. melanogaster* DSF protein at the HOLI domain. The tree is rooted at *D. melanogaster Dsf* B. The alignment sequence of HOLI protein domain used for the tree in A (Le and Gascuel, 2008; Kumar et al., 2016; Letunic and Bork, 2016; Felsenstein, 1985).

3.2.6.10 Investigating Fru orthologues

Fruitless (Fru) is a 'support' gene; found in *D. melanogaster* sex determination cascade (more information in Section 1.3.6). FRU is required for proper development of several anatomical structures needed for courtship (Demir and Dickson, 2005). FRU is a direct target of TRA (Yamamoto et al., 1998). RBBH analysis (methodology; fig. 2.1) identified FRU-PHO in all 11 hemipteran genomes. The E-value is low (Table 3.33), and similarity scores range from 72.8% (*D. citri*) to 77.39% (*R. prolixus*). Protein identity scores from the pairwise comparison analysis is lower than BLASTP protein similarity scores (Table 3.34).

Hemipteran species	Subject Seq-id	e-value	Length	Query Coverage	Identity (%)	Similarity (%)
B. tabaci MED	BTA006255.2	2.00E-40	124	13	56.45	72.58
B. tabaci MEAM1	Bta13492	2.00E-39	124	26	56.45	72.58
D. citri	D_citri_rna1194	9.00E-42	125	14	57.6	72.8
A. glycine	A_glycine_012334	1.00E-38	113	13	54.87	73.45
A. pisum	ACYPI006076-RA	2.00E-39	115	13	57.39	74.78
M. persicae G006	MYZPE13164_G006_v1.0_000040020.1	2.00E-39	115	13	59.13	75.65
M. persicae Clone O	MYZPE13164_0_v1.0_000092010.1	2.00E-39	115	13	59.13	75.65
N. lugens	NLU012995.5	1.00E-45	129	14	62.79	75.97
S. furcifera	Sfur-188.24	3.00E-44	129	14	62.79	75.97
C. lectularius	CLEC008059	9.00E-42	121	13	61.98	75.21
R. prolixus	RPRC014183-RA	7.00E-42	115	13	59.13	77.39

Table 3. 33 Results of BLASTP analysis of the full-length* D. melanogaster *FRU protein (Query) against all hemipteran FRU-PHO (Subject).

Species	Subject Seq-id	Percentage identity of full protein against the query
B. tabaci MED	BTA006255.2	11.28
B. tabaci MEAM1	Bta13492	14.71
D. citri	D_citri_rna1194	17.30
A. glycine	A_glycine_012334	17.07
A. pisum	ACYPI006076-RA	28.33
M. persicae G006	MYZPE13164_G006_v1.0_000040020.1	12.16
M. persicae Clone O	MYZPE13164_0_v1.0_000092010.1	12.16
N. lugens	NLU012995.5	14.34
S. furcifera	Sfur-188.24	10.33
C. lectularius	CLEC008059	18.27
R. prolixus	RPRC014183-RA	46.31

Table 3. 34 Percentage identity of the pairwise comparison of the FRU PHO against the full-length D. melanogaster FRU protein

D. melanogaster FRU contains one BTB (Broad- complex, Tramtrack and Bric a Brac) and two Zinc finger domains. Section 3.2.31 contains more information on zinc finger protein domains. BTB is also known as the Poxvirus and Zinc finger (POZ) domain. This domain occurs at the N terminus of proteins that contain zinc fingers. The domain function is to help facilitate homodimerization. The proteins that have these domains are normally transcriptional regulators, and they are thought to act through the control of the chromatin structure (Letunic and Bork, 2018). All FRU-PHO (identified in Table 3.33) underwent SMART protein analysis for protein domain identification (Letunic and Bork, 2018). Figure 3.29 is a schematic of the protein domains in *D. melanogaster* FRU and FRU-PHO (not to scale). All 11 PHO contain the BTB protein domain. The varying part of the protein depends on the number of zinc fingers that are present, which ranges from 0 to 4.

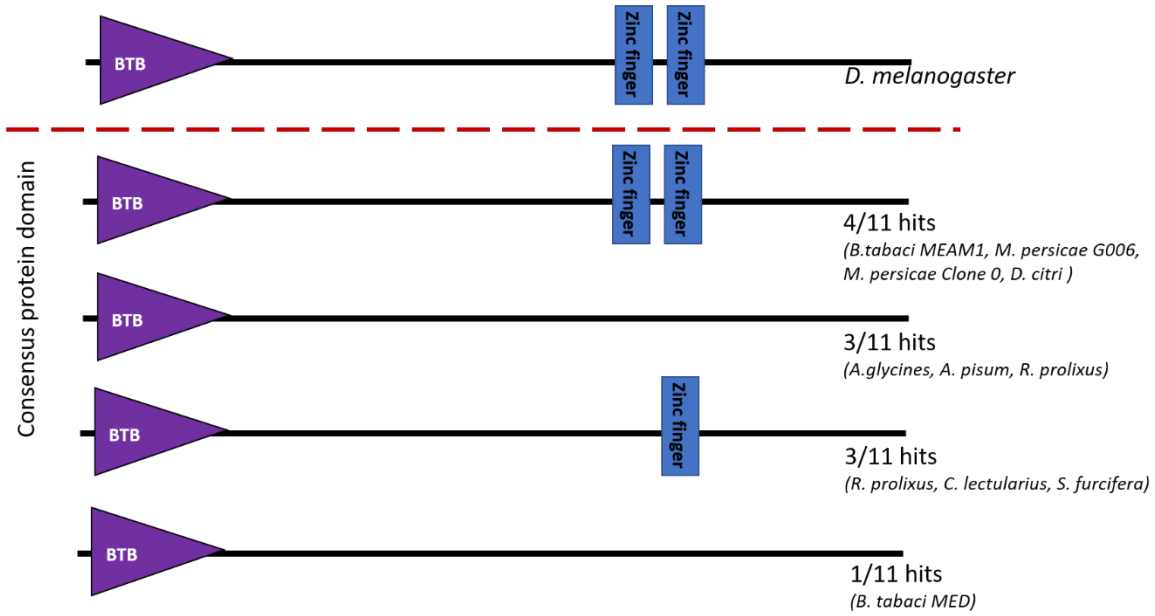

Figure 3. 29: A simplified schematic drawing of* D. melanogaster *FRU protein structure.

FRU BTB protein domain may be more conserved than full-length proteins. Full-length FRU-PHOs were aligned to *D. melanogaster* FRU using CLC, and the BRB domain identified (Figure 3.30b). The BTB domain is conserved; blue indicates same amino acid in that position and red means dissimilar positions. A maximum-likelihood phylogeny tree analysis was conducted to see the evolutionary relationship (Figure 3.30a). The branches range from being highly supported (bootstrap value 1, between *B. tabaci* MED and MEAM1) to being very poorly supported (bootstrap value of 0.028).

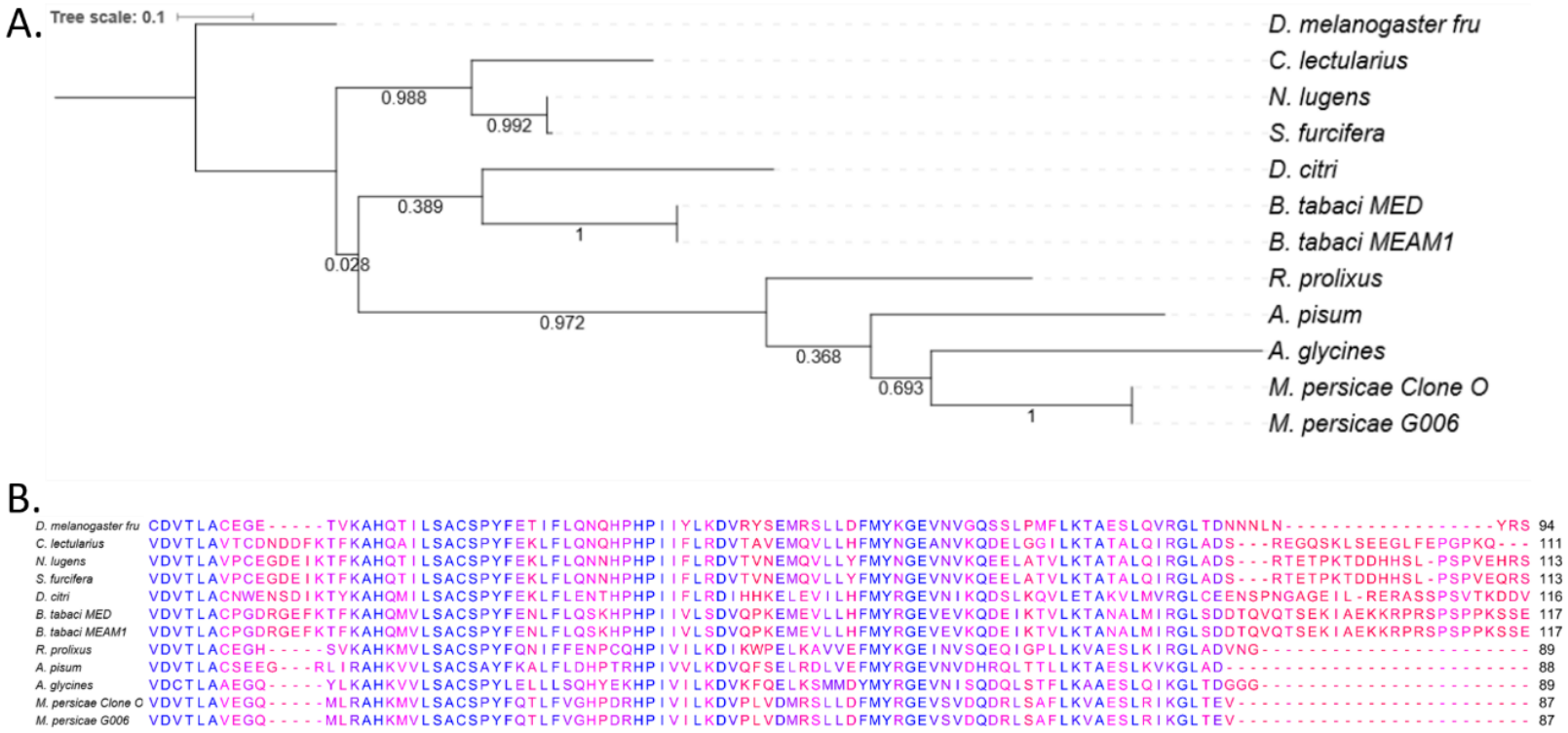

Figure 3. 30: Molecular Phylogenetic analysis of the fru at BTB protein domain for the hemipteran orthologues of FRU and* D. melanogaster *FRU protein

A. The evolutionary history was inferred by using the Maximum Likelihood method based on the Le_Gascuel_2008 model, which was calculated as the best model prediction by MEGA7. The bootstrap consensus tree inferred from 1000 replicates. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 12 amino acid sequences. Evolutionary analyses were conducted in MEGA7. It was edited in the Interactive tree of life. This tree is from the RBBH of the FRU protein. The tree is rooted at *D. melanogaster* FRU. B. The alignment of the FRU PHO and *D. melanogaster* FRU at the BTB protein domain. Blue indicates conserved regions on the sequence, red is the unconserved regions of the sequence (Le and Gascuel, 2008; Kumar et al., 2016; Letunic and Bork, 2016; Felsenstein, 1985).

3.2.7 Are the *M. persicae* SDGs found on the X chromosome?

The sex-determining chromosome typically contains the sex determination genes. T. Mathers in the Hogenhout laboratory has assembled the X chromosome of *M. persicae* Clone 0 (Mathers et al., 2018). I analysed the sex chromosome database (TBLASTN; protein sequences against nucleotide database) using the SDGs previously identified in this chapter to see if any SDG-PHO was found on the sex chromosome in *M. persicae* Clone 0, along with the location of the genes on this chromosome. Any SDG-PHO found on the sex chromosome will provide more evidence that these are true SDGs. The X-chromosome analysis for the *M. persicae* Clone 0 PHO, revealed that not all the PHO was present on the X chromosome. Initially, the *M. persicae* Clone 0 PHOs were from genome assembly version 1. Version 2 of the genome assembly had the X chromosome characterised. PHO location analysis used version 2 (Table 3.35).

	Query Seq-id	**Subject Seq-id**	**e-value**	**Start of alignment in subject**	**End of alignment in subject**
Da	MYZPE13164_0_v1.0_000036540.1	scaffold_1	5.00E-50	73950151	73950005
Dsf	MYZPE13164_0_v1.0_000032910.1	scaffold_1	1.00E-14	72328183	72327872
Fru	MYZPE13164_0_v1.0_000092010.1	scaffold_1	4.00E-23	52138318	52138151
Hemiptera-specific Dsx	MYZPE13164_0_v1.0_000064550.1	scaffold_1	7.00E-96	73373822	73373223
Sxl	MYZPE13164_0_v1.0_000097000.3	scaffold_1	2.00E-20	70632122	70632298
Tra2	MYZPE13164_0_v1.0_000075270.1	scaffold_1	2.00E-81	84847066	84847386

Table 3. 35 A table of all the sex determination genes in* M. persicae *Clone 0.

Figure 3.31 shows the location of the *M. persicae* clone 0 PHO identified in Table 3.35, on the X chromosome. There is clustering of *Sxl, Dsf, Dsx* and *Da* on the X chromosome.

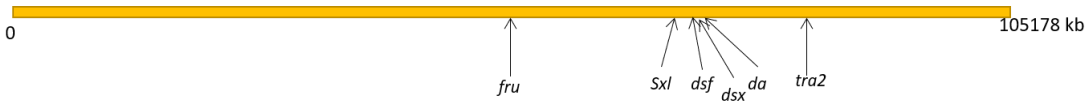

Figure 3. 31: The position of the orthologue* M. persicae *clone 0 SDGs found on the X chromosome (Scaffold1 in Myzus_persicae_O_v2.0.scaffolds.fa). The Dsx in this figure represents the Hemiptera-specific Dsx

In the Hogenhout laboratory, an RNA-seq experiment took place with different developmental and sexual stages of the *M. persicae* Clone 0 (Mathers et al., 2018). T. Mathers mapped the reads onto the g006 models and produced the FPKM for this. The genes that were found in Table 3.35 were investigated using this data set to see if they were differentially expressed. Sam Mugford (Hogenhout laboratory) provided the 1:1 orthologue from the G006 and Clone 0 gene set.

			Mean expression (FPKM)				Significance (PADJ)					
SDG	**Clone 0 id**	**1:1 Orthologue in g006**	**Male**	**Female asexual**	**Female winged**	**Nymphs**	**Male vs Female Asexual**	**Male vs Female Winged**	**Male vs Nymph**	**Female Asexual vs Female Winged**	**Female Asexual vs Nymph**	**Female Winged vs Nymph**
Fru	MYZPE13164_0_v1.0_000092010.1	MYZPE13164_G006_v1.0_000040020	73.68167	65.26	65.74	36.56333	1	1	5.73E-09	1	1	1
Sxl	MYZPE13164_0_v1.0_000097000.3	MYZPE13164_G006_v1.0_000089020	124.4383	137.7667	140.4633	110.4733	1	1	1	1	1	1
Dsf	MYZPE13164_0_v1.0_000032910.1	MYZPE13164_G006_v1.0_000188390	0.4	0.583333	0.745	0.508333	1	1	1	1	1	1
Hemiptera-specific Dsx	MYZPE13164_0_v1.0_000064550.1	MYZPE13164_G006_v1.0_000116760	11.67667	2.541667	8.063333	1.425	3.83E-18	1	1.24E-31	8.08E-06	1	1.50E-15
Da	MYZPE13164_0_v1.0_000036540.1	MYZPE13164_G006_v1.0_000018360	32.855	46.51833	45.41333	27.21	1	1	1	1	1	1
Tra2	MYZPE13164_0_v1.0_000075270.1	MYZPE13164_G006_v1.0_000099530	10.03333	16.30667	13.94667	11.06833	1	1	1	1	1	1

Table 3. 36: The differential expression data of the M. persicae SDGs that are found on the X chromosome at different life stages. (Mathers et al., 2018)

The FPKM and PADJ was calculated by T. Mathers. The DSX seems to be differentially expressed in the Hemiptera-specific '*Dsx'*. There is upregulation in male and winged females in the *Dsx.* There also seems to be a downregulation in nymphs *Fru*. This portion of the chapter gives more evidence that the SDGs found in *M. persicae* Clone 0 (*Da, Dsf, Fru, Dsx, Sxl and Tra2*) are true SDGs orthologues, as these were found on the X chromosome.

3.3 Discussion

Sex determination among insects typically follows the same structure, see Figure 1.8 in introduction. The primary signalling elements in Hemiptera can vary widely; such as haplodiploid organisms to XX/XO systems (see 1.3 for more information). At the beginning of this chapter, the hypothesis was that the signalling element would determine the unique sex determination proteins found in the pathways.

RBBH analysis (Section 3.2.1), BLASTP alignment statistics, full-length protein comparisons and protein domain (Section 3.2.2-3.2.6) analysed the PHO (putative hemipteran orthologue). This investigation highlighted the PHOs with the higher probability of being true orthologues. The list will be taken forward in the future chapters.

3.3.1 Are sex determination genes found in Hemiptera, if so how conserved are they?

Figure 3.32 presents a summary of the chapter. If the SDP had an RBBH PHO then it represented by a tick, if it did not have an RBBH PHO then it is represented by a cross. The ticks have two colours, yellow and green. Throughout this chapter, information on the PHO-SDGs were obtained, such as what protein domains were present or certain features that define the protein function, which allowed a generalisation on the protein as to whether I think the PHO is a true orthologue or not. The results are present in Figure 3.32, yellow means that there is a low probability of the PHO being true orthologues and green indicates a higher probability.

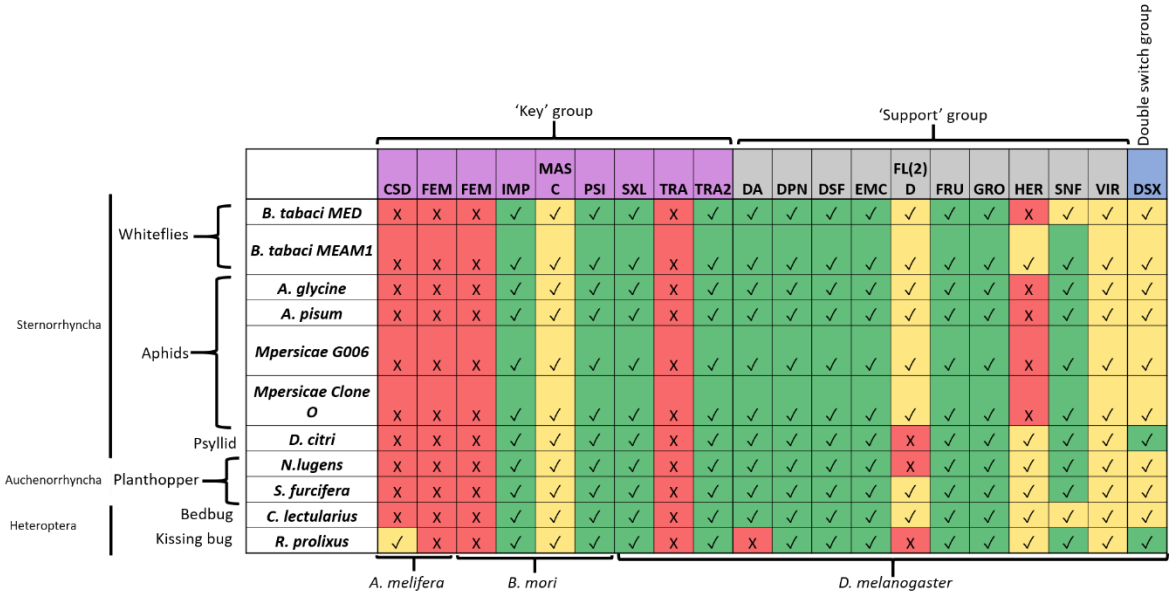

Figure 3. 32: A summary table of all the putative hemipteran SDG's.

Red is absence of any RBBH or blastp hit. Any presence of hemipteran putative SDG's was analysed further. Any green ticks suggest that there is high evidence of the putative SDG's being real orthologues, any yellow suggests that there is lower evidence of the putative SDG's being real orthologues. The genes are ordered by 'key genes', 'support genes' and the double switch gene. The organisms are ordered by Sternorryncha, Auchenorryncha and Heteroptera. On the bottom of the table shows which organism the SDG's originally came from. Please note there are two columns for FEM, however one is from *A. mellifera* whilst the other is from *B. mori*. The two FEMs have the same name but are not orthologues of each other.

Proteins from the 'key' group show the most variability in absence and presence. CSD, both FEMs and TRA are not present in any hemipteran (except CSD in *R. prolixus*). These proteins may have evolved in the holometabolous insects after their evolutionary divergence. During the RBBH analysis MASC-PHO was identified. However, there is a lack of masculizing residues; therefore, there is a low probability of this PHO having the same function as *B. mori*. Liu *et al* (2019) (Section 1.3.6, Table 1.3) also looked at the different *B. tabaci* MED SDGs to find PHO, there was no entry for *Masc*, so I conclude that they did not find a suitable MASC-PHO either.

The support genes vary a bit more in presence or absence. *Her* is an autoregulation gene of *Sxl*, it is not present in aphids and *B. tabaci* MED and the other SDGs do not give high support that these are true orthologues. Liu *et al* (2019) concurred with my results and did not find *Her* in *B. tabaci* MED. *Fl(2)d* also has variability of presence, however, unlike *Her* there is less of a pattern, with an absence in *D. citri, N. lugens* and *R. prolixus.* The putative FL(2)D protein that was found in the other Hemiptera had little evidence that they are orthologues. There was a VIR RBBH found in all the Hemiptera, however from further investigations there was little evidence that this was not a true orthologue. *Da* is not present in *R. prolixus*. These absences may be due to the annotation

quality of the genome. However, this one gene is missing from several distinct genomes speaks against this. The probability that poor assembly/ annotation would miss the same gene in several genomes is small.

Dsx is the double switch gene. The RBBH is present in all Hemiptera. There is high evidence that there are DM-like proteins present in all the Hemiptera. However, there is limited evidence that the putative DSX hemipteran proteins are DSX, and this is considering the protein domains and specifically the clades that all of the hemipteran proteins (when searched for DM protein domain) fall into.

All the RBBH and hits in Hemiptera only have DM protein domain and not the DSX dimer domain, see figure 3.1. The PHO RBBH cluster in the DMRT99B region; the RBBH PHO may be orthologues of DMRT99B. *D. melanogaster* larvae have DMRT99B expression; specifically, in the midline cells of the central nervous system (Fontana and Crews, 2012) and DMRT99B deficiency causes behavioural abnormalities in *B. mori* (Kasahara et al., 2018).

DSX in *Bemisia tabaci* has been tentatively characterised previously (Guo et al., 2018b). In this paper, the *B. tabaci 'Dsx'* contained 6 exons (details in Chapter 5), and the protein had a DM and a DSX domain. The genome was used for this analysis, which at the time was unpublished. My results disagree with this information, as previously mentioned about the DSX domain. The proteins that have been entered NCBI by this paper only have the DM domain and not the DSX domain- this is difficult to ascertain whether the information provided by the paper is true or not. The follow-up paper (Liu et al., 2019) suggested that *Dsx* was gene BTA004042.1 which was not my RBBH, but it was found in my general blast search. BTA004042.1 when analysed on the IGV is a mono-exonic protein, and when investigated further shows only a DM domain and no DSX. Regretfully, I have not been able to replicate the results from these two papers for finding a DSX protein with DM and DSX protein domains. However, I think my pipeline was rigorous to conclude that what I have found is correct, as it had more stringent cut-offs (i.e. E-value) and therefore limits the false-positives. It may be that the 'DSX' protein that was named as such by the previous paper. In my opinion this gene encodes a DM domain protein that is not DSX but does have sex differentiation properties. When this 'DSX' protein was knocked down by silencing the two tail pins of the male genitalia disappeared and the genitalia was malformed (Guo et al., 2018a). This knockdown result indicates that the 'DSX' is needed for genitalia development, which is more of a sex differentiation aspect of DM containing genes. In my opinion, if it was part of the sex determination cascade, the outcome would indicate a difference in sex ratios.

Liu *et al* (2019) (Section 1.3.6, Table 1.3) found other sex determination genes that I also identified. I found the same SDG-PHO for *Da, Dpn, Emc, Fru, Gro, Imp, Psi, Snf, Sxl, Tra2* and *Vir*, as the gene IDs were the same. My PHO differ in *Dsf, dsx* and *Fl(2)d* compared to Liu *et al* (2019) list, however due to the RBBH pipeline, I believe that my results are true.

3.3.2 Are the *M. persicae* clone 0 SDGs found on the X chromosome?

T. Mathers assembled the X chromosome of M. persicae Clone 0. I took the SDG's from *M. persicae* clone 0 and looked to see whether any was present on the X chromosome. Some SDGs have been found on the sex-determining chromosomes in the past, so the genes found on the sex chromosome will give better support to these genes being truly part of the sex determination pathway. The SDGs genes found on the X chromosome are *Da, Dsf, Fru, Dsx, Sxl* and *Tra2*. These genes also seem to cluster close on this X chromosome scaffold (figure 3.31).

I looked at the differential expression of these genes at different stages produced by T. Mathers The whole genome was analysed and provided by T. Mathers (Mathers et al., 2018), Table 3.36 was the extracted genes from the whole genome that was relevant to this thesis. *Dsx* seems to be upregulated in males and winged females. Both males and winged females have wings. The '*Dsx*' may be involved with sex determination, or possibly wing formation. The *DSX* has been repurposed in the Lepidoptera*, Papilio polytes*, as a colour pattern gene in the wings. Something similar may be happening in the winged aphids (Kunte et al., 2014).

3.4 Conclusion

Hemiptera do not have different cascades depending on the signalling element (as suggested in the introduction 3.1). However, the Hemiptera do seem to have a conserved sex determination pathway; as when one SDG is missing, it is typically missing in all. There is not one holometabolous sex-determination pathway that is an exact match for Hemiptera. *D. melanogaster* and *B. mori* seem similar but there are inconsistencies in the orthologue proteins, such as MASC-PHO not having the masculinizing effect.

I have found potential orthologues of sex determination genes. However, we do not know about the functionality of these genes. Sex determination genes should be expressed during the embryogenesis stage. To determine whether this is true, I will conduct single-embryo RNA-seq analysis in a later chapter. *B. tabaci* is an excellent model to get male samples, as all virgin females will produce the haploid male eggs so that sex-specific isoforms analysis can be explored. The single-egg RNA-seq experiment may not determine the functionality of the genes but will help find genes that are present in early embryogenesis. Therefore, genes that could be used in a self-limiting system.

Chapter 4: Investigations of embryogenesis in four developmental stages of *B. tabaci* eggs

4.1 Introduction

Embryogenesis (embryonic development) of insects usually happens in eggs after they are laid. Embryogenesis comprise a series of complex events that occur before an organism is born or hatches from an egg, and there is a commonality in the stages and sequences to events among animals, such as the formation of blastula and gastrula (Campos-Ortega and Hartenstein, 1997).

Drosophila melanogaster is a model organism for insect embryogenesis, which occurs in eggs after they are laid by a female. The embryogenic stages in *D. melanogaster* are well-defined. From the first zygotic nuclear divisions to the cellularization stage there are thirteen divisions of nuclei in the early embryo. Before the cellularization, in the pre-blastoderm stage, the nuclei are not separated by plasma membranes and lie within the same cytoplasm. The pre-blastoderm stage has 1-8 nuclear divisions, and the divisions occur synchronously (Zalokar and Erk, 1976; O'Farrell et al., 2004). During blastoderm formation, between nuclear divisions 8-9, the nuclei migrate outwards to form the shell and become part of a layer of cells that forms the wall of the blastula (Foe and Alberts, 1983). Germ-line cell formation occurs during cell cycles 9-10 (Farrell and O'Farrell, 2014) during which cytoplasmic bridges that connect cells of the blastoderm to the cells in the yolk sac are formed, and these bridges are pinched off during gastrulation. Gastrulation is the formation of germ layers (Campos-Ortega and Hartenstein, 1997).

In some agricultural pests, such as *B. tabaci*, there is a lack of fundamental knowledge on embryonic development (see Section 1.2.1 for more information). This knowledge is vital for developing new control methods. The best way to uncover the different embryonic stages of *B. tabac*i is to compare it to a well-known embryonic development system, like *D. melanogaster*. Using the comparative approach with *D. melanogaster*, new embryonic developmental stages have been described in the scuttle fly (*Megaselia abdita*) (Wotton et al., 2014) and the moth midge *(Clogmia albipunctata)* (Jimenez-Guri et al., 2014).

Current knowledge of *B. tabaci* embryogenesis is limited. However, what is known is that *B. tabaci* has telotrophic ovaries in which the nutritive chords are connected to the oocytes and the nurse cells (Ghanim et al., 2001). The *B. tabaci* ovaries contain oocytes with different developmental stages. At least four developmental stages of oocytes/eggs in the ovaries have been described, these are named the A, B, C and D stages with the A stage eggs being the most immature phase and D stage egg the fully mature stage (Guo et al., 2016).

This chapter aims to develop more knowledge of *B. tabaci* MED embryogenesis. The investigation includes optimisation of nuclei staining methods and nuclei counting. Confocal microscopy was used to count the nuclei of different *B. tabaci* MED developmental stages. Comparative analysis between nuclei from *B. tabaci* and *D. melanogaster* established the *B. tabaci* embryonic developmental stages. The hypothesis is that *B. tabaci* has similar embryogenesis and developmental stages as *D. melanogaster.*

This study provides evidence that embryogenesis and cellularization starts inside the adult female abdomen, not outside the abdomen like *D. melanogaster*. Germ-line cells, before the end of the blastoderm stage (between nuclear cycles 8-10,) will have to be transformed for obtaining stable transgenic insect lines. My data shows that transformation constructs will have to be introduced in egg stages that occur within whitefly females rather than into eggs that are already laid. This data, therefore, contributes crucial fundamental knowledge to achieve the goal to develop genetic methods to control whitefly populations.

4.2 Results

4.2.1 DAPI staining of *D. melanogaster* eggs images the nuclei

4',6-Diamidine-2'-phenylindole dihydrochloride (DAPI) is a fluorophore that detects nuclei in confocal microscopy. The DAPI protocol used is based on one designed for *D. melanogaster* (Rothwell and Sullivan, 2000). The goal of this section is to ensure that the DAPI protocol (described in Section 2.2.5) works before optimising it for *B. tabaci,* by testing the DAPI method on *D. melanogaster* eggs (provided by Prof. Tracey Chapman, School of Biological Sciences, UEA, Norwich, UK). Briefly, the DAPI protocol includes a Carnoy fixative that was used overnight to fix the eggs, and the addition of bleach to remove chorion from the eggs. Then, hybridisation buffer was added to the eggs to facilitate the spread of fluorophores. DAPI was added to the eggs before being washed with 1x PBS, and Vectashield (Vector Laboratories) was applied to the sample before confocal microscopy.

Figure 4.1 shows DAPI stained *D. melanogaster* eggs. The fluorescence image clearly shows discrete circular objects indicating the presence of cell nuclei, as expected. Therefore, the DAPI protocol worked for the detection of nuclei in *D. melanogaster* eggs, and it was decided to use this protocol to detect nuclei in *B. tabaci* eggs.

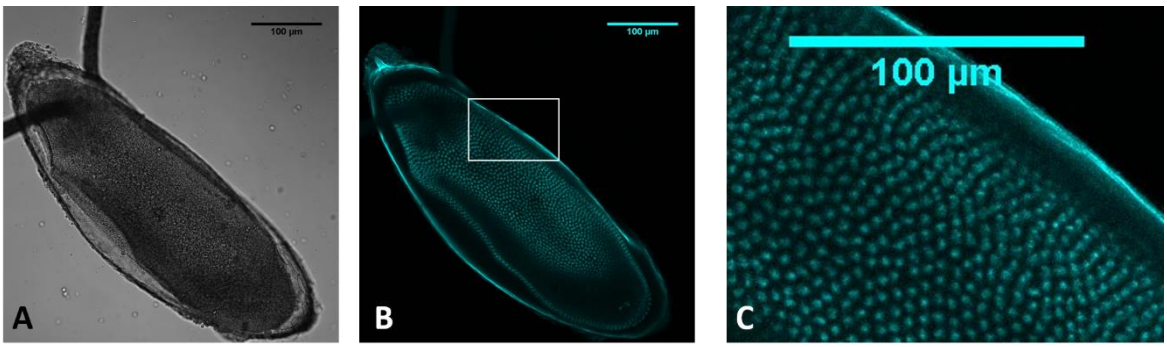

Figure 4. 1 0-5-hour old* D. melanogaster *eggs stained with DAPI viewed on a confocal microscope.
A.The brightfield image of the *D. melanogaster* egg at 20x magnification. B. The confocal image at 405 nm emission. C. Magnified image of the white rectangle in image B. The scale bars represemt 100 µm.

