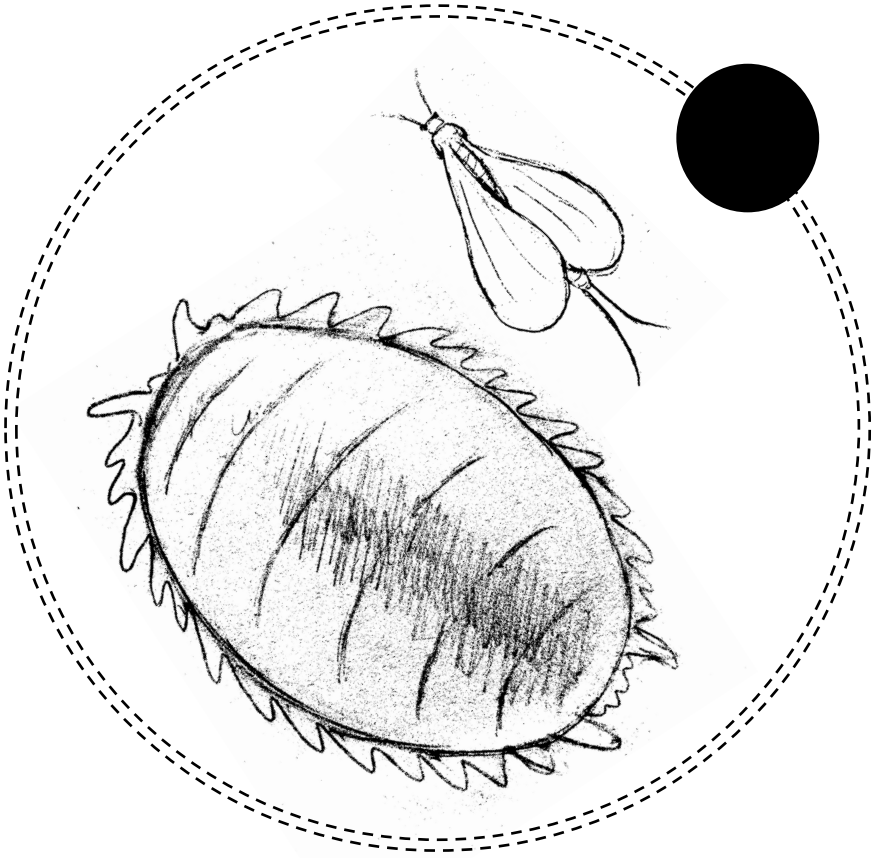
THE UNIVERSITY
of EDINBURGH

Epigenetic Mechanisms underlying Paternal Genome Elimination

Stevie Anne Bain



A thesis submitted for the degree of Doctor of Philosophy

School of Biological Sciences

University of Edinburgh

2018

Table of contents

Declaration

This thesis is of my own composition and has not been submitted for any other degree or professional qualification. I conducted all work carried out in this thesis except where explicitly stated.

In Chapter 3, A. G. de la Filia prepared bisulfite libraries and advised on microsatellite analysis. Dominik R Laestch provided support to develop a pipeline for allele-specific DNA methylation analysis.

Stevie Anne Bain

2018

Acknowledgements

There are many people who helped me through my PhD journey and I would like to take this opportunity to express my gratitude.

Firstly, I would like to thank my supervisor, Laura Ross. Building this project from the ground up was a challenge, with many ups and downs along the way. I want to thank you for the opportunity to work with you and for always believing that I was capable.

Many thanks to my thesis committee, Darren Obbard and Pedro Vale, for encouragement and feedback over the years.

A special thanks to my second supervisor, Patrick Ferree, for welcoming me into your lab in Claremont, CA. I appreciate you taking the time to mentor me over the years. Also, thank you to you, Elise, Claire and Keegan (and Nemo!) for treating me like one of the family.

To Silvia Bongiorni and Giorgio Prantera, thank you for inviting me to spend time with you in your lab and teaching me new cytological skills and techniques. Your research inspired so much of this thesis and it was a pleasure to work with you both.

Thanks to all members of the Ross Lab, past and present, including the many mealybugs who were sacrificed for this work. A special thank you to Dom, you are an excellent teacher. A massive thank you to Isabelle – you were by my side during the many, many immunostaining experiments and always positive, motivating and incredibly helpful. Thanks also to David Brown for dealing with many lab crises.

Thanks to my undergraduate supervisor, Steve Hubbard, for your wisdom and words of encouragement.

To my cohort – Ed, Rasmus, Tom B, Luiz, Nathan, Billy, Eileen, Lucy, Nora, Carlijn and Gabi – thank you for helping me celebrate the highs and cope with the lows. You are a great bunch of people!

A huge thank you to my office mate and friend, Lisa. Your endless supply of snacks and moral support helped me survive my final year. I will return the favour.

To Tom G, Amy MF, Kat, Cristina and Charlotte R: thanks for coffee break chats/life advice/PhD advice and generally being great.

A special thank you to my family for your love and support. To my mum and dad, Michelle and Steve, for always believing in me and allowing me to cause you such stress during my traveling adventures. You are the best people I know and the sacrifices you made to ensure I had every opportunity to succeed have not gone unnoticed. To my Auntie Tracy, for your endless encouragement (particularly before my PhD interview!) and for excellent road trips to and from interviews, internships, new houses etc. To my gran and grandad, for always reminding me to be proud of my achievements and to take time to enjoy life. To my late dog, Scrappy, who faithfully sat by my side as I studied for my exams.

Finally, thank you to Andrés. Where to begin? Both personally and professionally you have supported me over the last four years, and I am truly grateful. Thank you for our Pain and Panic days, for our conference adventures and for always making me laugh; but mostly thank you for our life outside of Ashworth, away from the mealybugs.

A special acknowledgement to Caroline Kennard (1827–1907) for bravely – and correctly – challenging Charles Darwin's assertion that women are intellectually inferior to men.

'Let the environment of women be similar to that of men and with his opportunities, before she be fairly judged, intellectually his inferior, please'

This thesis is dedicated to my wee granny, Anne, for your genuine belief that I am the best scientist in the world.

Thesis Abstract

For most sexually reproducing organisms, the two parentally inherited copies of a gene are equivalent in transmission and expression. However, there are exceptions to this rule. Genomic imprinting is an epigenetic process in which expression of one gene copy is favoured depending on its parental origin. One of the most striking cases of genomic imprinting is Paternal Genome Elimination (PGE). PGE is a genomic imprinting phenomenon found in thousands of insect species and involves the silencing and elimination of an entire haploid genome in a parent-of-origin specific manner. Under PGE, both sexes develop from fertilized eggs and initially possess a diploid euchromatic chromosome complement. However, males subsequently eliminate paternally-inherited chromosomes from their germline. Different PGE species vary in the timing of the elimination of the paternal genome, and in whether it becomes transcriptionally silenced or not. As a result, male gene expression varies from haploid to diploid with various intermediates. The recognition and silencing of paternally-inherited genes under PGE appear to be regulated by the same epigenetic machinery as silencing and imprinting in mammals, namely DNA methylation and histone modifications. However, the molecular details are poorly understood. Here, I investigate the epigenetic mechanisms underlying PGE using the citrus mealybug (*Planococcus citri*, Hemiptera: Pseudococcidae) a small plant-feeding insect, easily reared in laboratory conditions. I utilize molecular, cytogenetic and genomic techniques to address the following questions: i) do levels & patterns of global DNA methylation differ between the sexes and does this play a role in sex-specific gene expression? ii) are key histone-mediated heterochromatin pathways (H3K9me3-HP1 and H3K27me3-PRC2 pathways) involved in the recognition and silencing of the paternal genome in PGE males? iii) do DNA methylation marks differ between paternal and maternal alleles, potentially acting as a distinguisher of parental origin during PGE? Whole genome bisulfite sequencing and transcriptome sequencing reveal

evidence of sex-specific DNA methylation and gene expression. However, changes in gene methylation and expression between males and females are not correlated suggesting that this epigenetic modification may not mediate sex-specific expression. Cytogenetic studies in males show that both H3K9me3-HP1 and H3K27me3-PRC2 heterochromatin pathways are involved in the silencing of paternal alleles in PGE males but not in the recognition of paternal alleles during spermatogenesis. Finally, allele-specific bisulfite analysis in *Planococcus* hybrids suggests that differences in methylation on maternal and paternal alleles could potentially allow recognition of paternal alleles during PGE. My research provides insight into the putative roles of sex-specific and allele-specific epigenetic modifications in the recognition and silencing of the paternal genome during the process of PGE. Additionally, the methylome of a non-social, non-hymenopteran insect broadens understanding of the function(s) and evolution of DNA methylation within arthropods.

Lay summary

In most sexually reproducing organisms, the two copies of a gene - one from each parent - are interchangeable. However, in a process called genomic imprinting, this rule is violated and one gene copy is favoured depending on which parent it came from. This phenomenon was generally believed to be exclusive to mammals and flowering plants, where it has been extensively studied over the past decades. However, genomic imprinting has independently evolved at least seven times across the arthropods and is found in thousands of species. My thesis investigates arguably one of the most astounding cases of genomic imprinting - Paternal Genome Elimination (PGE). In PGE species, both sexes develop from fertilised eggs, but males only pass on maternally inherited genes to their offspring. Furthermore, in some PGE species, all the genes that a male inherits from his father are silenced. This is the case in the citrus mealybug, *Planococcus citri*. In this thesis, I focus on understanding the mechanisms involved in the recognition, silencing and elimination of specifically the paternally inherited genes in males. I present evidence for the involvement of DNA methylation and histone modifications in these processes.

Chapter 1: General Introduction

1.1 Asymmetric sexual reproduction

Sexual reproduction is almost universal amongst multicellular eukaryotes. Under standard Mendelian inheritance, a diploid individual receives a set of chromosomes from each of their parents that are functionally equivalent in expression and transmission (Mendel, 1865). However, there are exceptions to this rule. In ~15% of arthropods, sexual reproduction is asymmetric and gene expression and transmission is biased according to its parental origin (Normark, 2003; Bachtrog *et al.*, 2014). In particular, there are thousands of species in which males only transmit maternally inherited genes to their offspring (Burt and Trivers, 2008). This is a form of genomic imprinting known as paternal genome exclusion and can occur in three different forms, each of which has evolved several times (Herrick and Seger, 1999). The most familiar form of this genetic asymmetry is arrhenotoky, which is widespread across several invertebrate orders including the Hymenoptera (Normark, 2003; de la Filia, Bain and Ross, 2015), where males develop from unfertilised eggs and thus lack paternal genes at all stages of development. In the other forms, males develop from fertilised eggs, but their paternally inherited chromosomes are actively eliminated from gametes and, in some cases, from their somatic nuclei. These processes are known as Paternal Genome Elimination (PGE) and are a whole-genome form of meiotic drive. Germline elimination of paternal chromosomes was first discovered in the Dipteran fungus gnat, *Sciara* (Metz, 1928). However, studies in scale insects revealed that there are thousands of species in this group exhibiting every known variation of PGE (Herrick and Seger, 1999). Scale insects (Hemiptera: Sternorrhyncha: Coccoidea) are a group of plant feeding insects. There are approximately 8000 species of scale insect identified, belonging to 32 different families (Ross and Shuker, 2009), of which the mealybugs are the most species diverse. In scale insects, PGE is thought to have evolved only once and is present across a monophyletic clade of around 6000 species.

In this chapter, I will describe the processes of PGE and discuss its evolutionary origins and consequences. I will then move on to look at the mechanisms involved in genomic imprinting and how these may have a role in the recognition, silencing and elimination of the paternal genome in PGE males. Finally, I will outline the objectives of this thesis.

1.2 Paternal Genome Elimination

1.2.1 Meiotic drive and germline elimination of the paternal genome

The term genomic imprinting was first coined through study of the fungus gnat, *Sciara coprophila* to describe the parent-of-origin specific behaviour of their chromosomes (Crouse, 1960). During male meiosis I, the entire paternally derived chromosome set is discarded into a cytoplasmic bud and does not undergo further meiotic divisions (Metz, 1928). Then, during meiosis II, whilst maternal autosomes segregate conventionally, the two maternal X chromatids fail to disjoin and are packaged into a single sperm nucleus (Gerbi, 1986). Thus, males eliminate paternally inherited chromosomes during spermatogenesis and only transmit maternally inherited autosomes and two maternal X chromosomes to offspring. A similar process of germline elimination occurs in the citrus mealybug, *Planococcus citri*. As with *S. coprophila*, it is exclusively the paternally inherited chromosomes that are eliminated from the germline of males (Brown and Nelson-Rees, 1961). In both sexes of *P. citri*, meiosis differs from canonical mammalian meiosis. Meiotic divisions are inverted with the first division equational (separation of sister chromatids) and the second reductional (segregation of homologues) (Bongiorni *et al.*, 2004). With the exception of the inverted occurrence of divisions, in females, the remaining divisions occur canonically undergoing recombination and independent assortment. Male meiosis, however, is markedly different and is characterised by a striking non-independent assortment of chromosomes in the second division. A monopolar spindle

forms, as opposed to the bipolar spindle usually associated with meiosis, and separates the maternally inherited chromosomes from those that are paternally inherited (Brown and Nelson-Rees, 1961). The four haploid nuclei generated from the meiotic divisions form a quadrinucleate spermatid in which two nuclei contain exclusively maternally inherited chromosomes and the other two contain exclusively paternally inherited chromosomes (Brown and Nelson-Rees, 1961). Crucially, only the nuclei containing maternally inherited chromosomes elongate into mature sperm.

1.2.2 Silencing of the paternal genome

Although males of all PGE species fail to transmit paternally-derived chromosomes to their offspring, the presence and expression patterns of these chromosomes in their somatic tissues vary remarkably among different taxonomic groups (Herrick and Seger, 1999; Burt and Trivers, 2008; de la Filia, Bain and Ross, 2015). Once PGE has evolved, paternally derived genes are under strong selection to evolve counteradaptations (Herrick and Seger, 1999; Ross, Pen and Shuker, 2010). To avoid this, the maternally derived genes are selected to either disable the paternal genome, leading to the whole-genome heterochromatinization seen in scale insects (Hughes-Schrader, 1948), coffee borer beetles (Borsa and Kjellberg, 1996) and booklice (Hodson *et al.*, 2017); or completely eliminate the paternal genome, leading to embryonic PGE, as seen in mites and armoured scale insects (Herrick and Seger, 1999). In species with embryonic PGE, the complete paternally-derived genome is lost from male embryos shortly after fertilization. Consequentially, both somatic and germline tissue is completely haploid. The additional fitness costs associated with embryonic PGE suggests that it is the more derived form and germline PGE is the ancestral form (Nur, 1990; Ross, Pen and Shuker, 2010). In species with germline PGE, the complete paternally-derived genome is only lost from germline cells during male meiosis and all other tissues contain paternally-derived chromosomes. However, whether or not these paternal chromosomes are

expressed varies between species and as a result, male gene expression varies from haploid to diploid with various intermediates (de la Filia, Bain and Ross, 2015).

Both *S. coprophila* and *P. citri* are species with germline PGE. In *S. coprophila*, either one or two of the paternally inherited X chromosomes is eliminated from the somatic cells of the embryo during early embryogenesis. This leads to the development of a male or female embryo, respectively (Metz, 1938). Thus, the number of paternal X chromosomes in the somatic cells determines the sex of the offspring. It is most likely that maternal factors determine the number of X chromosomes eliminated, however, the molecular nature of this cytoplasmic factor is, as yet, unknown (Escribá and Goday, 2013). This embryonic X chromosome elimination occurs during the 7^{th}-9^{th}-cleavage division, which is around the same time that silencing of the paternally inherited chromosomes occurs in *P. citri*. In *P. citri,* transcriptional silencing of the paternally inherited chromosomes in males occurs through facultative heterochromatinization. This silencing of the paternal genome progresses from one pole of the embryo to the other and the heterochromatic paternal chromosomes move together to form a heterochromatic body within the male nucleus (Bongiorni and Prantera, 2001). In females, both maternal and paternal chromosomes remain euchromatic. Figure 1 shows DAPI stained embryos of both sexes, in which males can be clearly identified through the presence of a heterochromatic body.

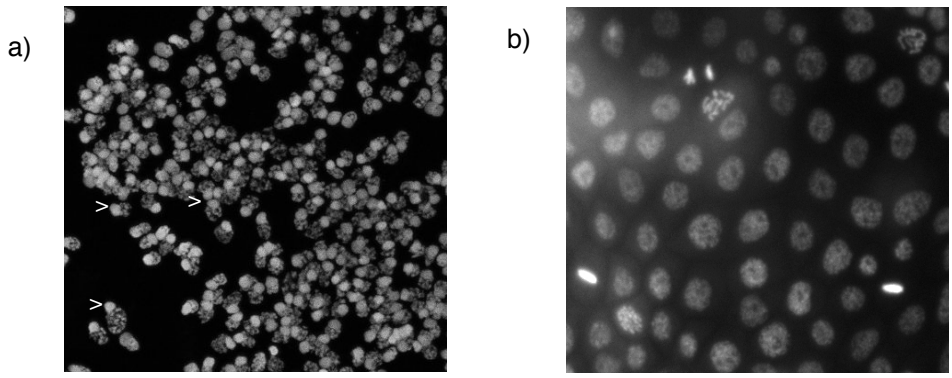


Figure 1: Female and male somatic nuclei. a) male nuclei with heterochromatic bodies (examples marked with >); b) female nuclei lacking heterochromatic bodies.

1.2.3 Evolution of Paternal Genome Elimination

PGE is found in approximately 20,000 species including scale insects, beetles, flies, springtails, lice and mites (de la Filia, Bain and Ross, 2015). It has at least seven independent origins across invertebrates and can often persist over considerable evolutionary time (Herrick and Seger, 1999). PGE represents an evolutionary response to underlying intragenomic conflict between maternal and paternal chromosomes in males (Herrick and Seger, 1999). All scale insects with PGE lack sex chromosomes, however, non-PGE coccid lineages retain XX-XO sex determination (Herrick and Seger, 1999). Haig (1993) uses Hamilton's theory of inclusive fitness (Hamilton, 1964) to show how a system of XX-XO sex determination can give rise to PGE. He suggests that a meiotically driving X chromosome during spermatogenesis would cause female-biased sex ratios and favour effective sex-linkage of maternal autosomes in males. In turn, female-biased sex ratios favour new mechanisms of sex determination and maternal control of offspring sex-ratio (Haig, 1993). Highly specialised spermatogenesis in mealybugs may predispose them to genomic drive as all four haploid products of meiosis are contained within a shared cytoplasm (Nur, 1962), giving the opportunity for one set of the chromatids to harm the other set and prevent it contributing to sperm production (Haig, 1993). An alternative theory suggests that conflict between the host and its endosymbiotic bacteria drove the evolution of PGE (Normark, 2004; Úbeda and Normark, 2006). Due to their nutrient poor diets, all scale insects harbour endosymbiotic bacteria – interestingly, mealybugs harbour two endosymbionts, one living inside the other (von Dohlen *et al.*, 2001)– to acquire essential nutrients. Endosymbiotic bacteria are maternally inherited and thus, from the bacteria's perspective, males are evolutionary dead ends. Normark's theory proposes that PGE arose from coevolution between male-killing endosymbiotic bacteria and their hosts (Normark, 2004).

Regardless of the origin, the evolutionary rationale of PGE can be easily explained from a selfish-gene perspective: maternal genes enjoy a two-fold transmission advantage in PGE males, as they are passed on to all of an individual's offspring – as opposed to only 50% (Bull, 1979). It has also been shown that a PGE causing gene expressed in females would spread rapidly throughout the population (Brown, 1964). Furthermore, subsequent theoretical studies of haplodiploidy suggest paternal genome exclusion is the result of maternal allele victory in a conflict with paternal alleles over transmission by males (Bull, 1979; Herrick and Seger, 1999). However, this two-fold transmission advantage has raised the question of why such biased inheritance is observed in only some species, and in some forms, but not in others (Gardner and Ross, 2014). It suggests that some necessary or predisposing conditions for PGE must be relatively uncommon (Herrick and Seger, 1999).