4.2.2 Assessment of autofluorescence of whitefly adults and eggs

Autofluorescence is fluorescence produced from endogenous materials or compounds present in a biological sample. The insect cuticle may produce autofluorescence. Cuticle properties in insects such as *Cimex lectularius* (bedbug) often produce autofluorescence (Reinhardt et al., 2017). To be able to count nuclei this study, the DAPI-stained nuclei in whitefly females need to be brighter than any autofluorescence. There is no information in the literature on the autofluorescence levels of *B. tabaci* MED adults or eggs. Therefore, before testing different staining methods, *B. tabaci* autofluorescence was investigated by confocal microscopy.

B. tabaci non-virgin adults were examined for autofluorescence levels. The adults were collected (Section 2.2.3), fixated (Section 2.2.5; without any DAPI and bleach), imaged (Section 2.2.8) and analysed (Section 2.2.9), and the excitation and emission wavelengths of standard fluorophore data points measured. Data are shown in Figure 4.2.

Whitefly adult bodies showed autofluorescence signals. The autofluorescence signals of adults was high at the wavelengths of DAPI fluorescence (excitation 358 nm; emission 461 nm) and were also seen at the wavelengths of GFP (excitation 488 nm; emission 510 nm), though no obvious autofluorescence was seen for wavelengths of CFP, YFP and RFP. Therefore, using DAPI and GFP fluorophores may mask the detection of signals from DAPI or GFP-stained nuclei in future experiments.

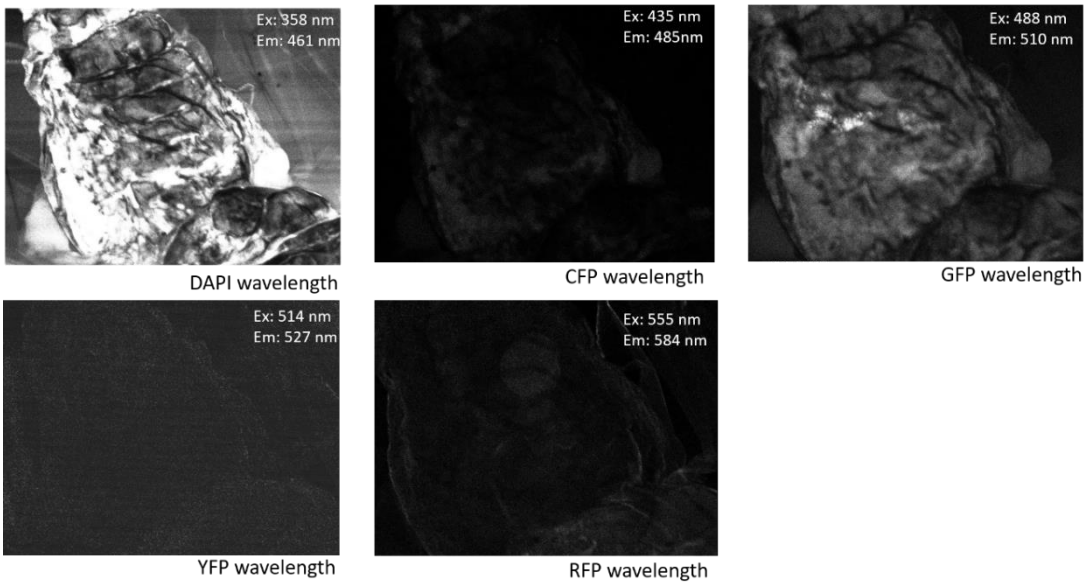

Figure 4. 2: Autofluorescence in the adult male abdomen throughout the different (Em) emission and (Ex) excitation wavelengths.

The emission and excitation wavelengths correspond with standard fluorophores; DAPI, cyan fluorescent protein (CFP), green fluorescent protein (GFP), yellow fluorescent protein (YFP) and red fluorescent protein (RFP).

In this chapter, eggs will be examined so eggs underwent autofluorescence testing. The adults were collected (Section 2.2.3) for dissection of D type eggs (Section 2.2.4), which were fixated (Section 2.2.5; without any DAPI and bleach), imaged (Section 2.2.8) and analysed (Section 2.2.9). The excitation and emission wavelengths used correlated with the standard fluorophore data points. The results are shown in Figure 4.3.

Eggs emitted autofluorescence signal (Figure 4.3). The autofluorescence signal was high at the DAPI wavelengths (excitation 358 nm; emission 461 nm), and at RFP wavelengths (excitation 555 nm; emission 584 nm). Using DAPI and RFP fluorophores may mask the detection of nuclei signal in D type eggs in experiments.

There is autofluorescence occurring in the DAPI wavelength. However, DAPI is a well-established nuclei fluorophore in insect embryogenesis. The eggs (Figure 4.3) or adult (Figure 4.2) samples were not bleached; therefore, it is unclear whether de-chorionating the samples will affect the autofluorescence.

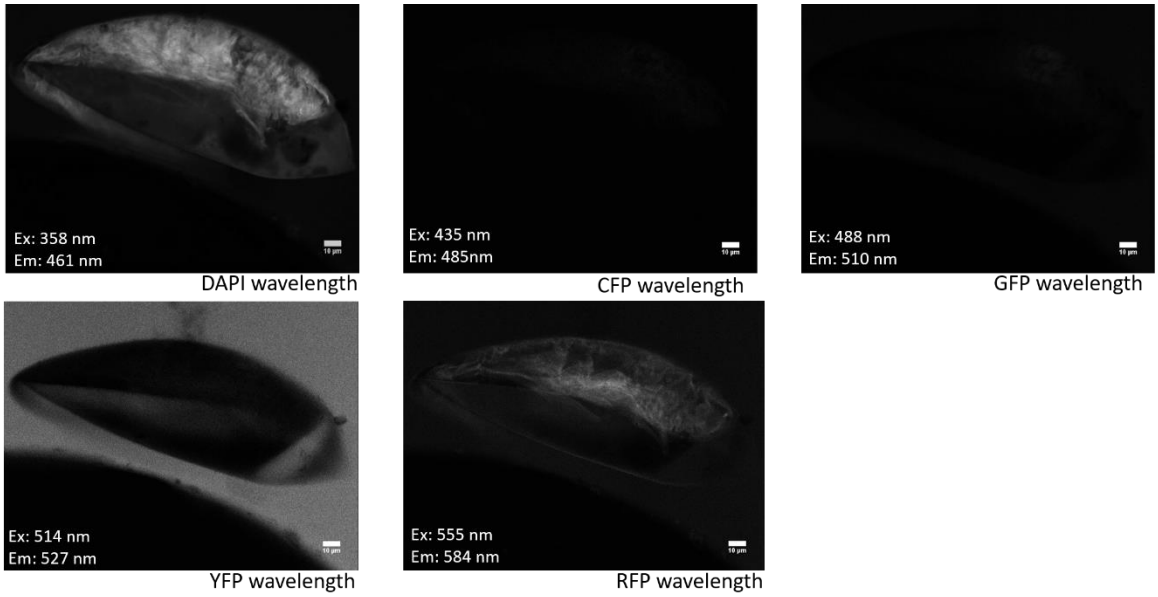

Figure 4. 3: autofluorescence in D type eggs throughout different (Em) emission and (Ex) excitation wavelengths.

The emission and excitation wavelengths correspond with the popular fluorophores; DAPI, cyan fluorescent protein (CFP), green fluorescent protein (GFP), yellow fluorescent protein (YFP) and red fluorescent protein (RFP). Scale bar represents 10 µm.

4.2.3 Autofluorescence is present at specific wavelengths in pre-oviposited eggs

There are three distinct morphological egg developmental stages in pre-oviposited eggs; A, B/C and D (see Section 1.2.1). I tested pre-oviposited eggs for autofluorescence signals at the DAPI wavelength, without DAPI staining (protocol from Section 2.2.7). Eggs were collected from non-virgin *B. tabaci* MED as described in Section 2.2.4.

Data is shown in Figure 4.3. Images with the suffix 1 in the upper right corners of images shown in Figure 4.3 denote brightfield images, and those with the suffix 2 denote confocal images (excitation 358nm; emission 461nm). Suffixes with 'A' denote A-stage eggs, 'B' B/C-stage eggs and 'C' D-type eggs. Fluorescence 'brioche' patterns, so-called because of the pattern showing aerated holes like a brioche pastry, were seen in the confocal images of D-type eggs and lightly in the B/C eggs. The 'brioche' pattern is probably derived from autofluorescence. The 'brioche' pattern was typical for the D eggs, and autofluorescence is likely to be extracellular material, such as chitin or something associated with chitin. Chitin is known to autofluorescence in insects. As it looks extracellular, cellularisation may have occurred stage at stages B/C and D. There is no autofluorescence in the A egg; therefore, DAPI used at this stage will not be masked by autofluorescence.

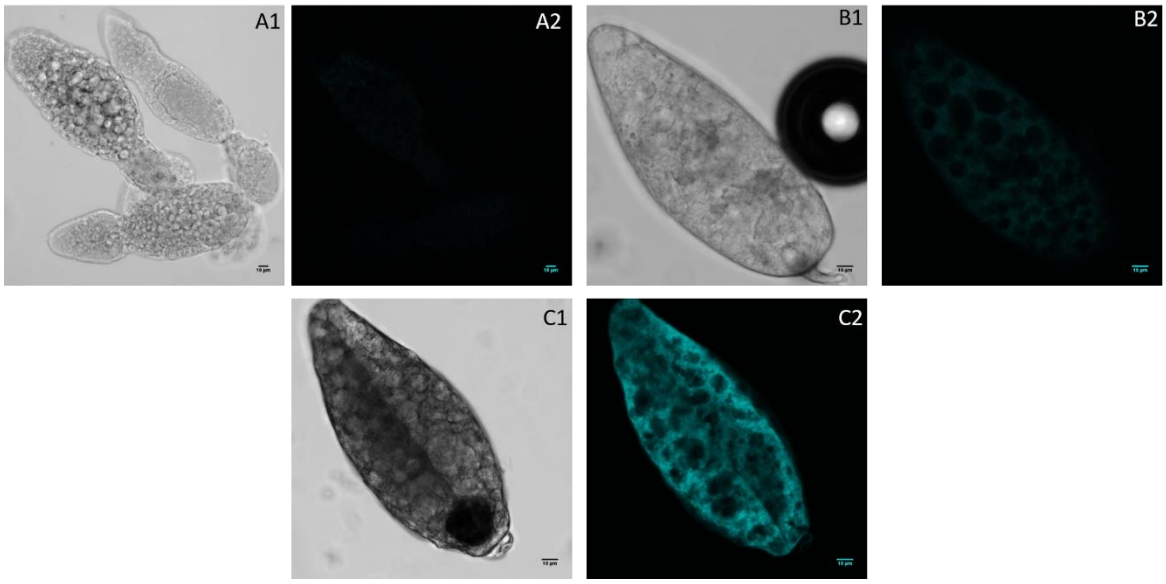

Figure 4. 3: Confocal images of A, B/C and D type eggs from potentially non-virgin Bemisia tabaci MED females.

Images with the 1 suffix are brightfield images. Images with 2 as the suffix are confocal images taken at wavelength 405 nm. This is the DAPI protocol without any DAPI. All images have a scale bar in the right bottom corner of 10 µm. A images are A type eggs, B are B/C type eggs and C is a D type egg. Scale bar represents 10 µm. Please refer to the electronic copy of thesis for pictures, if printed the contrast and quality will be poor.

4.2.4 Optimisation of the confocal microscopy protocol with DAPI and PI

DAPI is a fluorophore that binds to nuclei (see Figure 4.1). The DAPI protocol used on *D. melanogaster* in Section 4.2.1 must be optimised for *B. tabaci* MED. The following sections contain the optimisation process for *B. tabaci*.

4.2.4.1 DAPI protocol with D type eggs

There are three distinct morphological egg developmental stages; A, B/C and D (see Section 1.2.1). D eggs are numerous and easy to distinguish compared to the other stages. Therefore, D-type eggs will be used to optimise DAPI staining methods. The eggs were subjected to the protocol described in Section 2.2.7; without bleach. Figure 4.4 shows the DAPI stained D stage eggs; A pictures are brightfield images, and the B pictures are the confocal image (excitation 358nm; emission 461nm). A1 and B1 are stained with DAPI, while A2 and B2 underwent the same protocol without DAPI. Unlike Figure 4.1, there are no discrete circular objects representing cell nuclei; instead, there is the fluorescence 'brioche' pattern. The autofluorescence is likely to be extracellular material, such as chitin or compounds associated with chitin. Chitin is known to autofluorescence in insects. As it looks extracellular; cellularisation may have occurred at this stage.

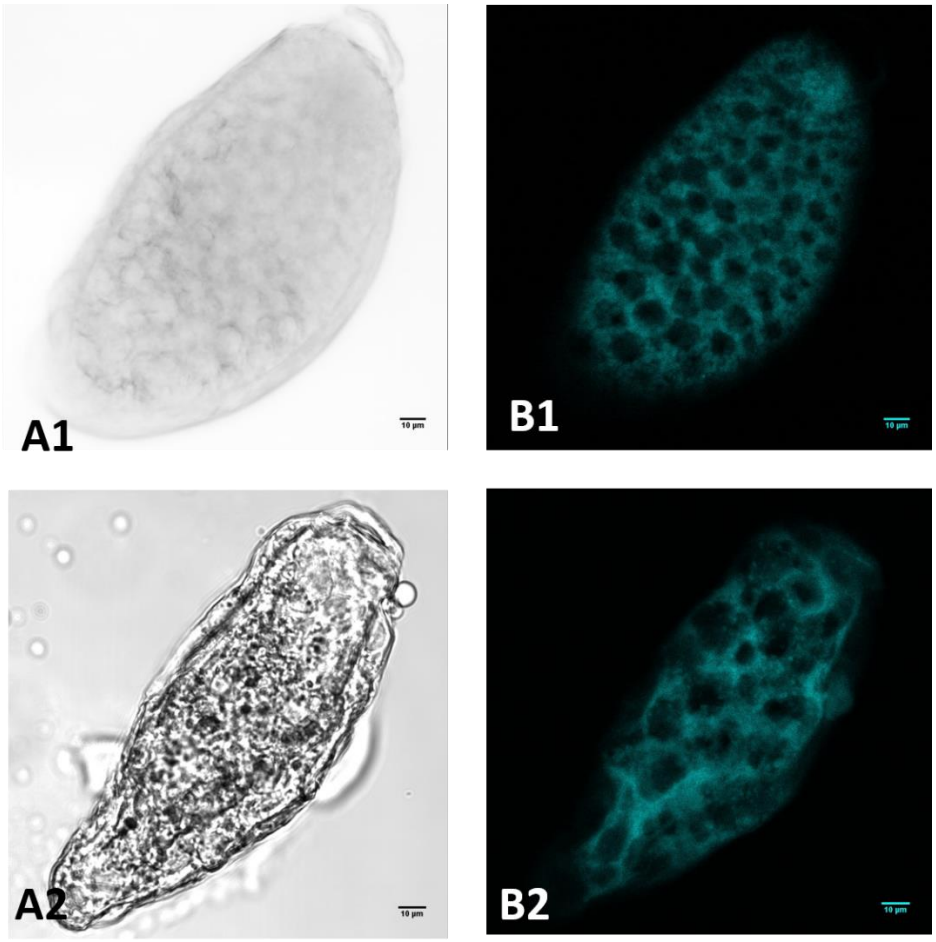

Figure 4. 4: The original DAPI protocol on D type eggs collected from the female abdomens of* B. tabaci *MED

The A pictures are brightfield images, and the B pictures are the confocal image (excitation 358 nm; emission 461 nm) falsely coloured with blue. A1 and B1 are stained with DAPI, while A2 and B2 underwent the same protocol without DAPI. The scale bar is 10 µm. Please refer to the electronic copy of thesis for pictures, if printed the contrast and quality will be poor.

4.2.4.2 Propidium Iodide protocol with D type eggs

In Figure 4.4, the autofluorescence seems to be occurring in the DAPI wavelength, whereas the eggs showed lower autofluorescence at the red fluorescent protein (RFP) wavelength, i.e. excitation at 555 nm and emission at 585 nm (Figure 4.3). With excitation at 535 nm and emission at 617 nm, propidium iodide (PI) falls close to the wavelength of RFP. Therefore, the PI stain was used to look for nuclei in eggs using the protocol described in Section 2.2.6.

Figure 4.5 shows PI-stained eggs. Images labelled with A1 and A2 are the controls, i.e. no PI was added to the mounting medium. These controls show that there still some auto-fluorescence occurring, and this may be explained by the requirement for a higher exposure to the emitting light of the confocal to see the fluorescence of PI compared to RFP.

The images in C2 of Figure 4.5 revealed a bright circular object that is likely to be the bacteriocyte, which is filled with symbiotic bacteria, of whitefly embryos, indicating that the PI was able to penetrate deeply into the egg. However, it was difficult to dissect other structures within the eggs with the PI-staining protocol, which was also long and labour-intensive leading to loss of many samples. The DAPI-staining protocol is easier to complete, can be done at higher throughput with more eggs, and enables easier nuclei counts. Consequently, the DAPI protocol was used for the research described in the rest of this chapter (Section 2.2.7).

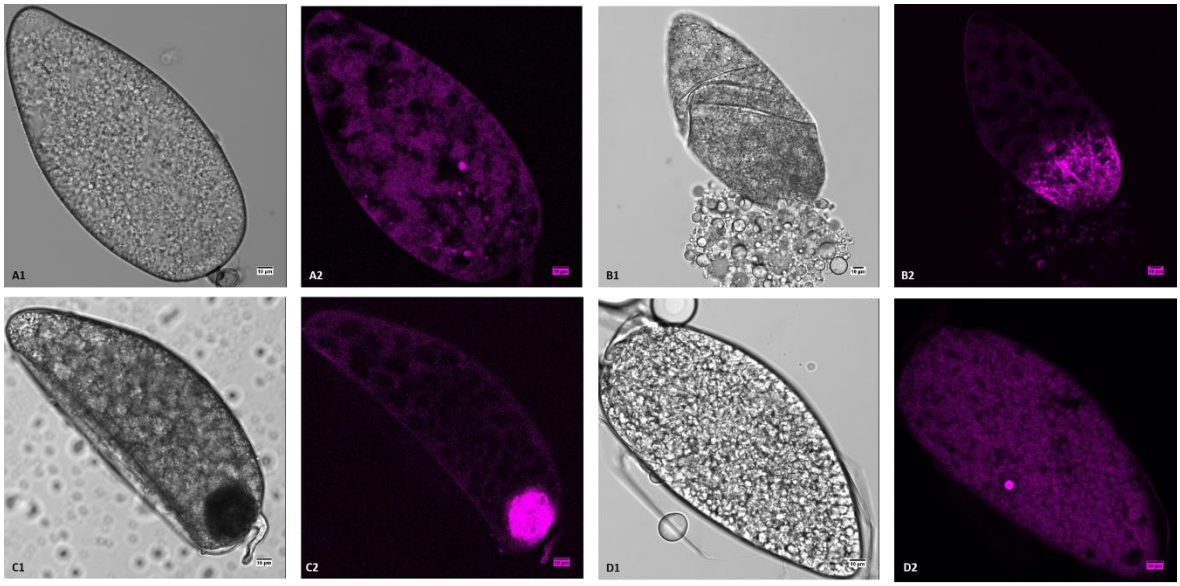

Figure 4. 5 **B. tabaci** ***MED D eggs, collected from the abdomen of a female adult.***
Whitefly type D eggs stained with PI. All the pictures labelled with suffix 1 are brightfield images, and those labelled with suffix 2 are confocal images taken at wavelength 561 nm. Eggs shown in A1 and A2 are controls that did not receive PI staining, and eggs shown in B1, B2, C1, C2 and D1, and D2 received PI staining. **The scale bar is 10 µm. Please refer to the electronic copy of thesis for pictures, if printed the contrast and quality will be poor.**

4.2.5 Male and female nuclear divisions of pre-oviposited eggs are similar

The DAPI staining protocol was best for counting nuclei in whitefly eggs and was henceforth used to also analyse nuclei in pre-oviposited eggs (A, B/C and D eggs). Mated females produce a mixture of males and females, whereas non-mated (virgin) females produce only males. Mated and non-mated females were collected (Section 2.2.3 and Section 2.2.1), dissected (Section 2.2.4), and the eggs categorised into type A, B/C and D developmental stages based on their morphology. The type A eggs showed positive staining of nuclei and were studied further; these eggs were fixated (Section 2.2.7), imaged (Section 2.2.8) and analysed (Section 2.2.9). Results are shown in Figure 4.6, and DAPI controls for type A eggs were shown previously in this chapter (Figure 4.3).

Figure 4.6 presents the results of the DAPI stained type A eggs obtained from mated (non-virgin) female adults and the eggs may contain female or male embryos. The confocal images show clear circular structures stained with DAPI and that look like cell nuclei, which were counted to assess the developmental stage of the embryo in the eggs.

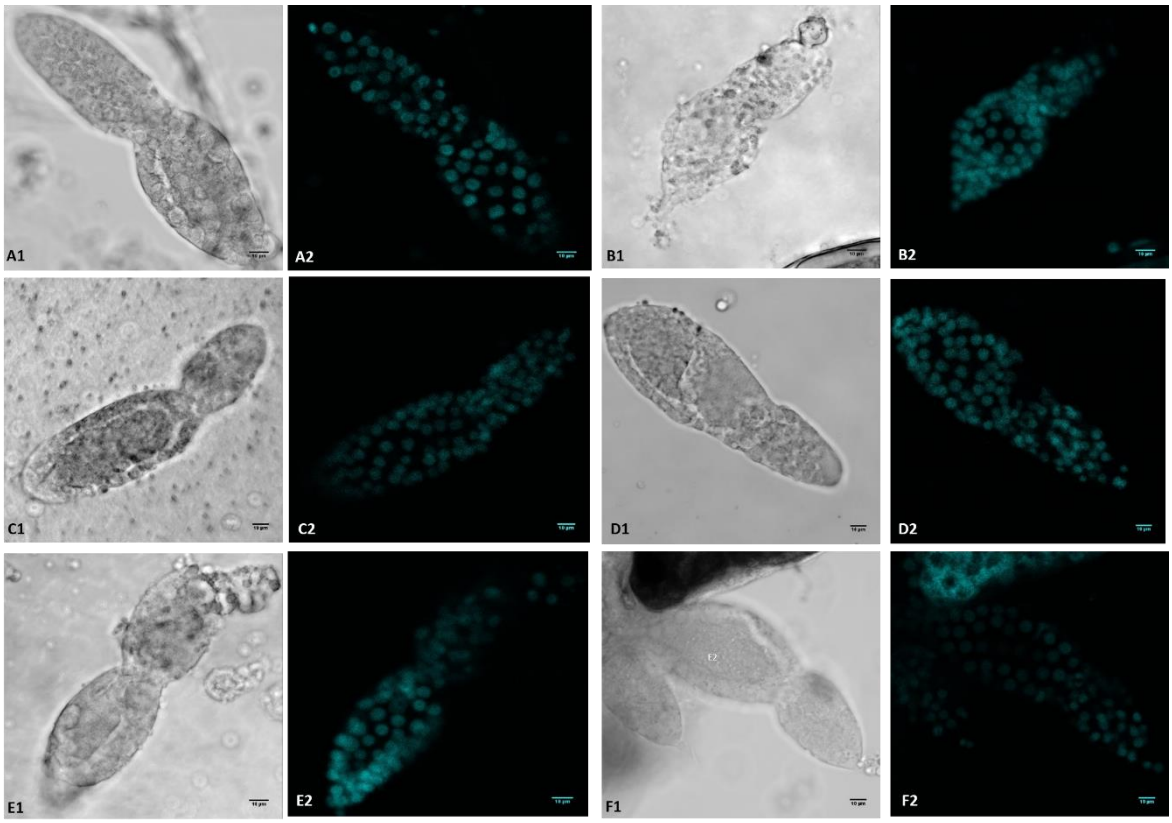

Figure 4. 6: Confocal images of A type eggs from potentially non-virgin* B. tabaci *MED females.
Images with the 1 suffix are brightfield images. Images with 2 as the suffix are confocal images taken at excitation 358 nm and emission 461. The eggs have been stained with DAPI, and the nuclei are visible as discrete circular objects. 2. All images have a scale bar in the right bottom corner of 10 µm. Please refer to the electronic copy of thesis for pictures, if printed the contrast and quality will be poor.

Figure 4.7 shows DAPI-stained type A eggs from non-mated (virgin) females' adults, and therefore the eggs will only contain male embryos. Like images shown in Figure 4.6, images of Figure 4.7 show clear DAPI-stained circular objects that are the cell nuclei, which were counted to assess the developmental stage of the embryo in the eggs.

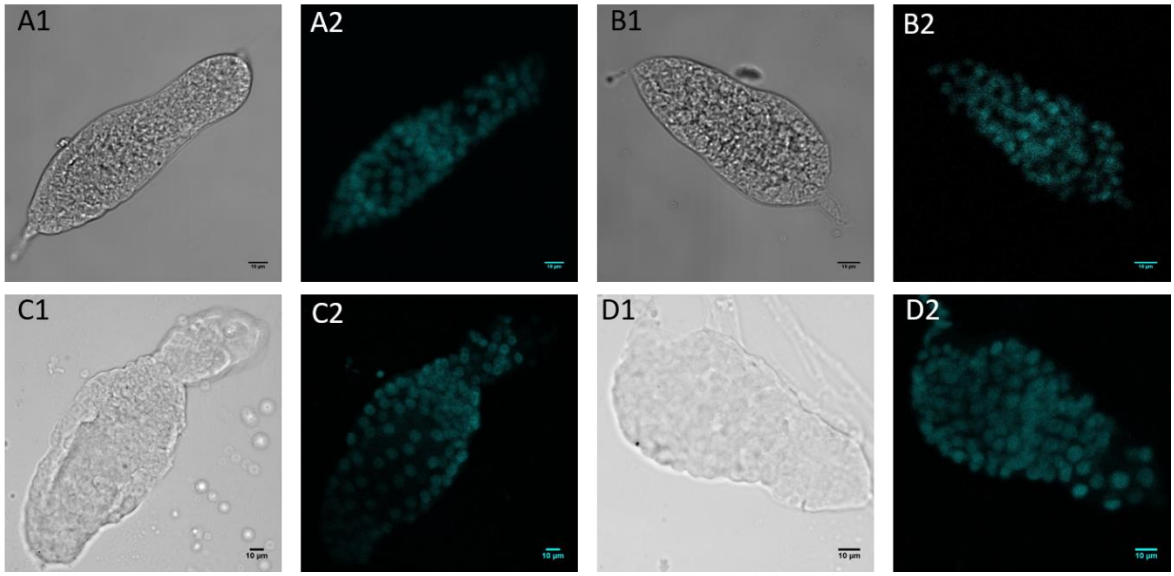

Figure 4. 7: Confocal images of A stage eggs from virgin* B. tabaci *MED females.
Images with the 1 suffix are brightfield images. Images with 2 as the suffix are confocal images taken at excitation 358 nm and emission 461 nm. The eggs have been stained with DAPI, and the nuclei are visible as discrete circular objects. 2. All images have a scale bar in the right bottom corner of 10 µm. Please refer to the electronic copy of thesis for pictures, if printed the contrast and quality will be poor.

Nuclei counting of the confocal images of Figures 4.6 and 4.7 was done with the maximum projection function of FIJI. In this function, the Z stacks are flattened to generate a 2D image. The produced 2D image shows all the nuclei in a sample. Figure 4.8 is an example of a 2D image obtained from the maximum projection function with the image labelled A2 in Figure 4.7.

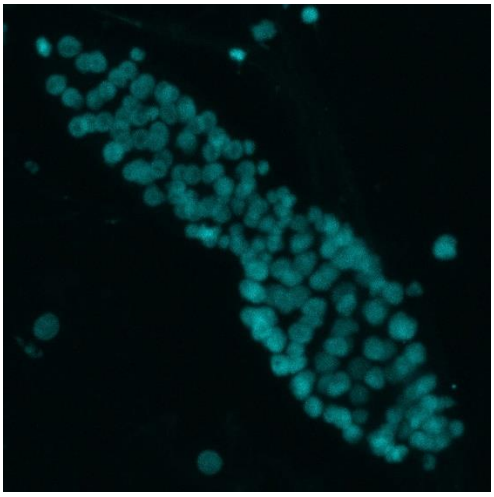

Figure 4. 8: Confocal images of A egg from potentially non-virgin* B. tabaci *MED females.
A 2D image generated from a Z-stack image of A2 in Figure 4.7 using the FIJI maximum projection analysis (National Institutes of Health, USA; v1.52c). The blue-stained circular objects within the image are DAPI-stained nuclei. Please refer to the electronic copy of thesis for pictures, if printed the contrast and quality will be poor.

All eggs shown in Figure 4.6 and 4.7 underwent FIJI maximum projection analyses, and the nuclei seen in each 2D image were counted. The results of the nuclei counts are shown in Table 4.1. The mean number of nuclei of type A eggs of non-virgin females was 118.3 (n=6), and the nuclear division cycle was therefore between 7-8. The mean number of nuclei of type A eggs of virgin females was 95.8 (n=5), and a nuclear division cycle of 7-8. Therefore, both in virgin and non-virgin females the embryo development of type A eggs had already arrived at stage 7-8, which is the pre-blastoderm stage just prior to the start of the blastoderm developmental programme (cleavage cycle 8-9). The type A eggs are in the abdomen of females, whitefly eggs apparently already started embryogenesis before oviposition, unlike *D. melanogaster* eggs where embryogenesis starts after oviposition. The embryo development does not show major differences in eggs from virgin and non-virgin females (Welch t-test score $p>0.05$), indicating that whitefly male and female embryos develop similarly.

	Sample type	Number of nuclei	Cleavage cycle
Non-virgin females	Female or male eggs	113	7
	Female or male eggs	87	7
	Female or male eggs	148	8
	Female or male eggs	124	8
	Female or male eggs	122	7
	Female or male eggs	116	7
	Mean of female or male eggs	118.3	7-8
Virgin females	Male eggs	86	7
	Male eggs	66	7
	Male eggs	145	8
	Male eggs	100	7
	Male eggs	82	7
	Mean of male eggs	95.8	7-8

Table 4. 1: Nuclei number for each sample from Figure 4.6 and 4.7 calculated after maximum projection. **Each type of sample underwent the welch t-test to compare whether there is a difference in development time of male eggs compared to potentially female eggs. Each count of nuclei represents a new sample.**

Each sample, in Figure 4.6 and 4.7, underwent Z stack confocal microscopy. In Figure 4.8, maximum projection was used to flatten a 3D image into 2D; however, this loses important morphological detail. The Z stack allows analysis on slices throughout the eggs. The egg sample analysed in Figure 4.6 (A2) had 49 stacks/ slices. Figure 4.7 contains different slices of the egg, from 1 to 49 so the interior can be shown in more detail.

In *D. melanogaster,* before blastoderm formation, the nuclei start to migrate from the centre of the egg to the periphery. The blastoderm is a single layer around the periphery of the egg. The A2 sample in Figure 4.7 has a hollow interior, and this suggests that the migration stage has occurred in the A stage egg. The nuclei are at 7-8 nuclear division stage (the division before blastoderm formation); however, the morphological analysis of Figure 4.7 reveals that the blastoderm has already formed.

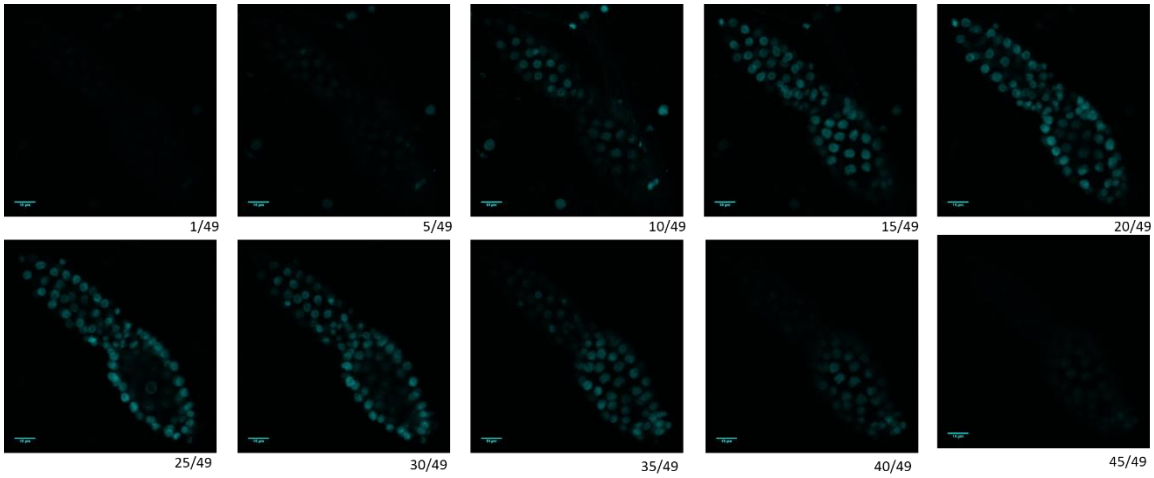

Figure 4. 7: The Z stacks of A2 in Figure 4.6 at different Z stack points.
Taken at excitation 358 nm and emission 461 nm. Please refer to the electronic copy of thesis for pictures, if printed the contrast and quality will be poor.

4.2.6 Nuclei cannot be counted in eggs after the A stage

Section 4.2.6 uses the protocol in Section 2.2.7 to count nuclei in B/C eggs. Figure 4.3 shows less autofluorescence in the B/C eggs than the D eggs. Therefore, nuclei may be bright enough to be counted in B/C eggs. Non-virgin adults were collected (Section 2.2.3 and Section 2.2.1), dissected (Section 2.2.4), and the eggs categorised into the different morphological stages. For this section, B/C eggs were fixated (Section 2.27), imaged (Section 2.2.8) and analysed (Section 2.29). The image in Figure 4.8 had the same exposure as Figure 4.6; therefore, Figure 4.3 is the control for 4.6. All B/C eggs show signs of the 'brioche' effect and no nuclei are seen outside the autofluorescence, as the autofluorescence may be masking the signal (n=20+).

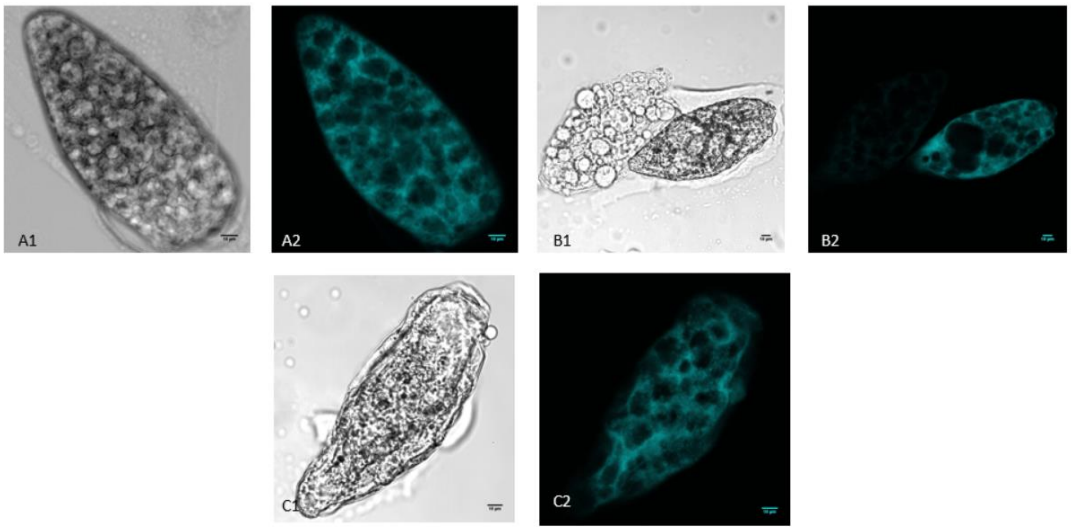

Figure 4. 8: Confocal images of B/C type eggs from potentially non-virgin* B. tabaci *MED females.
Images with the 1 suffix are brightfield images. Images with 2 as the suffix are confocal images taken at excitation wavelength 358 nm and emission 461 nm. The eggs have been stained with DAPI, and the nuclei have been successfully stained as seen in images with suffix 2. All images have a scale bar in the right bottom corner of 10 µm. Please refer to the electronic copy of thesis for pictures, if printed the contrast and quality will be poor.

4.3 Discussion

Embryogenesis in *D. melanogaster* is well understood (Campos-Ortega and Hartenstein, 1997); however, knowledge of embryogenesis in *B. tabaci* is lacking. Early embryo development knowledge is vital for germline transformation in genetic control methods. For germline transformation to be successful, eggs need to be transformed before the blastoderm formation. This chapter aims to identify when blastoderm transformation takes place in *B. tabaci*, and therefore determine whether traditional approaches for insect germline transformation are feasible for *B. tabaci*.

4.3.1 Embryogenesis occurs before pre-oviposition

D. melanogaster embryogenesis occurs after the eggs are laid, and this happens in other insects as well (Campos-Ortega J.A. and V., 1997). One would think that this would be the same for *B. tabaci*; however, this chapter has shown embryogenesis occurring in eggs dissected from the abdomen (pre-oviposited eggs), Figure 4.6 and 4.7.

4.3.2 Blastoderm formation occurs in A stage egg

Figure 4.6 and 4.7 (A stage eggs in non-virgin and virgin female abdomen) already have nuclei to count. Table 4.1 contain many samples at the A stage egg. The nuclei count analysis reveal that the A-stage eggs are at cell cycle 7-8. The blastoderm formation in *D. melanogaster* occurs between cell cycle 8-9. Before blastoderm formation in *D. melanogaster,* the nuclei start to migrate to the outer edges of the egg. If *B. tabaci* follow the same embryogenesis pattern as *D. melanogaster,* then the average cell cycle stage (7-8) would indicate the migration of the nuclei.

Blastoderm formation is when the nuclei start to form a single layer of nuclei around the outer edge of the egg. The morphology of the A egg in Figure 4.9 (z stacks of the A-sample eggs), shows a hollow interior of the eggs, with a layer of nuclei already forming the around the edge of the eggs. If looking at just the morphology, blastoderm formation has already occurred in A eggs.

For germline transformation, the aim is to microinject before the blastoderm formation. If this chapter only looked at the nuclei count data in Table 4.1, then the A-sample eggs would be pre-blastoderm. However, the morphological analysis reveals that blastoderm formation has already occurred in the A-sample egg. Based on the nuclei counting alone, microinjection would have to occur in A stage egg. However, injection of pre-oviposited eggs is likely to be unfeasible, and an alternative system would need to be developed/ explored.

4.3.3 Male and female eggs may have similar developmental rates

One of the questions answered in this chapter was whether male and female eggs develop at different rates. A difference in development rates may occur as the female eggs are diploid and male eggs are haploid (Blackman and Cahill, 1998). Table 4.1 shows the nuclei number in non-virgin females (sex of eggs unknown) and virgin females (male eggs), there is no statistical difference between these two samples. Therefore, based on this data, the male and female eggs develop at the same rate in early embryogenesis.

However, the problem with this experiment is that it is challenging to guarantee female-only eggs. Potentially non-virgin female has not been confirmed to have successful copulation. The courtship ritual of *B. tabaci* is incredibly elaborate, and the success rate is low (Li et al., 1989). Due to time constraints it was not possible to constantly monitor the insects over the 3 days to ensure successful copulation. After three days it is presumed that successful mating had occurred; however, this was not guaranteed. Table 4.1 may be comparing male eggs against male (rather than females).

4.4 Conclusion

Pre-oviposited eggs have different morphological developmental stages; A, B/C and D. A-stage egg in *B. tabaci* are at either pre-blastoderm or early blastoderm formation stages. However, in B/C and D type egg it is difficult to determine what stage these are due to the auto-fluorescence. However, extracellular material is present (evidenced by the autofluorescence), therefore cellularization has occurred in B/C and D sample eggs.

Chapter 5 further investigates the embryogenesis stages in *B. tabaci* by single-embryo RNA-seq data. Conserved embryogenesis genes can further pinpoint the developmental stages in *B. tabaci.* Development of genetic control methods would be very difficult in *B. tabaci,* due to embryogenesis stages already occurring at the A sample egg. If post-blastoderm eggs are transformed, only somatic transformation will occur. Through personal communication, (International Whitefly Symposium, 2018- details of laboratory will be anonymised as the research has not been published) laboratories have tried microinjection of post-oviposition eggs have only achieved somatic transformation. Also, 1000 eggs have been injected and only 3 had successful somatic transformation.

Chapter 5: The discovery of transcripts of early embryogenesis and sex determination genes in four stages of whitefly eggs

5.1 Introduction

Transcriptomic analysis is an approach to identify differentially regulated transcripts of an organism over developmental time and in response to the environment. RNA-seq is one of the many available methods to analyse the transcriptome of an organism. Others are quantitative reverse transcriptase PCR, northern blotting plus hybridisation and microarrays. Many of these methods assess the quantity of a transcript at a single time point and can assess transcript changes over a period. However, samples will have to be analysed with the same method over a time period or at the same time upon exposure to different treatments (Wang et al., 2009).

In this Chapter, samples have been analysed over time to assess transcript abundance of genes involved in determining the sex of the organisms, and these genes are expressed in early stages of embryogenesis. Hence, for this chapter different development stages of *B. tabaci* eggs were collected and RNA-seq was performed.

Throughout this thesis, the aim was to obtain more knowledge about embryogenesis and sex determination genes of *B. tabaci* MED. Chapter 3 was focused on the identification of sex determination genes/ proteins (SDGs/ SDPs) in 11 hemipteran genomes which are publicly available, including that of *B. tabaci* MED. Chapter 4 focused on the characterisation of the embryogenesis stages in *B. tabaci* MED A, B/C and D type eggs. Nuclei counting in Chapter 4 concluded that the A-sample eggs are at the migration/ blastoderm stage. Given that SDGs are expressed, and alternatively spliced between the sexes during early embryogenesis, this chapter will be focused on the identification SDG transcripts in *B. tabaci* MED A, B/C and D type eggs.

This chapter consists of two parts. The first part describes how the single embryo RNA-seq data were analysed, with a focus on the bioinformatic pipeline used, the quality control steps taken and assessing the likelihood of finding SDG transcripts. The second part of the chapter focussed on identification of transcripts, and (sex-specific) splice variants, derived from genes which encode for DM containing proteins and other 'key' group SDGs in the *B. tabaci* MED A, B/C and D type eggs.

A gene located at a specific locus can produce different mRNAs (called isoforms), which may have different functions. Some SDGs have sex-specific isoforms; for example, *B. mori*, *A. mellifera* and *D. melanogaster Dsx* isoforms have sex-specific exon retentions (Xu et al., 2017). Therefore, the hypothesis is that some SDGs of *B. tabaci* MED may also produce sex-specific splice variants/isoforms.

To detect sex-specific splicing of SDGs in *B. tabaci* MED eggs, the sex of the embryos in the eggs will have to be assessed. Female and male embryos of whiteflies do not have apparent morphological differences. Fortunately, *B. tabaci* is a haplodiploid organism; i.e. males are haploid and develop from unfertilised eggs and females are diploid; developing from fertilised eggs. Given that male embryos are haploid, their transcripts are derived from a single copy of a given gene and will have no heterozygous loci in a single sample SNP analysis, unlike female embryos. Therefore, we can use SNP analyses to assess the levels of heterozygosity of embryos and assign the biological samples as male or female.

The expression time of conserved embryonic genes, such as those that are specifically expressed in the germ cells, can serve as markers for the developmental stage of the embryo. Identification of such transcripts will also enable confirmation of the developmental stage of the embryo, as determined in the previous chapter of this thesis. A gene that is expressed in *D. melanogaster* embryos is *Vasa*. This gene encodes for a DEAD-Box RNA helicase family protein and is essential for embryonic patterning, germ cell function and pole plasm assembly (Papathanos et al., 2009; Van Doren et al., 1998). *Vasa* expression starts at the time of pole cell formation, around nuclear divisions 8-10 (Lasko and Ashburner, 1988). Pole cells form germ-line cells, and their development start at about the time of the blastoderm stage (Wotton et al., 2014). Therefore, if transcripts derived from the *Vasa* gene can be identified in whitefly egg transcriptomes, these embryos are likely to be at the blastoderm stage, which is also the time where some of the SDGs are expressed. Moreover, to obtain a germ-line transformation of whitefly, eggs will have to be transformed (via microinjection) at the blastoderm stage. Hence, the research described in this chapter serves two goals; one is to identify temporal expression patterns of SDGs and the other is to assess when whitefly eggs are at the blastoderm stage (confirming research in the previous chapter).

Results described in the first part of this chapter, show that I was successful in developing a bioinformatics pipeline for the RNA-seq data to identify male and female embryos. I then identified the *Vasa* gene transcripts in the egg transcriptomes. The *Vasa* gene was highly expressed in the whitefly A-stage eggs, indicating that these eggs are at the blastoderm developmental stage. This result agrees with results in the previous chapter. Whitefly A-type eggs will have to be transformed to obtain stable transgenic whiteflies. There were no sex-specific splice variants identified for the whitefly *Vasa* gene.

In the second part of the results, I investigated if sex-specific isoforms can be identified for genes encoding candidate *B. tabaci* MED DM proteins and 'key' group SDGs (identified in Chapter 3) in

early embryos. The files produced from the bioinformatic pipeline, described in the first part of this results chapter, were aligned to the gene model by a tool called Integrative Genomics Viewer (IGV) (Thorvaldsdottir et al., 2013; Robinson et al., 2011). I found potential candidates for the self-limiting system.

5.2 Results part 1

5.2.1 Collection of samples

Whitefly females were reared with and without the presence of males. The non-mated females are guaranteed to produce only haploid male eggs, whereas mated females may carry haploid male and diploid female eggs depending whether the eggs were fertilised. I dissected A, B/C and D type eggs from female abdomens as described in the protocol of the Materials and Methods chapter Section 2.3.1. Michael Giolai from the Hogenhout laboratory conducted the RNA extractions and RNA-seq libraries using the methods described in (Picelli et al., 2014), and I submitted the samples for sequencing. Table 5.1 provides a list of each egg dissected/collected, and whether the eggs were derived from virgin or non-virgin females (V or N) and the egg type (A, B/C or D). For each egg type, multiple replicate samples were obtained.

Sample ID	Collected from Virgin or Non-virgin adults	Type of egg	Sex of egg
VA3	Virgin	A	Male
VA5	Virgin	A	Male
VA6	Virgin	A	Male
VA7	Virgin	A	Male
VBC1	Virgin	B/C	Male
VBC3	Virgin	B/C	Male
VBC4	Virgin	B/C	Male
VBC6	Virgin	B/C	Male
VD2	Virgin	D	Male
VD3	Virgin	D	Male
VD4	Virgin	D	Male
VD5	Virgin	D	Male
NA1	Non	A	Unknown
NA2	Non	A	Unknown
NA3	Non	A	Unknown
NA4	Non	A	Unknown
NA5	Non	A	Unknown
NA8	Non	A	Unknown
NBC1	Non	B/C	Unknown
NBC2	Non	B/C	Unknown
NBC3	Non	B/C	Unknown
NBC5	Non	B/C	Unknown
NBC6	Non	B/C	Unknown
NBC7	Non	B/C	Unknown
NBC9	Non	B/C	Unknown
NBC11	Non	B/C	Unknown
ND1	Non	D	Unknown
ND2	Non	D	Unknown
ND3	Non	D	Unknown
ND6	Non	D	Unknown
ND8	Non	D	Unknown
ND11	Non	D	Unknown

Table 5. 1: Several stages of eggs collected from virgin or non-virgin (non) females of* B. tabaci *MED
The samples were processed for RNA extraction and paired-end RNA-sequencing on NextSeq500. Abbreviations used for 'Sample' column: V, Virgin female; N, Non-virgin/mated female; A, BC, D - A, B/C and D type eggs, respectively; numbers 1 through 11, biological replicates. See text above for further explanation.

5.2.2 Identification of sex in individual embryo RNA-seq samples

RNA-seq data of the samples listed in Table 5.1 were analysed to determine the sex of the embryos, as collected using the methods in Sections 2.2.2 and 2.2.3. Briefly, RNA-seq data of the individual eggs were aligned to the *B. tabaci* MED reference genome. The GATK haplotype (McKenna et al., 2010) caller was used to produce a variant call format (vcf) file, and the genotype (GT) values collected. The GT number indicates whether the allele is heterozygous or homozygous of a given sample, and the data are expressed as percentages (Table 5.2).

The male samples derived from the virgin/non-mated females had 22.66-25.11 % heterozygosity levels. This level of heterozygosity in the male eggs may be derived from; maternal egg tissues beyond the embryo, copy number variation (reads derived from members of a family with genes that have high identity in sequence may collapse into a single transcript), mismatch alignments, artefacts from adaptor contamination, and suboptimal parameters of the GATK programme (which was designed for analyses of human samples). The heterozygosity values from the eggs from non-virgin females fell into two distinct groups. Twelve had the same value as most male eggs (22.66%), and 9 had higher values ranging from 27% to 84%. Given that 9 eggs had higher than the 22.66-25.11 % heterozygosity levels of male embryos, these eggs were considered to harbour female embryos. A Z-test was conducted on all unknown samples against the known male samples and underwent Bonferroni correction, any samples that were statistically different ($p=<0.001$) has a star next to them in Table 5.2 and are classed as female. Male and female embryos were found for the A and B/C type eggs, but unfortunately, only male and no female type D eggs were found.

Parent	Sample ID	Sex	Heterozygosity %	Count of SNPs
VIRGIN	VA3	Male	22.66	86963
VIRGIN	VA5	Male	22.66	86963
VIRGIN	VA6	Male	22.66	86963
VIRGIN	VA7	Male	22.66	86963
VIRGIN	VBC6	Male	25.12	20040
VIRGIN	VBC1	Male	25.12	20040
VIRGIN	VBC3	Male	22.66	86963
VIRGIN	VBC4	Male	22.66	86963
VIRGIN	VD2	Male	22.66	86963
VIRGIN	VD3	Male	22.66	86963
VIRGIN	VD4	Male	22.66	86963
VIRGIN	VD5	Male	22.66	86963
POTENTIALLY NON-VIRGIN	NA1	Female *	35.71	173926
POTENTIALLY NON-VIRGIN	NA3	Female *	48.56	245677
POTENTIALLY NON-VIRGIN	NA2	Female *	45.68	260889
POTENTIALLY NON-VIRGIN	NA4	Female *	29.43	434815
POTENTIALLY NON-VIRGIN	NA5	Male	22.66	86963
POTENTIALLY NON-VIRGIN	NA8	Female *	36.45	521778
POTENTIALLY NON-VIRGIN	NBC1	Male	22.66	86963
POTENTIALLY NON-VIRGIN	NBC9	Female *	29.86	209373
POTENTIALLY NON-VIRGIN	NBC2	Male	22.66	86963
POTENTIALLY NON-VIRGIN	NBC3	Female *	84.12	64335
POTENTIALLY NON-VIRGIN	NBC4	Female *	27.93	87964
POTENTIALLY NON-VIRGIN	NBC5	Male	22.66	86963
POTENTIALLY NON-VIRGIN	NBC6	Male	22.66	86963
POTENTIALLY NON-VIRGIN	NBC7	Female *	39.00	100763
POTENTIALLY NON-VIRGIN	NBC11	Male	22.66	86963
POTENTIALLY NON-VIRGIN	NVD1	Male	22.66	86963
POTENTIALLY NON-VIRGIN	NVD11	Male	22.66	86963
POTENTIALLY NON-VIRGIN	NVD2	Male	22.66	86963
POTENTIALLY NON-VIRGIN	NVD3	Male	22.66	86963
POTENTIALLY NON-VIRGIN	NVD6	Male	22.66	86963
POTENTIALLY NON-VIRGIN	NVD8	Male	22.66	86963

Table 5. 2: Single nucleotide polymorphism (SNP) counts and heterozygosity levels of A, B/C and D type* B. tabaci *eggs leads to identification of male and female eggs.

Lanes with data from eggs with female embryos are highlighted in green. Abbreviations of sample IDs are the same as described in Table 5.1. * Z-score significantly different of <0.001

5.2.3 Genome-guided assembly of the *B. tabaci* MED RNA-seq data

5.2.3.1 Creation of files for isoform analysis

RNA-seq data consists of many short reads that must be assembled into transcript sequences. The *B. tabaci* MED published genome sequence was used to obtain a genome-guided transcriptome assembly for the *B. tabaci* MED eggs (using the pipeline described in the Material and Methods Section 2.3.4, Figure 2.3). Archana Singh (Swarbreck lab, EI, Norwich and member of the Hogenhout lab) helped to analyse the RNA-seq data and she also produced the expression level data (i.e. also referred to as count data) from Kallisto.

Reads from each egg sample were individually aligned to the gene models and assembled. The compositions and lengths of the assembled transcripts were then compared among the egg samples leading to the identification of potential isoform differences between male and female embryos. The expression levels of each transcript in each egg sample was determined by assessing the depths of unique reads to a given transcript.

Table 5.3 lists the results of the RNA-seq alignment to the *B. tabaci* MED genome. Low read alignment levels may be due to the presence of bacterial symbionts. The majority of the low alignment rates come from the later developmental stages (C and D stage eggs). The primary bacterial endosymbionts of whiteflies are transferred from mother to egg in the later developmental stages, and therefore the low alignments rates may come from the endosymbionts. Alternatively, RNA-seq data may be derived from contaminants that were picked up during the dissection, RNA isolation and RNA-seq library steps. Given that the eggs are tiny and produce small amounts of RNAs, contamination levels can be high for some samples. Nonetheless, the alignments for at least three samples of type A, B/C and D egg male and female embryos were considered good enough quality for further analyses.

	ID	Aligned reads	overall alignment rate (%)
Male	VA3	8,282,687	62.12
	VA5	12,404,496	71.32
	VA6	13,883,877	64.76
	VA7	13,274,733	67.93
	NA5	5,610,128	31.26
	VBC6	13,346,724	61.44
	VBC1	7,897,292	54.27
	VBC3	17,161,673	75.24
	VBC4	12,461,608	67.17
	NBC1	11,086,901	9.55
	NBC2	7,407,112	11.38
	NBC5	7,222,204	41.55
	NBC6	16,502,038	65.79
	NBC11	9,579,771	65.61
	VD2	11,169,231	56.16
	VD3	8,508,291	62
	VD4	10,175,965	68.15
	VD5	13,743,964	76.52
	NVD1	4,796,022	23.72
	NVD11	6,766,150	38.49
	NVD2	8,667,576	33.27
	NVD3	7,930,419	39.75
	NVD6	11,763,266	41.13
	NVD8	3,731,829	28.7
Female	NA1	9,666,883	36.45
	NA3	10,616,647	69.62
	NA2	14,664,343	62.12
	NA4	11,098,275	65.51
	NA8	5,475,957	51.22
	NBC9	13,057,560	35.16
	NBC3	10,650,369	37.34
	NBC4	7,417,688	43
	NBC7	8,065,851	44.48

Table 5. 3: Statistics of B. tabaci MED egg RNA-seq trimmed read alignments to* B. tabaci *MED reference genome

The table is produced using Figure 2.3 in the Materials and methods. Abbreviations of sample IDs are the same as described in Table 5.1.

5.2.4 Quantification of gene expression levels

Quantification of gene expression is essential for identifying SDGs because SDGs are expressed in early embryogenesis. The gene expression quantity requires counting the number of reads mapped to each locus in transcriptome assembly. The pipeline used for gene expression quantification is illustrated in Figure 2.4. Upon assembly of the transcripts, the expression level of each transcript was determined by realigning the total RNA-seq reads to each of the transcripts. Data are shown in Table 5.4. The percentages of RNA-seq reads aligned to the genome (Table 5.3), and transcriptome (Table 5.4) are comparable, generating confidence in the data analyses strategies. In general, the read alignment percentages are higher for the genome alignment than for the transcriptome.

	ID	Aligned reads	overall alignment rate (%)
Male	VA3	8,282,687	58.88
	VA5	12,404,496	67.30
	VA6	13,883,877	60.69
	VA7	13,274,733	63.35
	NA5	5,610,128	23.15
	VBC6	13,346,724	57.86
	VBC1	7,897,292	46.42
	VBC3	17,161,673	71.08
	VBC4	12,461,608	52.48
	NBC1	11,086,901	6.77
	NBC2	7,407,112	9.95
	NBC5	7,222,204	33.20
	NBC6	16,502,038	59.44
	NBC11	9,579,771	53.69
	VD2	11,169,231	47.13
	VD3	8,508,291	56.36
	VD4	10,175,965	62.28
	VD5	13,743,964	71.14
	NVD1	4,796,022	17.89
	NVD11	6,766,150	21.88
	NVD2	8,667,576	18.34
	NVD3	7,930,419	26.14
	NVD6	11,763,266	29.36
	NVD8	3,731,829	23.61
Female	NA1	9,666,883	27.27
	NA3	10,616,647	57.01
	NA2	14,664,343	50.38
	NA4	11,098,275	55.28
	NA8	5,475,957	37.43
	NBC9	13,057,560	25.13
	NBC3	10,650,369	28.66
	NBC4	7,417,688	35.75
	NBC7	8,065,851	29.23

Table 5. 4: Statistics of B. tabaci egg RNA-seq trimmed alignments to the* B. tabaci *MED transcriptome using kallisto

Abbreviations of sample IDs are the same as described in Table 5.1.

5.2.5 RNA-seq data for *B. tabaci* MED adult males and females

RNA-seq data generated for the eggs were compared to *B. tabaci* MED adults. Raw reads of the RNA-seq data published by Liu *et al* (2019) were downloaded and analysed via the bioinformatics pipeline (Material and Methods, Section 2.3 and 2.4). Between 78 to 81 % of the reads derived from the male and female adults aligned to the *B. tabaci* MED genome (Table 5.3). The RNA-seq reads from adult whiteflies align better to the whitefly genome than the eggs do. The adults may have a higher alignment because the abundance of endosymbiotic bacteria in the adults is lower than in the eggs. Therefore, it is easier to obtain sufficient high-quality RNA from adults than from eggs.

ID	Number of processed reads	Read alignment rates (%)
QF4	19096472	78.96
QF5	21943081	80.29
QF6	20320501	81.56
QM5	22564193	78.37
QM6	16673914	80.07
QM7	16545475	78.96

Table 5. 5: Statistics of alignment of raw RNA-seq reads from male and female* B. tabaci *MED adult samples to the B. tabaci genome.

Abbreviations in ID column: Q, MED (Q) biotype of *B. tabaci;* M, male; F, female; 4-7, independent biological replicates.

5.2.6 Embryonic conserved gene (*Vasa*) are present in all samples

The embryonic and adult RNA-seq data are needed to provide files for isoform and gene expression analysis. To test the data, a conserved embryonic gene (*Vasa*) went through isoform and gene expression analysis. The analysis will also provide secondary information of when germline cells form, as *D. melanogaster* germline cells express *Vasa*. The results of RBBH analysis between the *D. melanogaster Vasa* (from NCBI: CAA31405.1) and *B. tabaci* MED protein sequences (chapter 3, Section 3.2.2) ate shown in table 5.6. A *B. tabaci* VASA orthologue was found.

QUERY	SUBJECT	E-VALUE	QUERY COVERAGE	IDENTITIY (%)	SIMILARITY (%)
DM_VASA	BTA009465.2	1.00E-180	69	54.68	72.55

Table 5. 6: The blastp results of full-length* D. melanogaster *DSX (query) against the* B. tabaci *MED genome (subject)

The VASA protein sequence of *D. melanogaster* contains a Dead-like helicases superfamily (DEXDc) domain and a helicase superfamily c-terminal domain (HELICc)(Figure 5.2A) (Letunic and Bork, 2018). I found a protein sequence of *B. tabaci* that has both the DESXDc and HELICc domains in the same order and approximate distance (Figure 5.2B), indicating that *B. tabaci* MED has a VASA orthologue.

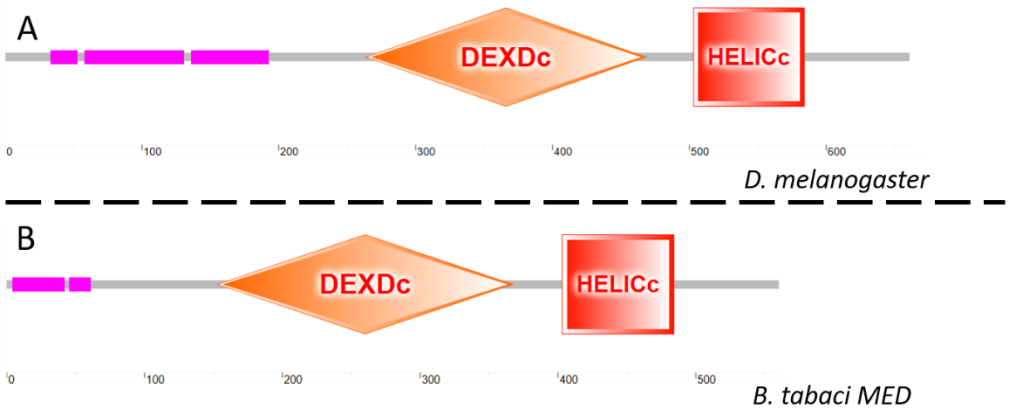

Figure 5. 1 Domains found in the VASA proteins of A.* D. melanogaster *B.* B. tabaci *MED.
Protein domains were discovered by the SMART protein analysis (Letunic and Bork, 2018). The purple/pink rectangles are regions of low complexity. The orange diamond indicates the location of the Dead-like helicases superfamily (DEXDc) region and the Red square the helicase superfamily c-terminal domain (HELICc) region within the VASA protein sequences. Numbers at the bottom represent protein lengths in amino acids.

A *B. tabaci* MED VASA orthologue was identified, so the next step was to assess the transcript expression levels of the corresponding gene in different whitefly developmental stages. The Integrative Genomics Viewer (IGV) software was used to visualise the embryo and adult VASA transcript composition (Thorvaldsdottir et al., 2013; Robinson et al., 2011).The transcript analysis of the female embryos (Figure 5.3), male embryos (Figure 5.4) and adults (Figure 5.5) are shown below. There is at least one transcript per sample for the *Vasa* gene. *Vasa* was on scaffold 2059 of the reference genome. The original gene model in A of Figure 5.2, has 10 exons. The mRNA spans across 16081 nt. There is a UTR at the 3' end spanning 1556nt, and a UTR at the 5' end spanning 2108nt.

A comparison of the *B. tabaci* MED gene model of *Vasa* (A in Figure 5.2, 5.3 and 5.4), and individual embryo and adult samples (B-D in Figure 5.2, 5.3, and 5.4) revealed some isoform differences. The variation of the transcripts in female and male single-egg RNA-seq replicates occur mainly at the UTR regions.

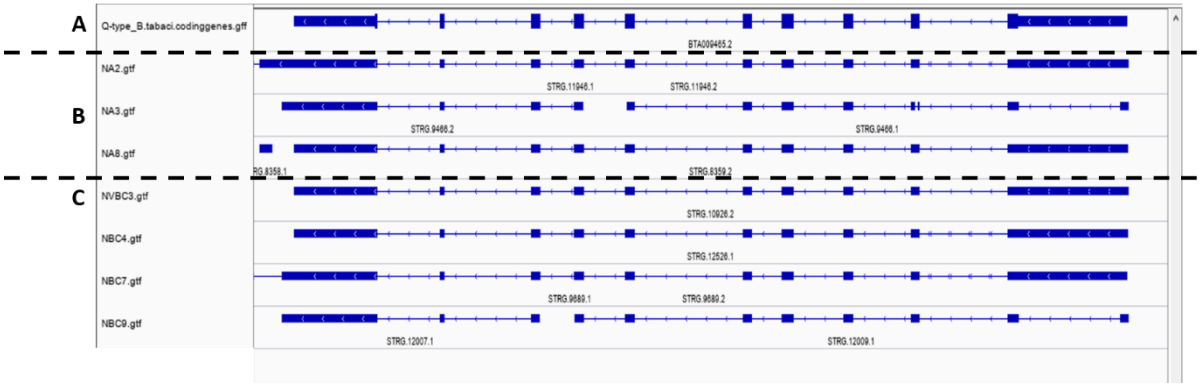

Figure 5. 2: Transcript of the Vasa hit the female single-embryo RNA-seq replicates.
The left column has the identifier for the sample type. All are from the RNA-seq samples except for Q-type_B.tabaci.codinggenes.gff. VASA is located on Scaffold_2059:122,176-144,492. A is the annotated VASA gene from the reference genome. B are the female A sample eggs. C are the female B/C samples eggs. The blue boxes and lines are the transcripts present at that locus in the samples. Under the transcripts there is the STRG number, which is a unique number for that sample. Viewed on IGV version 2.5.0.