PGE shares several key features with arrhenotoky (Burt and Trivers, 2006). In particular, all genes transmitted by males derive from their mothers, and are passed on only by their daughters. Both genetic systems often co-occur in closely related taxonomic groups including scale insects, mites and beetles suggesting that similar selection pressures may underlie the evolution of both genetic systems (Herrick and Seger, 1999). A recent theoretical study suggests that species mating ecology is an important factor in predisposing it to genome elimination (Gardner and Ross, 2014). In particular, inbreeding can facilitate the evolution of PGE through its resulting selection for female-biased sex ratios. Furthermore, haploid gene expression in PGE males could promote inbreeding by purging recessive deleterious alleles and reducing the costs of homozygosity (Gardner and Ross, 2014).

1.2.4 Evolutionary consequences of PGE

PGE affects the evolution of a species in a number of ways. Male beneficial traits that reduce female fitness are unlikely to spread under PGE as males

only obtain reproductive success through daughters (Kraaijeveld *et al.*, 2009). In males with haploid gene expression, recessive mutations are exposed to selection. This reduces the genetic load and increases the rate at which rare recessive beneficial mutations can spread. Therefore, these species are expected to adapt more readily to environmental changes (de la Filia, Bain and Ross, 2015). Of course, the evolution of male-limited traits is more convoluted, as sons do not inherit these from their fathers. Their closest male progenitors are their maternal grandfathers, and so selection on male traits skips generations (Boulton, Collins and Shuker, 2014). This has implications for sexual selection of exaggerated male traits that arise through competition for females. Such traits evolve if females choose to mate with a male carrying them because either i) the trait is indicative of high genetic quality (Handicap Principle) (Zahavi, 1975) or ii) because their sons will inherit this trait and thus have increased mating success (Fisherian Runaway selection) (Fisher, 1930). Models show that the effect of skipping a generation of male inheritance means rare alleles encoding male ornaments are particularly vulnerable to loss through genetic drift (Reeve and Pfennig, 2003). Furthermore, the haplodiploid transmission of genes in males promotes sexual selection through the Handicap principle and impedes Fisherian Runaway section (Kirkpatrick and Hall, 2004).

As a consequence of germline elimination of paternal chromosomes, the sperm produced by PGE males are genetically identical (barring mutations). Therefore, like haplodiploid males, they are more likely to evolve mechanisms of sperm cooperation than diploid males in which each individual sperm carries a unique haploid genome (Immler, 2008). Indeed, a striking feature of spermatogenesis in PGE scale insects is the formation of sperm bundles. In these species, individual sperm cells have lost motility. This motility is regained when they are assembled into motile sperm bundles towards the end of spermatogenesis (Jamieson, 1987).

As previously discussed, there is a strong association between the occurrence of inbreeding and PGE. PGE – and some haplodiploid – species are more resistant to inbreeding depression due to a reduced genetic load (Werren, 1993; Henter, 2003). However, PGE species in which the paternal genome is expressed in males are more likely to suffer from substantial inbreeding depression (Gardner and Ross, 2014). In terms of indirect genetic effects, PGE females produce broods that are on average genetically less diverse than broods produced by diplodiploid females. They might therefore be under selection to compensate for this through multiple mating. This hypothesis is supported by studies on haplodiploid eusocial insects (Oldroyd and Fewell, 2007).

The ability to control the sex ratio of offspring is well documented in haplodiploids, where, unlike genetic sex determination there is no default sex ratio of 50:50 (Bull, 1983). The ability to control the sex ratio of offspring increases conflicts over sex allocation. Haplodiploid mothers generally favour an investment into each sex (Gardner, 2014) and this is likely to be the case with PGE mothers. Whereas haplodiploid fathers, because they are not related to male offspring, favour a strongly female-biased sex ratio and may evolve ways to manipulate sex determination mechanisms (under PGE) or persuade their partner to increase fertilisation rates (under arrhenotoky) (de la Filia, Bain and Ross, 2015). Support for the possibility that arrhenotokous males can, under certain conditions, manipulate sex allocation in their mates comes from studies in parasitoid wasps (Shuker *et al.*, 2005) and spider mites (Macke, Olivieri and Magalhães, 2014). Although no studies have investigated PGE male influence on sex allocation, which may in fact be more likely to occur as fathers' genes are present in their sons (Shuker, Moynihan and Ross, 2009).

1.3 Genomic imprinting

As well as being connected evolutionarily, embryonic and germline elimination of the paternally inherited chromosomes in males are connected mechanistically. Both systems are directed by genomic imprints – differences between maternal and paternal homologs that influence chromosome behaviour (Nur, 1980). These imprints are established in the germline and must persist from the parent, through to the zygote to gametogenesis. Thus, they must be perpetuated through multiple cell divisions. Furthermore, because an allele that is paternally derived in one generation may be maternally derived in the subsequent generation, the imprint must be erased and reset in the germline (Herrick and Seger, 1999). In this way, genomic imprinting is a classic example of an epigenetic phenomenon. Indeed, imprinting has been extensively studied in mammals where findings have revealed a key role for epigenetic modifications in the parent-of-origin specific marking and behaviour exhibited by imprinted genes (Ferguson-Smith, 2011). Imprinted genes in mammals have many roles, predominantly in embryonic development and placental function (Weaver and Bartolomei, 2014). Furthermore, there are a number of rare congenital disorders caused by defects in imprinting, such as Prader-Willi syndrome and Angelman syndrome (Peters, 2014). Despite its biological importance, imprinting in mammals occurs on a relatively small scale, affecting few genes (~150 in mice and ~70 in humans discovered so far) (Peters, 2014). However, in insects with PGE, the scale of imprinting is much larger as the entire paternally-inherited haploid genome is imprinted. The molecular mechanisms underlying the process of PGE remain elusive but preliminary studies suggest a role for epigenetic mechanisms similar to those responsible for imprinting in mammals (Bongiorni *et al.*, 2007, 2009).

1.3.1 Epigenetic modifications

The field of epigenetics, and usage of the term 'epigenetic', has increased rapidly since it was first coined by C. H. Waddington in 1942 (Waddington, 1942). Initially defined as '*the branch of biology that studies the causal interactions between genes and their products which bring phenotype into being',* the definition of the term 'epigenetic' has evolved over the past decades and remains generally ill defined. The central idea of an epigenetic process is the decoupling of genotype and phenotype and the existence of a regulatory process between the two. In this thesis, the term 'epigenetic' is defined according to Deans and Maggert (2015): '*the study of phenomena and mechanisms that cause chromosome-bound, heritable changes to gene expression that are not dependent on changes to DNA sequence'*. DNA methylation and histone modifications are two different, but not mutually exclusive, epigenetic modifications that are known to have a role in genomic imprinting in mammals (Ferguson-Smith, 2011). These epigenetic markers are highly conserved across taxa and are well represented in insects (Burggren, 2017). In this section, I will discuss the diverse roles of epigenetic modifications in vertebrate and invertebrate genomes before moving on to outline their established and hypothesised functions in imprinting.

1.3.2 DNA methylation

DNA methylation is a well-characterised epigenetic modification that is associated with a number of regulatory and developmental processes in plants and animals. It is present in all three domains of life (Klose and Bird, 2006; Suzuki and Bird, 2008), suggesting a common ancestor of Metazoa and possibly, all multicellular life (Glastad *et al.*, 2011). DNA methylation patterns provide epigenetic markings of the genome that are inherited through mitotic, and sometimes meiotic, cell divisions and hence constitute a form of cellular 'memory' (reviewed by Bird, 2002). For this reason, DNA methylation has represented the archetypal mechanism of epigenetic inheritance (Li and Zhang, 2014). Although heritable, DNA methylation

patterns are not permanent, and changes can occur throughout the lifetime of an individual. These changes can be a physiological response to environment (examples in Schrey *et al.*, 2016; Boyko and Kovalchuk, 2011) whereas others can be associated with aging and development (Bird, 2002).

Extensive studies in vertebrate systems have shown that DNA methylation occurs throughout the genome (Suzuki and Bird, 2008), most frequently at 5'-CG-3' dinucleotides, known as CpG dinucleotides (Gonzalgo and Jones, 1997). In most mammals, between 60-90% of all CpG dinucleotides are methylated (Lister *et al.*, 2009; Li *et al.*, 2010). The functional effects of DNA methylation are dependent on its genomic context (Drewell *et al.*, 2014). Methylation in mammalian promoter regions can lead to gene silencing through transcriptional repression (Bird and Wolffe, 1999). In contrast, gene body methylation is often associated with active transcription and differential gene splicing (Elango *et al.*, 2009; Foret *et al.*, 2012). DNA methylation also appears to block mobile element transcription and many methylated cytosines in mammalian genomes are found within mobile elements (Bird, 2002). Indeed, a similar function is described in the fungus, *Neurospora crassa* (Selker *et al.*, 2003).

Until relatively recently, the low or absent levels of DNA methylation in the model organisms, *Drosophila melanogaster* (Rae and Steele, 1979) and *Caenorhabditis elegans* (Simpson, Johnson and Hammen, 1986) suggested a reduced functional significance for DNA methylation in the invertebrates. However, studies are revealing the persistence and functional relevance of DNA methylation in a variety of invertebrate taxa (Bewick *et al.*, 2016). In contrast to the pattern of globally methylated DNA found in vertebrates, DNA methylation in invertebrates is relatively sparse (Bird, 1980; Suzuki and Bird, 2008) and methylated genes are generally found in clusters throughout the genome (Wang *et al.*, 2013; Glastad *et al.*, 2016). Several insect studies have also demonstrated an association between DNA methylation and the

long-term evolutionary conservation of genes (Flores *et al.*, 2012; Sarda *et al.*, 2012; Hunt *et al.*, 2010; Wang *et al.*, 2013; Glastad *et al.*, 2016).

Amongst the insect species known to possess DNA methylation, there is great variation in their genomic methylation levels (Glastad, Hunt and Goodisman, 2014). However, the types of genes targeted by DNA methylation are conserved across insects (Glastad, Hunt and Goodisman, 2014). Methylated genes tend to have housekeeping functions and are ubiquitously expressed throughout development, suggesting that DNA methylation is marking genes for constitutive expression (Foret *et al.*, 2009; Bonasio *et al.*, 2012; Sarda *et al.*, 2012; Wang *et al.*, 2013; Glastad *et al.*, 2016). In contrast, unmethylated genes tend to have more dynamic developmental expression patterns, for example, tissue specific functions (Wang *et al.*, 2013; Glastad *et al.*, 2016). This suggests that an important role of DNA methylation in insects is to stabilise gene expression across development and tissues. This epigenetic modification also appears to demarcate exon-intron boundaries with high levels over exons and near sites of translational initiation and termination (Wang *et al.*, 2013). Additionally, DNA methylation of transposable and repetitive elements, common in mammals, has been observed only at modest levels in basal invertebrates (Feng *et al.*, 2010) and, in insects, is almost non-existent (Regev, Lamb and Jablonka, 1998; Feng *et al.*, 2010; Zemach *et al.*, 2010).

The majority of research into invertebrate DNA methylation focuses on social insects where epigenetic processes play a role in modulating caste development (Lyko *et al.*, 2010; Bonasio *et al.*, 2012). Investigations in eusocial Hymenopteran and Isopteran species have suggested that DNA methylation is associated with alternative splicing (Bonasio *et al.*, 2012; Flores *et al.*, 2012; Glastad *et al.*, 2016). However, no such relationship is found in the non-social hymenopteran, *Nasonia vitripennis* (Wang *et al.*, 2013). Despite the differences in presence of DNA methylation within the

genome, genic methylation appears most likely specialized to function in the regulation of transcription and mRNA splicing in both vertebrates and insects (Glastad, Hunt and Goodisman, 2014). However, at this stage, the universal or diverse roles of DNA methylation in insects remain unclear.

1.3.3 DNA methylation and imprinting

In mammals, methylation at CpG sites is arguably the most important epigenetic modification involved in gene imprinting. Parent-of-origin specific differential methylation of imprinting control regions (ICRs) regulates expression of imprinted genes (Delaval and Feil, 2004). The establishment of these imprints occurs in the germline and in most cases, ICR methylation originates from the egg (Delaval and Feil, 2004). In the primordial germ cells (PGCs), inherited methylation is erased and new methylation marks are added, which are inherited by the next generation. The mechanism by which DNA methylation is removed is not completely understood, but TET dioxygenases, enzymes expressed in PGCs that convert methylated cytosines to hydroxymethylcytosines are thought to be involved (Niemitz, 2013). Alternatively, as PGCs are still undergoing mitosis at this stage, demethylation could be passive (newly synthesised strands are not methylated) (Kagiwada *et al.*, 2013). Indeed, these methods are not mutually exclusive and so demethylation could occur through a combination of both (Weaver and Bartolomei, 2014). The time at which new methylation imprints are established differs between males and females. In males, they are established prenatally and in females they are established after birth (Weaver and Bartolomei, 2014). In the preimplantation embryo, the majority of methylation marks are stripped from the genome (Richards, 2006), however the ICRs retain their parental imprints (Hajkova *et al.*, 2002), which go on to mediate the imprinted expression of genes (Delaval and Feil, 2004).

In insects, the relationship between DNA methylation and genomic imprinting is not clear. This is largely due to the fact that in species for which there is

genomic DNA methylation data, there is no clear evidence of parent-of-origin specific imprinting. There have, however, been a handful of studies investigating the epigenetic regulation of PGE in mealybugs. One study in *P. citri* suggests that the recognition and silencing of paternally inherited alleles under PGE may be regulated by DNA methylation as paternally inherited alleles have lower levels of methylation than maternally inherited alleles in both sexes (Bongiorni, Cintio and Prantera, 1999). This suggests an evolutionarily conserved mechanism for the recognition of parental origin. However, this enzymatic study could not provide insight into the level or distribution of DNA methylation across male and female genomes. Furthermore, a subsequent study yielded contradictory results (Buglia, Predazzi and Ferraro, 1999). Thus, the extent to which DNA methylation plays a direct role in the recognition, transcriptional suppression and germline elimination of paternally inherited alleles is unknown.

Parallels can, however, be drawn between specific silencing of the paternal chromosomes in PGE males and the imprinted inactivation of the paternal X chromosome that occurs in some female mammals. Sex chromosome dosage compensation in mammals is a well-studied phenomenon in which whole chromosome parent-of-origin effects are apparent. Genetic sex determination in mammals generates females with two copies of an X chromosome and males with one copy. To correct the X-linked gene expression imbalance between males and females, females transcriptionally silence one of their X chromosomes (Lyon, 1961). In marsupials, the paternally inherited X chromosome is specifically targeted for inactivation in somatic cells (Cooper *et al.*, 1993). In mice, X chromosome inactivation is similarly imprinted during pre-implantation stages and in extra-embryonic tissues including the placenta (Takagi and Sasaki, 1975). This is comparable to the process by which the paternally inherited alleles in mealybug males are specifically targeted for transcriptional silencing in embryogenesis. The molecular mechanisms involved in distinguishing the parental origin of X

chromosomes in mammals are not conclusive; however, differences in levels and patterns of DNA methylation between the inactive X chromosomes and the active X chromosomes have been identified (Bernardino *et al.*, 2000; Hellman and Chess, 2007; Rens *et al.*, 2010). This lends support to the hypothesis that DNA methylation may have a similar role in PGE.

1.3.4 DNA methylation toolkit

In order to methylate DNA, organisms must possess the components of a DNA methylation toolkit. A family of evolutionarily conserved enzymes, DNA methyltransferases (DNMTs), is responsible for the methylation of DNA (Lyko, 2017). There are three DNMT classes, categorized by their activity in mammalian systems and it is assumed that insect DNMTs have the same function as their mammalian orthologs (first demonstrated in *Apis mellifera* (Wang *et al.*, 2006)). The DNMT1 family of enzymes catalyzes the maintenance of DNA methylation by preferentially methylating hemi-methylated DNA substrates, while the DNMT3 family catalyzes the de novo synthesis of DNA methylation (Klose and Bird, 2006). The DNMT2 family is involved in the methylation of tRNA (Goll *et al.*, 2006). It was generally accepted that one or more copies of DNMT1 and DNMT3 were pre-requisites for functional DNA methylation to occur (Yi and Goodisman, 2009). However, it is becoming clear that the absence of a full complement of DNMTs is not indicative of the absence of DNA methylation within a species (reviewed by Glastad, Hunt and Goodisman, 2014; Bewick *et al.,* 2016). For example, both the silkmoth, *Bombyx mori* and the desert locust, *Schistocerca gregaria*, can methylate their DNA despite lacking a copy of *DNMT3* (Xiang *et al.*, 2010; Falckenhayn *et al.*, 2013). Both species possess a copy of *DNMT1* suggesting that this gene may be sufficient to produce enzymes that possess both *de novo* and maintenance functions or, indeed, that other as yet unclassified genes may be involved. Indeed, a comparative study of DNA methylation in 123 insect species across 11 orders shows that the presence of DNA methylation in insects is associated with the presence of

maintenance *DNMT1* rather than *de novo DNMT3* (Bewick *et al.*, 2016). *DNMT1* is found in all orders of insects investigated in this study with the exception of Diptera, which lack DNA methylation. *DNMT3* is the least conserved of the *DNMT* family (Figure 2). This loss of DNA methylation co-occurred with the loss of *DNMT1* and *DNMT3*. (~200mya) (Misof *et al.*, 2014).

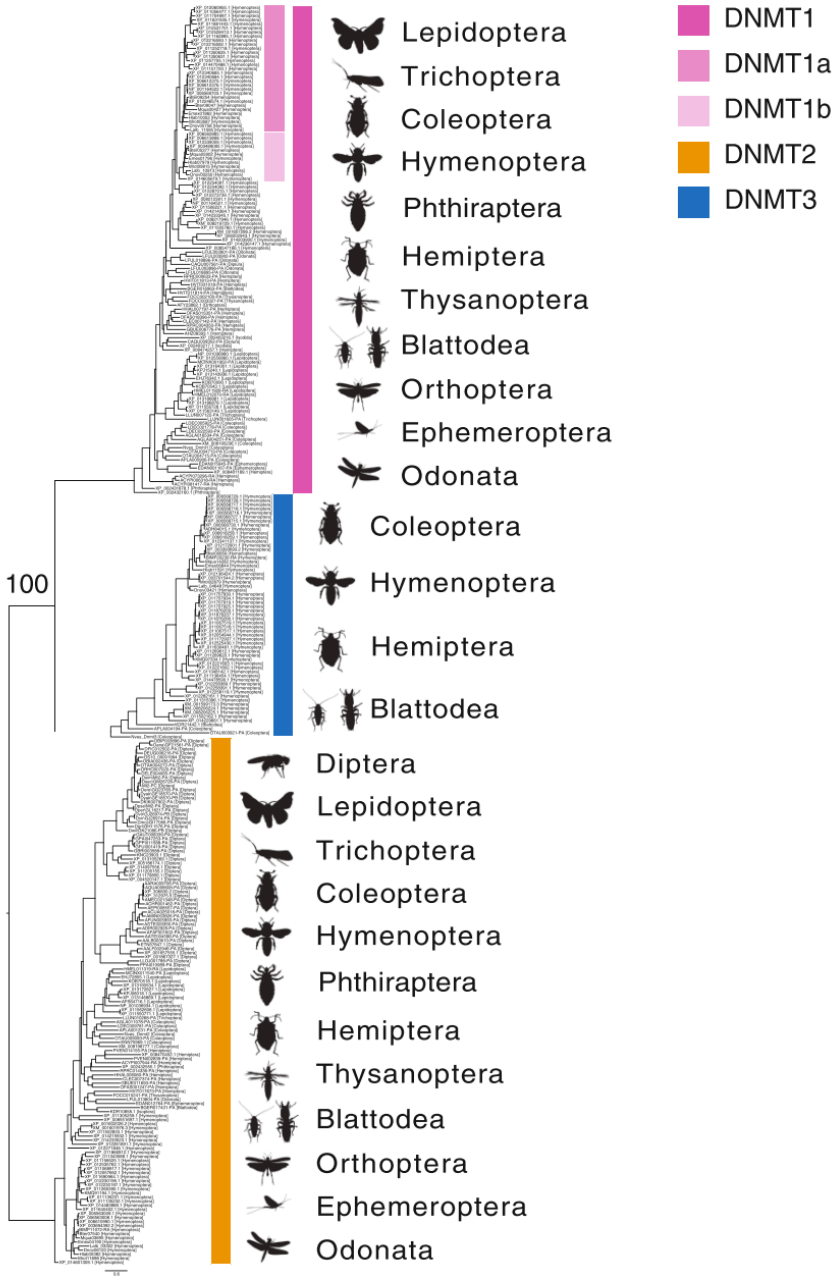


Figure 2: Evolution of DNMT1, DNMT2 and DNMT3 across Insecta and other Arthopoda. DNMT3 is the most order-poor of the DNMTs. Figure taken from Bewick *et al.,* 2016.