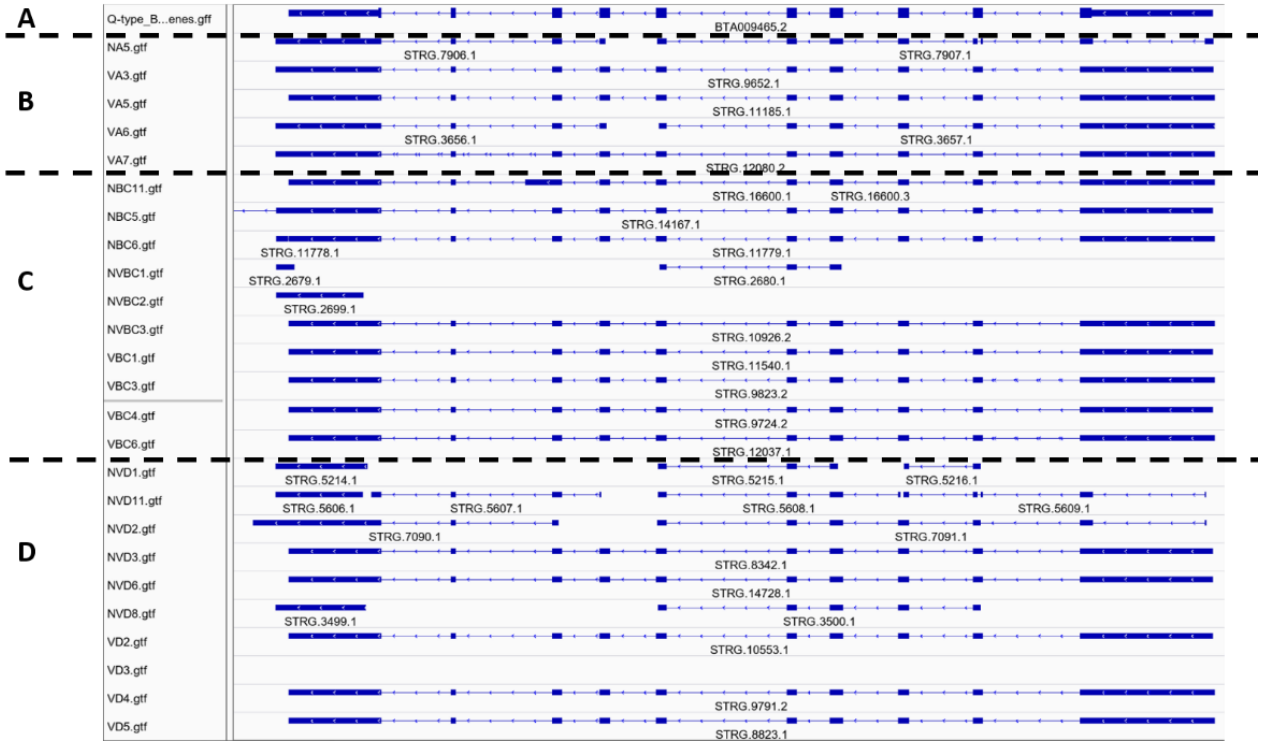

Figure 5. 3: Transcript of the Vasa hit against all the male single-embryo RNA-seq replicates.
The left column has the identifier for the sample type. All are from the RNA-seq samples except for Q-type_B.tabaci.codinggenes.gff. VASA is located on Scaffold_2059:122,176-144,492. A is the annotated VASA gene from the reference genome. B are the male A sample eggs. C are the male B/C samples eggs. D are the male D eggs. The blue boxes and lines are the transcripts present at that locus in the samples. Under the transcripts there is the STRG number, which is a unique number for that sample. Viewed on IGV version 2.5.0.

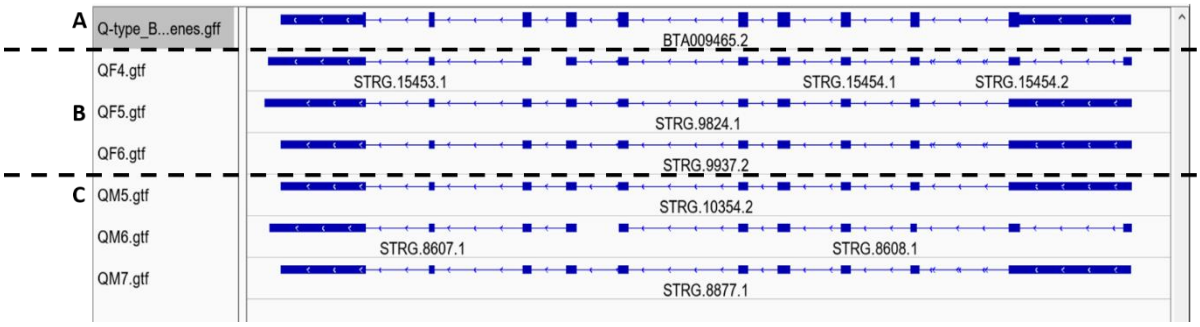

Figure 5. 4: Transcript of the Vasa hit against all the female and male adults RNA-seq replicates.
The left column has the identifier for the sample type. All are from the RNA-seq samples except for Q-type_B.tabaci.codinggenes.gff. VASA is located on Scaffold_2059:122,176-144,492. A is the annotated VASA gene from the reference genome. B are the adult female biological replicates. C are the male adult biological replicates. D are the male D eggs. The blue boxes and lines are the transcripts present at that locus in the samples. Under the transcripts there is the STRG number, which is a unique number for that sample. Viewed on IGV version 2.5.0.

Gene expression from male and female embryos and adults requires a dataCountTable (Materials and Methods, Figure 2.3). A. Singh created the dataCountTable for the expression values of the RNA-seq data set, in both embryonic and adult *B. tabaci* MED samples. The expression data units were in transcripts per million (TPM). The expression data for *Vasa* is shown in Figure 5.5.

Both male and female eggs are both expressing *Vasa* at the same rate. Unfortunately, due to the lack of data for the D type egg, it is not known if this pattern is similar in the D type egg in females. The male and female have similar expression. In female adults, the mean expression of *Vasa* is a lot higher than any of the other samples, including the corresponding male samples. *Vasa* expression in many species starts around the blastoderm formation (Braat et al., 2000) (Dearden et al., 2003). Based on this data, and in Chapter 3, A-stage eggs are at blastoderm formation.

A T-test 2 sample with unequal variance (Figure S2) showed only male BC eggs and male adults were statistically different (<0.001), female A eggs are statistically different from female adult samples (0.009), female B/C eggs are statistically different from female adults (0.009). Female adults had a statistically different TPM mean against male adults (0.017).

The observed high levels of expression of *Vasa* in adult females might be due to the contribution of pre-A stage eggs in the ovaries of the females. A way to test whether the high *Vasa* expression in adults were due to egg contamination would be to gather female body and dissect the ovaries, containing the eggs out of the sample.

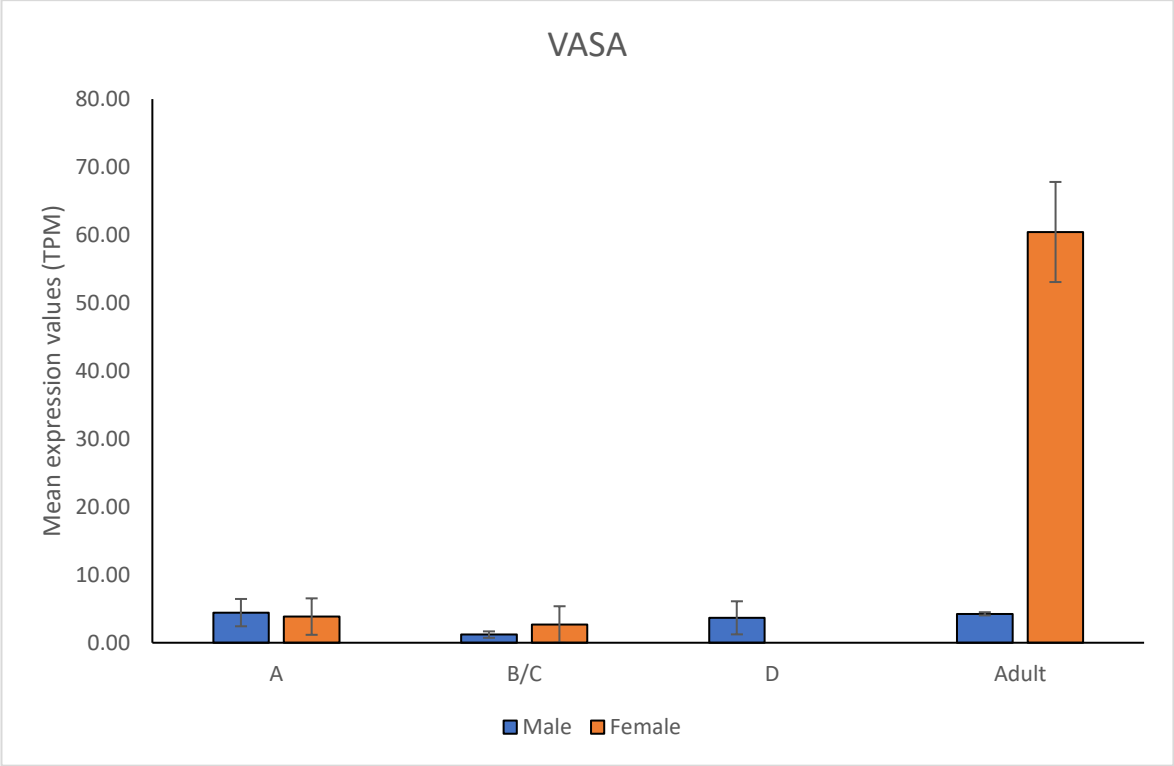

***Figure 5. 5: The expression data for the Vasa gene in* B. tabaci *MED embryos and adults*.**
Male A, B/C, D eggs and adults are represented in blue, females A, B/C eggs and adults are in orange. The error bars represent the standard error.

5.3 Results part 2

In the results part 2 of chapter 5, I have investigated the isoform and gene expression of the DM genes that were identified in chapter 3. I also looked at the gene expression analysis of some 'key' genes of interest identified in chapter 3.

5.3.1 Where are the DM genes in the genome?

There is no DSX orthologue in *B. tabaci* MED; however, there are DM-domain containing proteins (DMRT). The locations of whitefly DM-domain containing proteins were identified within the scaffolds of the whitefly genome (Table 5.7). The locations of the proteins are needed because it is a requirement of the programme used in this chapter; Integrative Genomics Viewer (IGV). IGV is a tool which allows exploration of large-scale genomic data sets on a desktop computer. IGV allows the user to zoom and pan across the genome to base-pair level (Thorvaldsdottir et al., 2013; Robinson et al., 2011).

The DM protein research is described in chapter 3, Section 3.2.1. Figure 3.3 shows a phylogenetic tree of the different hemipteran DM proteins. The *B. tabaci* MED DM proteins fell into four different clades; an unknown/Hemiptera specific clade, DMRT99B, DMRT11E and DMRT93B. For ease of nomenclature in the future, I will call the *B. tabaci* MED *dmrt* gene (BTA004042.1) found in the Hemiptera-specific clade of the tree *Btdmrt1.* BTA013024 found in the dmrt99b clade; *Btdmrt2*, BTA011988 found in the dmrt11E clade; *Btdmrt3,* and BTA021616 found in the dmrt93b clade; *Btdmrt4.* Only *Btdmrt1-3* was analysed in this chapter, as *Btdmrt4* transcripts from the RNA-seq data were fragmented in the isoform analysis and had zero expression in the gene expression. The genome locations of the *B. tabaci* MED DM proteins are shown in table 5.7.

Protein Query Seq-id (aa)	Subject Seq-id	Gene ID from Genome	Start of alignment in subject (nt)	End of alignment in subject (nt)
*Bt*dmrt1	Scaffold_1383	BTA004042.1	50409	50284
*Bt*dmrt2	Scaffold_265	BTA013024.1	111750	111625
*Bt*dmrt3	Scaffold_2466	BTA011988.1	87164	87039
Btdmrt4	Scaffold_471	BTA021616.1	370209	443894

***Table 5. 7: The scaffold locations of the sex determination gene orthologues found in* B. tabaci MED.**
Protein sequences of *B. tabaci* MED genes identified in Chapter 3 (column: Query seq-ID) were searched via TBLASTN against the *B. tabaci* MED genome assembly (Table 2.1) and the scaffold with the best hit selected (column: Subject Seq-id). Then, the gene ID was identified (column: Gene id from Genome). aa means amino acid sequence, nt means nucleotide sequence.

5.3.2 Btdmrt1 are present in *B. tabaci MED* male and female eggs and adults but are not differentially spliced

The first DM gene I investigated was the *Btdmrt1* in male and female *B. tabaci* MED eggs and adults. An example of the IGV analysis is present below, of male eggs (Figure 5.6). Fourteen RNA-seq samples had a transcript present at the *Btdmrt1* locus for male eggs; four replicates for A eggs (NA5, VA5, VA6 and VA7), five replicates for B/C eggs (NBC11, NBC6, NVBC2, VBC1 and VBC3), and five replicates for D type eggs (VD5, NVD2, NVD3, NVD6 and VD4). *Btdmrt1* is mono-exonic and spans 650nt; the transcripts present in the male egg samples also span 650nt.

In the female eggs there were four RNA-replicates have a transcript present at the *Btdmrt1* locus for female eggs; two biological replicates for A eggs (NA2 and NA8), and two biological replicates for B/C eggs (NBC4 and NBC7). The transcripts present in the female eggs' samples all span 650nt. In the female and male adults there were RNA-replicates for all the samples and the transcripts present in the eggs samples also span 650nt.

Some samples did not have transcripts present at the *Btdmrt1* locus; in the male egg these are VA3, VBC4, VBC5, VBC5, VD2, VD3, NVD1 and NVD8, and in the female samples these are NA8, NBC9 and NVBC3.

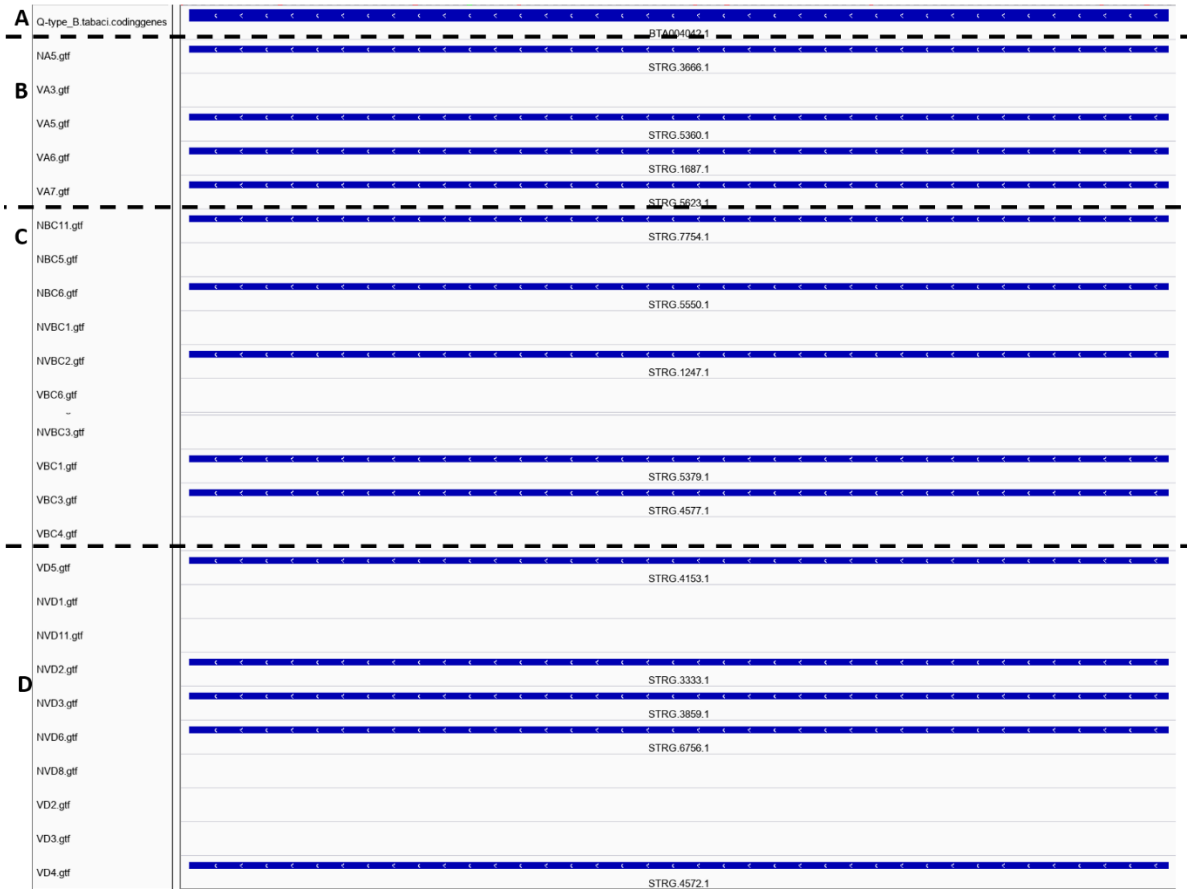

Figure 5. 6: Transcript of the Btdmrt1 hit against all the male egg RNA-seq replicates.
The left column has the identifier for the sample type. All are from the RNA-seq samples except for Q-type_B.tabaci.codinggenes.gff. *Btdmrt1* is located on scaffold Scaffold_1383. A is the *Btdmrt1* gene from the reference genome. B are the male A sample eggs. C are the male B/C samples eggs at the *Btdmrt1* locus. D are the male D sample eggs. The blue boxes and lines are the transcripts present at that locus in the samples. Under the transcripts there is the STRG number, which is a unique number for that sample. Viewed on IGV version 2.5.0.

The transcripts in Figure 5.6 all look identical, as they have the same length transcripts and the same mono-exonic characteristics. From this IGV view, it is difficult to determine whether the transcripts are identical at the nucleotide level. To analyse this, I extracted the sequences that are in the *Btdmrt1* locus from all the RNA-seq samples. Next, I conducted a pairwise comparison in CLC of all the transcripts and the *Btdmrt1* from the reference genome. The analysis for this is in Figure 5.7. All 24 transcripts were identical, as they had 100% percentage identity and were coloured red. Therefore, male and female adults and eggs only have one transcript of *Btdmrt1*, and *Btdmrt1* does not undergo differential splicing. One of the problems with this data set is the lack of female D -sample eggs. Therefore, conclusions cannot be made about this stage of egg in the female samples.

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
NA2_dsx_1383	1		100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
NA3_dsx_1383	2	100.00		100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
NA5_dsx_1383	3	100.00	100.00		100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
NBC11_dsx_1383	4	100.00	100.00	100.00		100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
NBC4_dsx_1383	5	100.00	100.00	100.00	100.00		100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
NBC6_dsx_1383	6	100.00	100.00	100.00	100.00	100.00		100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
NBC7_dsx_1383	7	100.00	100.00	100.00	100.00	100.00	100.00		100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
NVBC2_dsx_1383	8	100.00	100.00	100.00	100.00	100.00	100.00	100.00		100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
NVD2_dsx_1383	9	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00		100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
NVD3_dsx_1383	10	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00		100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
NVD8_dsx_1383	11	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00		100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
VA5_dsx_1383	12	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00		100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
VA6_dsx_1383	13	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00		100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
VA7_dsx_1383	14	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00		100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
VBC1_dsx_1383	15	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00		100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
VBC3_dsx_1383	16	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00		100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
VD4_dsx_1383	17	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00		100.00	100.00	100.00	100.00	100.00	100.00	100.00
VD5_dsx_1383	18	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00		100.00	100.00	100.00	100.00	100.00	100.00
QF1_1383_STRG.7149.1	19	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00		100.00	100.00	100.00	100.00	100.00
QF5_1383_STRG.4503.1	20	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00		100.00	100.00	100.00	100.00
QF6_1383_STRG.4558.1	21	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00		100.00	100.00	100.00
QM5_1383_STRG.4774.1	22	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00		100.00	100.00
QM8_1383_STRG.3890.1	23	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00		100.00
QM7_1383_STRG.4064.1	24	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	

Figure 5. 7: Btdmrt1 transcripts found in the male and female egg and adult samples are 100% identical.

A pairwise comparison using the full transcript alignment of the transcripts found at the BTA004042 position. The alignment and the pairwise comparison were conducted by CLC. Both the bottom and lower comparisons show the percentage identity scoring. The left is the male and female egg and adult identifiers for the transcripts.

I present a summary of the transcripts of *Btdmrt1* found in male and female eggs and adults. This summary is in Figure 5.8. This summary took the information of the transcripts found in the IGV along with the information from the pairwise comparison. All the transcripts are represented with a rectangle. In Figure 5.8, all the transcripts were identical to the *Btdmrt1*, and therefore all the transcripts were the same orange colour.

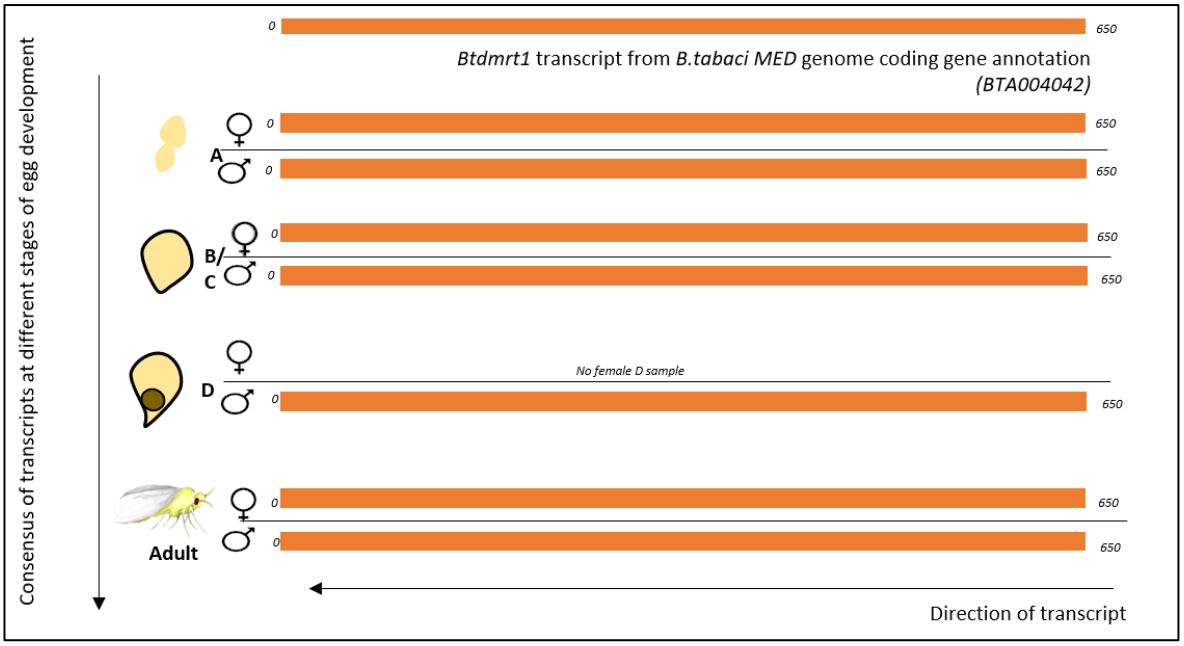

***Figure 5. 8*: The transcript consensus of *Bdmrt1* in all male and female eggs and adults.**

RNA-seq is one of the methods that can assess transcript changes over time. Kallisto provides the quantity of the transcripts in a form called dataCountTable (Materials and Methods, Figure 2.3). A. Singh created the dataCountTable for the expression values of the RNA-seq data set, in both embryonic and adult *B. tabaci* MED samples. The expression data units were in transcripts per million (TPM). The expression data for *Vasa* is shown in Figure 5.9.

Each RNA-seq sample had a *Btdmrt1* (except those already mentioned above), from this the mean expression for the different sample type (Male A eggs, Female A eggs etc.) was calculated. Figure 5.9 is the *Btdmrt1* mean expression for the different samples. Figure 5.9 reveals the *Btdmrt1* expression patterns across the different life stages and sexes. Both male and female eggs show the same downward trend from A to B/C type eggs. The expression of *Btdmrt1* increases from B/C to D male eggs (it is unknown whether this trend continues in female D eggs due to the lack of data). A T-test 2 sample with unequal variance was conducted on the mean expression data between all samples (Figure S2). The only samples that were statistically different was between Male BC eggs and Male adults (0.012), and therefore the only samples we can comment of being different between life stages.

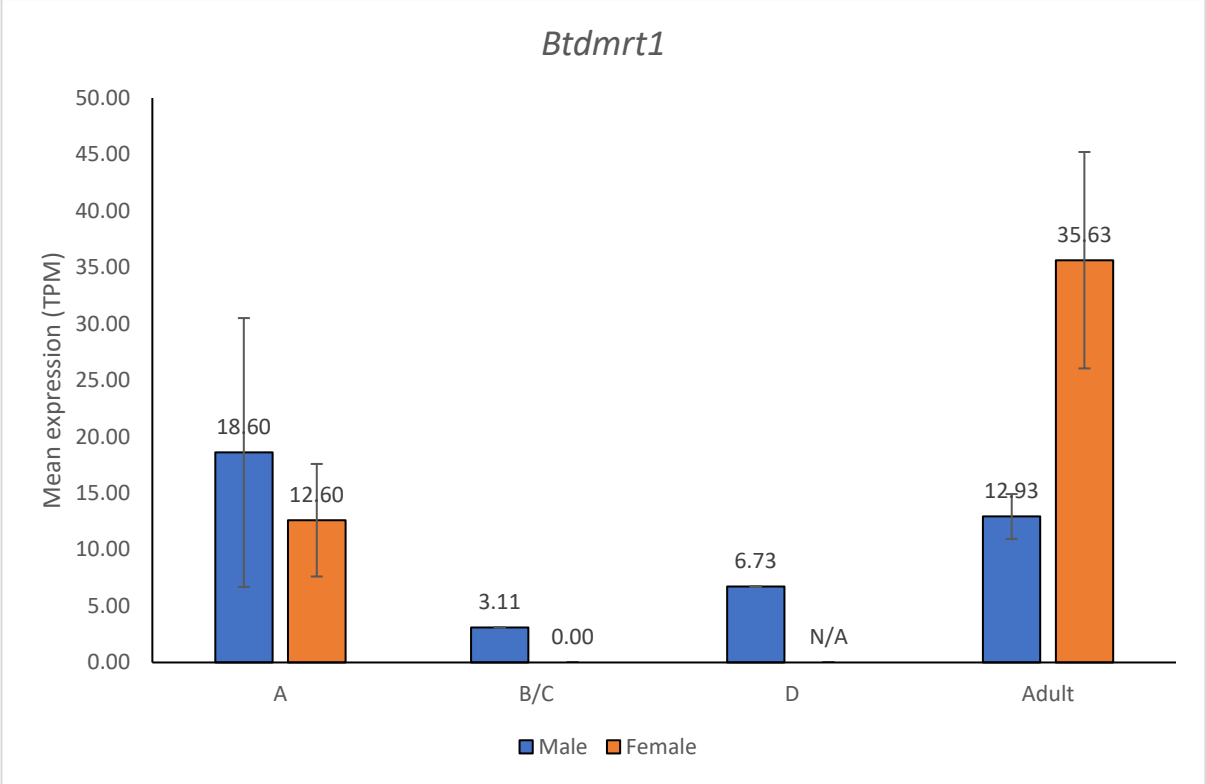

Figure 5. 9: **Btdmrt1** ***expression data in*** **B. tabaci** ***MED embryos and adults***
Males are represented in blue, and the females are in orange. The error bars represent the standard error. The numbers on the top of the bars represent the mean expression (TPM) for each type of sample.

5.3.3 Btdmrt2 transcripts are present in *B. tabaci* MED late-stage eggs and adults but are not differentially spliced

Btdmrt2 is the second *B. tabaci* MED DM gene I investigated in males and female at different developmental times. Isoform analysis protocol was the same as 5.3.2, and therefore only the summary figure will be shown in this section (Figure 5.10). There are only two male D samples that have transcripts located in *Btdmrt2*, and both are identical to each other. In the adult samples all the samples had transcripts. The *Btdmrt2* spans 413 nt, and this is the same in the transcripts found. All the transcripts are represented with a rectangle. In Figure 5.10, all the transcripts were identical to the *Btdmrt1*, and therefore all the transcripts were the same orange colour.

Btdmrt2 did not undergo expression analysis, as there was a lack of data for comparative analysis through the developmental stages. The lack of data suggests that *Btdmrt2* transcript is rare in the egg stage.

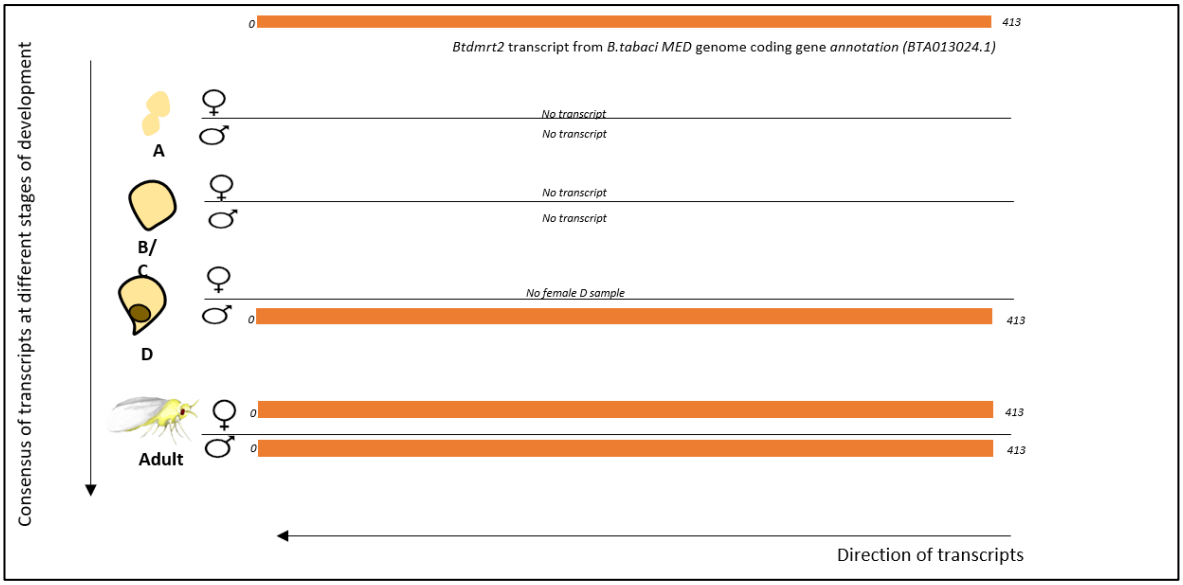

Figure 5. 10: The transcript consensus of Bdmrt2 in all male and female eggs and adults. A, B/C, D and adult male and female.

5.3.4 Btdmrt3 is present mainly in *B. tabaci* MED male eggs and male and female adults and are differentially spliced

Btdmrt3 is the third *B. tabaci* MED DM gene I investigated in male and females at different developmental stages. Isoform analysis protocol was the same as 5.3.2. In the isoform analysis protocol, there was different isoforms present, and therefore I will show the pairwise comparison (Figure 5.11). The pairwise comparison shows that not all transcripts were identical, anything less than 100% means that there are differences. Figure 5.11 shows the differences in the percentage identity, and the identity varies from 5% to 100%; therefore, *Btdmrt3* has more than one isoform. The original gene ranges over 7200 nt. Some transcripts were the same length. The summary figure will be shown in this section (Figure 5.12). *Btdmrt3* differs from *Btdmrt1* and *Btdmrt2*, as there are different transcripts present among the different samples.