Another important component of the DNA methylation toolkit is a family of proteins that encodes a highly conserved methyl-binding domain, known as methyl-CpG binding domain proteins (MBDs). MBDs are attracted and bind to methylated CpG sites in a DNA molecule and through this selective binding can localize chromatin-remodeling complexes to the areas of DNA methylation (Hendrich and Tweedie, 2003). Therefore, by 'reading' the DNA methylation conferred by DNMTs they can affect epigenetic modifications at multiple levels. All organisms with functional DNA methylation have been found to have MBDs. However, MBDs are also found in taxa with no substantial DNA methylation, suggesting another function for these proteins (Glastad, Hunt and Goodisman, 2014). In PGE species, there is a complete lack of knowledge about the presence or absence of DNA methylation machinery.

1.3.5 Histone modifications and chromatin remodelling

Histone proteins form the nucleosome, the structure around which DNA molecules are coiled, and are essential for packaging DNA into chromosomes (Lawrence, Daujat and Schneider, 2016). Histone modifications are an evolutionarily conserved mechanism involved in the regulation of gene expression and silencing (Kouzarides, 2007). The histone tails that protrude from the nucleosome core can be chemically modified in several ways, including via methylation and acetylation, which alter the compaction of chromatin structure and determine the transcriptional state of the genes they interact with (Dong and Weng, 2013). Actively transcribed genes tend to be in loose chromatin structures (euchromatin) so that essential transcription factors and polymerases can access them (Dong and Weng, 2013). Repression of gene transcription can occur via histone modifications that highly condense the chromatin in which genes are located. This form of highly condensed chromatin is called heterochromatin and prevents access to factors involved in gene transcription (Kouzarides, 2007). Histone modifications are known to play a key role in the regulation of gene

expression and chromosome segregation through the formation of heterochromatin (Vermaak and Malik, 2009). A classic example of gene regulation through heterochromatin is X chromosome inactivation in mammals. To compensate for differences in X-linked dosage, one of the X chromosomes in females is inactivated through increased compaction of DNA and formation of facultative heterochromatin (Barr and Carr, 1962; Heard and Disteche, 2006). There are two histone modification pathways involved in the formation of heterochromatin: i) the H3K9me3-HP1 pathway and ii) the H3K27me3-PRC2 pathway (Schotta *et al.*, 2002; Schwartz and Pirrotta, 2007). In mammals, both of these histone modifications are associated with the inactive X chromosomes suggesting a potential role in their transcriptional silencing (Heard, 2005). Indeed, histone modifications that cause the formation of heterochromatin play a crucial role in silencing the paternal genome in PGE (Bongiorni *et al.*, 2007). In *P. citri,* HP1 and H3K9me3 precede the onset of heterochromatinization of the paternal chromosomes in PGE males and, furthermore, HP1 knockouts show a loss of heterochromatinization and associated histone modifications in male embryos (Bongiorni and Prantera, 2001). This suggests a causative role in the silencing of the paternal genome in PGE.

1.3.6 Histone modifications and imprinting

Despite their well-described roles in the regulation of gene expression, less is known about the role of histone modifications in imprinting. However, both the histone modification H3K27me3 and protein complex PRC2 have been observed in imprinted loci in mammalian embryonic stem cells (ESCs), embryos and placental tissues (Weaver and Bartolomei, 2014). Based on this localisation it is likely that H3K27me3 and PRC2 play a role in imprinted gene expression at some level, however, this still remains to be confirmed. In mealybugs, it has been suggested that histone modification H3K9me3 acts as the molecular marker that distinguishes the parental origin of chromosomes in the male germline allowing for the non-independent

assortment of parental chromosomes during meiosis (Bongiorni *et al.*, 2009). It is also suggested that H3K9me3 is carried into the ooplasm on the nuclei of mature sperm and acts as the imprint that identifies the paternally inherited genome in male embryos (Bongiorni *et al.*, 2009). Furthermore, different levels of H3K9me3 found on the two sperm derived from the same meiotic division are believed to have a role in sex determination (Buglia and Ferraro, 2004). However, evidence for these hypotheses is limited. Consequently, the role(s) of histone modifications and heterochromatin in PGE remains unclear.

1.4 Thesis Outline

1.4.1 Study species

In this thesis, I use the citrus mealybug, *Planococcus citri* and the closely related vine mealybug, *Planococcus ficus* to investigate the epigenetic mechanisms underlying PGE. Mealybugs have served as a model organism for cytogenetic analyses of PGE for decades and are easily reared in the laboratory. *Planococcus citri* and *Planococcus ficus* are polyphageous pest species that feed on a large range of economically important crop plants including citrus, cocoa, banana and coffee. These mealybug species are naturally located in warmer climates such as the Mediterranean, the Americas, South Africa and the Middle East (Daane *et al.*, 2012). However, they are frequently found in greenhouses worldwide. In both species, sexual dimorphism is extreme, to the extent that the sexes could easily be confused as different species. Males metamorphose after the fifth-instar and emerge as winged adults, whereas females are neotenous so remain wingless into adulthood. This strong sexual dimorphism is found in all sexually reproducing scale insects and it has been argued that these differences can lead to a shortage of males (because of their fragility and short lifespan) and thus, making it beneficial to evolve reproductive systems that do not rely on males (Hughes-Schrader, 1948). However, it would likely be easier to evolve more

robust males than eliminate males completely (Ross, Pen and Shuker, 2010).

1.4.2 Key outstanding questions

The key outstanding questions regarding the epigenetic mechanisms underlying Paternal Genome Elimination addressed in this thesis are: 1) how is the parental origin of a chromosome distinguished; 2) what epigenetic mechanisms are involved in the silencing of the paternal genome in male somatic nuclei; 3) what epigenetic mechanisms are involved in the germline elimination of the paternal chromosomes during spermatogenesis? Understanding the mechanisms of imprinting in this species will allow for a better understanding of the evolution of this phenomenon, which is found in thousands of insects and has evolved repeatedly in different insect orders. There have been just a handful of studies on genomic imprinting in mealybugs and even less in other insects with this unusual reproductive system.

1.5 Chapter summaries

In Chapter 2, I analyse sex-specific methylome and transcriptome data of the citrus mealybug, *Planococcus citri*. I investigate the relationship between gene methylation and expression in this species. I discuss sex-specific methylation and gene expression patterns and relate this to sexual dimorphism and PGE.

In Chapter 3, I use whole genome bisulfite sequencing (WGBS) on the hybrid offspring of two closely related PGE species, *Planococcus citri* and its sister species, *Planococcus ficus* to investigate the role of parent-of-origin specific DNA methylation in PGE.

In Chapter 4, I investigate the role of two evolutionarily conserved heterochromatin pathways, H3K9me3-HP1 and H3K27me3-PRC2, in the two key processes of PGE: 1) silencing of the paternal genome in somatic tissues and 2) recognition, elimination and imprinting of paternal chromosomes during spermatogenesis.

In Chapter 5, I identify and characterise key genes in the evolutionarily conserved H3K9me3-HP1 heterochromatin pathway: The Heterochromatin Protein 1 (*HP1*) gene family and histone methyltransferase *SU(VAR)3-9* genes. I study expression profiles of these genes throughout development in both sexes and evaluate their similarities to extensively studied *HP1* family and *SU(VAR)3-9* genes in Drosophila.

In Chapter 6, I summarise the main findings of my work and present thoughts on future studies inspired by the results described in this thesis.

Chapter 2: Sex-specific DNA methylation and gene expression in *Planococcus citri*

2.1 Chapter Summary

In this study, I analyse the methylome and transcriptome of the citrus mealybug, *Planococcus citri*. This species exhibits extreme sexual dimorphism and has an unusual reproductive strategy — Paternal Genome Elimination (PGE) — where paternally inherited alleles are silenced in male tissues and then subsequently eliminated from the germline. Males and females of this species lack sex chromosomes and are genetically identical; morphological differences between the sexes must therefore be a consequence of sex-biased gene expression. DNA methylation is an epigenetic modification known to have a key role in the regulation of gene expression in mammals and plants but its functional role in invertebrates remains elusive. Furthermore, few studies have directly tested the role of sex-specific methylation in the regulation of gene expression biases. Here, I describe the DNA methylation machinery encoded within the genome of *P. citri* and present a base-pair resolution analysis of cytosine methylation across the genome. I then compare the methylation landscapes between males and females and relate this to sex-specific gene expression differences detected in the transcriptome.

2.2 Introduction

DNA methylation is a well-characterised epigenetic modification that is associated with a number of regulatory and developmental processes in plants and animals. In vertebrates, DNA methylation occurs extensively throughout the genome (Suzuki and Bird, 2008) most frequently at 5'-CG-3' dinucleotides, known as CpG dinucleotides (Gonzalgo and Jones, 1997): In

human somatic cells, ~60-90% of all CpG sites are methylated (Li *et al.*, 2010). In contrast, CpG methylation in invertebrates is diverse and relatively sparse, from 0% in Diptera to 14% in Blattodea (Bewick *et al.*, 2018), and is almost exclusively restricted to CpG sites in gene bodies (Zemach *et al.*, 2010). In order for an organism to methylate its DNA, it is suggested that at least one copy of both maintenance and *de novo* DNA methyltransferases (DNMT1 and DNMT3, respectively) are required. DNMTs are proteins that catalyse the addition of a methyl group to a cytosine residue in a DNA molecule (Goll and Bestor, 2005). However, recent studies have revealed that insects may be an exception to this rule, requiring only DNMT1 (Xiang *et al.*, 2010; Bewick *et al.*, 2016).

The functional role of DNA methylation has been extensively studied in mammals and plants and is often associated with suppression of gene, or transposable element, expression. Accordingly, sex-specific DNA methylation is implicated in the regulation of sex-specific and sex-biased gene expression in vertebrates (Hall *et al.*, 2014; Maschietto *et al.*, 2017). Sex-biased gene expression also plays a fundamental role in sexual dimorphism. In many species, males and females often differ dramatically in morphology, behaviour and physiology, despite being almost genetically identical. Most of these phenotypic differences are mediated by the differential expression of genes present in both sexes, which evolves as a consequence of different selection pressures acting on males and females (Ellegren and Parsch, 2007).

In invertebrates, the role of DNA methylation in the regulation of gene expression, particularly sex-specific gene expression, is less clear. Insect studies show that DNA methylation in gene bodies is associated with elevated and stable gene expression (Foret *et al.*, 2009; Bonasio *et al.*, 2012; Wang *et al.*, 2013; Glastad *et al.*, 2016). However, many unmethylated genes are also highly expressed, thus, its role in regulation remains elusive (Wang

et al., 2013). Even less is known about DNA methylation differences between the sexes and how these are associated with the sex-biased expression of genes that mediate sexual dimorphism. Sex-biased gene expression has been widely studied in invertebrates and the proportion of genes affected varies amongst species, tissues and developmental stages. In *Drosophila melanogaster* and *D. pseudobscura*, more than 75% of genes show sex-biased expression, with the majority of these differences attributed to genes expressed in gonads (Assis, Zhou and Bachtrog, 2012). However, all *Drosophila* species lack DNA methylation and so do not provide any insight into its role in the regulation of this sex-biased expression. A key study in the wasp species, *Nasonia vitripennis,* which does have DNA methylation, revealed that although over 75% of expressed genes show sex-biased expression, DNA methylation patterns between the sexes are similar and do not explain gene expression patterns (Wang, Werren and Clark, 2015). In contrast, a study in the peach aphid, *Myzus persicae*, in which 19% of genes exhibit sex-specific expression bias, does reveal a correlation between sex-specific gene expression and sex-specific methylation, particularly for genes located on the sex chromosomes (Mathers *et al.*, 2018). Thus, the role of sex-specific gene methylation in regulating sex-biased expression in insects remains unclear.

The citrus mealybug, *Planococcus citri* (Hemiptera: Pseudococcidae), is a unique and enticing insect model in which to study the functional role of DNA methylation in sex-specific gene expression. *P. citri* is a sexually-reproducing species in which sexual dimorphism is extreme, both in terms of morphology and patterns of gene expression: Whilst the sexes are indistinguishable as nymphs, adult males and females are so morphologically distinct they could be mistaken as members of different species (Figure 1 & Table 1). Males undergo metamorphosis after the second instar and develop into winged adults (Vea *et al.*, 2016). Females do not metamorphose, so remain wingless, and grow much larger than the males (Sutherland, 1932). There is

also a marked difference in life history between the sexes: In contrast to females, males do not feed after their second instar. Consequentially, there is a large difference in lifespan between the sexes; with males only living up to 3 days after eclosion, while females can live several weeks after sexual maturity (Nelson-Rees, 1960). Crucially, *P. citri* males and females are genetically identical (no sex chromosomes) (Hughes-Schrader, 1948); therefore, the observed sexual dimorphism is solely a consequence of gene expression differences between the sexes. In addition to extreme sexual dimorphism, this species also has an unusual reproductive strategy, known as Paternal Genome Elimination (PGE). PGE is a genomic imprinting phenomenon in which paternally-inherited alleles are silenced in early development and subsequently eliminated from the germline of males. As such, males are functionally haploid in terms of gene expression and only transmit maternally inherited alleles to their offspring (Brown and Nelson-Rees, 1961). Females, on the other hand, do not undergo the process of PGE and both maternally and paternally-derived chromosomes remain euchromatic throughout development.

Figure 1: Sexual dimorphism in *Planococcus citri*. This figure shows the extreme sexual dimorphism present in *Planococcus citri*: a) shows a winged adult male, b) shows a wingless, neotenous female and c) shows a male and female mating, where size difference between the sexes is apparent.

Due to ploidy and extreme sexual dimorphism, it is expected that *P. citri* will have sex-specific gene expression and that DNA methylation may be

involved in regulating these expression biases. However, there is no empirical data to support this hypothesis. Furthermore, the functional role(s) of DNA methylation in the silencing and loss of paternal chromosomes that takes place in *P. citri* males remains elusive. Previous studies provide evidence for DNA methylation in the *P. citri* genome (Achwal, Iyer and Chandra, 1983), but the extent and distribution of methylation throughout the genome are unknown. Enzymatic studies of DNA methylation provide no evidence for sex-specific methylation in *P. citri* (Bongiorni, Cintio and Prantera, 1999); but the technique used cannot detect methylation patterns at a base-pair resolution and so further studies are required to confirm this result. Furthermore, it is not yet known whether *P. citri* possesses the full complement of DNA methyltransferase genes required for DNA methylation.

In order to address these questions, I describe the DNA methylation machinery encoded within the *P. citri* genome by identifying DNA methyltransferase (*DNMT)* genes and associated methyl-binding domain containing proteins. Using quantitative expression analysis of DNA methyltransferase 1 (*DNMT1)* genes throughout development in both sexes, I examine the functional role of sex-specific DNA methylation in this species. Using whole genome bisulfite sequencing (WGBS) and transcriptome sequencing analyses of adult male and female *P. citri*, I describe the methylome of this species and analyse the relationship between gene methylation and expression. I then compare whole genome levels and patterns of DNA methylation between the sexes to identify key differences that may regulate sex-biased gene expression and mediate their strong dimorphism. I identify sex-biased gene expression patterns and describe their association with sex-specific methylation.

Together, these different approaches provide insight into the functional role of DNA methylation and sex-biased gene expression in sexual dimorphism. Additionally, the methylome and transcriptome of a non-social, non-

Hymenopteran insect broadens understanding of the function(s) and evolution of DNA methylation within insects, and also sheds light on the role of DNA methylation in the process of sex-biased gene expression and genomic imprinting.

2.3 Methods

2.3.1 Insect husbandry

Mealybugs were cultured on sprouting potatoes in sealed plastic bottles at 25°C and ~70% relative humidity. Under these conditions, *P. citri* has a generation time (time from oviposition until sexual maturity) of approximately 30 days. Experimental isofemale lines originate from natural populations and are reared in the laboratory under a sib-mating regime. In each generation, one mated female is taken from culture and transferred to a new container to give rise to the next generation. The *P. citri* line used (WYE 3-2) was obtained from pest control company, WyeBugs and had undergone 32 generations of sib-mating prior to this experiment. The highly inbred line of *P. citri* used here allows for precise mapping of Whole Genome Bisulfite-seq (WGBS) reads and CpG methylation calls to the reference genome and reduces complications caused by SNP variation found in various populations.

2.3.2 Isolating insects for WGBS and transcriptome sequencing

Virgin females were isolated after becoming sexually distinguishable from males (3rd-4th instar) and kept in separate containers until sexual maturity (>35-days old). Males were isolated at pupal stage and kept in separate containers until eclosion (~27 days). Insects were stored at -80°C until DNA and RNA extraction.

2.3.3 Identification of DNA methylation machinery

Amino acid sequences of DNA methyltransferases 1, 2 and 3 and methyl-binding domain (MBD) homologs from *Acyrthosiphon pisum* (DNMT1:

XP_008184506.1; DNMT2: NP_001280303.1; DNMT3: XP_016663751.1; MBD: NP_001156167.1), *Bombyx mori* (DNMT1: NP_001036980.1; DNMT2: NP_001036934; MBD: XP_004929675.1), *Apis mellifera* (DNMT1: NP_001036980.1; DNMT2: XP_006563008.1; DNMT3: NP_001177350.1; MBD: XP_006565475.1) and *Bemisia tabaci* (DNMT1: XP_018908714.1; DNMT3: ATN96644.2; MBD: XP_018906111) were collected from the NCBI database and used to BLAST search against the *P. citri* genome (mealybug.org, version v0). tBLASTn searches were carried out in Geneious R8.1.5 (Kearse *et al.*, 2012) using default settings. Hits were discarded if they did not meet the following threshold values: E-value ≤ 1e-10 and query coverage ≥ 50%. *P. citri* protein sequences meeting threshold criteria were then used as queries in BLAST-searches against the NCBI database (NCBI Resource Coordinators, 2016) to identify the presence of conserved domains and relevant orthologs from other species. A *P. citri* gene was considered to be a full, functional *DNMT* or *MBD* homolog if it fulfilled the following criteria: (1) top NCBI blast hit was the relevant gene in another species and (2) contained all conserved domains required for functionality.

2.3.4 Expression analysis of DNA methylation machinery

2.3.4.1 PCR validation

RT-PCR was used to confirm expression of *P. citri DNMT1* transcript in adult males and females. PCR was performed on cDNA using two biological replicates for each group. Negative controls were used to identify contamination and primer dimers. Products were amplified in 25µl reactions using MyTaq™ Red PCR mix (Bioline, UK). This was performed using the following program: [1] 1 min at 95°C for initial denaturation, [2] 35 cycles of 15 secs at 95°C, 15 secs at 65°C and 10 secs at 72°C and [3] 5 mins at 72°C for final extension.

2.3.4.2 RNA extraction for qPCR

Total RNA was extracted from inbred male and female *P. citri* at key developmental stages: male- and female-biased embryos, 3rd instar males and females, adult males, virgin and mated females (Table 1). Females' first broods are known to be male-biased while embryos laid on day 3 are female-biased (Ross *et al.*, 2012). As there is no way to identify the sex of an embryo without destroying it, I used first broods as a proxy for male embryos and third-day broods as a proxy for female embryos. To avoid bottle effects, each sample included insects from at least 3 breeding bottles. 10 biological replicates were prepared for each sample where 1 biological replicate equals: 10 females (for mated females, virgin females, 3rd instar females); 20 males (for adult males and 3rd instar males); 3 egg masses (for embryo groups). RNA extraction was performed using TRIzol® reagent (Thermo Fisher Scientific, USA) according to the manufacturer's instructions. RNA was treated with DNase to remove any DNA (Thermo Scientific DNase I, RNase-free kit) according to manufacturer instructions. Quantity and quality of extracted genetic material was assessed using NanoDrop ND-1000 Spectrophotometer (Thermo Scientific, USA) and Qubit (Thermo Fisher Scientific, USA) assays. A260/A280 and A260/A230 ratios were calculated for all samples and only samples with A260/A280 of 1.7 > 2.0 and A260/A230 of >1.0 were processed.

2.3.4.3 Reverse Transcriptase and cDNA preparation

RNA was reverse transcribed using M-MLV RT and random hexamers according to manufacturer's instructions (Promega, USA). Negative RT controls for use in qPCR were set up. In total, 10 cDNA samples for each group were prepared.

2.3.4.4 Identification of RP49 and RP17 housekeeping genes in Planococcus citri

Reference genes *Ribosomal Protein 49* (*RP49)* and *RP17*, were identified in the *P. citri* genome using the methods described above for identification of *DNMT* and *MBD* genes. These genes were chosen as *RP49* is used in qPCR analysis in the Japanese mealybug*, P. kraunhiae* and other Hemiptera (Sugahara *et al.*, 2017) and *RP17* is used to normalize gene expression data in *P. citri* (Duncan *et al.*, 2014).

2.3.4.5 qRT-PCR

Quantitative real-time PCR was carried out on the samples in Table 1 to compare expression levels of *DNMT1* in both sexes throughout development. qRT-PCR reaction was performed on an Applied Biosystems StepOnePlus system using Fast SYBR® Green master mix (Applied Biosystems, Thermo Fisher Scientific, USA). Reference genes are ribosomal protein *RP49* gene and ribosomal protein *RP17* gene. *RP49* and *RP17* primers were designed based on predicted *RP49* and *RP17* genes in the *P. citri* genome. Primers for reference sequences and *DNMT1* target sequence were designed such that amplicons produced for reference and target genes were similar in length (161-231bp) in order to minimize differential effects of RNA degradation or PCR inhibition (Table S1). PCR cycle using StepOne Real-Time PCR systems (Thermofisher) was as follows: [1] Holding stage: 2 mins at 60°C, [2] Cycling stage: 40 cycles of 10 secs at 95°C and 30 secs at 60°C, [3] melt curve stage (step and hold): 15 secs at 95°C, 1 min at 60°C and 15 secs at 95°C. All other settings were left as default. Applied Biosystems StepOnePlus system was used to validate amplification efficiency and specificity. 10-fold serial dilution standard curves were run with each primer pair on representatives from each sample group to ensure reaction efficiencies in the range of 90-100% and r^2 values of >0.9. Melt curves were also visualized to ensure specificity of reactions. All cDNA samples were

diluted 1:7 and run in triplicate to account for technical variation. For each sample, all target and reference genes were assayed on a single plate.

2.3.4.6 Analysis of DNMT1 expression

The relative expression levels of each gene were calculated using the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001). One-way ANOVAs and Tukey pairwise comparisons were used to identify significant differences between groups. Statistical analyses were performed in R Studio v3.5.0 (R Core Team, 2013).

Table 1: Expression of *DNMT1* gene was analysed in 7 *Planococcus citri* groups, which represent both sexes throughout development: embryos, 3rd-instar juveniles and adults. Illustrations are provided to detail morphology at different stages.

Group	Morphology	Description
Adult male		1-2 days after pupation
Mated female		~35 days old, immediately after first egg deposit
Virgin female		35 days old
3rd instar female		Sexually immature
3rd instar male		Sexually immature
Embryos		Male-biased embryos – 1st brood Female-biased embryos – 3rd brood

2.3.5 Whole Genome Bisulfite Sequencing

2.3.5.1 DNA extraction

Genomic DNA was extracted from pools of ~60 whole adult males and 15 whole virgin adult females (~35 days old) using DNeasy Blood and Tissue kit (Qiagen, CA) and Promega DNA Clean and Prep Kit (Promega) in a custom DNA extraction protocol. Individual adult males have less body mass than their female counterparts; therefore, a higher number of males were required for each pooled sample. Five independent biological replicates were set up for each sex. DNA samples were cleaned and concentrated using Zymo DNA Clean and Concentrator Kit according to manufacturer's instructions. DNA A260/A280 absorption ratios were measured with a NanoDrop ND-1000 Spectrophotometer (Thermo Scientific, USA) and concentrations were measured with a Qubit Fluorometer (Life Technologies, CA). Although five samples for each sex were prepared, two male samples had to be merged in order to collect adequate DNA for bisulfite conversion and library preparation processes. Therefore, there are only four male replicates.

2.3.5.2 Bisulfite conversion and library preparation

Bisulfite conversion and library preparation was carried out on adult male and virgin adult female DNA samples (500ng input) by Beijing Genomics Institute (BGI). Bisulfite treatment converts unmethylated cytosines to uracils which then become thymines in the subsequent PCR amplification, methylated cytosines remain unchanged (Grunau, Clark and Rosenthal, 2001). The efficiency of bisulfite conversion can be limited by a number of factors including reaction conditions and so conversion of methylated cytosines may not be 100% effective (Ehrich *et al.*, 2007). Therefore, the bisulfite conversion rate is estimated based on non-methylated *Escherichia coli lambda DNA* (provided by BGI; isolated from a heat-inducible lysogenic *E. coli* W3110 strain. GenBank/EMBL accession numbers J02459, M17233, M24325, V00636, X00906), which was added to *P. citri* DNA samples.

Sequencing of bisulfite libraries was carried out on an Illumina HiSeq4000 instrument to generate 150b paired-end reads at 30X coverage.

2.3.5.3 Quality control and bioinformatic analyses

Initial QC of Illumina reads was carried out using FastQC v.0.11.7 (Andrews, 2010). Quality and adapter trimming were carried out by BGI. *E. coli* and *P. citri* reference genomes (*P. citri* version v0, publicly available on mealybug.org) were converted to bisulfite format using Bismark Genome Preparation v0.19.0 (Kruger and Andrews, 2011). This process creates versions of the reference genome in which cytosines are converted to thymines and guanines are converted to adenines, allowing alignment of bisulfite converted reads. Illumina reads were first aligned to the converted unmethylated lambda *E. coli* control DNA sequence using Bismark v0.19.0 (Kruger and Andrews, 2011) to estimate the error rate of the C to T conversion. On average, a total of 400,000 reads (0.8%) were uniquely mapped to the *E. coli* genome, generating coverage of 6230X. The average methylation level in *E. coli* for cytosines (Cs) in any sequence context was 0.46% ±0.065 (mean, ±SD), indicating that bisulfite treatment of the *P. citri* DNA was 99.54% efficient and consistent across all samples. Illumina reads that did not map to the *E. coli* DNA sequence were then aligned to the converted *P. citri* genome using Bismark v0.19.0 and Bowtie2 at default alignment mismatch settings. Reads derived from PCR duplicates and reads that mapped to multiple locations in the genome were removed from downstream analyses. An average of 27.2 million reads per sample (~55%) uniquely mapped to the reference genome using a paired-end mapping approach.

Coverage for each of the samples was then calculated using CGmapTools (Guo *et al.*, 2017). Both overall coverage and coverage at CG sites (methylation effective coverage) were calculated. The average coverage per male sample is 17.0X and 18.0X per female sample. The average coverage

at CG sites for males is 9.4X and is 9.1X for female samples. Full details of per sample coverage are provided in Table S2. Overall levels of methylation in various cytosine (C) contexts were calculated for each sample using CGmapTools v0.1.0 (Guo *et al.*, 2017). Average CpG methylation levels of gene bodies (introns and exons) were measured with CGmapTools mtr function (Guo *et al.*, 2017). Genes were considered methylated if average CpG methylation level across the length of the gene is ≥1% and coverage in each sample is ≥5.0X. Gene bodies with <1% methylation level were considered unmethylated. CpG methylation levels of exons, introns, promoters and intergenic regions were calculated using CGmap files and BedTools v2.27.1 (Quinlan and Hall, 2010). Only sites common to all biological replicates were considered in subsequent analyses.

Methylation differences between males and females were assessed using principal component analysis (PCA) and by identifying differentially methylated (DM) sites, regions and genes. PCA was carried out using MethylKit (Akalin *et al.*, 2012). DM sites and regions were identified using CGmapTools v0.1.0 (Guo *et al.*, 2017) and only CpG sites and regions with a minimum coverage of 5 reads per sample were considered for analyses. I used an approach similar to other insect DNA methylation studies (Wang *et al.*, 2013; Wang, Werren and Clark, 2016; Mathers *et al.*, 2018) to detect differential levels of DNA methylation between the sexes. To be considered differentially methylated, a site had to have at least a 15% methylation difference at a 1% FDR ($Q < 0.01$) (Benjamini and Hochberg, 1995). Differentially methylated regions were calculated using the Dynamic Fragment Strategy in CGmapTools. Regions were calculated using the following criteria: [1] maximum fragment size is 1000b, [2] fragment must have at least 5 cytosines, and [3] the maximum distance between two adjacent common cytosines is 100b. I considered regions to be differentially methylated if they had at least a 10% methylation difference at a 1% FDR ($Q < 0.01$). A less stringent percent methylation difference is used at region-level

analysis as the signal of differential methylation may be diluted over the length of the region.

2.3.5.4 Genomic context of differentially methylated sites and GO term enrichment analysis

GO term enrichment analysis of DM gene sets was performed using GOAtools version v0.6.10 (Klopfenstein *et al.*, 2018). Redundant terms for molecular functions and biological processes were then removed using REVIGO (Supek *et al.*, 2011).

2.3.6 Transcriptome sequencing

2.3.6.1 RNA extraction

RNA (3 biological replicates per sample) was extracted using TRIzol® reagent (Thermo Fisher Scientific, USA) according to the manufacturer's instructions and PureLink RNA purification kit (including DNase I digestion). Samples were further purified with RNA Clean and Concentrator™-5. Quantity and quality of extracted genetic material was assessed using NanoDrop ND-1000 Spectrophotometer (Thermo Scientific, USA) and Qubit (Thermo Fisher Scientific, USA) assays. A260/A280 and A260/A230 ratios were calculated for all samples and only samples with A260/A280 of 1.7 > 2.0 and A260/A230 of >1.0 were processed. All RNA samples were sequenced by Edinburgh Genomics. Two of the samples (one male and one female) were sequenced on the Illumina HiSeq 4000 platform (75b paired-end reads). The remaining samples were prepared on the Illumina NovaSeq S2 platform (50b paired-end reads). All samples generated between 66.9 million and 84.1 million paired-end reads.

2.3.6.2 Gene expression analysis

Raw RNA-seq reads for each sample were trimmed for low quality bases and adapters using Fastp for paired-end reads (Chen *et al.*, 2018) (Table S3). Fastp was used as it allows removal of poly-G tails from NovaSeq reads. Gene-level expression quantification was performed for each sample using RSEM v1.2.31 (Li and Dewey, 2011) with STAR v2.5.2a (Dobin *et al.*, 2013) based on the *P. citri* reference genome and annotation (mealybug.org, version v0). Average expression and coefficient of variation was calculated per gene for individual male and female samples using FPKM (fragments per kilobase of transcript per million) values estimated by RSEM. Only genes with an average of > 1 FPKM in at least one of the sexes were retained for downstream analyses. Differentially expressed genes between the sexes were identified using EbSeq (Leng *et al.*, 2013) based on gene-level expected counts produced by RSEM. A gene was considered differentially expressed if it had a fold-change (FC) ≥1.5 and a *p*-value < 0.05 after adjusting for multiple testing using Benjamini-Hochberg (BH) procedure (Benjamini and Hochberg, 1995).

2.3.6.3 Testing for correlation between differences in methylation and gene expression

To investigate the relationship between changes in gene expression and methylation, the methylation and expression levels of genes in adult males and females were compared. Using average expression levels (FPKM) and methylation levels across replicates, $\log_2$ fold-change in expression and methylation was calculated between the sexes. The correlation was then tested using Spearman's ρ (*rho*).

2.4 Results

2.4.1 Annotation of DNA methylation machinery

In order to gain insight into the DNA methylation machinery present within the *Planococcus citri genome*, I used a computational approach to identify copies of DNA methyltransferase (*DNMT)* genes and methyl-binding domain (*MBD)* genes in the genome. DNMT1 and DNMT3 are an evolutionarily conserved group of enzymes involved in the maintenance and establishment of DNA methylation, respectively (Klose and Bird, 2006). DNMT2 is involved in the methylation of tRNA (Goll *et al.*, 2006). MBDs are another important component of the DNA methylation toolkit as they contain a recognition motif that selectively binds methylated DNA (Klose and Bird, 2006). All organisms with functional DNA methylation activity have at least one copy of a *DNMT* gene and an *MBD* gene (Glastad *et al.*, 2011).

2.4.2 DNA methyltransferases

It appears that *Planococcus citri* possesses one copy of DNMT1, involved in the maintenance of DNA methylation, and one copy of DNMT2. Interestingly, *P. citri* appears to lack a copy of *DNMT3*, which is involved in *de novo* DNA methylation. A similar search in the Hemipteran sister species, *Planococcus ficus* (version v0, mealybug.org) reveals that this species also lacks a copy of DNMT3 but possesses copies of both DNMT1 and DNMT2.

The putative DNMT1 protein identified in *P. citri* is 1343 amino acids in length and genome annotation (mealybug.org) identifies 4 conserved domains: Cytosine-specific DNA methyltransferase replication loci domain (IPR022702), 2 BAH domains (IPR001025) and C-5 cytosine-specific DNA methylase (IPR018117) across 26 exons (Figure S1). This putative DNMT protein sequence was then BLAST-searched against the NCBI database, which identified conserved domains (Table S4). The top hits from this BLAST-search were DNMT1 protein sequences from a number of different insect species including *Bemisia tabaci*, *Bombus impatiens* and *Nasonia*

vitripennis. The average identity match was 44%, query coverage in all hits was >90% and all E-values were 0.0.

Although a copy of DNMT3 appears to be absent from the *P. citri* genome, a gene containing a PWWP domain (IPR000313) is present. PWWP domains are ubiquitous eukaryotic protein modules frequently found in proteins associated with chromatin, including DNMT3 (Slater, Allen and Bycroft, 2003). The PWWP domains in DNMT3 interact with DNA and histone lysine-modified nucleosomes and are required for *de novo* DNA methylation (Dhayalan *et al.*, 2010). However, the *P. citri* gene containing this PWWP domain does not possess a C-5 cytosine-specific DNA methylase domain, which is required for DNA methylation (Bewick *et al.*, 2016). Therefore, it is unlikely that the occurrence of a PWWP domain in this case indicates the presence of DNMT3.

2.4.3 CpG-methyl binding domain (MBD) protein

Genes encoding for MBDs in insects remain poorly studied but their presence is phylogenetically widespread with copies found in Hemiptera, Hymenoptera, Coleoptera, Diptera and Lepidoptera (Glastad *et al.*, 2011). Their presence in species without DNA methylation supports the suggestion that MBD function extends beyond this role (Hendrich and Tweedie, 2003). Two copies of a CpG-methyl binding domain (*MBD*) gene are identified in the *P. citri* genome*,* both of which contain the highly conserved methyl-CpG binding domain. *P. citri MBD1* is 1339 amino acids in length and the transcript has 15 exons. *MBD2* is 798 amino acids in length and transcript has 12 exons.

2.4.4 DNMT1 expression analysis

In order to gain insight into the sex-specific function of DNA methylation throughout development, expression levels of the *P. citri DNMT1* transcript in both sexes throughout development were analysed (Figure 2). *DNMT1* is

expressed at all stages of development in both sexes. Within each sex, the highest expression is found in adults, with adult mated females having the highest expression levels of all groups. 3rd-instar males and 3rd-instar females have the lowest levels of expression. However, there is no significant difference in *DNMT1* expression between the sexes in comparable development groups (Table 2; adult males v virgin adult females, *p*-value = 0.99; adult males v mated females, *p*-value = 0.906; 3rd-instar males v 3rd-instar females, *p*-value = 0.99; male embryos v female embryos, *p*-value = 0.99, Tukey HSD test).

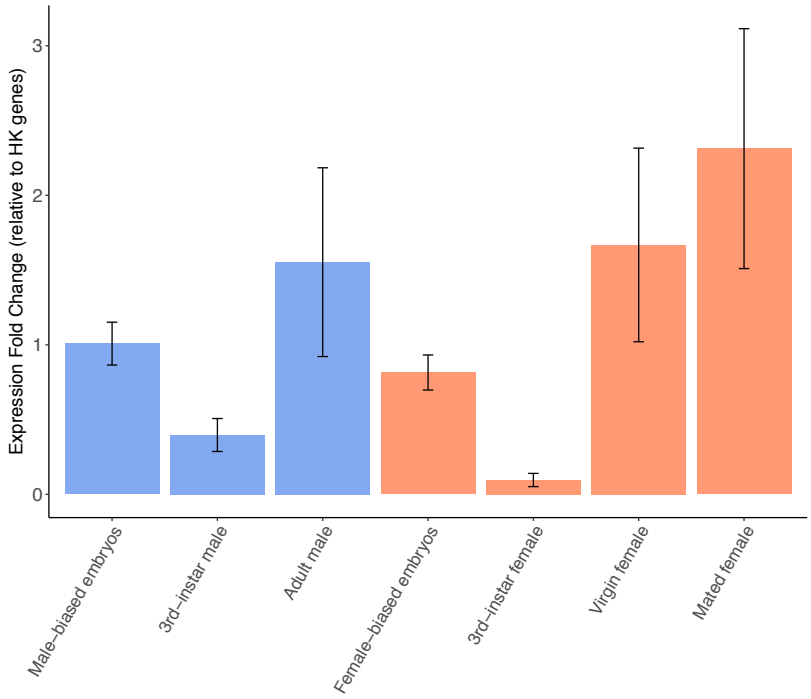


Figure 2: Expression analysis of DNA methyltransferase 1 (*DNMT1*) throughout development in males and females. Expression levels are averaged across 10 biological replicates per sample and are normalised to 2 housekeeping genes (*RP17* and *RP49*). Bar shows average relative expression along with standard error of the mean. There are no significant sex-specific differences in *DNMT1* expression at any of the developmental stages (ANOVA with Tukey HSD test, see Table 2).

Table 2: *DNMT1* expression level comparisons between males and females at key developmental stages. ANOVA and Tukey HSD pairwise comparison tests were carried out and corrected *p*-values show no significant differences in *DNMT1* expression between the sexes.

Comparison	Tukey HSD corrected *p*-value
Adult male x Virgin female	0.99
Adult male x Mated female	0.91
Male-biased embryo x Female-biased embryo	0.99
3rd-instar male x 3rd-instar female	0.99

2.4.5 Patterns of DNA methylation in the Planococcus citri genome

The role(s) of DNA methylation in insects remains unclear and studies of insect DNA methylation beyond the Holometabola are limited. Therefore, I first characterise genome-wide patterns of DNA methylation in *Planococcus citri* before investigating sex-specific differences in DNA methylation. I assessed methylation levels in adult males and virgin adult females in the following cytosine (C) contexts: C, CG, CHG, CHH, CA, CC, CT, CH and CW (where H = A, T or C and W = A or T). Across all *P. citri* samples, I find that only Cs in a CpG context have methylation levels higher than the false positive rate in *E. coli*, indicating that DNA methylation in *P. citri* is predominantly in a CG context (Figure 3). Overall, the global CpG methylation level in *P. citri* is 6.9% ±0.9% (averaged across all samples,

±SD), which is higher than levels reported in other Hemipteran insects (2-4%, Bewick *et al.*, 2016). There are 8,634,723 methylated CpG sites based on the criteria that the site has at least 5X coverage per sample and >10% methylation (Figure 4). My analysis reveals that 37% (3,212,590) of these methylated sites occur in gene bodies (exons = 1,620,275 sites; introns = 1,592,315 sites). The proportion of methylated Cs in exons (40%) is higher than in introns (25.6%) (Figure 4). This pattern is similar to the enrichment of methylation commonly found in the exons of Holometabolous insects (Lyko *et al.*, 2010; Glastad *et al.*, 2011; Bonasio *et al.*, 2012; Hunt *et al.*, 2013). Promoters also show signifcant CG methylation enrichment as 31.7% of all CG sites in these regions are methylated. 1,069,521 methylated CG sites (12.4%) are located within promoter regions and the remainder (4,352,612 sites) are found in intergenic regions.