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
QF4.2_2466	1		100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	80.90	46.09	43.03	36.46	25.41	11.88	8.44	8.13	8.23	8.13	6.74
QF5_2466	2	100.00		100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	80.90	46.09	43.03	36.46	25.41	11.88	8.44	8.13	8.23	8.13	6.74
QF6_2466	3	100.00	100.00		100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	80.90	46.09	43.03	36.46	25.41	11.88	8.44	8.13	8.23	8.13	6.74
QM6_2466	4	100.00	100.00	100.00		100.00	100.00	100.00	100.00	100.00	100.00	100.00	80.90	46.09	43.03	36.46	25.41	11.88	8.44	8.13	8.23	8.13	6.74
QM7_2466	5	100.00	100.00	100.00	100.00		100.00	100.00	100.00	100.00	100.00	100.00	80.90	46.09	43.03	36.46	25.41	11.88	8.44	8.13	8.23	8.13	6.74
NBC6_STRG.14946.1	6	100.00	100.00	100.00	100.00	100.00		100.00	100.00	100.00	100.00	100.00	80.90	46.09	43.03	36.46	25.41	11.88	8.44	8.13	8.23	8.13	6.74
VA3_STRG.12208.2	7	100.00	100.00	100.00	100.00	100.00	100.00		100.00	100.00	100.00	100.00	80.90	46.09	43.03	36.46	25.41	11.88	8.44	8.13	8.23	8.13	6.74
VA5_STRG.14174.2	8	100.00	100.00	100.00	100.00	100.00	100.00	100.00		100.00	100.00	100.00	80.90	46.09	43.03	36.46	25.41	11.88	8.44	8.13	8.23	8.13	6.74
VBC3_STRG.12448.1	9	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00		100.00	100.00	80.90	46.09	43.03	36.46	25.41	11.88	8.44	8.13	8.23	8.13	6.74
VD4_STRG.12322.1	10	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00		100.00	80.90	46.09	43.03	36.46	25.41	11.88	8.44	8.13	8.23	8.13	6.74
VD5_STRG.11168.1	11	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00		80.90	46.09	43.03	36.46	25.41	11.88	8.44	8.13	8.23	8.13	6.74
QM5_2466	12	80.90	80.90	80.90	80.90	80.90	80.90	80.90	80.90	80.90	80.90	80.90		37.29	34.82	29.55	20.56	9.61	6.82	6.58	6.66	6.58	5.45
QF4_2466	13	46.09	46.09	46.09	46.09	46.09	46.09	46.09	46.09	46.09	46.09	46.09	37.29		55.71	66.60	53.46	25.00	17.31	16.67	16.88	16.67	13.76
VA5_STRG.14174.1	14	43.03	43.03	43.03	43.03	43.03	43.03	43.03	43.03	43.03	43.03	43.03	34.82	55.71		66.18	51.09	23.89	16.98	16.34	16.55	16.34	13.55
VA3_STRG.12208.1	15	36.46	36.46	36.46	36.46	36.46	36.46	36.46	36.46	36.46	36.46	36.46	29.55	66.60	66.18		67.49	31.56	22.40	21.58	21.86	21.58	17.90
NVD11_STRG.7117.1	16	25.41	25.41	25.41	25.41	25.41	25.41	25.41	25.41	25.41	25.41	25.41	20.56	53.46	51.09	67.49		46.76	25.14	24.03	24.40	24.03	19.04
NBC4_STRG.15888.1	17	11.88	11.88	11.88	11.88	11.88	11.88	11.88	11.88	11.88	11.88	11.88	9.61	25.00	23.89	31.56	46.76		11.64	11.81	11.75	11.81	9.79
VA3_STRG.12207.1	18	8.44	8.44	8.44	8.44	8.44	8.44	8.44	8.44	8.44	8.44	8.44	6.82	17.31	16.98	22.40	25.14	11.64		97.87	98.58	97.87	81.56
VBC3_STRG.12447.1	19	8.13	8.13	8.13	8.13	8.13	8.13	8.13	8.13	8.13	8.13	8.13	6.58	16.67	16.34	21.58	24.03	11.81	97.87		99.28	100.00	83.33
VBC6_STRG.15238.1	20	8.23	8.23	8.23	8.23	8.23	8.23	8.23	8.23	8.23	8.23	8.23	6.66	16.88	16.55	21.86	24.40	11.75	98.58	99.28		99.28	82.73
VD2_STRG.13333.1	21	8.13	8.13	8.13	8.13	8.13	8.13	8.13	8.13	8.13	8.13	8.13	6.58	16.67	16.34	21.58	24.03	11.81	97.87	100.00	99.28		83.33
STRG.10585.1	22	6.74	6.74	6.74	6.74	6.74	6.74	6.74	6.74	6.74	6.74	6.74	5.45	13.76	13.55	17.90	19.04	9.79	81.56	83.33	82.73	83.33	

Figure 5. 11: Btdmrt3 has more than one isoform

A pairwise comparison of the Btdmrt3 transcripts from *B. tabaci* embryos and adults. The numbers indicate the percentage identity, the colours are the blue to red scale, where blue is the low percentage and red is higher percentage. The full-length proteins were aligned in CLC.

Figure 5.12 is a summary of the *Btdmrt3* isoforms found in adults and embryos. *Btdmrt3* has 3 exons and 2 UTR sites. In the transcript of the *Btdmrt3* from the reference genome, the UTR's are represented by the blue rectangles on the second line in Figure 5.12.

The isoforms differ mainly at the UTR sites, whether the UTR 5' or 3' are present or absent. For example, the orange transcript has all the sequence, including the UTR ends. The same colour blocks represent isoforms with 100% identity to each other. All developmental stages have a transcript present; however, in egg stages transcripts are not present in the female A egg. In *Btdmrt3* all males have an identical isoform, coloured orange. The orange isoform is not in female A or B/C eggs, and it is unknown whether it is present in female D eggs. There is an orange isoform in adult females. However, this may be a false positive, as there could be eggs in the

female samples, and this is what is sequenced. There is a female-specific isoform (red) in female B/C eggs. The female-specific isoform is a truncated at the 3' end and may only be present in early embryogenesis.

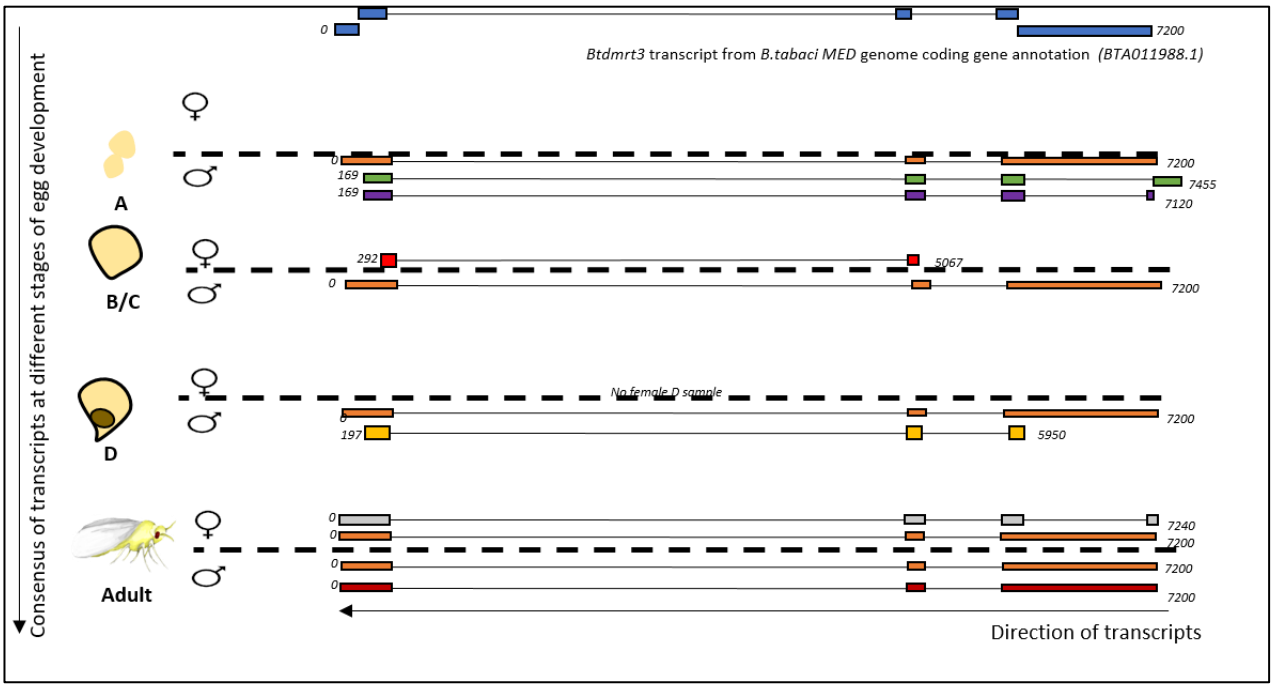

Figure 5. 12: The Btdmrt3 isoform consensus from* B. tabaci *MED adults and eggs.
The transcript consensus of Btdmrt3 transcripts that align with BTA011988.1 in *B. tabaci* in all A, B/C, D and adult male and female.

The quantity of *Btdmrt3* was evaluated for all developmental stages and sexes, this is present in Figure 5.13, conducted as Section 5.3.2. The calculation of the mean expression for each sample type was from the expression data from each RNA-seq sample that has a *Btdmrt3*. Figure 5.13 reveals the *Btdmrt3* expression patterns across the different life stages and sexes. The *Btdmrt3* expression in male eggs are always higher than the female eggs. Male A eggs have a high expression of *Btdmrt3*, whereas there is zero expression of *Btdmrt3* in female A eggs. *Btdmrt3* expression in female B/C eggs is at 0.33 TPM, higher than 0 TPM in female A eggs. *Btdmrt3* expression in adults is higher than in eggs. *Btdmrt3* expression in female adults is higher than male adults.

Statistical testing of the different mean (TPM) was conducted using T-test 2 sample with unequal variance (see Figure S2). The statistically significant means were between male A eggs vs. male adults (0.037), male BC eggs vs male adults (0.037), male D eggs vs. male adults (0.038), female B/C eggs between female adults (<0.001) and female adults vs. male adults (0.047).

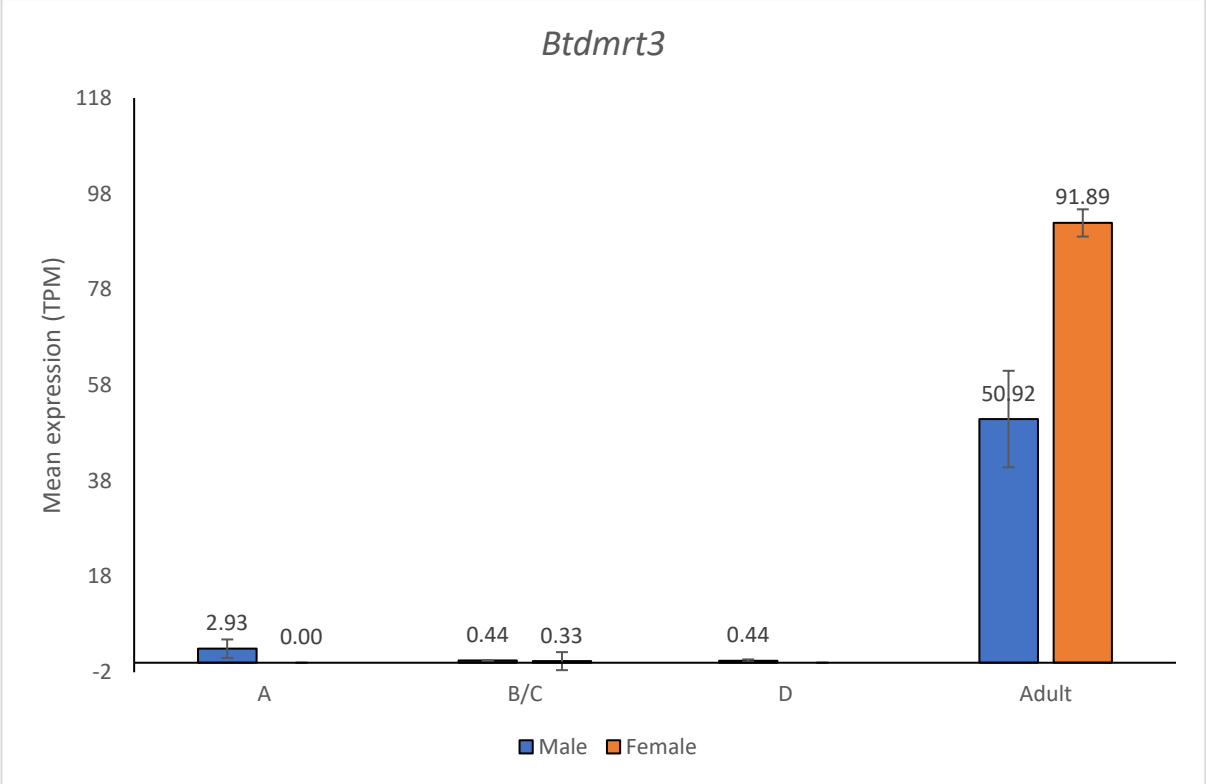

Figure 5. 13: **the expression data for the *Btdmrt3* gene in *B. tabaci* MED embryos, for male A, B/C and D type eggs, in females A and B/C type eggs.**

5.3.5 Some 'Key' SDG are expressed at early embryogenesis

The summary of the 'Key' genes found in Hemiptera is in Chapter 3, Figure 3.32. Overall, PSI, IMP, SXL and TRA2 are the 'key' genes with the highest probability of being true orthologues. Gene expression analysis will be conducted on these 'key' genes, to see if there is high expression in early embryogenesis. If there is high expression in early embryogenesis, this may provide a list of genes of interest for further isoform analysis in the future.

5.3.5.1 PSI isoforms are found in early embryogenesis

The quantity of *B. tabaci Psi* orthologue was evaluated for all developmental stages and sexes, and this is present in Figure 5.14, conducted as Section 5.3.2. The calculation of the mean expression for each sample type was from the expression data from each RNA-seq sample that has *Psi.* There is zero gene expression of *BtPSI* in eggs, however there is expression in adults of 1.12 TPM in males and 1.22 TPM in females.

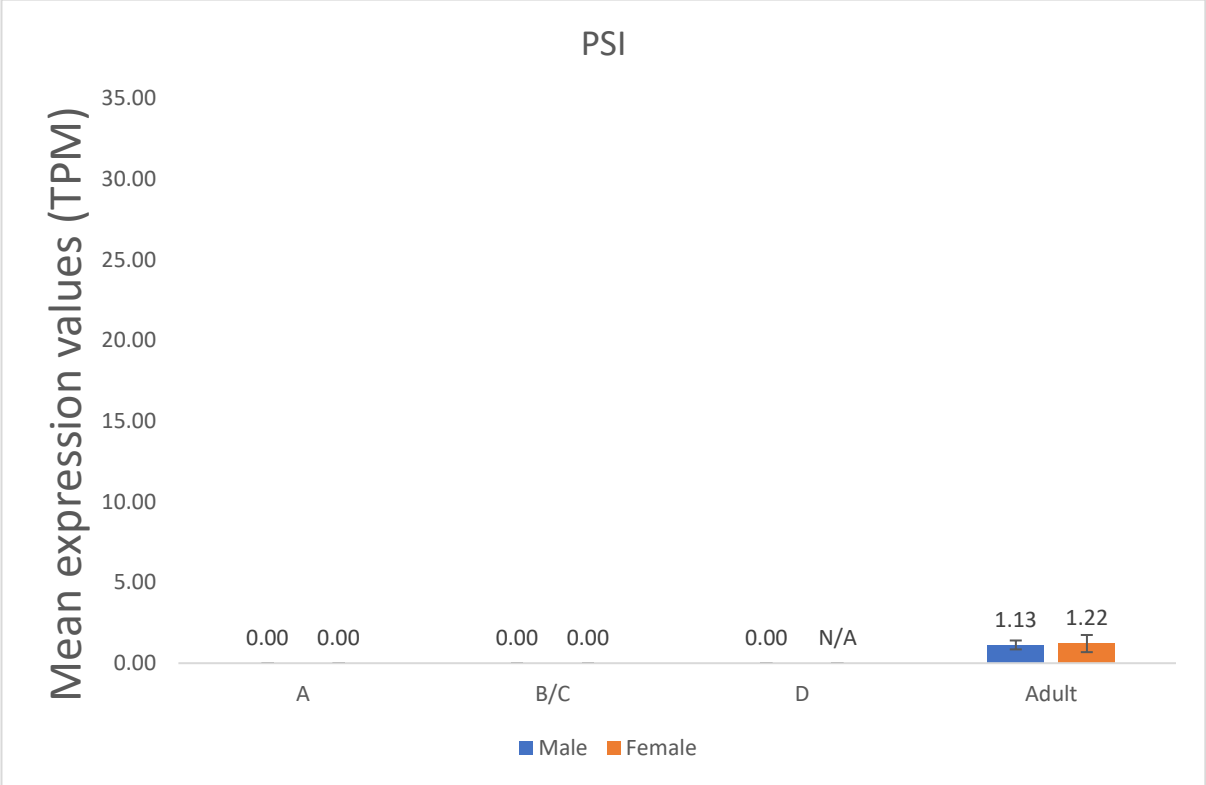

Figure 5. 14: the expression data for the PSI gene in* B. tabaci *MED embryos, for male A, B/C and D type eggs, in females A and B/C type eggs.

5.3.5.2 Sxl is present in early embryogenesis

The quantity of *B. tabaci Sxl* orthologue was evaluated for all developmental stages and sexes, this is present in Figure 5.15, conducted as Section 5.3.2. *Sxl* expression was conducted for each RNA-seq sample, and the mean expression was produced for the different sample type. Figure 5.15 shows the mean expression values for *Sxl. Sxl* expression is highest in the eggs. *Sxl* expression is at the lowest in female B/C eggs, and it is unknown whether this trend continues to D eggs. T-test 2 sample with unequal variance (Figure S2) shows that the samples that are statistically different are male A eggs vs. male adults (0.041) and female adults vs. male adults (0.001).

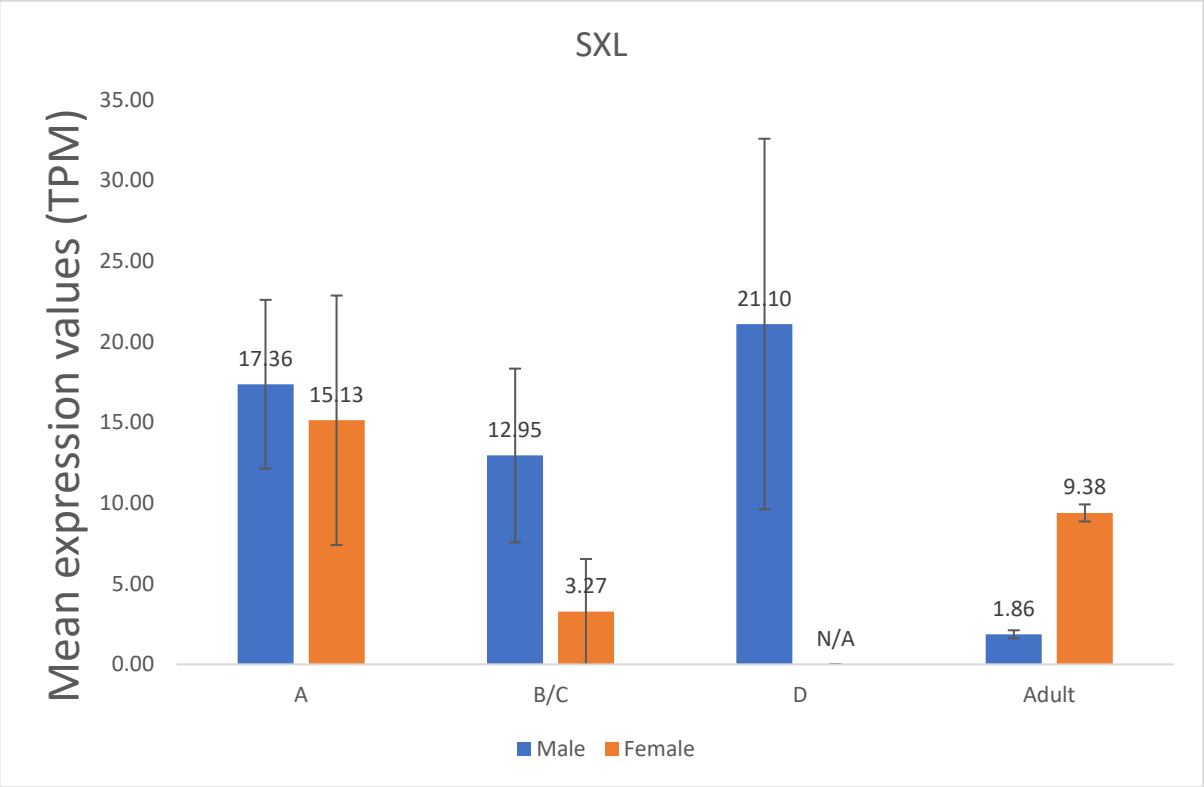

Figure 5. 15: the expression data for the Sxl gene in* B. tabaci *MED embryos, for male A, B/C and D type eggs, in females A and B/C type eggs.

5.3.5.3 Imp is present in early embryogenesis

The quantity of *B. tabaci Imp* orthologue was evaluated for all developmental stages and sexes, and this is present in Figure 5.16, conducted as Section 5.3.2. *Imp* expression was conducted for each RNA-seq sample and the mean expression was produced for the different sample type. Figure 5.16 shows the mean expression values for *Imp. Imp* expression is highest at A eggs. *Imp* expression is at the lowest in female B/C eggs, and it is unknown whether this trend continues to D eggs. *Imp* expression is higher in the male eggs than female.

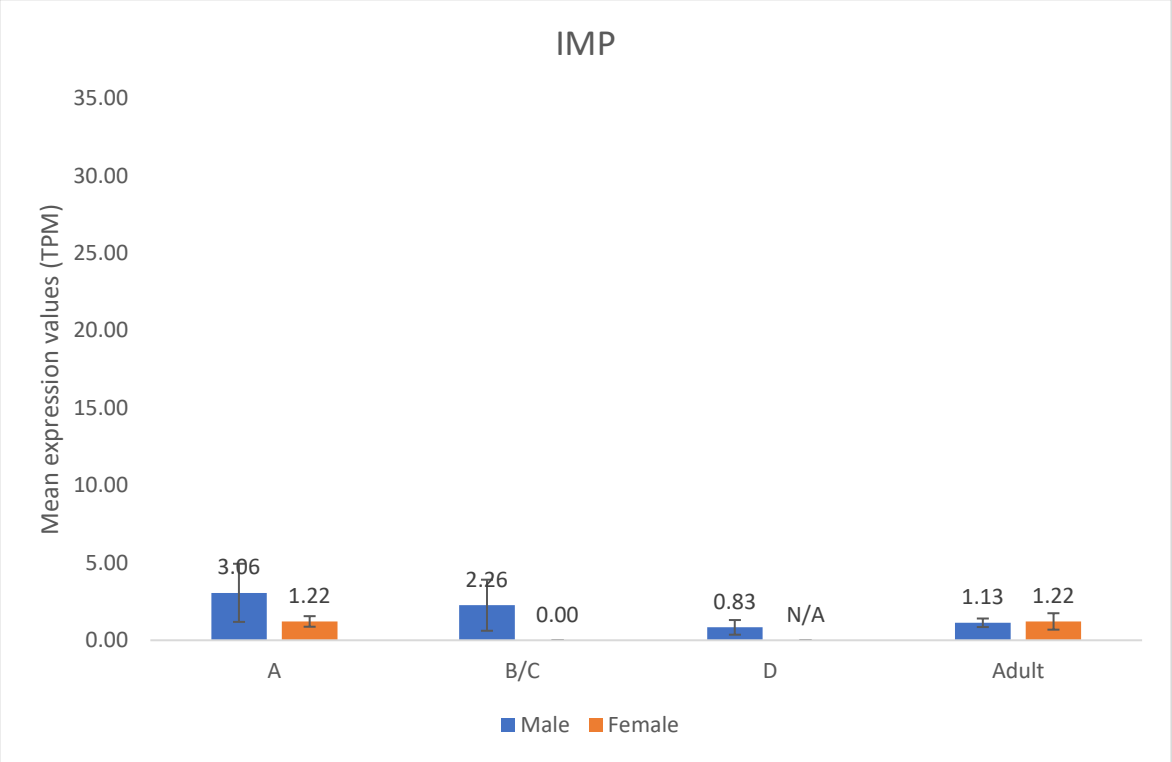

Figure 5. 16: the expression data for the* Imp *gene in* B. tabaci *MED embryos, for male A, B/C and D type eggs, in females A and B/C type eggs.

5.3.5.4 Tra2 is present in early embryogenesis

The quantity of *B. tabaci Tra2* orthologue was evaluated for all developmental stages and sexes, and this is present in Figure 5.17, conducted as Section 5.3.2. *Tra2* expression was conducted for each RNA-seq sample, and the mean expression was produced for the different sample type. Figure 5.17 shows the mean expression values for *Tra2. Tra2* expression peaks in male and female A eggs. *Tra2* expression is higher in the female eggs than male.

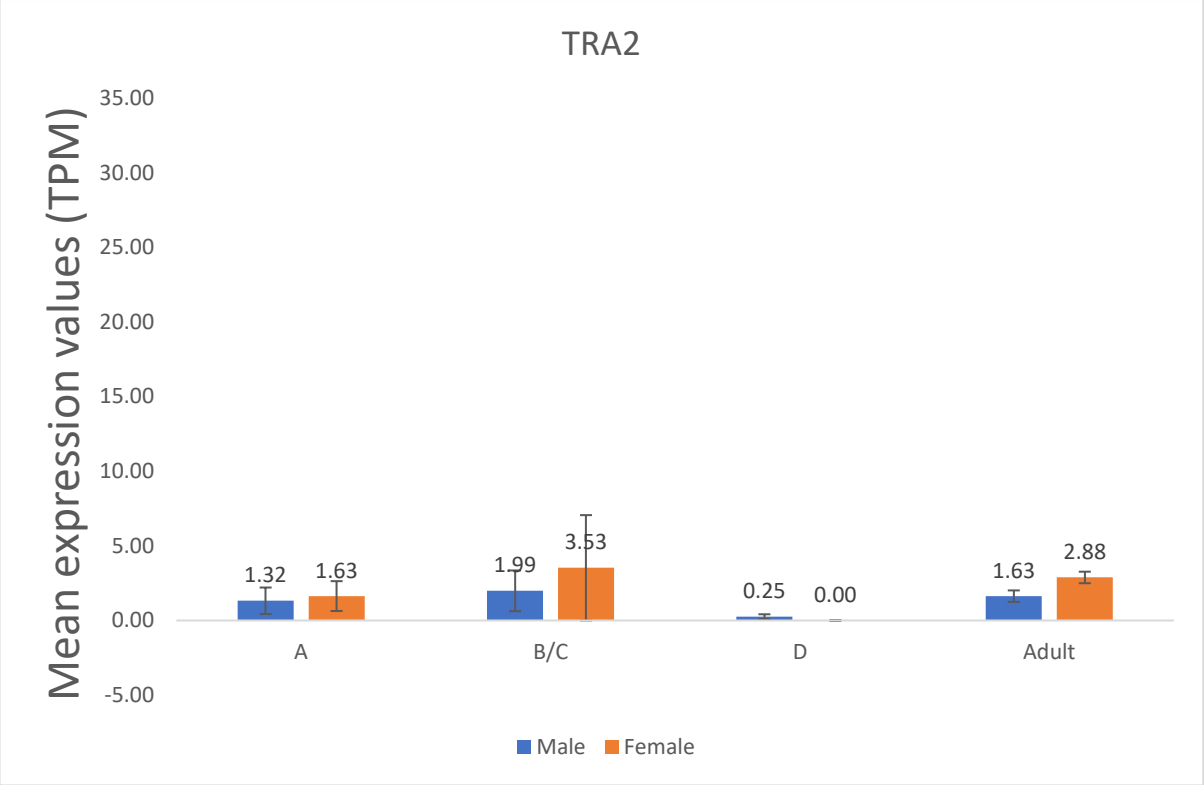

Figure 5. 17: the expression data for the* tra2 *gene in* B. tabaci *MED embryos, for male A, B/C and D type eggs, in females A and B/C type eggs

5.3.5.5 Summary of 'key' genes

Figure 5.18 is a heatmap of the gene expression of *PSI, Sxl, Tra2* and *Imp,* and this summary includes the data shown from Sections 5.3.5.1 to 5.3.5.4. This figure helps to visualise the patterns of the gene expression in various developmental stages. Red indicates low expression (TPM) and green represents high expression.

SDGs are known to be expressed during embryogenesis. *Psi* is not expressed in egg stages. This data suggests that this gene is not involved in sex determination at these stages, because it is not being expressed. *Sxl* is highly expressed in egg developmental stages, and this suggests that there is a high probability that this gene is part of the sex determination cascade. Both *Tra2* and *Imp* have a similar expression in embryos as in adults, which suggests that these may be a possibility that these genes are part of the sex determination pathway.

	A	B/C	D	Adult	
PSI	0.00	0.00	0.00	1.13	Male
	0.00	0.00	N/A	1.22	Female
SXL	17.36	12.95	21.10	1.86	Male
	15.13	3.27	N/A	9.38	Female
TRA2	1.32	1.99	0.25	1.63	Male
	1.63	3.53	N/A	2.88	Female
IMP	3.06	2.26	0.83	1.13	Male
	1.22	0.00	N/A	1.22	Female

Figure 5. 18: A heatmap summary of the gene expression of* PSI, SXL, Tra2 *and* Imp *in male and female eggs and adults

5.4 Discussion

This results chapter was in two parts. The first part involves the description and performance of the method used to determine the sex of single embryos from RNA-seq data, and this could apply to any haplodiploid organism. Along with this, the bioinformatic pipeline used for isoform and gene expression analysis was validated and tested with a conserved germ-line gene, *Vasa*. The *Vasa* test was successful and provided more evidence for embryonic stages. The second part of this results chapter is the isoform and gene analysis of the *Dmrt* genes, and the gene expression of interesting 'key' SDGs.

5.4.1 Eggs can be separated into sex by GATK analysis

The embryonic samples underwent the GATK haplotype pipeline (Chapter 2, Section 2.2.3). The results for this are in Table 5.2. Virgin females produce only male progeny, so eggs sampled from virgin females are male. Eggs collected from potentially non-virgin females have unknown sex. Most male eggs had a heterozygosity of 22.66%. The male eggs were identified from the unknown sex eggs by using the 22.66% percentage. The female samples have 5 A type eggs, 4 B/C type eggs and 0 D type eggs.

5.4.2 The bioinformatic pipeline can produce files for isoform and gene expression analysis

The bioinformatic pipeline for isoform and gene expression analysis is in Figure 2.3 and 2.4. The alignment of the individual files against the *B. tabaci* genome is in Figure 5.1. Which may mean that some of the conserved genes may not be detected. The low alignment percentage was low in the later egg stages. The bacteriocyte is transferred from mother to egg in the later egg stages. Therefore, the low alignment rates maybe because of the bacteriocyte.

5.4.3 A conserved germ-line gene is expressed in all *B. tabaci* egg stages tested

A conserved embryonic gene, *Vasa,* went through the gene expression and isoform analysis to test whether the bioinformatic pipeline (Figure 2.3 and 2.4) is reliable. The *Vasa* gene expression would also provide more information about the embryogenesis stages of the eggs.

Vasa is a conserved embryonic gene that has been found in *B. tabaci* MED (Figure 5.3, 5.4 and 5.5). *Vasa* is a gene that has been used to distinguish when the germline cells occur, and therefore should be expressed in early embryogenesis stages. When analysing the transcripts present in *B. tabaci* MED adults and differing embryo stages, most samples express the transcript. The expression of vasa in these samples is consistent with blastulation occurring before the A-stage, consistent with observations in Chapter 4.

Figure 5.6 shows the data expression values (TPM) of the embryos and adults, split into male and female samples. The mean expression of *Vasa* in female adults is a lot higher than any of the other samples, including the corresponding male samples. It is possible that as *Vasa* expression is expected to peak during blastulation, that the high levels of expression in adult females might be due to the present of pre-A-stage embryos inside. A way to test this would be to gather female body and dissect the ovaries, containing the eggs out of the sample. The conserved embryo gene used (*Vasa*) was expressed as early as A egg in both male and female samples. A type egg must be when blastoderm has already formed, as this is when *Vasa* expression peaks in *D. melanogaster*. Overall, the findings show that blastulation is likely to occur in *Bemisia* embryos before they are released from the ovaries. Therefore, traditional microinjection methods for insect transformation to create germline transformation would likely be impossible, and alternative methods of transformation are more likely to yield results (discussed in Section 6).