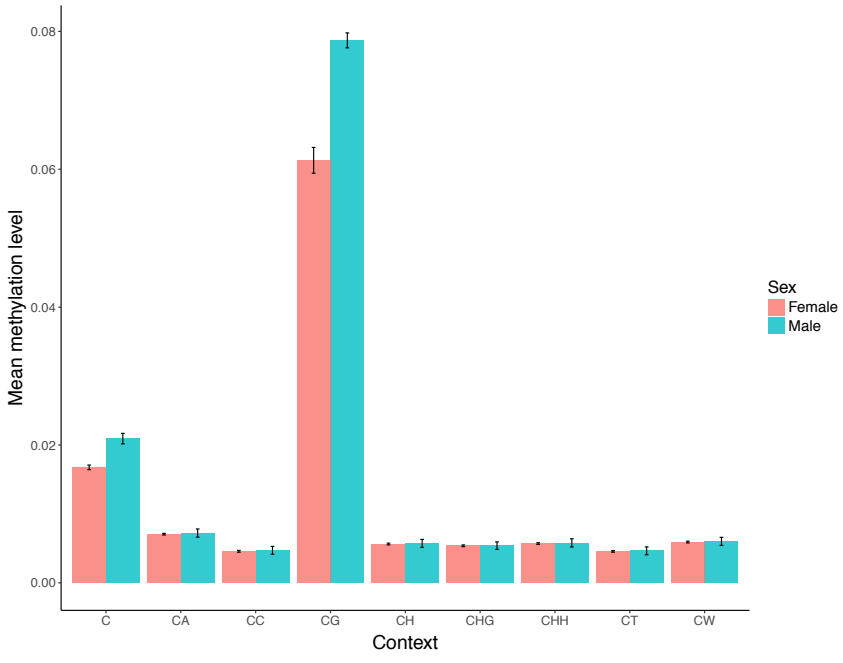


Figure 3: Average level of global methylation in *P. citri* in all C contexts (C, CG, CHG, CHH, CA, CC, CT, CH, CW). In both males and females, the majority of methylation is present in a CpG context.

Genomic Location	Total number of CG sites	Number of methylated CG sites (% of all CGs in region) [% of all mCGs]
Exon	4,062,368	1,620,275 (40%) [18.7%]
Intron	6,207,972	1,592,315 (25.6%) [18.4%]
Promoter	3,374,576	1,069,521 (31.7%) [12.4%]
Gene body (intron & exon)	10,270,340	3,212,590 (31.3%) [37%]

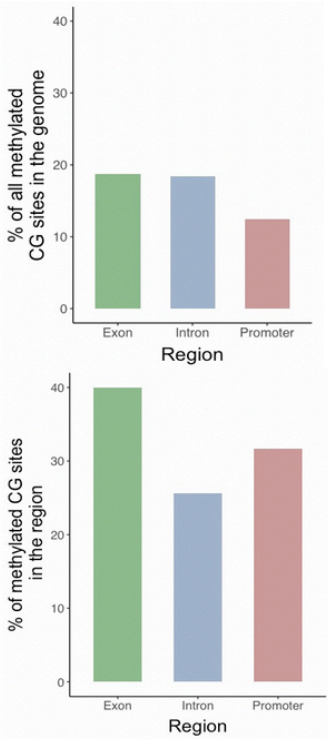


Figure 4: Total number of CG sites within defined genomic regions in the *P. citri* genome (exons, introns and promoters) and the number of methylated CG sites within these regions. The percentages of methylated CG sites within a region are given in brackets (). The percentages of all methylated CGs that fall into regions are given in square brackets [].

Levels of gene body (introns and exons) methylation throughout the genome were calculated. Genes with less than 1% methylation across the length of the gene body are classed as *unmethylated*; genes with > 1% methylation across the length of the gene body are classed as *methylated*. Out of all genes covered in the methylation analysis (n = 37,106), 5399 (14.5%) have less than 1% methylation and are thus, unmethylated. Of these, 304 have 0% methylation. The remainder (31,707 genes) have methylation levels between 1-44%, with the majority of genes in the *P. citri* genome (29,266 genes; 78.9%) having less than 10% methylation (Table 3). Both methylated and unmethylated genes are enriched for GO terms related to core biological processes such as metabolism and biosynthesis (Figure 5, also S5 & S6).

Table 3: Number of genes in different methylation level categories. This table contains the number of genes present in WGBS-seq and RNA-seq data sets within defined categories of DNA methylation levels.

	Methylation level					
	0-1%	1-10%	10-20%	20-30%	30-40%	40-50%
Number of genes	5399	29266	1618	738	78	7

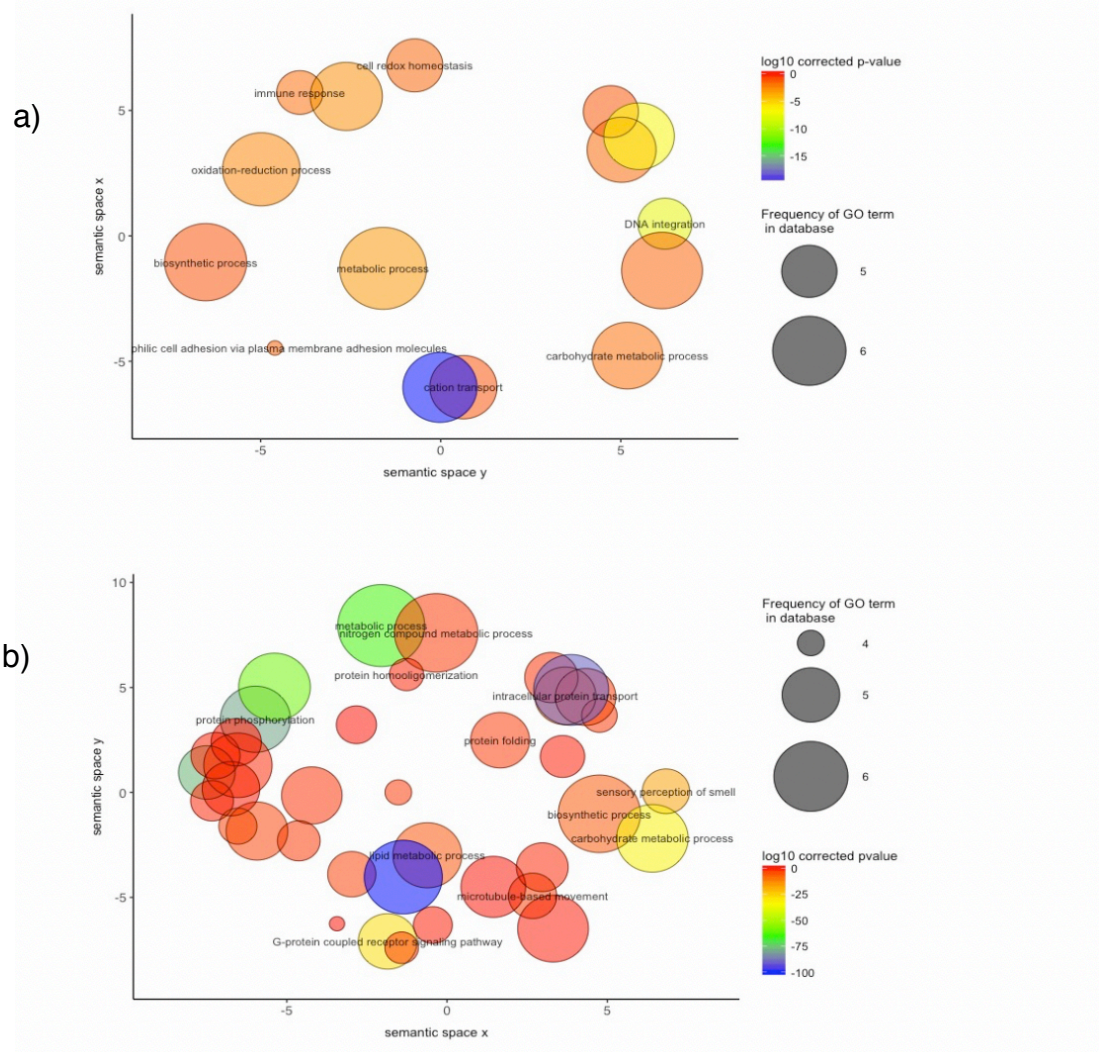


Figure 5: Enriched GO terms related to biological processes plotted in semantic space for a) unmethylated genes and b) methylated genes in *P. citri.* For terms related to molecular functions see Tables S5 & S6.

2.4.6 Sex-specific DNA methylation in Planococcus citri

I compared the methylomes of adult males and females. A PCA analysis based on CpG methylation levels shows that male and female samples cluster together indicating shared variation in global patterns of CpG methylation between the sexes (Figure 6). The level of DNA methylation at CpG sites differs significantly between males and females (Figure 7, T-test, *p*-value = 0.00016): on average, global CpG methylation levels in males (7.9%, se = ±0.2%) are higher than in females (6%, se = ±0.01%). To further characterise sex-specific methylation differences, site-wise and region-wise differential methylation analyses were conducted. I identified 12,962 sites in the *P. citri* genome with significant differences in DNA methylation between males and females. Of these sites, 89% (11,520) are hypermethylated (have higher methylation levels) in females relative to males and 11% (1442) are hypomethylated (have lower methylation levels) in females relative to males (Table 4). DM regions (DMRs) are fragments of the genome in which methylation is significantly different between males and females. DMRs offer a broader analysis of methylation levels than DM sites as DM regions must contain at least 5 cytosine residues and can vary in length from 5bp to 1000bp. In total, 38,848 sex-specific differentially methylated regions were identified in the *P. citri* genome. In 59% of these regions, females have higher methylation levels relative to males and 41% show the opposite pattern (Table 4). As with DM sites, the majority of DM regions are hypermethylated in females compared to males. However, sex-specific methylation differences across regions are less striking than those found at DM sites.

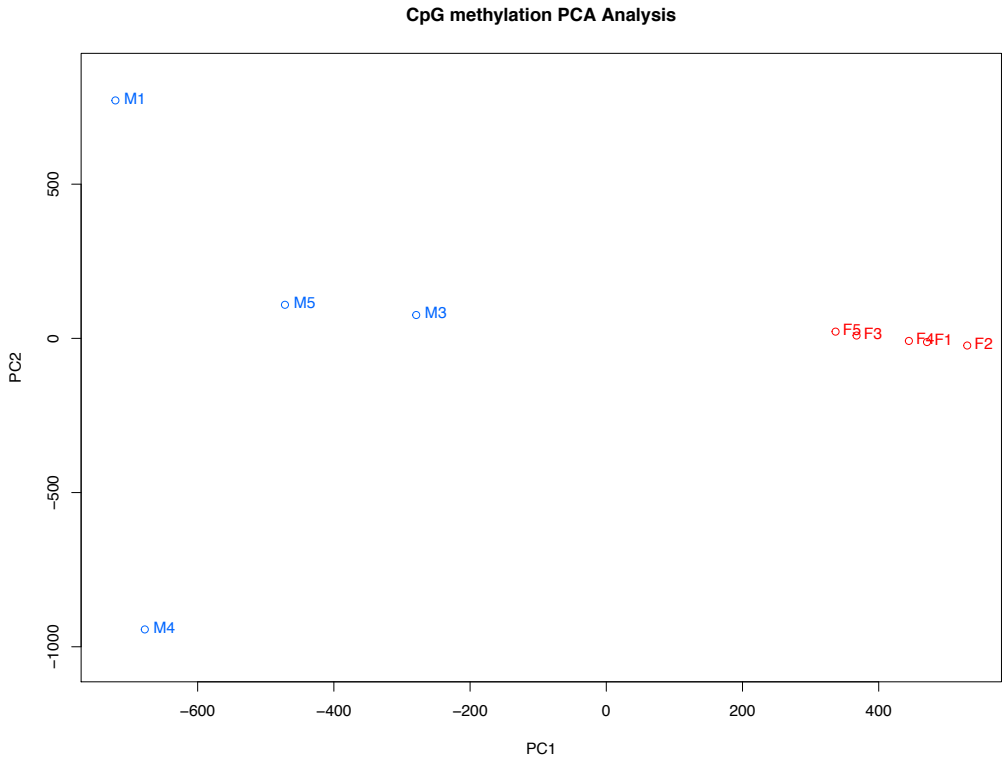


Figure 6: Principle component analysis (PCA) based on methylation levels at CpG sites in all samples. Female samples are red, male samples are blue. Distinct clustering of the samples based on sex suggests reproducible differences in CpG methylation between the two.

Therefore, although males have higher global CpG methylation levels, at the majority of differentially methylated sites and regions, females have higher methylation levels than males. Although, sex-specific methylation differences across regions are less striking than those found at DM sites. This may be due to the fact that regions vary in size from 5 base pairs in length to 1000 base pairs; therefore, the signal of differential methylation may be diluted over the length of the region (Figure S2). As DNA methylation in insects tends to be associated with high and stable gene expression (Glastad, Hunt and Goodisman, 2014), it may be the case that these sites correspond with genes that have elevated expression in females. These results are congruent with the hypothesis that CpG methylation may have a role in regulating sex-specific gene expression in *Planococcus citri.*

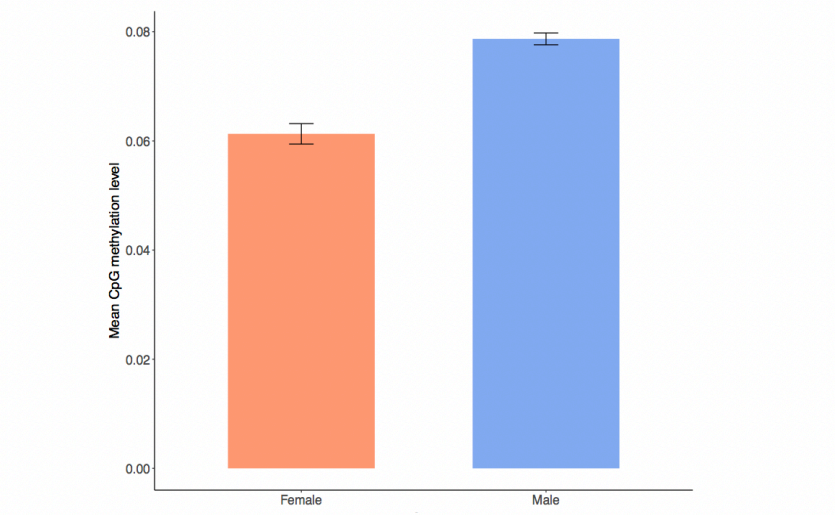


Figure 7: Average global CpG methylation levels of *P. citri* males and females. Female DNA methylation level (6%, se = +/- 0.01%) is significantly lower than male methylation level (7.9%, se = +/- 0.2%), *p*-value = 0.0002 (Welch Two-sample T-test).

Table 4: Number of differentially methylated (DM) sites (methylation difference > 15%, FDR < 1%) and regions (methylation difference > 10%, FDR < 1%) between males and females. A total of 12,962 DM sites and 38,848 DM regions are identified between the sexes. 89% of DM sites are hypermethylated in females and only 11% are hypermethylated in males. At regions, 59% show female hypermethylation, the remainder (41%) show male hypermethylation.

Methylation bias	Sites	Regions
Male hypermethylated	1442 (11%)	15738 (41%)
Female hypermethylated	11520 (89%)	23110 (59%)
Total number	12962	38848

2.4.7 Genomic context of differentially methylated sites

To investigate the function of DNA methylation in mealybugs, the genomic context of sex-specific DM sites was analysed. The majority of both male hypermethylated and female hypermethylated sites are located in gene bodies: 813 sites (55%) and 6260 sites (54%), respectively. The proportion of DM CpGs in gene bodies is higher than the proportion of methylated Cs found in exons (40%), suggesting DM sites may be targeted to regions where regulatory control of gene expression may occur. There is a striking difference in the distributions of DM sites between exons and introns of male and female hypermethylated sites. In males, the majority of hypermethylated sites are located in introns (605 in introns, 208 in exons; binomial test, *p*-value = 2.2×10^{-16}). In females, the opposite is found (1102 in introns and 5158 in exons; binomial test, *p*-value = 2.2×10^{-16}). The proportion of hypermethylated intronic sites is significantly higher in males than in females (binomial test, *p*-value = 2.2×10^{-16}). Additionally, I find that 2023 female hypermethylated DM sites (18%) and 166 male hypermethylated DM sites (12%) are found in promoter regions (2000bp upstream of gene body). The remainder of DM sites fall into intergenic regions (463 male hypermethylated DM sites (32%), 3228 female hypermethylated sites (28%)). These results are summarised in Figure 8.

Genomic Location	Number of male hypermethylated sites	Number of female hypermethylated sites
Exon	208	5158
Intron	605	1102
Promoter	166	2032
Intergenic	463	3228
Gene body (intron + exon)	813	6260

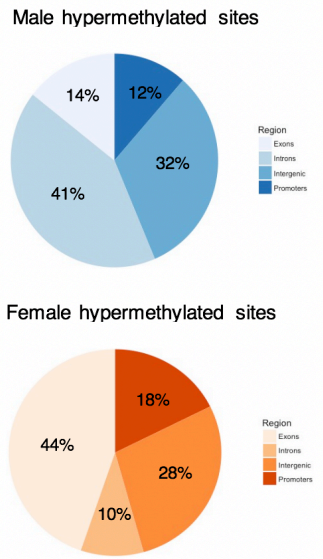


Figure 8: Genomic context of differentially methylated sites that are hypermethylated in males (blue) and females (orange).

2.4.8 Genes with sex-specific DNA methylation

I identified 7953 genes with a sex-specific methylation difference, of which 4863 (61%) were hypermethylated in females relative to males and 3090 (39%) were hypermethylated in males relative to females (Table 5). This is 25% of the 31,713 genes that are methylated in the *P. citri* genome. The number of female hypermethylated genes is significantly higher than the number of male hypermethylated genes (binomial test, *p*-value = 2.2×10^{-16}). GO term enrichment analysis was carried out to identify molecular functions and biological processes that are enriched in differentially methylated genes in males and females. There is no apparent difference between functions and processes enriched in different gene categories. Both male-hypermethylated and female-hypermethylated genes are enriched for GO terms relating to core molecular functions and biological processes, including metabolism and transport (Figure 9) (Tables S7 & S8).

Total number of DM genes	Number hypermethylated in males	Number hypermethylated in females
7953	3090	4863

Table 5: Number of unique genes at differentially methylated sites and regions in *P. citri*. A total of 7953 genes are found within DM regions and the number of genes hypermethylated in females is higher than the number hypermethylated in males.

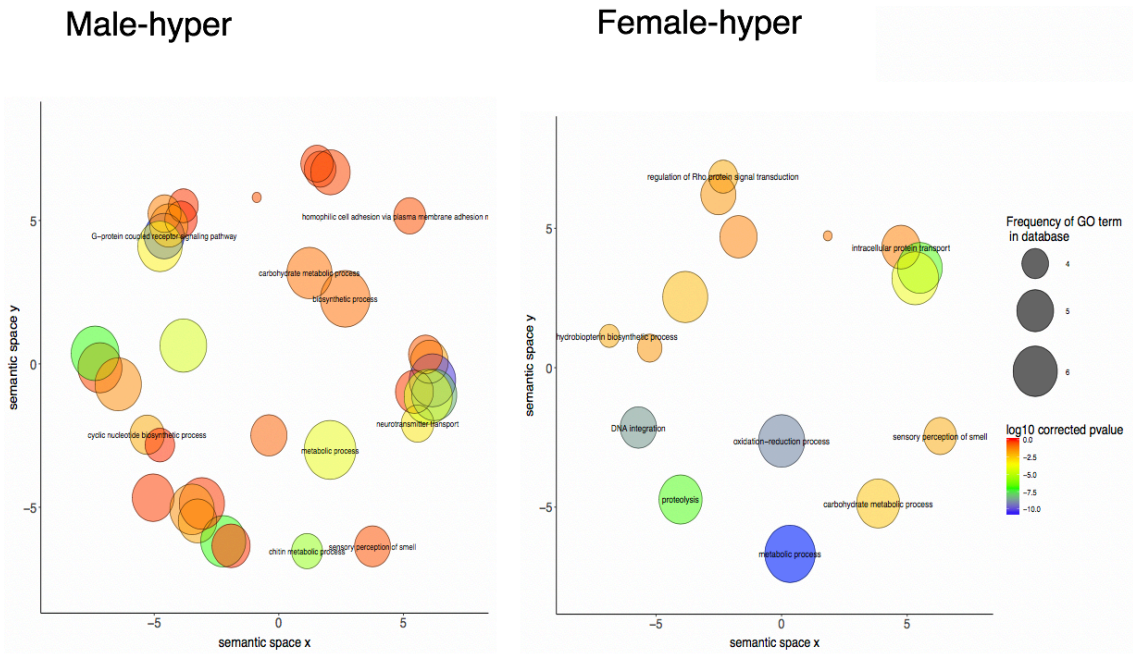


Figure 9: Enriched GO terms related to biological processes plotted in semantic space for differentially methylated genes (for terms relating to molecular functions see Tables S7 & S8).

2.4.9 Sex-specific gene expression

Using an FDR < 0.05 and an expression fold change > 1.5 FPKM between sexes as cut-offs, I identify a total of 10,548 differentially expressed (DE) genes between *P. citri* males and females (Figure 10a). This is approximately 26.5% of the estimated 39,801 genes in the *P. citri* genome (mealybug.org, *P. citri* version v0) and ~40% of all genes with detectable expression in RNA-sequence data (n=19,282; greater than 1 FPKM on average per sample). Genes showing different levels and patterns of sex-bias are likely subject to different evolutionary processes modulating their expression and sex-specificity (Wang, Werren and Clark, 2015). Therefore, in this study I distinguish two general categories of sex-biased genes. The first category contains sex-biased genes, defined as having >1.5-fold difference in expression between the sexes (FDR < 0.05). The second contains extremely sex-biased genes, which are those that show >10-fold difference in

expression between the sexes (FDR < 0.05). Of all DE genes identified, 5448 have female-biased expression and 5100 have male-biased expression (Figure 10b, binomial test, *p*-value = 1.652×10^{-6}). Out of all sex-biased genes, 178 showed extreme female-bias and 344 showed extreme male-biased expression (Figure 10c, binomial test, *p*-value = 1.663×10^{-13}). GO term enrichment analysis of sex-biased genes show that both female and male biased genes are enriched for core biological processes such as biosynthetic processing and carbohydrate metabolism (Figure 11; for molecular functions see Tables S8 & S9). In both sexes, these GO terms overlap with those enriched in methylation-biased genes and include metabolism and sensory perception. GO term enrichment analysis of extremely male-biased genes showed enrichment of genes involved in sensory perception of smell (odorant binding and olfactory receptors) and proteolysis. Female *P. citri* are known to produce pheromones to attract males (Bierl-Leonhardt *et al.*, 1981), therefore it may be that these extremely male-biased genes are involved in pheromone response. Additionally, two of the most highly expressed genes in males are involved in mitochondrial protein transport (IPR018108) and myosin production (IPR001609 & IPR002928), both of which are involved in insect flight (Sacktor, 1970; Bullard, Dabrowska and Winkelman, 1973) These genes are expressed 421 and 342.9 times more in males than in females, respectively. In females, the most highly expressed gene is a lipase (IPR013818).

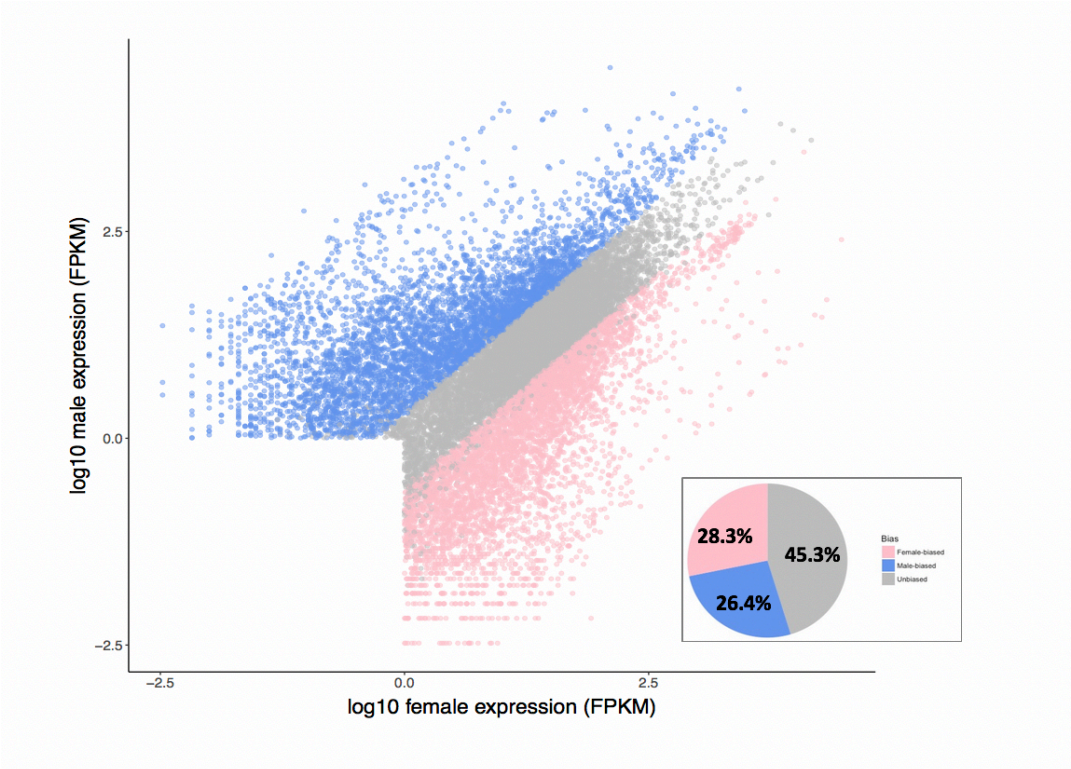


Figure 10a: Male (y-axis) and female (x-axis) gene expression expressed as $\log_{10}$ fragments per kilobase of transcript per million mapped reads (FPKM) averaged over 3 biological replicates. Only genes with >1 FPKM in at least one of the samples were retained for analysis (n=19,282). Differential expression analysis was conducted using EbSeq and DE genes are coloured according to direction of sex-bias (female-bias = pink, male-bias = blue, unbiased = grey). The pie chart insert shows the percentage of genes that fall into each expression bias category. The majority of genes show no significant expression bias (n=8734), 5100 genes show male-biased expression and 5448 show female-biased expression.

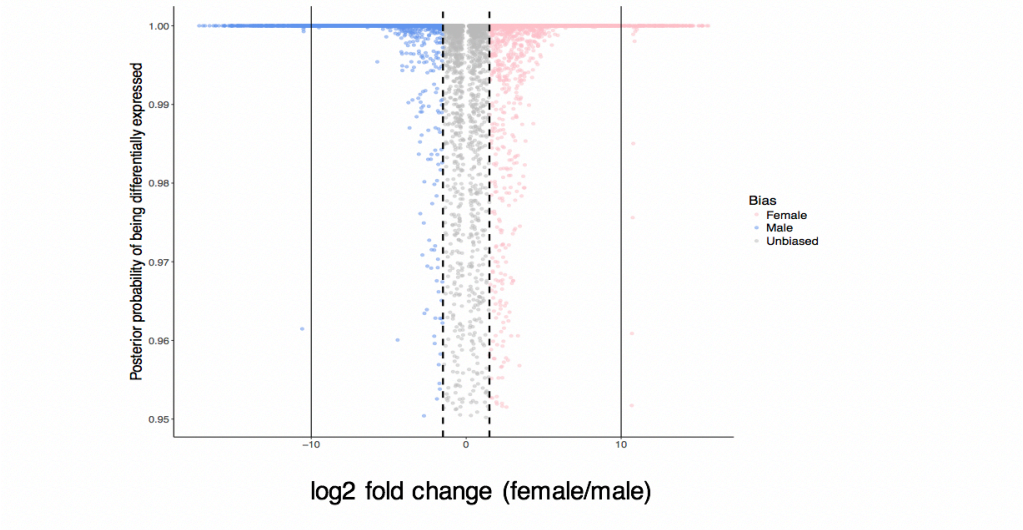


Figure 10b: Fold-change of expression of differentially expressed (DE) genes between males and females (as $\log_2$ fold change female/male) at an FDR < 5% (posterior probability of being differentially expressed > 0.95). DE genes were identified using EbSeq. Only DE genes with > 1.5-fold change are considered sex-biased and genes with > 10-fold change are considered extremely sex-biased. Genes are coloured according to direction of sex-bias (female-bias = pink, male-bias = blue, unbiased = grey).

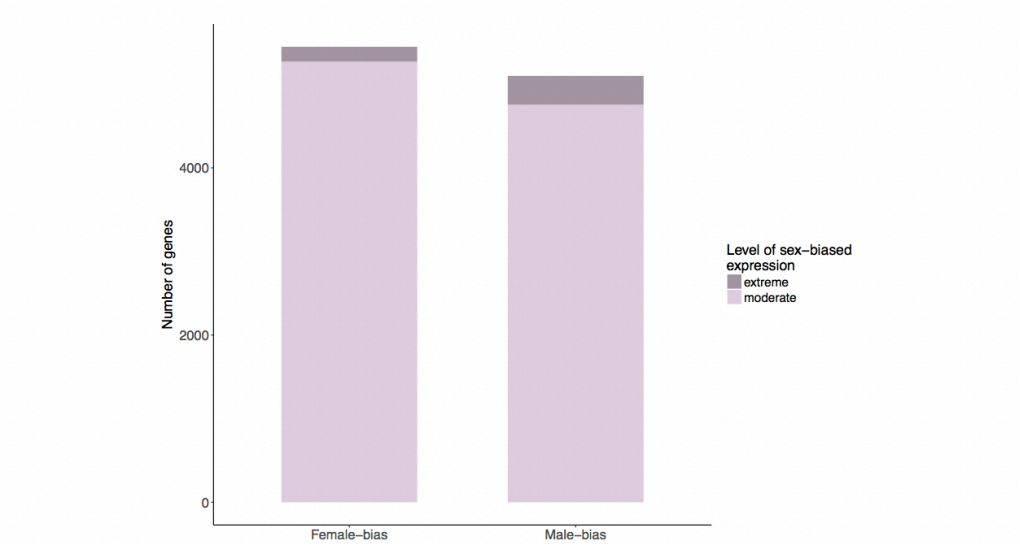


Figure 10c: Number of differentially expressed genes in *P. citri* with male and female expression bias. In total, 10,548 genes are differentially expressed between the sexes (≥1.5-fold change in expression and FDR < 5%). 5448 have female-biased expression and 5100 have male-biased expression. Of these, 178 show extreme female bias (>10-fold change in expression, FDR <5%) and 344 show extreme male bias.

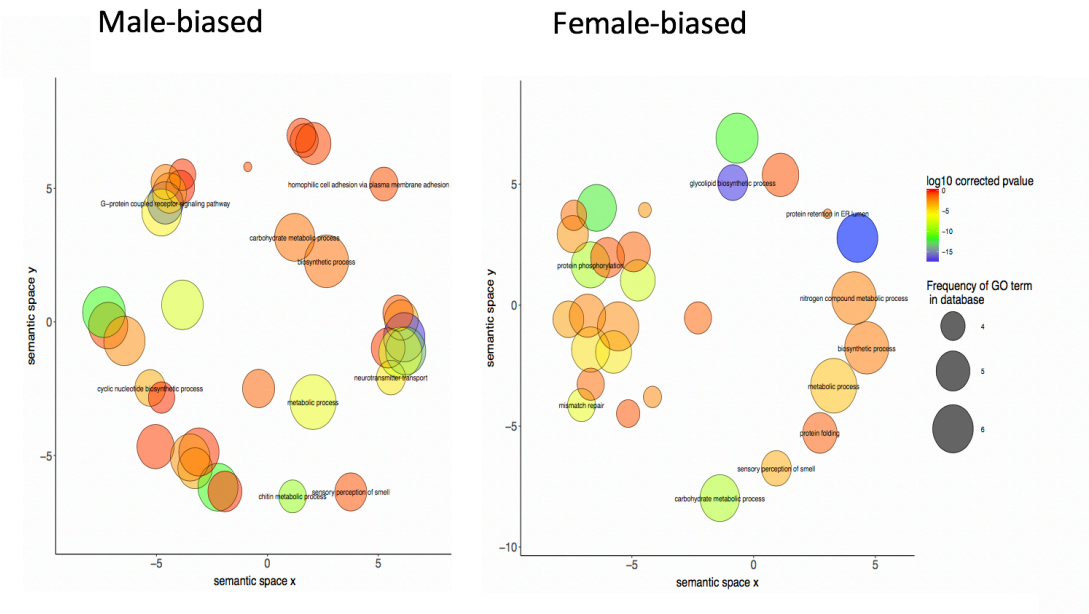


Figure 11: Enriched GO terms related to biological processes plotted in semantic space for sex-biased genes in *P. citri* (for terms relating to molecular functions see Tables S8 & S9).

2.4.10 Relationship between gene expression and methylation

DNA methylation is known to play a role in the regulation of gene expression across a wide range of taxa. Here, I analyse the relationship between the methylation status of a gene and its expression level (FPKM) by concatenating the RNA-seq and WGBS-seq datasets (number of genes covered by both datasets = 23,099, 58% of all predicted genes). Overall, 84% of expressed genes (FPKM > 1, averaged across replicates) are methylated (methylation level > 1% across the length of the gene body) in the *P. citri* genome (number of methylated genes = 19,914; number of unmethylated genes = 3185). However, I find no association between methylation status of a gene and whether it is expressed or not (expressed > 1FPKM; not expressed < 1FPKM; chi-squared test, *p*-value = 0.47). Furthermore, there is no significant association between gene methylation

status (methylated or unmethylated) and expression level (*p*-value = 0.62, Mann-Whitney U Test).

Next, I analyse the association between gene expression and gene body methylation in all genes in both sexes and find no significant correlation (Figure 12a; Spearman's correlation, *rho* = -0.01, *p*-value = 0.06). This analysis was repeated for males and females separately to detect evidence of a sex-specific relationship between gene expression and methylation. In males, I find no correlation between levels of gene expression and levels of gene body methylation (Figure 12b; Spearman's correlation, *rho* = 0, *p*-value = 0.8287). In females, I find a weak but significant negative correlation (Figure 12c; Spearman's correlation, *rho* = -0.02, *p*-value = 0.0043). I also conducted these correlation analyses on only genes with >10% methylation (n=2447) to investigate whether higher levels of methylation may be associated with gene expression. The results show no correlation between gene methylation levels and expression in male, female and combined datasets (Figure S3).

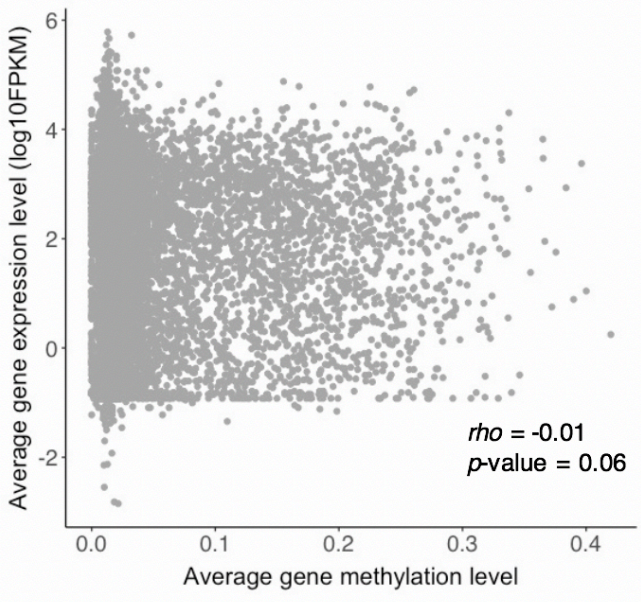


Figure 12a: Relationship between average methylation level across a gene (%) and its average expression level (log10FPKM). Methylation and expression levels were averaged across all samples. There is no significant association between gene methylation and expression levels (*rho* = -0.01, *p*-value = 0.06).

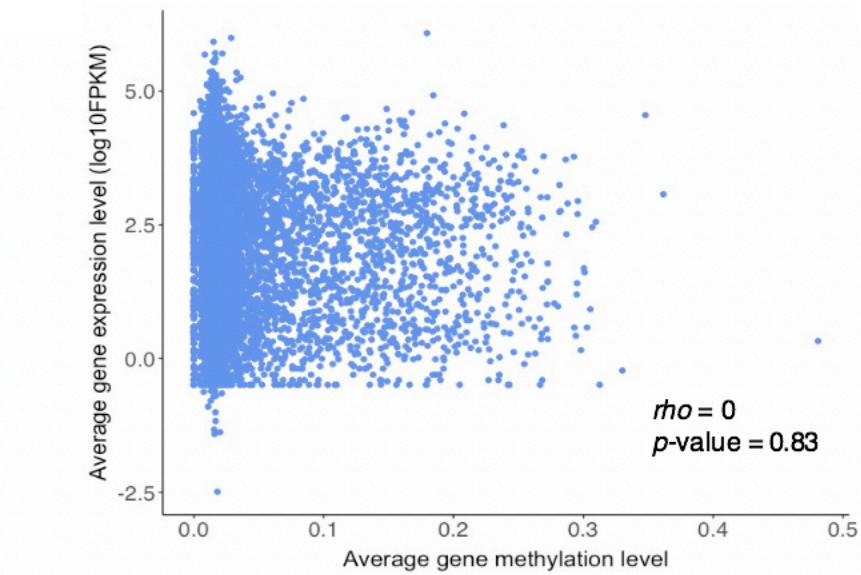


Figure 12b: Relationship between average methylation level across a gene in males (%) and its average expression level in males (log10FPKM). Methylation and expression levels were averaged across all male samples. There is no significant association between gene methylation and expression levels (*rho* = 0, *p*-value = 0.83).

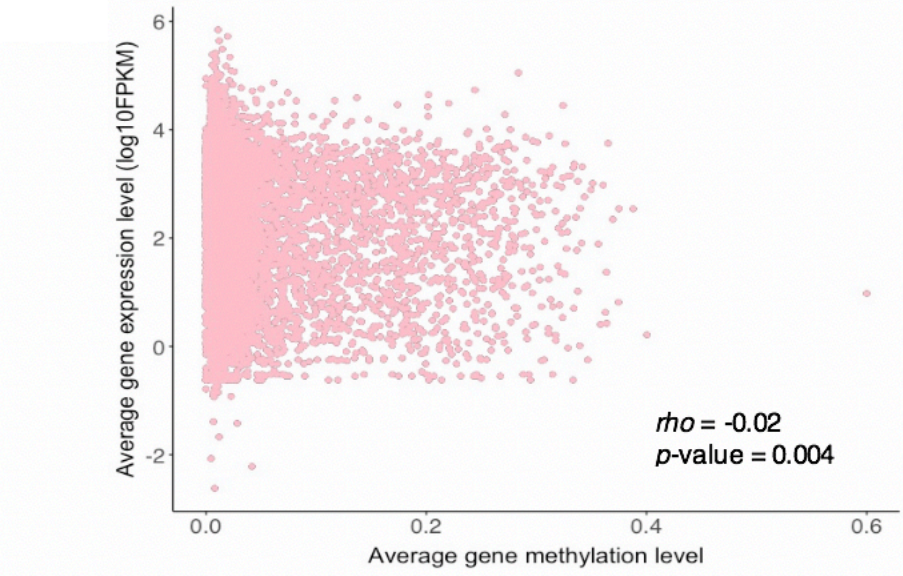


Figure 12c: Relationship between average methylation level across a gene in females (%) and its average expression level in females (log10FPKM). Methylation and expression levels were averaged across all female samples. There is a weak negative but significant association between gene methylation and expression levels (*rho* = -0.02, *p*-value = 0.004).

In many insect species, high levels of gene body methylation are associated with stable gene expression (Glastad, Hunt and Goodisman, 2012; Wang *et al.*, 2013; Mathers *et al.*, 2018). To investigate this hypothesised relationship in *P. citri,* an analysis of the relationship between the coefficient of variation of gene expression and average gene methylation level was conducted. The results suggest that level of gene methylation positively correlates with the stability of gene expression (Figure 13, *rho* = 0.02, *p*-value = 0.01). This correlation is, however, weak.