5.4.4 DM proteins are expressed in early embryo stages of *B. tabaci*

Three *Dmrt* genes have been identified in *B. tabaci* (Figure 3.3). These genes have been given the names; *Btdmrt1 (Hemiptera-specific clade), Btdmrt2 (dmrt99B), Btdmrt3 (dmrt93b)* respectively.

The isoform analysis of *Btdmrt1* shows that most of the samples have a transcript hit (Figure 5.7). The isoform consensus figure (Figure 5.9), shows all the samples have the same isoform. There are no sex-specific isoforms in this gene. In gene expression analysis for *Btdmrt1* shows that the *Btdmrt1* peaks at A egg in the embryogenesis stage. In the embryo stages, the male samples have a higher expression than the females (although this is unknown in D egg). The Male A samples have a higher expression than the male adults; however, female adults have a higher expression over all the samples. The female adults would have embryos inside the abdomen so that the RNA-seq data expression could be falsely increased. To stop this inflation of data expression values, it would be beneficial to gain RNA-seq with females without the abdomens. In the Liu *et al* (2019) paper, the identified *Dsx* gene was my *Btdmrt1*. A gene expression analysis was conducted in this paper on *Dsx* in post-oviposited eggs, larvae and adults. The gene expression heatmap in Liu *et al* (2019) shows the high expression in all stages except lower expression in male adults and first and second instar nymphs. Figure 5.9 shows the gene expression of *Btdmrt1* in pre-oviposited eggs and the adult samples. The only statistically different expression difference is between male B/C eggs and male adults, with the male B/C eggs have a lower expression.

Overall, *Btdmrt1* may be a sex differentiation gene as it has high expression in the A eggs; but it is not statistically different so more samples may need to be taken to ensure a statistically significant result. Therefore, *Btdmrt1* has a high probability of being a sex differentiation gene.

Btdmrt1 would be a good candidate for genetic control technologies; however, it would not be a good candidate for self-limiting technologies as there is no sex-specific isoform.

Btdmrt2 in *B. tabaci* is a mono-exonic transcript. There are not many hits for the eggs, so a reproducible data expression analysis could not be conducted for this gene. This may indicate that the gene is not expressed at all in the early embryos, but it is expressed in the adults. The adults and eggs have the same isoform.

Btdmrt3 is not a mono-exonic gene, which differs from *Btdmrt1* and *Btdmrt2. Btdmrt3* has 7 unique isoforms (Figure 5.13). There is a male-specific isoform present at the embryo stages, and this is coloured orange on Figure 5.13. The male isoform seems to be identical to the standard gene model. In both adult samples both sexes have the orange transcript. To ensure that the orange isoform is male-only, RNA-seq samples need to be collected of females without the abdomen, as we do not know if the transcript that is being picked up is from the male eggs in the abdomen. In female embryo samples there is one isoform that can be analysed, this is coloured red. The red isoform is truncated at 2nd exon, and it is unknown whether this isoform is also in the female D samples. The gene expression analysis shows small expression in the embryos. The peak for *Btdmrt3* in embryos is in the male A egg, and it decreases over time. The female embryo samples do not have any expression in the A egg (which correlates with Figure 5.13), and a slight expression in B/C eggs, although still lower than males. The adult samples both have high expression levels, with the females having the highest expression level overall (which may be due to contamination of eggs in the sample). Overall, *Btdmrt3* has a high probability of being a sex-differentiation gene and could be a potential candidate for genetic control. It is expressed in early embryogenesis (although smaller quantities than adults) and has a potentially male-specific and female-specific isoform in the embryo stages.

5.4.5 Key genes are present in the early embryo stages in *B. tabaci* MED

The summary of the gene expression of *PSI, Imp, Tra2* and *Sxl* is in Figure 5.18. SDGs are known to be expressed during embryogenesis. All SDGs tested are expressed during early embryogenesis, except for *PSI*. *Sxl* is of special interest as it has high expression in the early embryo stages compared to the expression in the adults. *Imp, Tra2* and *Sxl* are potential candidates for self-limiting system. However, isoform analysis needs to be performed along with functionality tests. Liu *et al* (2019) (discussed in Section 1.3.6) explored the gene expression profile of putative SDGs in *B. tabaci* MED at different developmental stages (post-oviposited eggs, larvae, male and female adults). My work slightly compliments this, as it investigates the gene expression of the pre-oviposited eggs, which gives a broader picture of what is occurring in *B. tabaci* SDG through the

life cycle. My work concurs with the same patterns in the heatmap in Liu *et al* (2019) for *Sxl, Tra2* and *Imp*. The main difference between my work and Liu *et al* (2019) is in *PSI*; in pre-oviposited eggs there was no expression found in my transcripts but there was in adults. In the post-oviposited eggs *Psi* is high and decreases in the adults. If I combine both stories it describes a lack of PSI in pre-oviposited eggs, there is expression which peaks in the post-oviposited eggs and gradually decreases in the adults. This suggests that *PSI* starts to be expressed once the eggs are oviposited.

Chapter 6: Discussion

6.1 A summary of the aims of this thesis

B. tabaci is an agricultural pest, with a wide geographical range and resistant to many insecticides. New novel approaches need to be developed to tackle this increasing threat to food security. This thesis aimed to assess the feasibility of the self-limiting system in *B. tabaci* MED. Discussion of the self-limiting system is in Section 1.4.2, and briefly; it is the generation of transgenic insects with a sex-specific lethal transgene. Creation of a self-limiting system in a novel species requires particular knowledge of the organism's biology.

Traditionally, to create transgenic insects, eggs are microinjected with a transgenic element with the hope it will transfer into the germ-cells, and therefore pass the transgenic element into the progeny. These eggs need to be transformed before the blastoderm (and therefore the germ-cells) have been formed. In *B. tabaci,* there is limited knowledge on embryogenesis stages, including the time point in which the blastoderm has formed. One of the main questions that this thesis aimed to answer was; when does the blastoderm form in *B. tabaci*. The answer to this is in the pre-oviposited A eggs.

A construct is needed for the creation of a transgenic insect. An example of this construct is described in Section 1.4. Briefly, for the self-limiting system, this construct requires a sex-specific lethality element. Typically, this element is often a conserved gene, that undergoes sex-specific splicing early in the insect development. Previously this element has been a sex-determination gene. Knowledge of the sex-determination pathway in *B. tabaci* (along with other Hemiptera) is lacking. In this thesis, I aimed to explore the sex-determination genes in *B. tabaci*, and therefore discover potential candidate genes for the self-limiting system.

6.2 Traditional microinjection techniques would not be successful with *B. tabaci*

Traditional microinjection techniques require an injection of a construct into an egg before the germ-cells have formed, which is usually before the blastoderm formation. The construct needs to be injected into the eggs before germline transformation because it will promote DNA transfer into the germline cells. In *Anopheles gambiae* and *Aedes aegypti,* the microinjection period is 1-2.5 hours after oviposition of the eggs (Eggleston, 2014). In *D. melanogaster* this microinjection point is at 1 hour after oviposited eggs (Ringrose, 2009). This critical time point is unknown in *B. tabaci*.

Before this thesis, there was limited knowledge of embryo development in *B. tabaci*. There was knowledge of the *B. tabaci* ovary structure (formed by an ovariole mass) (Figure 1.2), and the morphology of the eggs within the ovaries. There are different morphological development stages of eggs within the ovaries. The different stages of eggs are designated A, B, C and D; with A being the earliest developmental stage and D being the latest (Figure 1.3) (Guo et al., 2010). I saw these stages in Chapter 4 analysis and therefore decided to investigate the developmental stages of these eggs. A comparative developmental stage approach compared the known *D. melanogaster* egg developmental stages to unknown *B. tabaci* egg stages.

D. melanogaster embryogenesis has been well researched as a model organism for insect embryogenesis. *D. melanogaster* embryogenesis starts after oviposition of the egg. There are eight pre-blastoderm nuclear division stages. From nuclear division 0-8, the nuclei are dividing synchronously and are occupying the same cytoplasm (Zalokar and Erk, 1976). In nuclear divisions 8-9, the nuclei start to migrate from the middle of the egg to the outer layer. The outer layer is called the blastoderm. The scuttle fly (*Megaselia abdita*) (Wotton et al., 2014) and the moth midge *(Clogmia albipunctata)* (Jimenez-Guri et al., 2014) had previous unknown developmental stages until a comparative approach with *D. melanogaster*. This success was the reason why I took this approach with *B. tabaci.*

The *B. tabaci* MED nuclei were stained with nuclei binding fluorophore and imaged with confocal microscopy, allowing the nuclei to be imaged and counted. I compared the nuclei numbers in *B. tabaci* against the nuclei numbers in specific nuclear division stages of *D. melanogaster*. The first difference between *D. melanogaster* and *B. tabaci* is that embryogenesis in *B. tabaci* occurs in pre-oviposited eggs. There was autofluorescence in B/C, D and post-oviposited eggs, so therefore only A sample eggs could be imaged. The A sample *B. tabaci* eggs had nuclei that varied from 7 to 9 nuclear divisions in *D. melanogaster* (Table 4.1); this stage is between pre-blastoderm and fully formed blastoderm. There was no difference in the embryogenesis developmental time between the haploid male and diploid female eggs.

I conducted a Z-stack project on the confocal microscope for every egg sample. The Z-stacks created a 3D image; this revealed the morphological features of the eggs. The A sample eggs (Figure 4.7) had a hollow interior, with a single layer of nuclei around the inner outer edge of the egg. This morphology indicates that the nuclei are migrating/ have migrated to form the blastoderm, which concurs with the nuclei counting experiment.

I conducted single egg RNA-seq data for each morphological group of pre-oviposited eggs. This data allowed quantification of expression of genes of interest at the different developmental time points. One gene of interest was the conserved germline marker, called *Vasa*. In *D. melanogaster Vasa* is essential for pole cell formation, germ cell formation and pole plasm formation. Therefore, expression of *Vasa* indicates that the germ cells have already formed. In Chapter 5 the orthologue of *Vasa* was found in *B. tabaci* MED (Figure 5.6). The BLASTP statistics revealed a low e-value between the *D. melanogaster Vasa* and *B. tabaci Vasa.* Protein domain analysis revealed that the *B. tabaci* hit had the same protein domains as *D. melanogaster* VASA*,* providing high probability that this was a true orthologue. The gene expression was quantified for *B. tabaci Vasa* in A, B/C and D sampled eggs. *Vasa* was expressed in all the early embryogenesis stages, which indicates that the blastoderm has already formed and therefore so have the germ-cells. However, for further confirmation of *Vasa* expression the eggs should be stained with antibodies against the *Vasa*. This has been successful in revealing exactly when the pole cells occur in previous papers (Wotton et al., 2014; Jimenez-Guri et al., 2014).

Germ-cell transformation requires the construct to be inside the egg before germ cells have formed. The nuclei counting revealed that A stage eggs are between pre-blastoderm and formation of blastoderm. However, both morphological analyses along with quantification of *Vasa* indicate the A eggs are likely to be at blastoderm formation. Therefore, most eggs that I have sampled would be too late for effective germ-line transformation.

Through personal communication, I know that there have been past attempts at creating germ-line transformation through the traditional microinjection technology. This research was trying to achieve transformation with the CRISPR/CAS9 system, on laid eggs. Only somatic transformation has been achieved, which concurs with my theory that any oviposited egg would be too late for germ-line transformation.

As discussed previously, traditional microinjection techniques would be inappropriate for *B. tabaci* eggs. There is a new CRISPR/CAS9 method that may be more appropriate for *B. tabaci*; this strategy is called the Receptor-Mediated Ovary Transduction of Cargo (ReMOT Control) and has been successful in mosquitoes. This method differs from the traditional microinjection, as the adults themselves are microinjected with a CAS9 ribonucleoprotein complex fused to the arthropod yolk protein. The technique successfully mediated the complex from the female haemolymph to the developing mosquito oocytes and resulted in a heritable gene editing of the

offspring (Chaverra-Rodriguez et al., 2018). I suggest that in future work this technology should be explored in *B. tabaci*.

6.3 Hemiptera do not have DSX but instead have DM containing proteins

Male or female development require double switch genes, these genes are the last in the sex determination pathway (Marin and Baker, 1998; Oliveira et al., 2009). The double-switch gene varies in name depending on the species, in many insects it is the *doublesex (Dsx)*, in the *C. elegans* it is the *male-abnormal-3* (*Mab-3*), and in humans it is *Dmrt1*. In insects, the DSX produces alternatively spliced mRNAs that encode sex-specific polypeptides. The proteins act as transcription factors on downstream sex determination genes, such as yolk protein gene transcription and neuroblast differentiation (Letunic and Bork, 2018).

The DSX in *D. melanogaster* contains two protein domains, the DM domain and DSX dimer. The DM domains are needed for functionality (Erdman and Burtis, 1993), and bind to the DNA's minor groove, which helps coordinate sex and tissue-specific signals (Zhu et al., 2000). The putative hemipteran orthologues of DSX, discovered by RBBH, only have a DM protein domain and not the DSX dimer. The *D. melanogaster* genome contains other genes (apart from *Dsx*) that encode DM (DMRT) protein domains; these genes are *dmrt99b, dmrt93b* and *dmrt11e.* Most DMRTs are related to sex determination and sexual differentiation, and it is suspected that the function of DMRTs are related to neural control of behaviour.

I conducted a BLASTP search in all the hemipteran protein database to find the DM containing proteins (Table 3.5). I took this list of proteins and conducted a phylogenetic tree at the DM domain, of all the hemipteran DM proteins and *D. melanogaster* proteins (Figure 3.3). The hemipteran DM proteins cluster in groups; Hemiptera-specific/ unknown clade, DMRT99B, DMRT11E and DMRT93B. The majority of the RBBH hit fell into the dmrt99B clade. Dmrt99B is expressed in the midline cells of the central nervous system in *D. melanogaster* larvae (Fontana and Crews, 2012) and also affects the mushroom body size in the adult brain (Zwarts et al., 2015). A deficiency in the *B. mori* orthologues of dmrt99b, which is expressed mainly in ovary and brain of the larvae, causes behavioural abnormalities (Kasahara et al., 2018).

The result in Chapter 3, differ from data already published. DSX in *Bemisia tabaci* has been tentatively characterised previously (Guo et al., 2018b). This paper shows that the *Btdsx* has six exons and the encoded protein has a DM and DSX domain. However, the same gene in this paper

describes and entered NCBI (which in this thesis is called *Btdmrt1*) only had one exon and only one protein domain; the DM domain. The *Btdmrt1* clustered in the Hemiptera-specific clade, not in the DSX clade. When this 'DSX' protein was knocked down by silencing the two tail pins of the male genitalia disappeared, and the genitalia was malformed (Guo et al., 2018a). The knockdown morphology indicates that the 'DSX' is needed for genitalia development, which is more of a sex differentiation aspect of DM containing genes. In my opinion, if it were part of the sex determination cascade, the outcome would indicate a difference in sex ratios.

In Chapter 5, the DM containing proteins were further explored. The SDGs are known to be expressed in early embryogenesis stage, and sometimes have sex-specific splicing (Xu et al., 2017). The *Dmrt* genes were investigated to see if they were expressed in the early embryogenesis stage and whether they have sex-specific splicing. Firstly, *Btdmrt1* was explored, this was the gene found in the Hemiptera- specific clade in Figure 3.3. *Btdmrt1* is mono-exonic and there were no splicing events between male and female during different developmental times. *Btdmrt1* was expressed in early embryogenesis. Overall, due to the DM protein domain and the expression in early embryogenesis, this gene has a high probability of being a sex-determination gene.

Btdmrt2 is the RBBH protein from Table 3.2. The *Btdmrt2* was present in the later stages of egg development and adults. It was mono-exonic, and no splicing events seemed to occur in the isoform analysis. The RBBH hit for DSX identified in Chapter 3 was named *Btdmrt3,* this fell into the DMRT11E clade. *Btdmrt3* was not mono-exonic, and there was splicing events occurring in this gene. Firstly, there was a male-specific splicing event occurring in the embryo stages, although this needs to be confirmed by future work. The male-specific event may be an exciting target for genetic control approaches, as *Btdmrt3* is also expressed in the early embryo stages. The *Btdmrt4* was part of the *Btdmrt93b* clade, unfortunately there was zero expression in the gene expression analysis and the RNA-seq data seemed fragmented at this locus, so further analysis was not continued.

6.4 Hemiptera have other sex determination gene orthologues

In the previous section, I explored the lack of DSX in the hemipteran sex determination pathway. In this section, I will explore the other 'key' and 'support' SDGs analysed and highlighted in Chapter 3. This analysis included BLASTP statistical analysis, pairwise comparison of the whole

protein against the original protein query, protein domain investigation and other investigations specialised to that SDG. The summary of all this analysis is presented in Figure 3.33.

Using my criteria in Section 2.1, no unique *A. mellifera* key gene orthologues were found in Hemiptera. This result was surprising as it was thought that the fellow haplodiploid organisms (*B. tabaci* MED and MEAM1) would have these SDGs. The *D. melanogaster* TRA was also not present in any hemipterans. TRA has not been reported in any hemipteran in past studies (Zhuo et al., 2019). The paper by Liu *et* al. did find TRA orthologue in *B. tabaci* MED. However, the e-value threshold was significantly raised, and therefore it was questionable as to whether this was a true orthologue or not. (Zhuo et al., 2019)

MASC was found in all hemipteran genomes, and it had low e-value and the same protein domains as *B. mori* MASC. In the summary table, it indicates that the hemipteran MASC have a low probability of it being an orthologue and having the same function as *B. mori* MASC. The hemipteran MASCs do not have the same two regions that are required for the masculizing effect, Cys-301 and Cys-304, in *B. mori* (Figure 3.6).

PSI, TRA2, IMP and SXL were the sex determination proteins most likely to be true orthologues, in Figure 3.33. *Imp* was investigated with expression analysis in early embryogenesis. There is higher expression in the embryos than there is in the adults. *PSI* is essential to *B. mori* sex determination pathway. IMP and PSI interact with each other. *PSI* has been investigated previously in *B. tabaci* adults, and different splicing event was found in this developmental stage, with 92 female-specific isoforms and 14 male-specific (Liu et al., 2016). Despite this previous research, *PSI* has no gene expression in the early embryogenesis stage, but there is gene expression in the adult stage. Isoform analysis should be performed on the *PSI* adults to see if the same sex-specific isoforms are present in other stages. *Tra2* is present in early embryogenesis, and there is some expression of the gene. However, it is the same in adults. Previous papers found that silencing of the *Tra2* in *B. tabaci* caused malformation in male genitalia (Guo et al., 2018a).

The most exciting 'key' gene studied so far is *Sxl. Sxl* is present in the *D. melanogaster* pathway and is turned off in haplo- X male individuals and turned on in diplo-X female individuals; this causes sexual differentiation and dosage compensation (Cline, 1993). *Sxl* is established and expressed in *D. melanogaster* blastoderm stage (Salz et al., 1989). In the early embryogenesis gene analysis, *Sxl* has the highest expression in embryos. The high expression gives a good

indicator that the gene has something to do with early embryogenesis formation and therefore, a high possibility of being an SDG.

In the 'key' genes, there are some potential targets for genetic control. Therefore, the functionality of these genes need to be explored further. If the *B. tabaci* SDGs are silenced, the loss of function phenotype can indicate a possible function.

6.5 Implications of research findings

The work outlined in this thesis contributes significantly to our understanding of the embryogenesis stages in *B. tabaci* along with potential candidates for the sex determination pathway. The mechanisms that underlie both the sex determination pathway and embryogenesis in *B. tabaci* were not well characterised before this thesis. This lack of knowledge is surprising given that the insect causes large amount of damage to crop species around the world.

The embryogenesis work suggests that the traditional microinjection work would not work for germ-line transformation. Traditional microinjection techniques are labour, time and economically expensive. Therefore, other techniques should be explored in *B. tabaci*. The work in this thesis also highlighted some potential candidates for genetic control. However, more work needs to be conducted to determine the actual function of these orthologues.

Bibliography

ABE, H., MITA, K., YASUKOCHI, Y., OSHIKI, T. & SHIMADA, T. 2005. Retrotransposable elements on the W chromosome of the silkworm, *Bombyx mori*. *Cytogenetic and Genome Research,* 110**,** 144-151.

AGEEP, T. B., DAMIENS, D., ALSHARIF, B., AHMED, A., SALIH, E. H. O., AHMED, F. T. A., DIABATE, A., LEES, R. S., GILLES, J. R. L. & EL SAYED, B. B. 2014. Participation of irradiated *Anopheles arabiensis* males in swarms following field release in Sudan. *Malaria Journal,* 13.

ALPHEY, L. 2014a. Genetic Control of Mosquitoes. *In:* BERENBAUM, M. R. (ed.) *Annual Review of Entomology, Vol 59, 2014.* 205-224

ALPHEY, L. 2014b. Genetic Control of Mosquitoes. *Annual Review of Entomology, Vol 59, 2014,* 59**,** 205-224.

ALPHEY, L., BENEDICT, M., BELLINI, R., CLARK, G. G., DAME, D. A., SERVICE, M. W. & DOBSON, S. L. 2010. Sterile-Insect Methods for Control of Mosquito-Borne Diseases: An Analysis. *Vector-Borne and Zoonotic Diseases,* 10**,** 295-311.

ALPHEY, N., BONSALL, M. B. & ALPHEY, L. 2009. Combining Pest Control and Resistance Management: Synergy of Engineered Insects With Bt Crops. *Journal of Economic Entomology,* 102**,** 717-732.

ALPHEY, N., COLEMAN, P. G., DONNELLY, C. A. & ALPHEY, L. 2007. Managing insecticide resistance by mass release of engineered insects. *Journal of Economic Entomology,* 100**,** 1642-1649.

AN, W. Q., CHO, S. Y., ISHII, H. & WENSINK, P. C. 1996. Sex-specific and non-sex-specific oligomerization domains in both of the doublesex transcription factors from *Drosophila melanogaster*. *Molecular and Cellular Biology,* 16**,** 3106-3111.

ARNÓ, J., GABARRA, R., LIU, T.-X., SIMMONS, A. M. & GERLING, D. 2009. Natural Enemies of *Bemisia tabaci*: Predators and Parasitoids. 385-421.

AYARS, G. H., ALTMAN, L. C., ONEIL, C. E., BUTCHER, B. T. & CHI, E. Y. 1986. Cotton Dust Mediated Lung Epithelial Injury. *Journal of Clinical Investigation,* 78**,** 1579-1588.

BACHTROG, D. 2013. Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. *Nature Review Genetics,* 14**,** 113-24.

BACHTROG, D., MANK, J. E., PEICHEL, C. L., KIRKPATRICK, M., OTTO, S. P., ASHMAN, T. L., HAHN, M. W., KITANO, J., MAYROSE, I., MING, R., PERRIN, N., ROSS, L., VALENZUELA, N., VAMOSI, J. C. & CONSORTIUM, T. S. 2014. Sex Determination: Why So Many Ways of Doing It? *Plos Biology,* 12.

BARBOSA, L. D., YUKI, V. A., MARUBAYASHI, J. M., DE MARCHI, B. R., PERINI, F. L., PAVAN, M. A., DE BARROS, D. R., GHANIM, M., MORIONES, E., NAVAS-CASTILLO, J. & KRAUSE-SAKATE, R. 2015. First report of *Bemisia tabaci* Mediterranean (Q biotype) species in Brazil. *Pest Management Science,* 71**,** 501-504.

BASS, C., PUINEAN, A. M., ZIMMER, C. T., DENHOLM, I., FIELD, L. M., FOSTER, S. P., GUTBROD, O., NAUEN, R., SLATER, R. & WILLIAMSON, M. S. 2014. The evolution of insecticide resistance in the peach potato aphid, *Myzus persicae. Insect Biochemistry and Molecular Biology,* 51**,** 41-51.

BAUMANN, P. 2005. Biology of bacteriocyte-associated endosymbionts of plant sap-sucking insects. *Annual Review of Microbiology,* 59**,** 155-189.

BEAMENT, J. W. L. 1946. The formation and structre of the Chorion of the egg in an Hemipteran, *Rhodnius prolixus*. *Journal of cell science***,** 393-439.

BEDFORD, I. D., BRIDDON, R. W., BROWN, J. K., ROSELL, R. C. & MARKHAM, P. G. 1994. Geminivirus-Transmission and Biological Characterization of *Bemisia-Tabaci* (Gennadius) Biotypes from Different Geographic Regions. *Annals of Applied Biology,* 125**,** 311-325.

BELLINI, R., BALESTRINO, F., MEDICI, A., GENTILE, G., VERONESI, R. & CARRIERI, M. 2013. Mating Competitiveness of *Aedes albopictus* Radio-Sterilized Males in Large Enclosures Exposed to Natural Conditions. *Journal of Medical Entomology,* 50**,** 94-102.

BEN AMI, E., YUVAL, B. & JURKEVITCH, E. 2010. Manipulation of the microbiota of mass-reared Mediterranean fruit flies *Ceratitis capitata* (Diptera: Tephritidae) improves sterile male sexual performance. *Isme Journal,* 4**,** 28-37.

BERTIN, S., LUIGI, M., PARRELLA, G., GIORGINI, M., DAVINO, S. & TOMASSOLI, L. 2018. Survey of the distribution of *Bemisia tabaci* (Hemiptera: Aleyrodidae) in Lazio region (Central Italy): a threat for the northward expansion of Tomato leaf curl New Delhi virus (Begomovirus: Geminiviridae) infection. *Phytoparasitica,* 46**,** 171-182.

BEYE, M., HASSELMANN, M., FONDRK, M. K., PAGE, R. E. & OMHOLT, S. W. 2003. The gene csd is the primary signal for sexual development in the honeybee and encodes an SR-type protein. *Cell,* 114**,** 419-429.

BLACKMAN, R. L. & CAHILL, M. 1998. The karyotype of *Bemisia tabaci* (Hemiptera : Aleyrodidae). *Bulletin of Entomological Research,* 88**,** 213-215.

BONATO, O., COUTON, L. & FARGUES, J. 2006. Feeding preference of *Macrolophus caliginosus* (Heteroptera : Miridae) on *Bemisia tabaci* and *Trialeurodes vaporariorum* (Homoptera : Aleyrodidae). *Journal of Economic Entomology,* 99**,** 1143-1151.

BOURTZIS, K. 2008. *Wolbachia*-Based technologies for insect pest population control. *Transgenesis and the Management of Vector-Borne Disease,* 627**,** 104-113.

BOURTZIS, K., LEES, R. S., HENDRICHS, J. & VREYSEN, M. J. B. 2016. More than one rabbit out of the hat: Radiation, transgenic and symbiont-based approaches for sustainable management of mosquito and tsetse fly populations. *Acta Tropica,* 157**,** 115-130.

BRAAT, A. K., VAN DE WATER, S., GOOS, H., BOGERD, J. & ZIVKOVIC, D. 2000. Vasa protein expression and localization in the zebrafish. *Mechanisms of Development,* 95**,** 271-274.

BRAY, N. L., PIMENTEL, H., MELSTED, P. & PACHTER, L. 2016. Near-optimal probabilistic RNA-seq quantification. *Nature Biotechnoly,* 34**,** 525-7.

BUCKNER, J. S., FREEMAN, T. P., RUUD, R. L., CHU, C. C. & HENNEBERRY, T. J. 2002. Characterization and functions of the whitefly egg pedicel. *Archives of Insect Biochemistry and Physiology,* 49**,** 22-33.

BURTIS, K. C., COSCHIGANO, K. T., BAKER, B. S. & WENSINK, P. C. 1991. The Doublesex Proteins of *Drosophila-Melanogaster* Bind Directly to a Sex-Specific Yolk Protein Gene Enhancer. *Embo Journal,* 10**,** 2577-2582.

BYRNE, D. N. & BELLOWS, T. S. 1991. Whitefly Biology. *Annual Review of Entomology,* 36**,** 431-457.

CALDER, G. 2017. Operation of Zeiss LSM780 confocal microscope. *In:* CENTRE, J. I. (ed.).

CALVITTI, M., REMOTTI, P. C. & CIRIO, U. 2000. The sterile insect technique in the integrated pest management of whitefly species in greenhouses, Pulau Pinang, Malaysia, Penerbit Universiti Sains Malaysia. *CABI.* 185-192

CALVO, F. J., KNAPP, M., VAN HOUTEN, Y. M., HOOGERBRUGGE, H. & BELDA, J. E. 2015. *Amblyseius swirskii*: What made this predatory mite such a successful biocontrol agent? *Experimental and Applied Acarology,* 65**,** 419-433.

CAMPBELL, B. C., STEFFEN-CAMPBELL, J. D. & GILL, R. J. 1994. Evolutionary origin of whiteflies (Hemiptera: Sternorrhyncha: Aleyrodidae) inferred from 18S rDNA sequences. *Insect Molecular Biology,* 3**,** 73-88.

CAMPOS-ORTEGA J.A. & V., H. 1997. A Summary of *Drosophila* Embryogenesis. *The Embryonic Development of Drosophila melanogaster.* Springer, Berlin, Heidelberg.

CAMPOS-ORTEGA, J. A. & HARTENSTEIN, V. 1997. A summary of *Drosophila* embryogenesis. *The Embryonic Development of Drosophila melanogaster.* Springer.

CASTLE, S. J., PALUMBO, J. C., PRABHAKER, N., HOROWITZ, A. R. & DENHOLM, I. 2009. Ecological Determinants of Bemisia tabaci Resistance to Insecticides. 423-465.

CAVALCANTE, A. C. C., BORGES, L. R., LOURENCAO, A. L. & DE MORAES, G. J. 2015. Potential of two populations of *Amblyseius swirskii* (Acari: Phytoseiidae) for the control of *Bemisia tabaci* biotype B (Hemiptera: Aleyrodidae) in Brazil. *Experimental and Applied Acarology,* 67**,** 523-533.

CHARLESWORTH, B. 1996. The evolution of chromosomal sex determination and dosage compensation. *Current Biology,* 6**,** 149-62.

CHAUBEY, R., ANDREW, R. J., NAVEEN, N. C., RAJAGOPAL, R., AHMAD, B. & RAMAMURTHY, V. V. 2015. Morphometric Analysis of Three Putative Species of *Bemisia tabaci* (Hemiptera: Aleyrodidae) Species Complex from India. *Annals of the Entomological Society of America,* 108**,** 600-612.

CHAVERRA-RODRIGUEZ, D., MACIAS, V. M., HUGHES, G. L., PUJHARI, S., SUZUKI, Y., PETERSON, D. R., KIM, D., MCKEAND, S. & RASGON, J. L. 2018. Targeted delivery of CRISPR-Cas9 ribonucleoprotein into arthropod ovaries for heritable germline gene editing. *Nature Communications,* 9.

CHEN, W., HASEGAWA, D. K., ARUMUGANATHAN, K., SIMMONS, A. M., WINTERMANTEL, W. M., FEI, Z. & LING, K.-S. 2015. Estimation of the Whitefly *Bemisia tabaci* Genome Size Based on k-mer and Flow Cytometric Analyses. *Insects,* 6**,** 704-15.

CHEN, W., HASEGAWA, D. K., KAUR, N., KLIOT, A., PINHEIRO, P. V., LUAN, J., STENSMYR, M. C., ZHENG, Y., LIU, W., SUN, H., XU, Y., LUO, Y., KRUSE, A., YANG, X., KONTSEDALOV, S., LEBEDEV, G., FISHER, T. W., NELSON, D. R., HUNTER, W. B., BROWN, J. K., JANDER, G., CILIA, M., DOUGLAS, A. E., GHANIM, M., SIMMONS, A. M., WINTERMANTEL, W. M., LING, K.-S. & FEI, Z. 2016. The draft genome of whitefly *Bemisia tabaci* MEAM1, a global crop pest, provides novel insights into virus transmission, host adaptation, and insecticide resistance. *BMC Biology,* 14**,** 110.

CHIEL, E., GOTTLIEB, Y., ZCHORI-FEIN, E., MOZES-DAUBE, N., KATZIR, N., INBAR, M. & GHANIM, M. 2007. Biotype-dependent secondary symbiont communities in sympatric populations of *Bemisia tabaci*. *Bulletin of Entomological Research,* 97**,** 407-413.