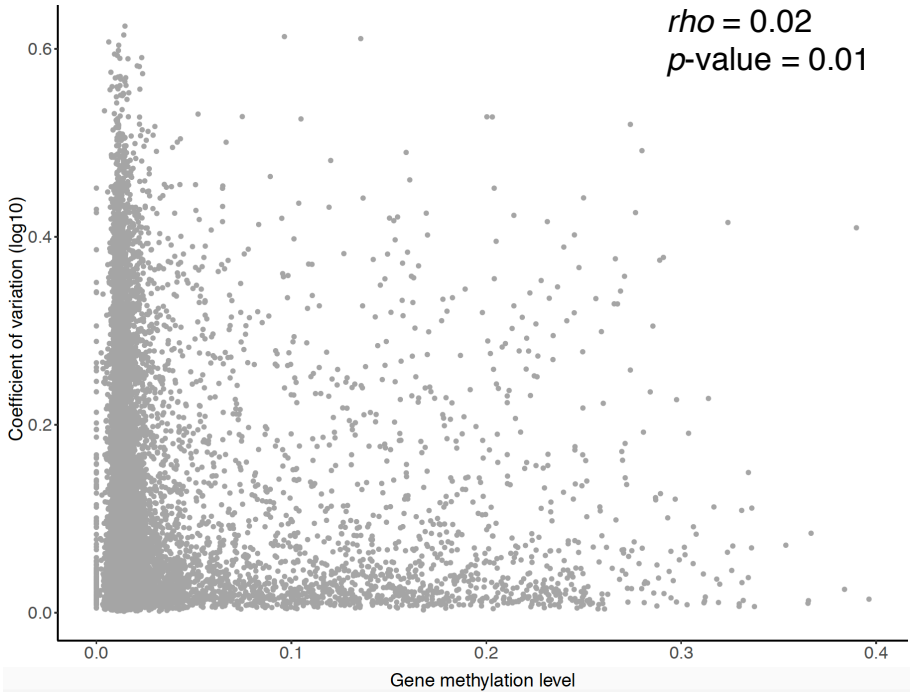


Figure 13: Relationship between average gene methylation level and stability of gene expression**.** This graph shows the (log10) coefficient of variation of gene expression between RNA-seq replicates across all genes at different methylation levels. There is a weak but significant positive association between gene methylation and co-efficient of variation levels (*rho* = 0.02, *p*-value = 0.01).

2.4.11 Changes in sex-specific methylation and sex-specific gene expression levels

As described previously, I find clear evidence for both sex-specific gene methylation and expression in this species. However, there is no significant correlation between changes in gene expression and methylation between males and females (Spearman's correlation, *rho* = 0.01, *p*-value = 0.06) (Figure 14). Out of 7953 DM genes identified in the *P. citri* genome, a total of 2031 genes (just under 25%) show sex-specific expression bias (1055 female-biased expression and 976 male-biased expression). The proportions of methylated and unmethylated genes in male-biased, female-biased and unbiased gene categories is similar and there is no significant association between expression bias of a gene and its methylation status (*p*-value = 0.3; Chi-squared test) (Figure 15). This suggests that DNA methylation may not be the sole mediator of the sex-specific gene regulation involved in sexual dimorphism.

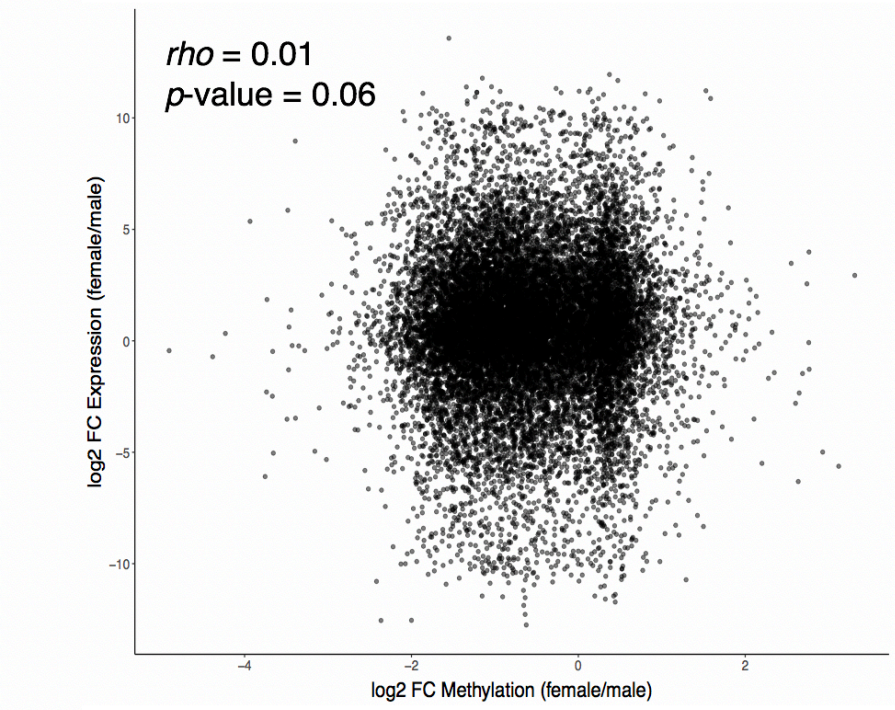
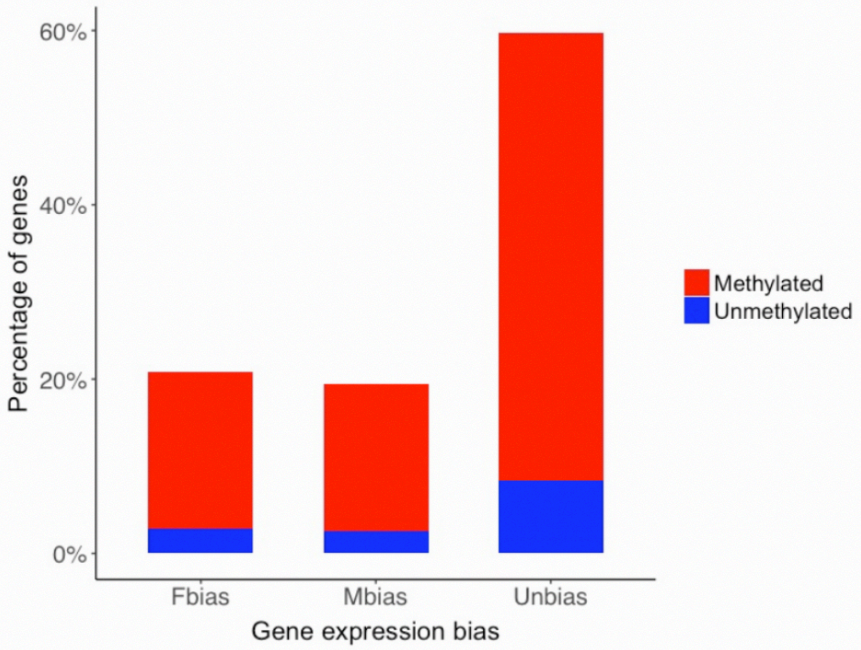


Figure 14: Relationship between fold change in gene expression and methylation between females and males for genes expressed (>1 FPKM) and methylated (>1%) in at least one of the sexes (n = 23,099). Methylation levels were averaged across replicates. Positive values indicate increased expression or methylation in females, relative to males; negative values indicate increased expression or methylation in males, relative to females. Spearman's ρ was used to test the strength and significance of the relationship between changes in gene expression and methylation. There is no significant correlation between gene expression changes and methylation changes between males and females (*rho* = 0.01, *p*-value = 0.06).



	Female-biased expression	Male-biased expression	Unbiased expression
Methylated	4151 (86%)	3898 (80%)	11865 (86%)
Unmethylated	660 (14%)	590 (20%)	1935 (14%)

Figure 15: Sex-biased expression and methylation. Bar plot shows the percentage of methylated genes in female-biased (Fbias), male-biased (Mbias) and unbiased (Unbias) genes. Genes with < 1% methylation are considered unmethylated (blue), genes with > 1% methylation are considered methylated (red). Only genes with an average FPKM > 1 are included in this analysis. Accompanying table provides a breakdown of the number of genes in each of the categories.

2.5 Discussion

Despite being genetically identical, *Planococcus citri* males and females exhibit striking levels of sexual dimorphism. These sex specific differences are therefore likely driven by differential gene expression between the sexes.

DNA methylation is reported in many species to have a key role in the regulation of gene expression. Here, I describe the DNA methylation machinery and patterns of DNA methylation across the genome of *P. citri* males and females and evaluate the relationship between sex-specific patterns of DNA methylation and sex-biased gene expression.

2.5.1 DNA methylation machinery

The *P. citri* genome encodes a single copy of the maintenance DNA methyltransferase, *DNMT1*, and two copies of a CpG methyl-binding domain (*MBD*). A putative *DNMT2* ortholog is also present in the *P. citri* genome; however, this is associated with the methylation of tRNA not DNA (Goll *et al.*, 2006). Interestingly, *P. citri* lacks the *de novo* DNA methyltransferase, *DNMT3*, which is also absent from its sister species *Planococcus ficus*. A recent comparative study of insect DNA methylation shows that *DNMT3* is the least conserved of the DNA methyltransferases, and was only identified in species belonging to Blattodea, Coleoptera, Hymenoptera *and* Hemiptera (Bewick *et al.*, 2016). Its presence in other Hemipteran species suggests that *DNMT3* has been lost in *P. citri* and *P. ficus*. This study clearly shows that *P. citri* has a functional DNA methylation system, thus suggesting that *DNMT3* may be dispensable for DNA methylation or, alternatively, that *DNMT1* or another protein can compensate for the *de novo* functions of *DNMT3*. Indeed, the functional roles of DNA methyltransferases in arthropods are inferred from mammalian studies (Wang *et al.*, 2006) and the discrete *de novo* and maintenance functions of arthropod *DNMT*s remain unclear. The lack of sex-specific *DNMT1* expression in *P. citri* is not unexpected as although global CpG methylation levels differ between males and females, this difference is rather small (~1.9%) and both sexes still have relatively high levels of DNA methylation compared to other insect species (Bewick *et al.*, 2016). These findings support the hypothesis that insects violate the convention that requires at least one copy of both maintenance and *de novo* DNA methyltransferase genes for a functional DNA methylation system.

However, it may be the case that *DNMT3* is just not sequenced, assembled or annotated in both *Planococcus* genomes.

2.5.2 Global CpG methylation in P. citri

Analysis of genome-wide methylation at base-pair resolution reveals that levels of CpG methylation in *P. citri* are higher than those reported in other Hemiptera and in most other insect taxa (Bewick *et al.*, 2016). I find that average global CpG methylation levels are significantly higher in males relative to females, although the difference is only ~1.9%. The slightly higher methylation levels in males may be related to the silencing of the paternal genome in their somatic tissues, as males with PGE have approximately half of their genome in a heterochromatic state. In mammals and plants, DNA methylation is associated with the formation of heterochromatin – albeit not the only mediator of its formation (Suzuki and Bird, 2008). DNA methylation in insects is generally associated with elevated, stable gene expression (reviewed by Glastad, Hunt and Goodisman, 2014); however, the function of invertebrate DNA methylation is far from clear and studies have shown a role in transcriptional silencing through the formation of heterochromatin (Brown and Nelson-Rees, 1961). There are 12,962 differentially methylated (DM) sites and 38,848 DM regions between males and females. The majority of DM sites (~55%) are located in gene bodies and there is a striking difference in methylation patterns of introns and exon between male- and female-hyper DM sites. The majority of DM sites with female-biased methylation are located in exons whereas the majority of those with male-biased methylation are located in introns. Global CpG methylation in *P. citri* and in most other insects is enriched in the exons of gene bodies (Lyko *et al.*, 2010; Hunt *et al.*, 2013; Wang *et al.*, 2013; Mathers *et al.*, 2018). Therefore, this may suggest a role specifically for intronic methylation in males. The majority of information about intronic methylation and its putative function(s) comes from studies in mammals. Intron methylation can influence gene expression and this is particularly of relevance in cancers and other diseases (Zhang *et al.*, 2010;

Yoshino *et al.*, 2017). Perhaps more relevant to *P. citri*, intron methylation is known to have a role in gene imprinting through regulation of anti-sense transcription in the mouse gene *Igf2r* (Wutz, Smrzka and Schweifer, 1997). In this example, intronic methylation of the maternal allele acts as the imprinting signal that maintains maternal expression of *Igf2r*. Further allele-specific methylation analysis in males is required to identify whether this intronic methylation is associated with the parental origin of alleles. However, the role of DNA methylation in *P. citri* – and indeed in the invertebrates – remains inconclusive so functional analyses are required to elucidate the relevance of these patterns of methylation.

2.5.3 Sex-biased gene expression

I also find 10,548 genes with sex-biased expression in the *P. citri* transcriptome, comprising 26,5% of the estimated total number of genes in this species (n = 39,801). *P. citri* has a rather small proportion of genes with sex-biased expression compared to *Nasonia vitripennis* and *Drosophila melanogaster* where 75% of genes show sex-biased expression (Assis, Zhou and Bachtrog, 2012; Wang, Werren and Clark, 2015). However, levels are similar to those in the peach aphid, *M. persicae*, in which 19% of genes show sex-biased expression (Mathers *et al.*, 2018). Males have an excess of extremely biased genes (fold change > 10) compared to extremely female-biased genes. Both female-biased and male-biased genes are enriched for core metabolic processes, similar to those enriched in methylation-biased genes. Interestingly, male-biased genes are enriched for myosin production and mitochondrial protein transport. Both of these processes are involved in insect flight (Sacktor, 1970; Bullard, Dabrowska and Winkelman, 1973) which could explain the high fold-expression differences in these genes between winged males and wingless females. However, it is essential to highlight that these observations are made in adult whole body samples and the extent of expression biases can vary greatly between tissues and developmental stages (Grath and Parsch, 2016). Thus, tissue-specific and stage-specific

analyses, particularly single cell RNA-seq, would allow the identification of the tissues contributing to sex-specific gene expression and developmental changes in expression patterns.

2.5.4 Relationship between gene methylation and expression

Paternal Genome Elimination is a form of pseudohaplodiploidy in which, although both sexes develop from fertilised eggs, males have haploid expression whilst females are diploid. The regulatory mechanisms underlying this difference in expression are poorly understood. Studies in true haplodiploid species, in which males develop from unfertilised eggs, have revealed striking differences in DNA methylation levels between diploid female and haploid male genomes that relate to gene expression differences (Glastad *et al.*, 2014). Elevated DNA methylation levels in haploid males are suggested to be indicative of regulatory pressures associated with the single-copy state of haploid loci (Glastad *et al.*, 2014). In *P. citri*, I find that levels of gene expression are not correlated with levels of gene methylation; however, the most highly methylated genes in *P. citri* show significantly lower expression than those with lower levels of methylation. Although, there are only seven genes methylated at this level (covered by both WGBS-seq and RNA-seq data) and so this makes results difficult to interpret. These are contrary to previous studies in insects that suggest high methylation levels are associated with elevated gene expression (Wang *et al.*, 2013; Glastad *et al.*, 2016). Furthermore, I find a weak but statistically significant positive association between stability of gene expression and level of gene methylation. This result is similar to findings in other insect species and suggests that increased levels of gene methylation may have a role in stabilising gene expression in *P. citri*.

However, the functional role(s) of DNA methylation in insects is far from conclusive and whilst many methylated genes are highly expressed in insects, many non-methylated genes have similar expression levels (Wang *et*

al., 2013). Therefore, it is unclear exactly how and if DNA methylation is involved in the regulation of gene expression. Indeed, an RNAi study in the Hemipteran, *Oncopeltus fasciatus*, revealed that depletion of DNA methylation did not result in changes in gene or transposable element expression but did lead to aberrant egg production and follicle development (Bewick *et al.*, 2018). Thus, suggesting a functional role for DNA methylation that is independent to gene expression. Development of an RNAi protocol in *P. citri* is currently underway that will allow a similar analysis to be carried out in this species.

Although many studies of the relationships between DNA methylation and gene expression have been conducted in insects, particularly the Hymenoptera, it is worth noting that the findings may not be reproducible (Libbrecht *et al.*, 2016). Thus, the patterns and relationships suggested may not be representative of one species, let alone an entire genus, order or class. Regarding sex-specific differences, I identify a number of sex-specific differentially methylated and expressed genes in *P. citri*. However, there is no significant correlation between sex-specific changes in gene expression and methylation. This suggests that DNA methylation is not solely responsible for regulation of gene expression and that it may not mediate the sex-specific gene expression that leads to the extreme sexual dimorphism found in this species. However, although a clear correlation between methylation and expression on a gene-by-gene basis is not identified, the evident sex-specific methylation patterns indicate that this epigenetic modification could still regulate gene expression differences through *trans* rather than *cis* effects. Intraspecific hybrid crosses would provide an excellent system in which to study *cis* versus *trans* effects as the parental origin of alleles can be recognised (Wang, Werren and Clark, 2016). If DNA methylation differences are due only to changes in *cis*-regulatory sequences, then allele-specific methylation in offspring will resemble parental methylation status. If changes are exclusively a result of *trans* factors (e.g. methylation status is remodelled

in every generation) then offspring allele-specific methylation will be ~50% on both parental alleles with no interspecific differences.

2.5.5 Conclusions

This study provides a preliminary analysis of DNA methylation and gene expression in an insect with PGE. There is much scope for future analyses, particularly as sequencing methods develop. These analyses were carried out on pooled whole adult samples and thus cannot detect variation between individuals, tissues or developmental stages. Therefore, a relationship between DNA methylation and expression cannot be ruled out in specific tissues or at specific periods in development. Of particular interest would be the germline, where sexual conflict between parental alleles occurs and crucially where elimination of the paternal genome occurs in PGE species. Additionally, analysis of early stages of development where key sex-specific gene expression occurs could provide information about the role of DNA methylation in sex determination.

2.6 SUPPLEMENTARY TABLES AND FIGURES

Table S1: qPCR primers and amplicon sizes

Locus	Amplicon size (bp)	Primer-F	Primer-R
DNMT1	161	GCCTCGTTACGTGATCATGG	TGGGCAACTTCAGCACAATA
RP49	165	AAGAAGGTTCAAGGGCCAGT	TGGGCAACTTCAGCACAATA
RP17	231	CTGCGAACCCTACATCACCT	TTGAAAGGCCAGAAGAATCG

Table S2: Bisulfite sequencing sample coverage

Sample	Coverage at all sites
F1	17x
F2	17x
F3	20x
F4	19x
F5	19x
M1	15x
M3	20x
M4	15x
M5	16x

Sample	Coverage at CG sites
F1	8.7x
F2	8.7x
F3	10.2x
F4	9.6x
F5	9.5x
M1	7.4x
M3	10x
M4	8.7x
M5	10.2x

Sample	Coverage at all C sites
F1	15.44x
F2	14.89x
F3	17.89x
F4	16.92x
F5	17.38x
M1	12.94x
M3	19.22x
M4	13.46x
M5	13.5x

Table S3: FastP settings

Parameter	**Value**
cut_by_quality	3
window_size	4
cut_mean_quality	20
trim_poly_g	10

Figure S1: DNMT1 ortholog in *Planococcus citri* (version v0, mealybug.org) contains 4 conserved domains across 26 exons.