CHOUCHANE, S. G., GORSANE, F., NAKHLA, M. K., MAXWELL, D. P., MARRAKCHI, M. & FAKHFAKH, H. 2007. First report of tomato yellow leaf curl virus-israel species infecting tomato, pepper and bean in Tunisia. *Journal of Phytopathology,* 155**,** 236-240.

CHRISTODOULOU, M. 2011. Biological vector control of mosquito-borne diseases. *Lancet Infectious Diseases,* 11**,** 84-85.

CLINE, T. W. 1988. Evidence That Sisterless-a and Sisterless-B Are 2 of Several Discrete Numerator Elements of the X/a Sex Determination Signal in *Drosophila* That Switch Sxl between 2 Alternative Stable Expression States. *Genetics,* 119**,** 829-862.

CLINE, T. W. 1993. The Drosophila Sex Determination Signal - How Do Flies Count to 2. *Trends in Genetics,* 9**,** 385-390.

CLINE, T. W. & MEYER, B. J. 1996. Vive la Difference: Males vs females in flies vs worms. *Annual Review of Genetics,* 30**,** 637-702.

COHEN, A. C., HENNEBERRY, T. J. & CHU, C. C. 1996. Geometric relationships between whitefly feeding behavior and vascular bundle arrangements. *Entomologia Experimentalis Et Applicata,* 78**,** 135-142.

COOK, J. M. 1993. Sex Determination in the Hymenoptera - a Review of Models and Evidence. *Heredity,* 71**,** 421-435.

COSTA, H. S., HENNEBERRY, T. J. & TOSCANO, N. C. 1997. Effects of antibacterial materials on *Bemisia argentifolii* (Homoptera: Aleyrodidae) oviposition, growth, survival, and sex ratio. *Journal of Economic Entomology,* 90**,** 333-339.

COSTA, H. S., TOSCANO, N. C. & HENNEBERRY, T. J. 1996. Mycetocyte inclusion in the oocytes of *Bemisia argentifolii* (Homoptera: Aleyrodidae). *Annals of the Entomological Society of America,* 89**,** 694-699.

CRANSTON, P. S. & GULLAN, P. J. 2009. Chapter 199 - Phylogeny of Insects. *In:* RESH, V. H. & CARDÉ, R. T. (eds.) *Encyclopedia of Insects (Second Edition).*

CRAVES, J. A. M. 2008. Weird Animal Genomes and the Evolution of Vertebrate Sex and Sex Chromosomes. *Annual Review of Genetics,* 42**,** 565-586.

CURTIS, C. F. 1985. GENETIC-CONTROL OF INSECT PESTS - GROWTH INDUSTRY OR LEAD BALLOON. *Biological Journal of the Linnean Society,* 26**,** 359-374.

CUTHBERTSON, A. G. S., BLACKBURN, L. F., EYRE, D. P., CANNON, R. J. C., MILLER, J. & NORTHING, P. 2011. *Bemisia tabaci*: The current situation in the UK and the prospect of developing strategies for eradication using entomopathogens. *Insect Science,* 18**,** 1-10.

CUTHBERTSON, A. G. S., BUXTON, J. H., BLACKBURN, L. F., MATHERS, J. J., ROBINSON, K. A., POWELL, M. E., FLEMING, D. A. & BELL, H. A. 2012. Eradicating *Bemisia tabaci* Q biotype on poinsettia plants in the UK. *Crop Protection,* 42**,** 42-48.

CUTHBERTSON, A. G. S. & VANNINEN, I. 2015. The Importance of Maintaining Protected Zone Status against *Bemisia tabaci*. *Insects,* 6**,** 432-441.

CZOSNEK, H. & BROWN, J. K. 2009. The Whitefly Genome – White Paper: A Proposal to Sequence Multiple Genomes of Bemisia tabaci. 503-532.

DAVIS, R. L. & TURNER, D. L. 2001. Vertebrate hairy and Enhancer of split related proteins: transcriptional repressors regulating cellular differentiation and embryonic patterning. *Oncogene,* 20**,** 8342-8357.

DE BARRO, P. J., LIU, S.-S., BOYKIN, L. M. & DINSDALE, A. B. 2011a. *Bemisia tabaci*: A Statement of Species Status. *In:* BERENBAUM, M. R., CARDE, R. T. & ROBINSON, G. E. (eds.) *Annual Review of Entomology, Vol 56.*

DE BARRO, P. J., LIU, S. S., BOYKIN, L. M. & DINSDALE, A. B. 2011b. *Bemisia tabaci*: A Statement of Species Status. *Annual Review of Entomology, Vol 56,* 56**,** 1-19.

DEARDEN, P., GRBIC, M. & DONLY, C. 2003. Vasa expression and germ-cell specification in the spider mite Tetranychus urticae. *Development Genes and Evolution,* 212**,** 599-603.

DEMIR, E. & DICKSON, B. J. 2005. fruitless splicing specifies male courtship behavior in *Drosophila. Cell,* 121**,** 785-794.

DIMMIC, M. W., REST, J. S., MINDELL, D. P. & GOLDSTEIN, R. A. 2002. rtREV: An amino acid substitution matrix for inference of retrovirus and reverse transcriptase phylogeny. *Journal of Molecular Evolution,* 55**,** 65-73.

DITTRICH, V., HASSAN, S. O. & ERNST, G. H. 1985. Sudanese Cotton and the Whitefly - a Case-Study of the Emergence of a New Primary Pest. *Crop Protection,* 4**,** 161-176.

DOBIN, A., DAVIS, C. A., SCHLESINGER, F., DRENKOW, J., ZALESKI, C., JHA, S., BATUT, P., CHAISSON, M. & GINGERAS, T. R. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics,* 29**,** 15-21.

DUBENDORFER, A., HEDIGER, M., BURGHARDT, G. & BOPP, D. 2002. *Musca domestica,* a window on the evolution of sex-determining mechanisms in insects. *International Journal of Developmental Biology,* 46**,** 75-79.

DYE, C. 1984. Models for the Population-Dynamics of the Yellow-Fever Mosquito, Aedes-Aegypti. *Journal of Animal Ecology,* 53**,** 247-268.

EGGLESTON, P. M., J. M. 2014. *Docking systems for site-directed transgene integration,* Wallingford, UK, CABI.

ELLIS, H. M., SPANN, D. R. & POSAKONY, J. W. 1990. Extramacrochaetae, a Negative Regulator of Sensory Organ Development in *Drosophila*, Defines a New Class of Helix-Loop-Helix Proteins. *Cell,* 61**,** 27-38.

ERDMAN, S. E. & BURTIS, K. C. 1993. The *Drosophila* Doublesex Proteins Share a Novel Zinc Finger Related DNA-Binding Domain. *Embo Journal,* 12**,** 527-535.

ERICKSON, J. W. & QUINTERO, J. J. 2007. Indirect effects of ploidy suggest X chromosome dose, not the X : A ratio, signals sex in Drosophila. *Plos Biology,* 5**,** 2821-2830.

EVANS, B. R., KOTSAKIOZI, P., COSTA-DA-SILVA, A. L., IOSHINO, R. S., GARZIERA, L., PEDROSA, M. C., MALAVASI, A., VIRGINIO, J. F., CAPURRO, M. L. & POWELL, J. R. 2019. Transgenic *Aedes aegypti* Mosquitoes Transfer Genes into a Natural Population. *Scientific Reports,* 9.

FARBOUD, B., NIX, P., JOW, M. M., GLADDEN, J. M. & MEYER, B. J. 2013. Molecular antagonism between X-chromosome and autosome signals determines nematode sex. *Genes & Development,* 27**,** 1159-1178.

FARRELL, J. A. & O'FARRELL, P. H. 2014. From Egg to Gastrula: How the Cell Cycle Is Remodeled During the *Drosophila* Mid-Blastula Transition. *Annual Review of Genetics, Vol 48,* 48**,** 269-294.

FELDHAAR, H. 2011. Bacterial symbionts as mediators of ecologically important traits of insect hosts. *Ecological Entomology,* 36**,** 533-543.

FELSENSTEIN, J. 1985. Confidence-Limits on Phylogenies - an Approach Using the Bootstrap. *Evolution,* 39**,** 783-791.

FINLEY, K. D., EDEEN, P. T., FOSS, M., GROSS, E., GHBEISH, N., PALMER, R. H., TAYLOR, B. J. & MCKEOWN, M. 1998. dissatisfaction encodes a tailless-like nuclear receptor expressed in a subset of CNS neurons controlling *Drosophila* sexual behavior. *Neuron,* 21**,** 1363-1374.

FOE, V. E. & ALBERTS, B. M. 1983. Studies of Nuclear and Cytoplasmic Behavior during the 5 Mitotic-Cycles That Precede Gastrulation in *Drosophila* Embryogenesis. *Journal of Cell Science,* 61**,** 31-70.

FONTANA, J. R. & CREWS, S. T. 2012. Transcriptome analysis of *Drosophila* CNS midline cells reveals diverse peptidergic properties and a role for castor in neuronal differentiation. *Developmental Biology,* 372**,** 131-142.

FU, G. L., CONDON, K. C., EPTON, M. J., GONG, P., JIN, L., CONDON, G. C., MORRISON, N. I., DAFA'ALLA, T. H. & ALPHEY, L. 2007. Female-specific insect lethality engineered using alternative splicing. *Nature Biotechnology,* 25**,** 353-357.

FU, G. L., LEES, R. S., NIMMO, D., AW, D., JIN, L., GRAY, P., BERENDONK, T. U., WHITE-COOPER, H., SCAIFE, S., PHUC, H. K., MARINOTTI, O., JASINSKIENE, N., JAMES, A. A. & ALPHEY, L. 2010. Female-specific flightless phenotype for mosquito control. *Proceedings of the National Academy of Sciences of the United States of America,* 107**,** 4550-4554.

FUKUI, T., KAWAMOTO, M., SHOJI, K., KIUCHI, T., SUGANO, S., SHIMADA, T., SUZUKI, Y. & KATSUMA, S. 2015. The Endosymbiotic Bacterium Wolbachia Selectively Kills Male Hosts by Targeting the Masculinizing Gene. *Plos Pathogens,* 11.

GAILEY, D. A., BILLETER, J. C., LIU, J. H., BAUZON, F., ALLENDORFER, J. B. & GOODWIN, S. F. 2006. Functional conservation of the fruitless male sex-determination gene across 250 Myr of insect evolution. *Molecular Biology and Evolution,* 23**,** 633-643.

GATEHOUSE, J. A. 2002. Plant resistance towards insect herbivores: a dynamic interaction. *New Phytologist,* 156**,** 145-169.

GELMAN, D. B., BLACKBURN, M. B., HU, J. S. & GERLING, D. 2002. The nymphal-adult molt of the silverleaf whitefly (*Bemisia argentifolii*): Timing, regulation, and progress. *Archives of Insect Biochemistry and Physiology,* 51**,** 67-79.

GEMPE, T., HASSELMANN, M., SCHIOTT, M., HAUSE, G., OTTE, M. & BEYE, M. 2009. Sex Determination in Honeybees: Two Separate Mechanisms Induce and Maintain the Female Pathway. *Plos Biology,* 7.

GHANIM, M., ROSELL, R. C., CAMPBELL, L. R., CZOSNEK, H., BROWN, J. K. & ULLMAN, D. E. 2001. Digestive, salivary, and reproductive organs of *Bemisia tabaci* (Gennadius) (Hemiptera : Aleyrodidae)B type. *Journal of Morphology,* 248**,** 22-40.

GIRALDO-CALDERON, G. I., EMRICH, S. J., MACCALLUM, R. M., MASLEN, G., DIALYNAS, E., TOPALIS, P., HO, N., GESING, S., MADEY, G., COLLINS, F. H., LAWSON, D. & CONSORTIUM, V. 2015. VectorBase: an updated bioinformatics resource for invertebrate vectors and other organisms related with human diseases. *Nucleic Acids Research,* 43**,** D707-D713.

GOLDMAN, V. & CZOSNEK, H. 2002. Whiteflies (Bemisia tabaci) issued from eggs bombarded with infectious DNA clones of Tomato yellow leaf curl virus from Israel (TYLCV) are able to infect tomato plants. *Archives of Virology,* 147**,** 787-801.

GOTTLIEB, Y., GHANIM, M., CHIEL, E., GERLING, D., PORTNOY, V., STEINBERG, S., TZURI, G., HOROWITZ, A. R., BELAUSOV, E., MOZES-DAUBE, N., KONTSEDALOV, S., GERSHON, M., GAL, S., KATZIR, N. & ZCHORI-FEIN, E. 2006. Identification and localization of a *Rickettsia* sp in *Bemisia tabaci* (Homoptera : Aleyrodidae). *Applied and Environmental Microbiology,* 72**,** 3646-3652.

GRAMATES, L. S., MARYGOLD, S. J., DOS SANTOS, G., URBANO, J. M., ANTONAZZO, G., MATTHEWS, B. B., REY, A. J., TABONE, C. J., CROSBY, M. A., EMMERT, D. B., FALLS, K., GOODMAN, J. L., HU, Y. H., PONTING, L., SCHROEDER, A. J., STRELETS, V. B., THURMOND, J., ZHOU, P. L. & CONSORTIUM, F. 2017. FlyBase at 25: looking to the future. *Nucleic Acids Research,* 45**,** D663-D671.

GRANADINO, B., CAMPUZANO, S. & SANCHEZ, L. 1990. The Drosophila-Melanogaster Fl(2)D Gene Is Needed for the Female-Specific Splicing of Sex-Lethal Rna. *Embo Journal,* 9**,** 2597-2602.

GUEGUEN, G., VAVRE, F., GNANKINE, O., PETERSCHMITT, M., CHARIF, D., CHIEL, E., GOTTLIEB, Y., GHANIM, M., ZCHORI-FEIN, E. & FLEURY, F. 2010. Endosymbiont metacommunities, mtDNA diversity and the evolution of the *Bemisia tabaci* (Hemiptera: Aleyrodidae) species complex. *Molecular Ecology,* 19**,** 4365-4378.

GUO, J. Y., WAN, F. H. & YE, G. Y. 2016. Oogenesis in the *Bemisia tabaci* MEAM1 species complex. *Micron,* 83**,** 1-10.

GUO, J. Y., YE, G. Y., DONG, S. Z. & LIU, S. S. 2010. An Invasive Whitefly Feeding on a Virus-Infected Plant Increased Its Egg Production and Realized Fecundity. *Plos One,* 5.

GUO, L., W., XIE, W., LIU, Y., YANG, Z., YANG, X., XIA, J., WANG, S., WU, Q. & ZHANG, Y. 2018a. Identification and characterization of doublesex in *Bemisia tabaci*. *Insect molecular biology,* 00.

GUO, L., XIE, W., LIU, Y., YANG, Z., YANG, X., XIA, J., WANG, S., WU, Q. & ZHANG, Y. 2018b. Identification and characterization of doublesex in *Bemisia tabaci. Insect Molecular Biology,* 27**,** 620-632.

HANCOCK, P. A., SINKINS, S. P. & GODFRAY, H. C. J. 2011. Strategies for Introducing *Wolbachia* to Reduce Transmission of Mosquito-Borne Diseases. *Plos Neglected Tropical Diseases,* 5.

HANDA, N., NUREKI, O., KURIMOTO, K., KIM, I., SAKAMOTO, H., SHIMURA, Y., MUTO, Y. & YOKOYAMA, S. 1999. Structural basis for recognition of the tra mRNA precursor by the sex-lethal protein. *Nature,* 398**,** 579-585.

HANSEN, D. & PILGRIM, D. 1999. Sex and the single worm: sex determination in the nematode *C-elegans*. *Mechanisms of Development,* 83**,** 3-15.

HARGREAVES, K., KOEKEMOER, L. L., BROOKE, B. D., HUNT, R. H., MTHEMBU, J. & COETZEE, M. 2000. *Anopheles funestus* resistant to pyrethroid insecticides in South Africa. *Medical and Veterinary Entomology,* 14**,** 181-189.

HASSELMANN, M., GEMPE, T., SCHIOTT, M., NUNES-SILVA, C. G., OTTE, M. & BEYE, M. 2008. Evidence for the evolutionary nascence of a novel sex determination pathway in honeybees. *Nature,* 454**,** 519-U7.

HE, Y., CAO, X. L., LI, K., HU, Y. X., CHEN, Y. R., BLISSARD, G., KANOST, M. R. & JIANG, H. B. 2015. A genome-wide analysis of antimicrobial effector genes and their transcription patterns in Manduca sexta. *Insect Biochemistry and Molecular Biology,* 62**,** 23-37.

HERPIN, A. & SCHARTL, M. 2015. Plasticity of gene-regulatory networks controlling sex determination: of masters, slaves, usual suspects, newcomers, and usurpators. *Embo Reports,* 16**,** 1260-1274.

HODGKIN, J. 2002. The remarkable ubiquity of DM domain factors as regulators of sexual phenotype: ancestry or aptitude? *Genes & Development,* 16**,** 2322-2326.

HOFF, M. 2009. Male or Female? For Honeybees, a Single Gene Makes All the Difference. *Plos Biology,* 7.
HOGENHOUT, S. A. & BOS, J. 2011a. Effector proteins that modulate plant–insect interactions. *C Current Opinion in Plant Biology,* 14**,** 1-7.
HOGENHOUT, S. A. & BOS, J. I. 2011b. Effector proteins that modulate plant--insect interactions. *Current Opinion in Plant Biology,* 14**,** 422-8.
HOGENHOUT, S. A. & BOS, J. I. B. 2011c. Effector proteins that modulate plant-insect interactions. *Current Opinion in Plant Biology,* 14**,** 422-428.
HOGENHOUT, S. A., VAN DER HOORN, R. A., TERAUCHI, R. & KAMOUN, S. 2009. Emerging concepts in effector biology of plant-associated organisms. *Molecular Plant Microbe Interactions,* 22**,** 115-22.
HOROWITZ, A. R. & ISHAAYA, I. 2014. Dynamics of biotypes B and Q of the whitefly *Bemisia tabaci* and its impact on insecticide resistance. *Pest Management Science,* 70**,** 1568-1572.
HOROWITZ, A. R., KONTSEDALOV, S., KHASDAN, V. & ISHAAYA, I. 2005. Biotypes B and Q of Bemisia tabaci and their relevance to neonicotinoid and pyriproxyfen resistance. *Archives of Insect Biochemistry and Physiology,* 58**,** 216-225.
INSTITUTE, B. 2019. *Picard Tools, version 2.1.1* [Online]. [Accessed 05/03/2019].
IWASAKI, Y. W., SIOMI, M. C. & SIOMI, H. 2015. PIWI-Interacting RNA: Its Biogenesis and Functions. *Annual Review of Biochemistry, Vol 84,* 84**,** 405-433.
JIMENEZ-GURI, E., WOTTON, K. R., GAVILAN, B. & JAEGER, J. 2014. A Staging Scheme for the Development of the Moth Midge *Clogmia albipunctata*. *Plos One,* 9.
JOHNSON, M. W., TOSCANO, N.C., REYNOLDS, H.T., SYLVESTER, E.S., , KIDO, K., NATWICK, E.T 1982. Whiteflies cause problems for southern California growers. *California Agriculture*.
JONES, D. R. 2003. Plant viruses transmitted by whiteflies. *European Journal of Plant Pathology,* 109**,** 195-219.
JURSNICH, V. A. & BURTIS, K. C. 1993. A Positive Role in Differentiation for the Male Doublesex Protein of *Drosophila*. *Developmental Biology,* 155**,** 235-249.
KARANDIKAR, U. C., SHAFFER, J., BISHOP, C. P. & BIDWAI, A. P. 2005. Drosophila CK2 phosphorylates Deadpan, a member of the HES family of basic-helix-loop-helix (bHLH) repressors. *Molecular and Cellular Biochemistry,* 274**,** 133-139.
KARUT, K. & TOK, B. 2014. Secondary endosymbionts of Turkish Bemisia tabaci (Gennadius) populations. *Phytoparasitica,* 42**,** 413-419.
KASAHARA, R., AOKI, F. & SUZUKI, M. G. 2018. Deficiency in dmrt99B ortholog causes behavioral abnormalities in the silkworm, Bombyx moril. *Applied Entomology and Zoology,* 53**,** 381-393.
KATO, Y., KOBAYASHI, K., ODA, S., COLBOURN, J. K., TATARAZAKO, N., WATANABE, H. & IGUCHI, T. 2008. Molecular cloning and sexually dimorphic expression of DM-domain genes in *Daphnia magna*. *Genomics,* 91**,** 94-101.
KATSUMA, S., SUGANO, Y., KIUCHI, T. & SHIMADA, T. 2015. Two Conserved Cysteine Residues Are Required for the Masculinizing Activity of the Silkworm Masc Protein. *Journal of Biological Chemistry,* 290**,** 26114-26124.
KAWAOKA, S., KADOTA, K., ARAI, Y., SUZUKI, Y., FUJII, T., ABE, H., YASUKOCHI, Y., MITA, K., SUGANO, S., SHIMIZU, K., TOMARI, Y., SHIMADA, T. & KATSUMA, S. 2011. The silkworm W chromosome is a source of female-enriched piRNAs. *Rna-a Publication of the Rna Society,* 17**,** 2144-2151.
KIM, S. S., KETTLEWELL, J. R., ANDERSON, R. C., BARDWELL, V. J. & ZARKOWER, D. 2003. Sexually dimorphic expression of multiple doublesex-related genes in the embryonic mouse gonad. *Gene Expression Patterns,* 3**,** 77-82.
KIUCHI, T., KOGA, H., KAWAMOTO, M., SHOJI, K., SAKAI, H., ARAI, Y., ISHIHARA, G., KAWAOKA, S., SUGANO, S., SHIMADA, T., SUZUKI, Y., SUZUKI, M. G. & KATSUMA, S. 2014. A single

female-specific piRNA is the primary determiner of sex in the silkworm. *Nature,* 509**,** 633-+.

KIUCHI, T., SUGANO, Y., SHIMADA, T. & KATSUMA, S. 2019. Two CCCH-type zinc finger domains in the Masc protein are dispensable for masculinization and dosage compensation in *Bombyx mori*. *Insect Biochemistry Molecular Biology,* 104**,** 30-38.

KNIPLING, E. F. 1959. Sterile-Male Method of Population Control. *Science,* 130**,** 902-904.

KUMAR, S., STECHER, G. & TAMURA, K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular Biology and Evolution,* 33**,** 1870-1874.

KUNTE, K., ZHANG, W., TENGER-TROLANDER, A., PALMER, D. H., MARTIN, A., REED, R. D., MULLEN, S. P. & KRONFORST, M. R. 2014. doublesex is a mimicry supergene. *Nature,* 507**,** 229-+.

LABBE, G. M. C., SCAIFE, S., MORGAN, S. A., CURTIS, Z. H. & ALPHEY, L. 2012. Female-Specific Flightless (fsRIDL) Phenotype for Control of *Aedes albopictus*. *Plos Neglected Tropical Diseases,* 6.

LASKO, P. F. & ASHBURNER, M. 1988. The Product of the *Drosophila* Gene Vasa Is Very Similar to Eukaryotic Initiation Factor-4a. *Nature,* 335**,** 611-617.

LAUZON, C. R. & POTTER, S. E. 2012. Description of the irradiated and nonirradiated midgut of *Ceratitis capitata* Wiedemann (Diptera: Tephritidae) and *Anastrepha ludens* Loew (Diptera: Tephritidae) used for sterile insect technique. *Journal of Pest Science,* 85**,** 217-226.

LE, S. Q. & GASCUEL, O. 2008. An improved general amino acid replacement matrix. *Molecular Biology and Evolution,* 25**,** 1307-1320.

LEE, J., KIUCHI, T., KAWAMOTO, M., SHIMADA, T. & KATSUMA, S. 2015. Identification and functional analysis of a Masculinizer orthologue in *Trilocha varians* (Lepidoptera: Bombycidae). *Insect Molecular Biology,* 24**,** 561-569.

LETUNIC, I. & BORK, P. 2016. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Research,* 44**,** W242-W245.

LETUNIC, I. & BORK, P. 2018. 20 years of the SMART protein domain annotation resource. *Nucleic Acids Research,* 46**,** D493-D496.

LI, H. & BAKER, B. S. 1998. hermaphrodite and doublesex function both dependently and independently to control various aspects of sexual differentiation in *Drosophila*. *Development,* 125**,** 2641-2651.

LI, H., HANDSAKER, B., WYSOKER, A., FENNELL, T., RUAN, J., HOMER, N., MARTH, G., ABECASIS, G., DURBIN, R. & PROC, G. P. D. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics,* 25**,** 2078-2079.

LI, T. Y., VINSON, S. B. & GERLING, D. 1989. Courtship and Mating-Behavior of Bemisia-Tabaci (Homoptera, Aleyrodidae). *Environmental Entomology,* 18**,** 800-806.

LIU, Y., XIE, W., GUO, L., YANG, X., YANG, J., WANG, S., WU, Q., ZHOU, X. & ZHANG, Y. 2019. Genome-wide dissection of sex determination genes in the highly invasive whitefly species *Bemisia tabaci* Q/MED. *Insect Molecular Biology*.

LIU, Y. T., XIE, W., YANG, X., GUO, L. T., WANG, S. L., WU, Q. J., YANG, Z. Z., ZHOU, X. G. & ZHANG, Y. J. 2016. Molecular cloning of the sex-related gene PSI in Bemisia tabaci and its alternative splicing properties. *Gene,* 580**,** 104-110.

LUCAS, E. & ALOMAR, O. 2001. Macrolophus caliginosus (Wagner) as an intraguild prey for the zoophytophagous Dicyphus tamaninii Wagner (Heteroptera : Miridae). *Biological Control,* 20**,** 147-152.

MAEKI, K. 1981. Notes on the W-Chromosome of the Butterfly, *Graphium Sarpedon* (Papilionidae, Lepidoptera). *Proceedings of the Japan Academy Series B-Physical and Biological Sciences,* 57**,** 371-373.

MARIN, I. & BAKER, B. S. 1998. The evolutionary dynamics of sex determination. *Science,* 281**,** 1990-1994.

MAROIS, E., SCALI, C., SOICHOT, J., KAPPLER, C., LEVASHINA, E. A. & CATTERUCCIA, F. 2012. High-throughput sorting of mosquito larvae for laboratory studies and for future vector control interventions. *Malaria Journal,* 11.

MARUBAYASHI, J. M., YUKI, V. A., ROCHA, K. C. G., MITUTI, T., PELEGRINOTTI, F. M., FERREIRA, F. Z., MOURA, M. F., NAVAS-CASTILLO, J., MORIONES, E., PAVAN, M. A. & KRAUSE-SAKATE, R. 2013. At least two indigenous species of the *Bemisia tabaci* complex are present in Brazil. *Journal of Applied Entomology,* 137**,** 113-121.

MATHERS, T. C. 2019. Improved genome assembly and annotation of the soybean aphid (Aphis glycines Matsumura). *bioRxiv***,** 781617.

MATHERS, T. C., CHEN, Y. Z., KAITHAKOTTIL, G., LEGEAI, F., MUGFORD, S. T., BAA-PUYOULET, P., BRETAUDEAU, A., CLAVIJO, B., COLELLA, S., COLLIN, O., DALMAY, T., DERRIEN, T., FENG, H. L., GABALDON, T., JORDAN, A., JULCA, I., KETTLES, G. J., KOWITWANICH, K., LAVENIER, D., LENZI, P., LOPEZ-GOMOLLON, S., LOSKA, D., MAPLESON, D., MAUMUS, F., MOXON, S., PRICE, D. R. G., SUGIO, A., VAN MUNSTER, M., UZEST, M., WAITE, D., JANDER, G., TAGU, D., WILSON, A. C. C., VAN OOSTERHOUT, C., SWARBRECK, D. & HOGENHOUT, S. A. 2017. Rapid transcriptional plasticity of duplicated gene clusters enables a clonally reproducing aphid to colonise diverse plant species (vol 18, 27, 2017). *Genome Biology,* 18.

MATHERS, T. C., MUGFORD, S. T., PERCIVAL-ALWYN, L., CHEN, Y., KAITHAKOTTIL, G., SWARBRECK, D., HOGENHOUT, S. A. & VAN OOSTERHOUT, C. 2018. Sex-specific changes in the aphid DNA methylation landscape. *bioRxiv***,** 286302.

MCKENNA, A., HANNA, M., BANKS, E., SIVACHENKO, A., CIBULSKIS, K., KERNYTSKY, A., GARIMELLA, K., ALTSHULER, D., GABRIEL, S., DALY, M. & DEPRISTO, M. A. 2010. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research,* 20**,** 1297-1303.

MCLAREN, A. 1991. Sex Determination - the Making of Male-Mice. *Nature,* 351**,** 96-96.

MORGAN, T. H. 1909. Sex Determination and Parthenogenesis in Phylloxerans and Aphids *Science,* 29**,** 4.

MORIONES, E. & NAVAS-CASTILLO, J. 2000. Tomato yellow leaf curl virus, an emerging virus complex causing epidemics worldwide. *Virus Research,* 71**,** 123-34.

MORIONES, E. & NAVAS-CASTILLO, J. 2009. Tomato Yellow Leaf Curl Disease Epidemics. 259-282.

MUNOZ, D., JIMENEZ, A., MARINOTTI, O. & JAMES, A. A. 2004. The AeAct-4 gene is expressed in the developing flight muscles of female *Aedes aegypt*i. *Insect Molecular Biology,* 13**,** 563-568.

NAUEN, R. & DENHOLM, I. 2005. Resistance of insect pests to neonicotinoid insecticides: Current status and future prospects. *Archives of Insect Biochemistry and Physiology,* 58**,** 200-215.

NODA, H. T., R. 1990. Re-examination of chromosomes of three species of rice planthoppers (Homoptera: Delphacidae). *Applied Entomology and Zoology* 25**,** 538-540.

NORMARK, B. B. 2003. The evolution of alternative genetic systems in insects. *Annual Review of Entomology,* 48**,** 397-423.

O'FARRELL, P. H., STUMPFF, J. & SU, T. T. 2004. Embryonic cleavage cycles: How is a mouse like a fly? *Current Biology,* 14**,** R35-R45.

OHBAYASHI, F., SUZUKI, M. G., MITA, K., OKANO, K. & SHIMADA, T. 2001. A homologue of the *Drosophila* doublesex gene is transcribed into sex-specific mRNA isoforms in the silkworm, *Bombyx mori. Comparative Biochemistry and Physiology B-Biochemistry & Molecular Biology,* 128**,** 145-158.

OLIVEIRA, D. C. S. G., WERREN, J. H., VERHULST, E. C., GIEBEL, J. D., KAMPING, A., BEUKEBOOM, L. W. & VAN DE ZANDE, L. 2009. Identification and characterization of the doublesex gene of *Nasonia*. *Insect Molecular Biology,* 18**,** 315-324.

OLIVEIRA, M. R. V., HENNEBERRY, T. J. & ANDERSON, P. 2001. History, current status, and collaborative research projects for *Bemisia tabaci. Crop Protection,* 20**,** 709-723.

OTTO, S. P. 2009. The evolutionary enigma of sex. *American Naturalist,* 174 Suppl 1**,** S1-S14.

OXITEC. 2019. *Oxitec responds to article entitled 'Transgenic Aedes aegypti Mosquitoes Transfer Genes into a Natural Population'* [Online]. Available: https://www.oxitec.com/en/news/oxitec-response-scientific-reports-article [Accessed 18 September 2019].

PANE, A., SALVEMINI, M., BOVI, P. D., POLITO, C. & SACCONE, G. 2002. The transformer gene in Ceratitis capitata provides a genetic basis for selecting and remembering the sexual fate. *Development,* 129**,** 3715-3725.

PANZERA, Y., PITA, S., FERREIRO, M. J., FERRANDIS, I., LAGES, C., PEREZ, R., SILVA, A. E., GUERRA, M. & PANZERA, F. 2012. High Dynamics of rDNA Cluster Location in Kissing Bug Holocentric Chromosomes (Triatominae, Heteroptera). *Cytogenetic and Genome Research,* 138**,** 56-67.

PAPATHANOS, P. A., WINDBICHLER, N., MENICHELLI, M., BURT, A. & CRISANTI, A. 2009. The vasa regulatory region mediates germline expression and maternal transmission of proteins in the malaria mosquito Anopheles gambiae: a versatile tool for genetic control strategies. *Bmc Molecular Biology,* 10.

PAROUSH, Z., FINLEY, R. L., KIDD, T., WAINWRIGHT, S. M., INGHAM, P. W., BRENT, R. & ISHHOROWICZ, D. 1994. Groucho Is Required for *Drosophila* Neurogenesis, Segmentation, and Sex Determination and Interacts Directly with Hairy-Related Bhlh Proteins. *Cell,* 79**,** 805-815.

PARTRIDGE, L. 1983. The Masterpiece of Nature - the Evolution and Genetics of Sexuality - Bell,G. *Animal Behaviour,* 31**,** 961-961.

PENALVA, L. O. F., RUIZ, M. F., ORTEGA, A., GRANADINO, B., VICENTE, L., SEGARRA, C., VALCARCEL, J. & SANCHEZ, L. 2000. The *Drosophila* fl(2)d gene, required for female-specific splicing of Sxl and tra pre-mRNAs, encodes a novel nuclear protein with a HQ-rich domain (vol 155, pg 129, 2000). *Genetics,* 155**,** 2021-2021.

PERTEA, M., KIM, D., PERTEA, G. M., LEEK, J. T. & SALZBERG, S. L. 2016. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nature Protocols,* 11**,** 1650-1667.

PERTEA, M., PERTEA, G. M., ANTONESCU, C. M., CHANG, T. C., MENDELL, J. T. & SALZBERG, S. L. 2015. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature Biotechnology,* 33**,** 290-5.

PICELLI, S., FARIDANI, O. R., BJORKLUND, A. K., WINBERG, G., SAGASSER, S. & SANDBERG, R. 2014. Full-length RNA-seq from single cells using Smart-seq2. *Nature Protocols,* 9**,** 171-181.

RACHAMIM, T., MORGENSTERN, D., AHARONOVICH, D., BREKHMAN, V., LOTAN, T. & SHER, D. 2015. The Dynamically Evolving Nematocyst Content of an Anthozoan, a Scyphozoan, and a Hydrozoan. *Molecular Biology and Evolution,* 32**,** 740-753.

RAYMOND, C. S., SHAMU, C. E., SHEN, M. M., SEIFERT, K. J., HIRSCH, B., HODGKIN, J. & ZARKOWER, D. 1998. Evidence for evolutionary conservation of sex-determining genes. *Nature,* 391**,** 691-695.

REINHARDT, K., BREUNIG, H. G. & KONIG, K. 2017. Autofluorescence lifetime variation in the cuticle of the bedbug *Cimex lectularius*. *Arthropod Structure & Development,* 46**,** 56-62.

RENDON, P., MCINNIS, D., LANCE, D. & STEWART, J. 2004. Medfly (Diptera : Tephritidae) genetic sexing: Large-scale field comparison of males-only and bisexual sterile fly releases in Guatemala. *Journal of Economic Entomology,* 97**,** 1547-1553.

RICHARDS, S., GIBBS, R. A., GERARDO, N. M., MORAN, N., NAKABACHI, A., STERN, D., TAGU, D., WILSON, A. C. C., MUZNY, D., KOVAR, C., CREE, A., CHACKO, J., CHANDRABOSE, M. N., DAO, M. D., DINH, H. H., GABISI, R. A., HINES, S., HUME, J., JHANGIAN, S. N., JOSHI, V., LEWIS, L. R., LIU, Y. S., LOPEZ, J., MORGAN, M. B., NGUYEN, N. B., OKWUONU, G. O., RUIZ, S. J., SANTIBANEZ, J., WRIGHT, R. A., FOWLER, G. R., HITCHENS, M. E., LOZADO, R. J., MOEN, C., STEFFEN, D., WARREN, J. T., ZHANG, J. K., NAZARETH, L. V., CHAVEZ, D., DAVIS, C., LEE, S. L., PATEL, B. M., PU, L. L., BELL, S. N., JOHNSON, A. J., VATTATHIL, S., WILLIAMS,

R. L., SHIGENOBU, S., DANG, P. M., MORIOKA, M., FUKATSU, T., KUDO, T., MIYAGISHIMA, S. Y., JIANG, H. Y., WORLEY, K. C., LEGEAI, F., GAUTHIER, J. P., COLLIN, O., ZHANG, L., CHEN, H. C., ERMOLAEVA, O., HLAVINA, W., KAPUSTIN, Y., KIRYUTIN, B., KITTS, P., MAGLOTT, D., MURPHY, T., PRUITT, K., SAPOJNIKOV, V., SOUVOROV, A., THIBAUD-NISSEN, F., CAMARA, F., GUIGO, R., STANKE, M., SOLOVYEV, V., KOSAREV, P., GILBERT, D., GABALDON, T., HUERTA-CEPAS, J., MARCET-HOUBEN, M., PIGNATELLI, M., MOYA, A., RISPE, C., OLLIVIER, M., QUESNEVILLE, H., PERMAL, E., LLORENS, C., FUTAMI, R., HEDGES, D., ROBERTSON, H. M., ALIOTO, T., MARIOTTI, M., NIKOH, N., MCCUTCHEON, J. P., BURKE, G., KAMINS, A., LATORRE, A., KUDO, T., MORAN, N. A., ASHTON, P., CALEVRO, F., et al. 2010. Genome Sequence of the Pea Aphid *Acyrthosiphon pisum*. *Plos Biology,* 8.

RINGROSE, L. 2009. Transgenesis in *Drosophila melanogaster*. *In:* CARTWRIGHT, E. J. (ed.) *Transgenesis Techniques: Principles and Protocols.* Totowa, NJ: Humana Press.

ROBINSON, J. T., THORVALDSDOTTIR, H., WINCKLER, W., GUTTMAN, M., LANDER, E. S., GETZ, G. & MESIROV, J. P. 2011. Integrative genomics viewer. *Nature Biotechnology,* 29**,** 24-26.

ROSENFELD, J. A., REEVES, D., BRUGLER, M. R., NARECHANIA, A., SIMON, S., DURRETT, R., FOOX, J., SHIANNA, K., SCHATZ, M. C., GANDARA, J., AFSHINNEKOO, E., LAM, E. T., HASTIE, A. R., CHAN, S., CAO, H., SAGHBINI, M., KENTSIS, A., PLANET, P. J., KHOLODOVYCH, V., TESSLER, M., BAKER, R., DESALLE, R., SORKIN, L. N., KOLOKOTRONIS, S. O., SIDDALL, M. E., AMATO, G. & MASON, C. E. 2016. Genome assembly and geospatial phylogenomics of the bed bug *Cimex lectularius*. *Nature Communications,* 7.

ROTHWELL, W. F. & SULLIVAN, W. 2000. Fluorescent Analysis of *Drosophilia* Embryos. Cold Spring Harbor Laboratory Press.

RYNER, L. C., GOODWIN, S. F., CASTRILLON, D. H., ANAND, A., VILLELLA, A., BAKER, B. S., HALL, J. C., TAYLOR, B. J. & WASSERMAN, S. A. 1996. Control of male sexual behavior and sexual orientation in *Drosophila* by the fruitless gene. *Cell,* 87**,** 1079-1089.

SADILEK, D., STAHLAVSKY, F., VILIMOVA, J. & ZIMA, J. 2013. Extensive fragmentation of the X chromosome in the bed bug *Cimex lectularius* Linnaeus, 1758 (Heteroptera, Cimicidae): a survey across Europe. *Comparative Cytogenetics,* 7**,** 253-269.

SAKAI, H., SAKAGUCHI, H., AOKI, F. & SUZUKI, M. G. 2015. Functional analysis of sex-determination genes by gene silencing with LNA-DNA gapmers in the silkworm, *Bombyx mori*. *Mechanisms of Development,* 137**,** 45-52.

SALEH, D., LAARIF, A., CLOUET, C. & GAUTHIER, N. 2012. Spatial and host-plant partitioning between coexisting *Bemisia tabaci* cryptic species in Tunisia. *Population Ecology,* 54**,** 261-274.

SALVUCCI, M. E. 2000. Sorbitol accumulation in whiteflies: evidence for a role in protecting proteins during heat stress. *Journal of Thermal Biology,* 25**,** 353-361.

SALZ, H. K., MAINE, E. M., KEYES, L. N., SAMUELS, M. E., CLINE, T. W. & SCHEDL, P. 1989. The *Drosophila* Female-Specific Sex-Determination Gene, Sex-Lethal, Has Stage-Specific, Tissue-Specific, and Sex-Specific Rnas Suggesting Multiple-Modes of Regulation. *Genes & Development,* 3**,** 708-719.

SANCHEZ, J. A. 2008. Zoophytophagy in the plantbug *Nesidiocoris tenuis*. *Agricultural and Forest Entomology,* 10**,** 75-80.

SAUNDERS, P. A., NEUENSCHWANDER, S. & PERRIN, N. 2018. Sex chromosome turnovers and genetic drift: a simulation study. *Journal of Evolutionary Biology,* 31**,** 1413-1419.

SAWANTH, S. K., GOPINATH, G., SAMBRANI, N. & ARUNKUMAR, K. P. 2016. The autoregulatory loop: A common mechanism of regulation of key sex determining genes in insects. *Journal of Biosciences,* 41**,** 283-294.

SCHINDELIN, J., ARGANDA-CARRERAS, I., FRISE, E., KAYNIG, V., LONGAIR, M., PIETZSCH, T., PREIBISCH, S., RUEDEN, C., SAALFELD, S., SCHMID, B., TINEVEZ, J. Y., WHITE, D. J., HARTENSTEIN, V., ELICEIRI, K., TOMANCAK, P. & CARDONA, A. 2012. Fiji: an open-source platform for biological-image analysis. *Nature Methods,* 9**,** 676-682.

SCHUTT, C., HILFIKER, A. & NOTHIGER, R. 1998. Virilizer regulates Sex-lethal in the germline of *Drosophila melanogaster*. *Development,* 125**,** 1501-1507.
SPARKS, T. C. 2013. Insecticide discovery: An evaluation and analysis. *Pesticide Biochemistry and Physiology,* 107**,** 8-17.
STANSLY, P. A. & NATWICK, E. T. 2009. Integrated Systems for Managing *Bemisia tabaci* in Protected and Open Field Agriculture. 467-497.
STOUTHAMER, R., BREEUWER, J. A. J. & HURST, G. D. D. 1999. *Wolbachia pipientis*: Microbial manipulator of arthropod reproduction. *Annual Review of Microbiology,* 53**,** 71-102.
SU, Y. L., LI, J. M., LI, M., LUAN, J. B., YE, X. D., WANG, X. W. & LIU, S. S. 2012. Transcriptomic Analysis of the Salivary Glands of an Invasive Whitefly. *Plos One,* 7.
SUETSUGU, Y., FUTAHASHI, R., KANAMORI, H., KADONO-OKUDA, K., SASANUMA, S., NARUKAWA, J., AJIMURA, M., JOURAKU, A., NAMIKI, N., SHIMOMURA, M., SEZUTSU, H., OSANAI-FUTAHASHI, M., SUZUKI, M. G., DAIMON, T., SHINODA, T., TANIAI, K., ASAOKA, K., NIWA, R., KAWAOKA, S., KATSUMA, S., TAMURA, T., NODA, H., KASAHARA, M., SUGANO, S., SUZUKI, Y., FUJIWARA, H., KATAOKA, H., ARUNKUMAR, K. P., TOMAR, A., NAGARAJU, J., GOLDSMITH, M. R., FENG, Q. L., XIA, Q. Y., YAMAMOTO, K., SHIMADA, T. & MITA, K. 2013. Large Scale Full-Length cDNA Sequencing Reveals a Unique Genomic Landscape in a Lepidopteran Model Insect, *Bombyx mori. G3-Genes Genomes Genetics,* 3**,** 1481-1492.
SUN, D. B., LIU, Y. Q., QIN, L., XU, J., LI, F. F. & LIU, S. S. 2013. Competitive displacement between two invasive whiteflies: insecticide application and host plant effects. *Bulletin of Entomological Research,* 103**,** 344-353.
SUN, D. B., XU, J., LUAN, J. B. & LIU, S. S. 2011. Reproductive incompatibility between the B and Q biotypes of the whitefly *Bemisia tabaci* in China: genetic and behavioural evidence. *Bulletin of Entomological Research,* 101**,** 211-220.
SUZUKI, M. G., FUNAGUMA, S., KANDA, T., TAMURA, T. & SHIMADA, T. 2003. Analysis of the biological functions of a doublesex homologue in Bombyx mori. *Development Genes and Evolution,* 213**,** 345-354.
SUZUKI, M. G., IMANISHI, S., DOHMAE, N., ASANUMA, M. & MATSUMOTO, S. 2010. Identification of a Male-Specific RNA Binding Protein That Regulates Sex-Specific Splicing of Bmdsx by Increasing RNA Binding Activity of BmPSI. *Molecular and Cellular Biology,* 30**,** 5776-5786.
SUZUKI, M. G., IMANISHI, S., DOHMAE, N., NISHIMURA, T., SHIMADA, T. & MATSUMOTO, S. 2008. Establishment of a novel in vivo sex-specific splicing assay system to identify a trans-acting factor that negatively regulates splicing of Bombyx mori *dsx* female exons. *Molecular and Cellular Biology,* 28**,** 333-343.
SUZUKI, M. G., KOBAYASHI, S. & AOKI, F. 2014. Male-specific splicing of the silkworm *Imp* gene is maintained by an autoregulatory mechanism. *Mechanisms of Development,* 131**,** 47-56.
SUZUKI, M. G., OHBAYASHI, F., MITA, K. & SHIMADA, T. 2001. The mechanism of sex-specific splicing at the doublesex gene is different between *Drosophila melanogaster* and *Bombyx mori. Insect Biochemistry and Molecular Biology,* 31**,** 1201-1211.
TAJEBE, L. S., GUASTELLA, D., CAVALIERI, V., KELLY, S. E., HUNTER, M. S., LUND, O. S., LEGG, J. P. & RAPISARDA, C. 2015. Diversity of symbiotic bacteria associated with *Bemisia tabaci* (Homoptera: Aleyrodidae) in cassava mosaic disease pandemic areas of Tanzania. *Annals of Applied Biology,* 166**,** 297-310.
THAO, M. L. & BAUMANN, P. 2004. Evolutionary relationships of primary prokaryotic endosymbionts of whiteflies and their hosts. *Applied and Environmental Microbiology,* 70**,** 3401-3406.
THOMAS, D. D., DONNELLY, C. A., WOOD, R. J. & ALPHEY, L. S. 2000. Insect population control using a dominant, repressible, lethal genetic system. *Science,* 287**,** 2474-2476.
THORPE, P., COCK, P. J. A. & BOS, J. 2016. Comparative transcriptomics and proteomics of three different aphid species identifies core and diverse effector sets. *Bmc Genomics,* 17.

THORVALDSDOTTIR, H., ROBINSON, J. T. & MESIROV, J. P. 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in Bioinformatics,* 14**,** 178-192.

TIAN, M. & MANIATIS, T. 1993. A Splicing Enhancer Complex Controls Alternative Splicing of Doublesex Premessenger Rna. *Cell,* 74**,** 105-114.

TRAPNELL, C., WILLIAMS, B. A., PERTEA, G., MORTAZAVI, A., KWAN, G., VAN BAREN, M. J., SALZBERG, S. L., WOLD, B. J. & PACHTER, L. 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnoogyl,* 28**,** 511-5.

TSAI, J. H. & WANG, K. H. 1996. Development and reproduction of *Bemisia argentifolii* (Homoptera: Aleyrodidae) on five host plants. *Environmental Entomology,* 25**,** 810-816.

URBANEJA, A., SANCHEZ, E. & STANSLY, P. A. 2007. Life history of Eretmocerus mundus, a parasitoid of *Bemisia tabaci*, on tomato and sweet pepper. *Biocontrol,* 52**,** 25-39.

VAN DOREN, M., WILLIAMSON, A. L. & LEHMANN, R. 1998. Regulation of zygotic gene expression in *Drosophila* primordial germ cells. *Current Biology,* 8**,** 243-246.

VERHULST, E. C., VAN DE ZANDE, L. & BEUKEBOOM, L. W. 2010. Insect sex determination: it all evolves around transformer. *Current Opinion in Genetics & Development,* 20**,** 376-383.

VREYSEN, M. J. B., SALEH, K. M., ALI, M. Y., ABDULLA, A. M., ZHU, Z. R., JUMA, K. G., DYCK, V. A., MSANGI, A. R., MKONYI, P. A. & FELDMANN, H. U. 2000. *Glossina austeni* (Diptera : Glossinidae) eradicated on the Island of Unguja, Zanzibar, using the sterile insect technique. *Journal of Economic Entomology,* 93**,** 123-135.

VSN_INTERNATIONAL 2019. Genstat for Windows 20th edition Hemel Hempstead, UK.

WAINWRIGHT, S. M. & ISHHOROWICZ, D. 1992. Point Mutations in the *Drosophila* Hairy Gene Demonstrate Invivo Requirements for Basic, Helix-Loop-Helix, and Wrpw Domains. *Molecular and Cellular Biology,* 12**,** 2475-2483.

WALKER, G. P. & PERRING, T. M. 1994. Feeding and Oviposition Behavior of Whiteflies (Homoptera, Aleyrodidae) Interpreted from Ac Electronic Feeding Monitor Wave-Forms. *Annals of the Entomological Society of America,* 87**,** 363-374.

WALKER, G. P., PERRING, T. M. & FREEMAN, T. P. 2009. Life History, Functional Anatomy, Feeding and Mating Behavior. 109-160.

WALKER, G. P., PERRING, T. M. & FREEMAN, T. P. 2010. *[Life history, functional anatomy, feeding and mating behavior],*, Springer.

WANG, L., TANG, N., GAO, X., CHANG, Z., ZHANG, L., ZHOU, G., GUO, D., ZENG, Z., LI, W., AKINYEMI, I. A., YANG, H. & WU, Q. 2017. Genome sequence of a rice pest, the white-backed planthopper (*Sogatella furcifera*). *GigaScience,* 6**,** 1-9.

WANG, Z., GERSTEIN, M. & SNYDER, M. 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics,* 10**,** 57-63.

WEINSTOCK, G. M., ROBINSON, G. E., GIBBS, R. A., WORLEY, K. C., EVANS, J. D., MALESZKA, R., ROBERTSON, H. M., WEAVER, D. B., BEYE, M., BORK, P., ELSIK, C. G., HARTFELDER, K., HUNT, G. J., ZDOBNOV, E. M., AMDAM, G. V., BITONDI, M. M. G., COLLINS, A. M., CRISTINO, A. S., LATTORFF, H. M. G., LOBO, C. H., MORITZ, R. F. A., NUNES, F. M. F., PAGE, R. E., SIMOES, Z. L. P., WHEELER, D., CARNINCI, P., FUKUDA, S., HAYASHIZAKI, Y., KAI, C., KAWAI, J., SAKAZUME, N., SASAKI, D., TAGAMI, M., ALBERT, S., BAGGERMAN, G., BEGGS, K. T., BLOCH, G., CAZZAMALI, G., COHEN, M., DRAPEAU, M. D., EISENHARDT, D., EMORE, C., EWING, M. A., FAHRBACH, S. E., FORET, S., GRIMMELIKHUIJZEN, C. J. P., HAUSER, F., HUMMON, A. B., HUYBRECHTS, J., JONES, A. K., KADOWAKI, T., KAPLAN, N., KUCHARSKI, R., LEBOULLE, G., LINIAL, M., LITTLETON, J. T., MERCER, A. R., RICHMOND, T. A., RODRIGUEZ-ZAS, S. L., RUBIN, E. B., SATTELLE, D. B., SCHLIPALIUS, D., SCHOOFS, L., SHEMESH, Y., SWEEDLER, J. V., VELARDE, R., VERLEYEN, P., VIERSTRAETE, E., WILLIAMSON, M. R., AMENT, S. A., BROWN, S. J., CORONA, M., DEARDEN, P. K., DUNN, W. A., ELEKONICH, M. M., FUJIYUKI, T., GATTERMEIER, I., GEMPE, T., HASSELMANN, M.,

KADOWAKI, T., KAGE, E., KAMIKOUCHI, A., KUBO, T., KUCHARSKI, R., KUNIEDA, T., LORENZEN, M. D., MILSHINA, N. V., MORIOKA, M., OHASHI, K., OVERBEEK, R., ROSS, C. A., SCHIOETT, M., SHIPPY, T., TAKEUCHI, H., TOTH, A. L., WILLIS, J. H., WILSON, M. J., GORDON, K. H. J., LETUNIC, I., HACKETT, K., et al. 2006. Insights into social insects from the genome of the honeybee *Apis mellifera. Nature,* 443**,** 931-949.

WERREN, J. H., BALDO, L. & CLARK, M. E. 2008. Wolbachia: master manipulators of invertebrate biology. *Nature Reviews Microbiology,* 6**,** 741-751.

WEXLER, J., DELANEY, E. K., BELLES, X., SCHAL, C., WADA-KATSUMATA, A., AMICUCCI, M. J. & KOPP, A. 2019. Hemimetabolous insects elucidate the origin of sexual development via alternative splicing. *Elife,* 8.

WHELAN, S. & GOLDMAN, N. 2001. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Molecular Biology and Evolution,* 18**,** 691-699.

WILL, T., FURCH, A. C. U. & ZIMMERMANN, M. R. 2013. How phloem-feeding insects face the challenge of phloem-located defenses. *Frontiers in Plant Science,* 4.

WOTTON, K. R., JIMENEZ-GURI, E., MATHEU, B. G. & JAEGER, J. 2014. A Staging Scheme for the Development of the Scuttle Fly *Megaselia abdita. Plos One,* 9.

WU, J. Q. & BALDWIN, I. T. 2010. New Insights into Plant Responses to the Attack from Insect Herbivores. *Annual Review of Genetics, Vol 44,* 44**,** 1-24.

XIE, W., CHEN, C., YANG, Z., GUO, L., YANG, X., WANG, D., CHEN, M., HUANG, J., WEN, Y., ZENG, Y., LIU, Y., XIA, J., TIAN, L., CUI, H., WU, Q., WANG, S., XU, B., LI, X., TAN, X. & ZHANG, Y.-J. 2017a. *Genome sequencing of the sweetpotato whitefly Bemisia tabaci MED/Q*.

XIE, W., CHEN, C. H., YANG, Z. Z., GUO, L. T., YANG, X., WANG, D., CHEN, M., HUANG, J. Q., WEN, Y. A., ZENG, Y., LIU, Y. T., XIA, J. X., TIAN, L. X., CUI, H. Y., WU, Q. J., WANG, S. L., XU, B. Y., LI, X. C., TAN, X. Q., GHANIM, M., QIU, B. L., PAN, H. P., CHU, D., DELATTE, H., MARUTHI, M. N., GE, F., ZHOU, X. P., WANG, X. W., WAN, F. H., DU, Y. Z., LUO, C., YAN, F. M., PREISSER, E. L., JIAO, X. G., COATES, B. S., ZHAO, J. Y., GAO, Q., XIA, J. Q., YIN, Y., LIU, Y., BROWN, J. K., ZHOU, X. G. & ZHANG, Y. J. 2017b. Genome sequencing of the sweetpotato whitefly Bemisia tabaci MED/Q. *Gigascience,* 6.

XIE, W., GUO, L. T., JIAO, X. G., YANG, N. N., YANG, X., WU, Q. J., WANG, S. L., ZHOU, X. G. & ZHANG, Y. J. 2014. Transcriptomic dissection of sexual differences in Bemisia tabaci, an invasive agricultural pest worldwide. *Scientific Reports,* 4.

XU, J., ZHAN, S., CHEN, S. Q., ZENG, B. S., LI, Z. Q., JAMES, A. A., TAN, A. J. & HUANG, Y. P. 2017. Sexually dimorphic traits in the silkworm, *Bombyx mori*, are regulated by doublesex. *Insect Biochemistry and Molecular Biology,* 80**,** 42-51.

XUE, J., ZHOU, X., ZHANG, C.-X., YU, L.-L., FAN, H.-W., WANG, Z., XU, H.-J., XI, Y., ZHU, Z.-R., ZHOU, W.-W., PAN, P.-L., LI, B.-L., COLBOURNE, J. K., NODA, H., SUETSUGU, Y., KOBAYASHI, T., ZHENG, Y., LIU, S., ZHANG, R., LIU, Y., LUO, Y.-D., FANG, D.-M., CHEN, Y., ZHAN, D.-L., LV, X.-D., CAI, Y., WANG, Z.-B., HUANG, H.-J., CHENG, R.-L., ZHANG, X.-C., LOU, Y.-H., YU, B., ZHUO, J.-C., YE, Y.-X., ZHANG, W.-Q., SHEN, Z.-C., YANG, H.-M., WANG, J., WANG, J., BAO, Y.-Y. & CHENG, J.-A. 2014. Genomes of the rice pest brown planthopper and its endosymbionts reveal complex complementary contributions for host adaptation. *Genome Biology,* 15**,** 521.

YAMAMOTO, D., FUJITANI, K., USUI, K., ITO, H. & NAKANO, Y. 1998. From behavior to development: genes for sexual behavior define the neuronal sexual switch in *Drosophila*. *Mechanisms of Development,* 73**,** 135-146.

YAMAMOTO, D. & KOGANEZAWA, M. 2013. Genes and circuits of courtship behaviour in *Drosophila* males. *Nature Reviews Neuroscience,* 14**,** 681-692.

YI, W., ROSS, J. M. & ZARKOWER, D. 2000. mab-3 is a direct tra-1 target gene regulating diverse aspects of *C-elegans* male sexual development and behavior. *Development,* 127**,** 4469-4480.

YU, X. D. & KILLINY, N. 2018. Effect of parental RNA interference of a transformer-2 homologue on female reproduction and offspring sex determination in Asian citrus psyllid. *Physiological Entomology,* 43**,** 42-50.

ZALOKAR, M. & ERK, I. 1976. Division and Migration of Nuclei during Early Embryogenesis of *Drosophila-Melanogaster. Journal De Microscopie Et De Biologie Cellulaire,* 25**,** 97-+.

ZCHORI-FEIN, E. & BROWN, J. K. 2002. Diversity of prokaryotes associated with *Bemisia tabaci* (Gennadius) (Hemiptera : Aleyrodidae). *Annals of the Entomological Society of America,* 95**,** 711-718.

ZEH, D. W., ZEH, J. A. & BONILLA, M. M. 2005. Wolbachia, sex ratio bias and apparent male killing in the harlequin beetle riding pseudoscorpion. *Heredity,* 95**,** 41-49.

ZHONG, Y. & LI, Z. X. 2013. Influences of tetracycline on the reproduction of the B biotype of *Bemisia tabaci* (Homoptera: Aleyrodidae). *Applied Entomology and Zoology,* 48**,** 241-246.

ZHONG, Y. & LI, Z. X. 2014. Bidirectional Cytoplasmic Incompatibility Induced by Cross-Order Transfection of *Wolbachia:* Implications for Control of the Host Population. *Microbial Ecology,* 68**,** 463-471.

ZHOU, X. F. & LI, Z. X. 2016. Establishment of the cytoplasmic incompatibility-inducing *Wolbachia* strain wMel in an important agricultural pest insect. *Scientific Reports,* 6.

ZHU, L. Y., WILKEN, J., PHILLIPS, N. B., NARENDRA, U., CHAN, G., STRATTON, S. M., KENT, S. B. & WEISS, M. A. 2000. Sexual dimorphism in diverse metazoans is regulated by a novel class of intertwined zinc fingers. *Genes & Development,* 14**,** 1750-1764.

ZHUO, J.-C., ZHANG, H.-H., XIE, Y.-C., LI, H.-J., HU, Q.-L. & ZHANG, C.-X. 2019. Identification of a female determinant gene for the sexual determination of a hemipteran insect, the brown planthopper. *bioRxiv***,** 775577.

ZHUO, J. C., LEI, C., SHI, J. K., XU, N., XUE, W. H., ZHANG, M. Q., REN, Z. W., ZHANG, H. H. & ZHANG, C. X. 2017. Tra-2 Mediates Cross-Talk Between Sex Determination and Wing Polyphenism in Female *Nilaparvata lugens*. *Genetics,* 207**,** 1067-1078.

ZWARTS, L., VANDEN BROECK, L., CAPPUYNS, E., AYROLES, J. F., MAGWIRE, M. M., VULSTEKE, V., CLEMENTS, J., MACKAY, T. F. C. & CALLAERTS, P. 2015. The genetic basis of natural variation in mushroom body size in *Drosophila melanogaster*. *Nature Communications,* 6.

Supplementary Figures

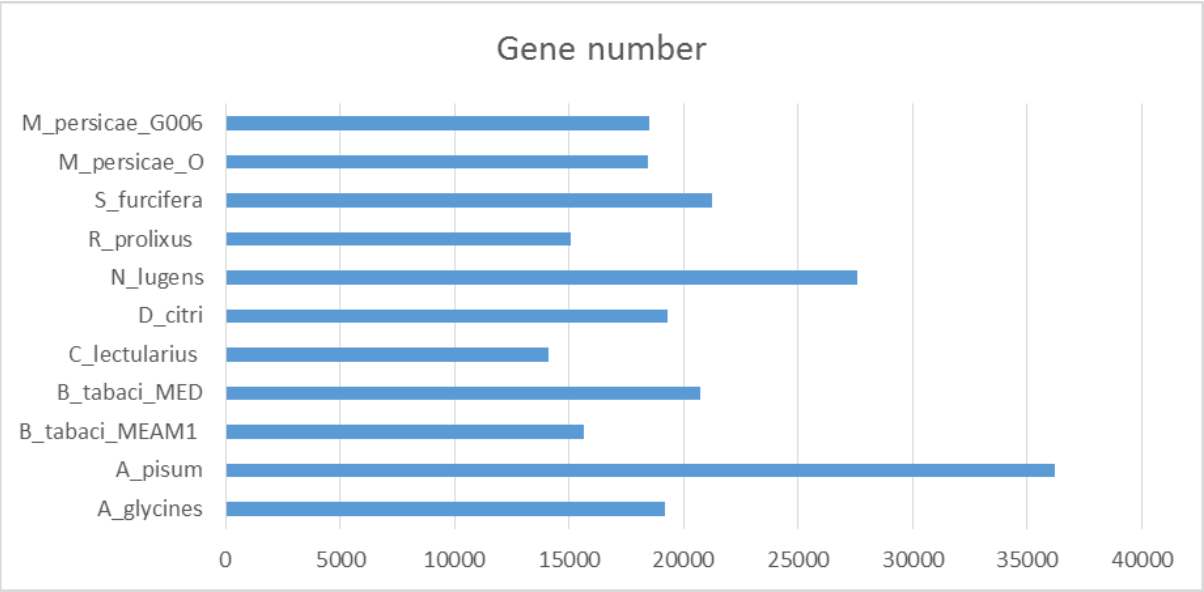

Figure S. 1: The gene numbers present in each of the published hemipteran genomes used throughout the thesis

A.

	Male A eggs Vs Male BC eggs	Male A eggs Vs Male D eggs	Male A eggs Vs Male adults	Male B/C eggs Vs Male D eggs	Male BC eggs Vs Male adults	Male D eggs Vs Male adults
SXL	0.569	0.773	0.041 *	0.535	0.074	0.138
TRA2	0.689	0.296	0.765	0.24	0.805	0.053
IMP	0.755	0.307	0.363	0.426	0.516	0.603
BTDMRT1	0.265	0.489	0.662	0.178	0.012 *	0.412
BTDMRT3	0.264	0.277	0.037 *	0.877	0.037 *	0.038 *
VASA	0.197	0.918	0.934	0.423	<0.001 *	0.947

B.

	Female A eggs Vs Female BC eggs	Female A eggs Vs Female Adults	Female BC eggs Vs Female Adults
SXL	0.373	0.685	0.199
TRA2	0.981	0.919	0.935
IMP	-	0.478	-
BTDMRT1	-	0.119	-
BTDMRT3	-	-	<0.001 *
VASA	0.77	0.009 *	0.009 *

C.

	Female A eggs Vs Male A eggs	Female BC eggs Vs Male BC eggs	Female Adults Vs Male Adults
SXL	0.762	0.156	0.001
TRA2	0.441	0.834	0.083
IMP	0.28	-	0.895
BTDMRT1	0.66	-	0.136
BTDMRT3	-	0.901	0.047
VASA	0.867	0.661	0.017

Figure S. 2: T- test 2 sample with unequal variance results for the gene expression of SDGs in chapter 5. Test were conducted in Genstat 20th edition, version 20.1.0.23823. Table A; the comparison results of different male samples. Table B; the comparison results of different female samples. Table C; the comparison results of same developmental stage in different sexes (VSN_International, 2019)

*statistically different

- the statistical tests could not be conducted as the TPM for all samples was 0.