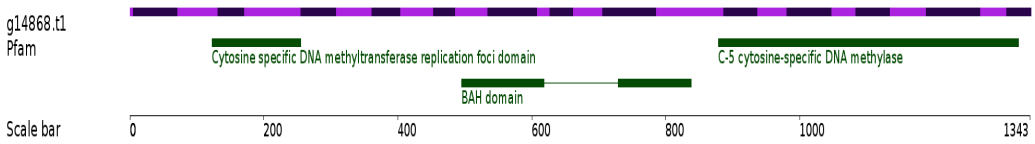


Table S4: Conserved domains identified after BLAST-searching putative *Planococcus citri* DNMT1 ortholog against NCBI database

Name	InterPro Number	Description	e-value
Dcm	IPR018117	Site-specific DNA-cytosine methylase [Replication, recombination and repair]	8.04e-14
Cyt_C5_DNA_methylase	IPR001525	Cytosine-C5 specific DNA methylase	2.68e-13
DNA_methylase	IPR018117	C-5 cytosine-specific DNA methylase	2.69e-12
dcm	IPR001525	DNA-methyltransferase (dcm)	5.20e-09
BAH_Dnmt1_II	IPR001025	BAH, or Bromo Adjacent Homology domain, second copy present in DNA (Cytosine-5)-methyltransferases	1.13e-38
DNMT1-RFD	IPR022702	Cytosine specific DNA methyltransferase replication foci domain	1.64e-25
BAH_Dnmt1_I	IPR001025	BAH, or Bromo Adjacent Homology domain, second copy present in DNA (Cytosine-5)-methyltransferases	3.97e-19

Table S5: Molecular function GO terms for unmethylated genes

GO number	Molecular function	Corrected *p*-value
GO:0003824	catalytic activity	0.015381546
GO:0005216	ion channel activity	4.42996E-05
GO:0022857	transmembrane transporter activity	4.42996E-05
GO:0005248	voltage-gated sodium channel activity	0.007689534
GO:0005515	protein binding	8.81049E-28
GO:0008528	G-protein coupled peptide receptor activity	6.68036E-05
GO:0016624	oxidoreductase activity, acting on the aldehyde or oxo group of donors, disulfide as acceptor	0.007309708
GO:0042302	structural constituent of cuticle	0.037110773
GO:0004725	protein tyrosine phosphatase activity	0.007689534
GO:0004672	protein kinase activity	0.008943345
GO:0016874	ligase activity	0.039819885
GO:0005509	calcium ion binding	3.29003E-11
GO:0051539	4 iron, 4 sulfur cluster binding	0.003041585
GO:0046983	protein dimerization activity	0.006487838
GO:0020037	heme binding	0.005750426
GO:0005544	calcium-dependent phospholipid binding	0.037110773
GO:0008234	cysteine-type peptidase activity	0.007689534
GO:0008237	metallopeptidase activity	0.021286291
GO:0008146	sulfotransferase activity	0.042933866
GO:0004553	hydrolase activity, hydrolyzing O-glycosyl compounds	0.049579255
GO:0005524	ATP binding	6.89922E-09
GO:0000166	nucleotide binding	0.01264154
GO:0003677	DNA binding	0.040841342
GO:0003676	nucleic acid binding	3.67029E-05
GO:0016758	transferase activity, transferring hexosyl groups	0.027739588
GO:0016705	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen	0.033036954
GO:0008270	zinc ion binding	3.30979E-05
GO:0005506	iron ion binding	0.040841342
GO:0046872	metal ion binding	6.68036E-05
GO:0005184	neuropeptide hormone activity	0.021522861
GO:0004198	calcium-dependent cysteine-type endopeptidase activity	0.019892987

Table S6: Molecular function GO terms for methylated genes

GO number	Molecular function	Corrected *p*-value
GO:0001104	RNA polymerase II transcription cofactor activity	1.69981E-06
GO:0003713	transcription coactivator activity	0.012353784
GO:0003712	transcription cofactor activity	0.000275043
GO:0003700	transcription factor activity, sequence-specific DNA binding	3.93007E-25
GO:0003735	structural constituent of ribosome	3.92013E-21
GO:0042302	structural constituent of cuticle	4.93969E-10
GO:0003824	catalytic activity	5.36043E-39
GO:0004672	protein kinase activity	2.73023E-81
GO:0003887	DNA-directed DNA polymerase activity	3.41979E-06
GO:0016301	kinase activity	0.004653717
GO:0016773	phosphotransferase activity, alcohol group as acceptor	0.015324983
GO:0004674	protein serine/threonine kinase activity	2.78997E-05
GO:0003899	DNA-directed 5'-3' RNA polymerase activity	0.000393459
GO:0004871	signal transducer activity	0.024757113
GO:0005089	Rho guanyl-nucleotide exchange factor activity	7.50931E-07
GO:0005085	guanyl-nucleotide exchange factor activity	0.001965622
GO:0005086	ARF guanyl-nucleotide exchange factor activity	0.001965622
GO:0005096	GTPase activator activity	0.003696579
GO:0005198	structural molecule activity	8.16959E-05
GO:0005215	transporter activity	0.000273401
GO:0005328	neurotransmitter:sodium symporter activity	0.000124997
GO:0015078	hydrogen ion transmembrane transporter activity	0.023067472
GO:0005351	sugar:proton symporter activity	0.012229255
GO:0005272	sodium channel activity	0.008020473
GO:0005515	protein binding	1E-300
GO:0008528	G-protein coupled peptide receptor activity	1.46994E-34
GO:0005234	extracellular-glutamate-gated ion channel activity	0.00310599

GO:0005230	extracellular ligand-gated ion channel activity	5.23962E-12
GO:0005216	ion channel activity	1.29003E-08
GO:0004970	ionotropic glutamate receptor activity	4.01976E-05
GO:0004930	G-protein coupled receptor activity	1.50003E-27
GO:0004984	olfactory receptor activity	1.63005E-19
GO:0009055	electron carrier activity	2.52988E-05
GO:0016849	phosphorus-oxygen lyase activity	9.09913E-05
GO:0004252	serine-type endopeptidase activity	1.61994E-32
GO:0004185	serine-type carboxypeptidase activity	0.003696579
GO:0004181	metallocarboxypeptidase activity	0.016687835
GO:0008236	serine-type peptidase activity	4.12003E-09
GO:0008234	cysteine-type peptidase activity	4.00037E-08
GO:0008237	metallopeptidase activity	0.000160103
GO:0036459	thiol-dependent ubiquitinyl hydrolase activity	2.25996E-07
GO:0070008	serine-type exopeptidase activity	1.80011E-05
GO:0004222	metalloendopeptidase activity	1.88018E-09
GO:0004843	thiol-dependent ubiquitin-specific protease activity	0.001965622
GO:0004812	aminoacyl-tRNA ligase activity	6.72977E-06
GO:0016705	oxidoreductase activity	2.13993E-33
GO:0005549	odorant binding	3.8699E-25
GO:0003682	chromatin binding	0.001965622
GO:0016787	hydrolase activity	1.3499E-27
GO:0051537	2 iron, 2 sulfur cluster binding	0.012229255
GO:0051536	iron-sulfur cluster binding	0.022532013
GO:0030246	carbohydrate binding	0.046494331
GO:0046983	protein dimerization activity	3.48017E-13
GO:0008017	microtubule binding	8.22053E-08
GO:0003779	actin binding	7.41993E-09
GO:0051082	unfolded protein binding	2.52988E-05
GO:0005102	receptor binding	0.021897776
GO:0019904	protein domain specific binding	0.028523315
GO:0008083	growth factor activity	0.00129509
GO:0043015	gamma-tubulin binding	0.001965622
GO:0020037	heme binding	7.12033E-35
GO:0016491	oxidoreductase activity	8.80035E-22
GO:0008270	zinc ion binding	3.2802E-77

GO:0030145	manganese ion binding	0.028523315
GO:0005506	iron ion binding	1.93999E-36
GO:0005507	copper ion binding	0.012353784
GO:0046872	metal ion binding	6.5298E-56
GO:0043565	sequence-specific DNA binding	1.9602E-17
GO:0001733	galactosylceramide sulfotransferase activity	6.02005E-06
GO:0008146	sulfotransferase activity	0.005588561
GO:0003676	nucleic acid binding	1.29E-120
GO:0004523	RNA-DNA hybrid ribonuclease activity	2.90001E-29
GO:0008408	3'-5' exonuclease activity	0.000216172
GO:0004519	endonuclease activity	0.012229255
GO:0004518	nuclease activity	0.000832339
GO:0005525	GTP binding	9.46019E-19
GO:0000166	nucleotide binding	8.05008E-10
GO:0005524	ATP binding	3.29E-121
GO:0035091	phosphatidylinositol binding	0.001965622
GO:0003743	translation initiation factor activity	0.006935855
GO:0000049	tRNA binding	0.021897776
GO:0008061	chitin binding	6.74062E-05
GO:0004842	ubiquitin-protein transferase activity	6.72977E-11
GO:0016887	ATPase activity	3.05985E-23
GO:0003774	motor activity	1.27997E-08
GO:0004386	helicase activity	0.000293089
GO:0003777	microtubule motor activity	2.52E-05
GO:0003755	peptidyl-prolyl cis-trans isomerase activity	0.039930063
GO:0003924	GTPase activity	1.16011E-05
GO:0003918	DNA topoisomerase type II (ATP-hydrolyzing) activity	0.021897776
GO:0003678	DNA helicase activity	1.00995E-07
GO:0046933	proton-transporting ATP synthase activity, rotational mechanism	0.028523315
GO:0042626	ATPase activity, coupled to transmembrane movement of substances	8.1508E-07
GO:0009982	pseudouridine synthase activity	0.028523315
GO:0008483	transaminase activity	0.008491805
GO:0016758	transferase activity, transferring hexosyl groups	2.51015E-29
GO:0003950	NAD+ ADP-ribosyltransferase	0.010120454

	activity	
GO:0008378	galactosyltransferase activity	0.039930063
GO:0050660	flavin adenine dinucleotide binding	5.5195E-16
GO:0050661	NADP binding	0.008491805
GO:0030170	pyridoxal phosphate binding	1.82012E-09
GO:0051287	NAD binding	4.25011E-08
GO:0070403	NAD+ binding	0.013626992
GO:0004553	hydrolase activity, hydrolyzing O-glycosyl compounds	2.49E-08
GO:0003796	lysozyme activity	0.046494331
GO:0004555	alpha,alpha-trehalase activity	0.001965622
GO:0004571	mannosyl-oligosaccharide 1,2-alpha-mannosidase activity	0.000832339
GO:0016810	hydrolase activity, acting on carbon-nitrogen (but not peptide) bonds	0.001965622
GO:0016702	oxidoreductase activity	0.021897776
GO:0016757	transferase activity, transferring glycosyl groups	0.004150496
GO:0016747	transferase activity, transferring acyl groups other than amino-acyl groups	0.000144744
GO:0004402	histone acetyltransferase activity	9.09913E-05
GO:0008080	N-acetyltransferase activity	0.012229255
GO:0003995	acyl-CoA dehydrogenase activity	0.000275043
GO:0008137	NADH dehydrogenase (ubiquinone) activity	0.021897776
GO:0008168	methyltransferase activity	1.74985E-09
GO:0016746	transferase activity, transferring acyl groups	8.41977E-06
GO:0016788	hydrolase activity, acting on ester bonds	0.000164021
GO:0016627	oxidoreductase activity, acting on the CH-CH group of donors	4.08037E-05
GO:0005337	nucleoside transmembrane transporter activity	0.028523315
GO:0005509	calcium ion binding	3.76964E-15
GO:0016817	hydrolase activity, acting on acid anhydrides	0.001965622
GO:0003677	DNA binding	2.4598E-80
GO:0003993	acid phosphatase activity	3.2802E-07
GO:0004114	3',5'-cyclic-nucleotide phosphodiesterase activity	0.006640488

GO:0004435	phosphatidylinositol phospholipase C activity	0.046494331
GO:0008138	protein tyrosine/serine/threonine phosphatase activity	0.004150496
GO:0016614	oxidoreductase activity, acting on CH-OH group of donors	1.77992E-12
GO:0004383	guanylate cyclase activity	0.004415704
GO:0016616	oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor	4.56983E-07
GO:0008484	sulfuric ester hydrolase activity	5.41003E-05
GO:0000287	magnesium ion binding	0.001965622
GO:0017048	Rho GTPase binding	0.028523315
GO:0019901	protein kinase binding	0.046494331
GO:0008440	inositol-1,4,5-trisphosphate 3-kinase activity	0.028523315
GO:0003723	RNA binding	1.13006E-23
GO:0015171	amino acid transmembrane transporter activity	0.006640488
GO:0008509	anion transmembrane transporter activity	0.012229255
GO:0022857	transmembrane transporter activity	2.08978E-28
GO:0008565	protein transporter activity	0.046494331

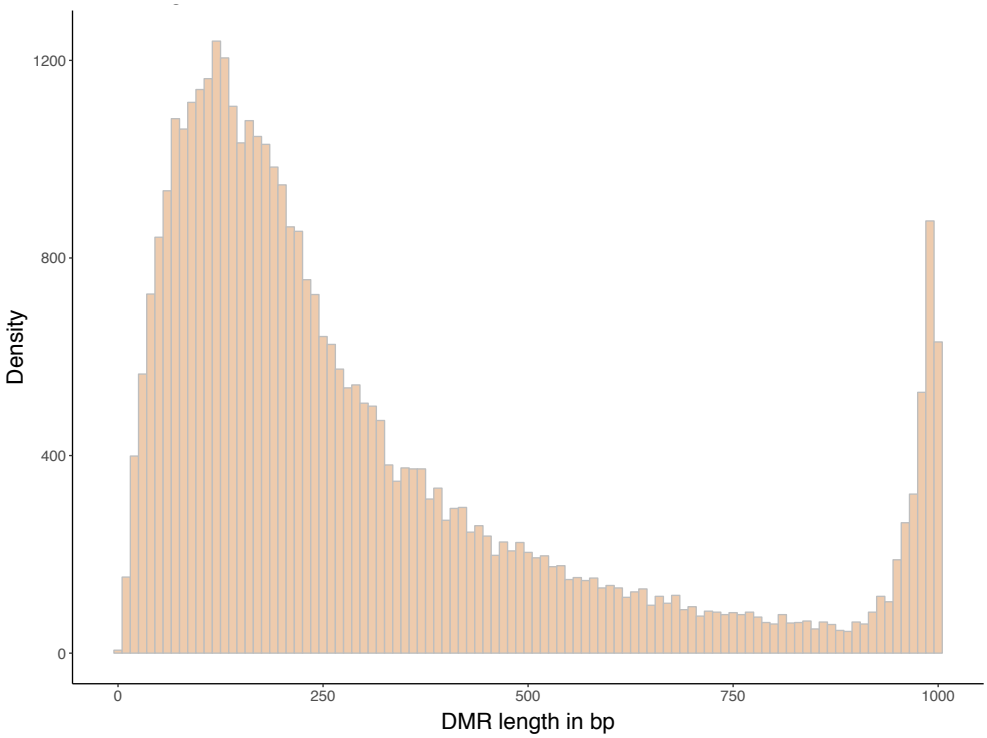


Figure S2: Histogram of variation in length of differentially methylated regions (DMRs) generated in CGmapTools DMR analysis. The shortest is 5bp; the longest is 1000bp.

Table S7: Molecular function GO terms for male hypermethylated genes

GO number	Molecular function	Corrected *p*-value
GO:0005515	protein binding	2.40E-16
GO:0005524	ATP binding	1.59E-13
GO:0003774	motor activity	1.59E-13
GO:0003676	nucleic acid binding	2.09E-05
GO:0004672	protein kinase activity	0.00053183
GO:0005198	structural molecule activity	0.000639335
GO:0008270	zinc ion binding	0.000834874
GO:0020037	heme binding	0.003111634
GO:0003677	DNA binding	0.003111634
GO:0046872	metal ion binding	0.003540024
GO:0003755	peptidyl-prolyl cis-trans isomerase activity	0.00958619
GO:0022857	transmembrane transporter activity	0.014484348
GO:0005215	transporter activity	0.019866993
GO:0050660	flavin adenine dinucleotide binding	0.022900958
GO:0008528	G-protein coupled peptide receptor activity	0.030073158
GO:0005506	iron ion binding	0.039998372
GO:0003735	structural constituent of ribosome	0.04302498
GO:0004298	threonine-type endopeptidase activity	0.045276175
GO:0016705	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen	0.047323546

Table S8: Molecular function GO terms for female hypermethylated genes

GO number	Molecular function	Corrected *p*-value
GO:0005506	iron ion binding	0.006257098
GO:0004984	olfactory receptor activity	0.006409157
GO:0008124	4-alpha-hydroxytetrahydrobiopterin dehydratase activity	0.006409157
GO:0005388	calcium-transporting ATPase activity	0.006409157
GO:0016705	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen	0.006554497
GO:0005509	calcium ion binding	0.007556249
GO:0008270	zinc ion binding	0.007556249
GO:0016758	transferase activity, transferring hexosyl groups	0.007799698
GO:0005525	GTP binding	0.008907392
GO:0003899	DNA-directed 5'-3' RNA polymerase activity	0.014462426
GO:0004523	RNA-DNA hybrid ribonuclease activity	0.016046054
GO:0008092	cytoskeletal protein binding	0.016046054
GO:0008527	taste receptor activity	0.016046054
GO:0051287	NAD binding	0.02176634
GO:0016616	oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor	0.02176634
GO:0008528	G-protein coupled peptide receptor activity	0.026215676
GO:0004812	aminoacyl-tRNA ligase activity	0.033490915
GO:0016787	hydrolase activity	0.036373352
GO:0005515	protein binding	1.12E-09
GO:0005524	ATP binding	6.67E-05
GO:0005085	guanyl-nucleotide exchange factor activity	0.002637782
GO:0003917	DNA topoisomerase type I activity	0.003930258
GO:0003951	NAD+ kinase activity	0.003930258
GO:0004252	serine-type endopeptidase activity	0.003930258
GO:0005089	Rho guanyl-nucleotide exchange factor activity	0.004966
GO:0003676	nucleic acid binding	0.005123637
GO:0008061	chitin binding	0.007206694
GO:0003677	DNA binding	0.014488702
GO:0046923	ER retention sequence binding	0.018454942

GO:0003824	catalytic activity	0.025836826
GO:0008324	cation transmembrane transporter activity	0.029937013
GO:0046872	metal ion binding	0.043905213
GO:0016491	oxidoreductase activity	0.043905213
GO:0003873	6-phosphofructo-2-kinase activity	0.043905213

Table S9: Molecular function GO terms for male biased genes

GO number	Molecular function	Corrected *p*-value
GO:0005515	protein binding	1.22E-47
GO:0004930	G-protein coupled receptor activity	1.11E-46
GO:0003700	DNA binding transcription factor activity	5.73E-26
GO:0008061	chitin binding	6.28E-26
GO:0042302	structural constituent of cuticle	1.21E-25
GO:0005216	ion channel activity	9.54E-22
GO:0004252	serine-type endopeptidase activity	3.20E-21
GO:0003774	motor activity	1.90E-19
GO:0020037	heme binding	4.89E-19
GO:0005509	calcium ion binding	2.14E-18
GO:0005524	ATP binding	3.56E-17
GO:0016705	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen	6.40E-17
GO:0005506	iron ion binding	1.69E-16
GO:0005328	neurotransmitter:sodium symporter activity	3.28E-14
GO:0043565	sequence-specific DNA binding	9.20E-14
GO:0005549	odorant binding	1.35E-13
GO:0003677	DNA binding	2.85E-13
GO:0016887	ATPase activity	4.52E-13
GO:0022857	transmembrane transporter activity	7.72E-12
GO:0016849	phosphorus-oxygen lyase activity	1.68E-10
GO:0004970	ionotropic glutamate receptor activity	3.10E-09
GO:0005230	extracellular ligand-gated ion channel activity	3.80E-09
GO:0004672	protein kinase activity	9.55E-09
GO:0005215	transporter activity	1.05E-08
GO:0005234	extracellularly glutamate-gated ion channel activity	1.09E-08
GO:0016831	carboxy-lyase activity	1.68E-08
GO:0004725	protein tyrosine phosphatase activity	2.42E-08
GO:0003924	GTPase activity	3.10E-08
GO:0016788	hydrolase activity, acting on ester bonds	4.60E-08
GO:0005089	Rho guanyl-nucleotide exchange factor activity	7.22E-07

GO:0050660	flavin adenine dinucleotide binding	7.34E-07
GO:0004222	metalloendopeptidase activity	1.66E-06
GO:0004114	3',5'-cyclic-nucleotide phosphodiesterase activity	2.34E-06
GO:0008137	NADH dehydrogenase (ubiquinone) activity	1.07E-05
GO:0008270	zinc ion binding	1.10E-05
GO:0016614	oxidoreductase activity, acting on CH-OH group of donors	1.14E-05
GO:0003777	microtubule motor activity	1.67E-05
GO:0004553	hydrolase activity, hydrolyzing O-glycosyl compounds	2.60E-05
GO:0005085	guanyl-nucleotide exchange factor activity	7.44E-05
GO:0016746	transferase activity, transferring acyl groups	7.44E-05
GO:0005523	tropomyosin binding	0.000134206
GO:0016758	transferase activity, transferring hexosyl groups	0.000146712
GO:0005525	GTP binding	0.000207191
GO:0004984	olfactory receptor activity	0.000216286
GO:0008017	microtubule binding	0.00023133
GO:0030170	pyridoxal phosphate binding	0.000231646
GO:0003824	catalytic activity	0.000306067
GO:0016747	transferase activity, transferring acyl groups other than amino-acyl groups	0.000321953
GO:0004871	signal transducer activity	0.000360936
GO:0000166	nucleotide binding	0.000369445
GO:0003779	actin binding	0.000408954
GO:0008146	sulfotransferase activity	0.000505429
GO:0016627	oxidoreductase activity, acting on the CH-CH group of donors	0.000934987
GO:0004181	metallocarboxypeptidase activity	0.001017831
GO:0003872	6-phosphofructokinase activity	0.001058192
GO:0005245	voltage-gated calcium channel activity	0.001058192
GO:0005184	neuropeptide hormone activity	0.001058192
GO:0015293	symporter activity	0.001058192
GO:0004383	guanylate cyclase activity	0.001412391
GO:0005507	copper ion binding	0.001559696
GO:0016616	oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor	0.002719131

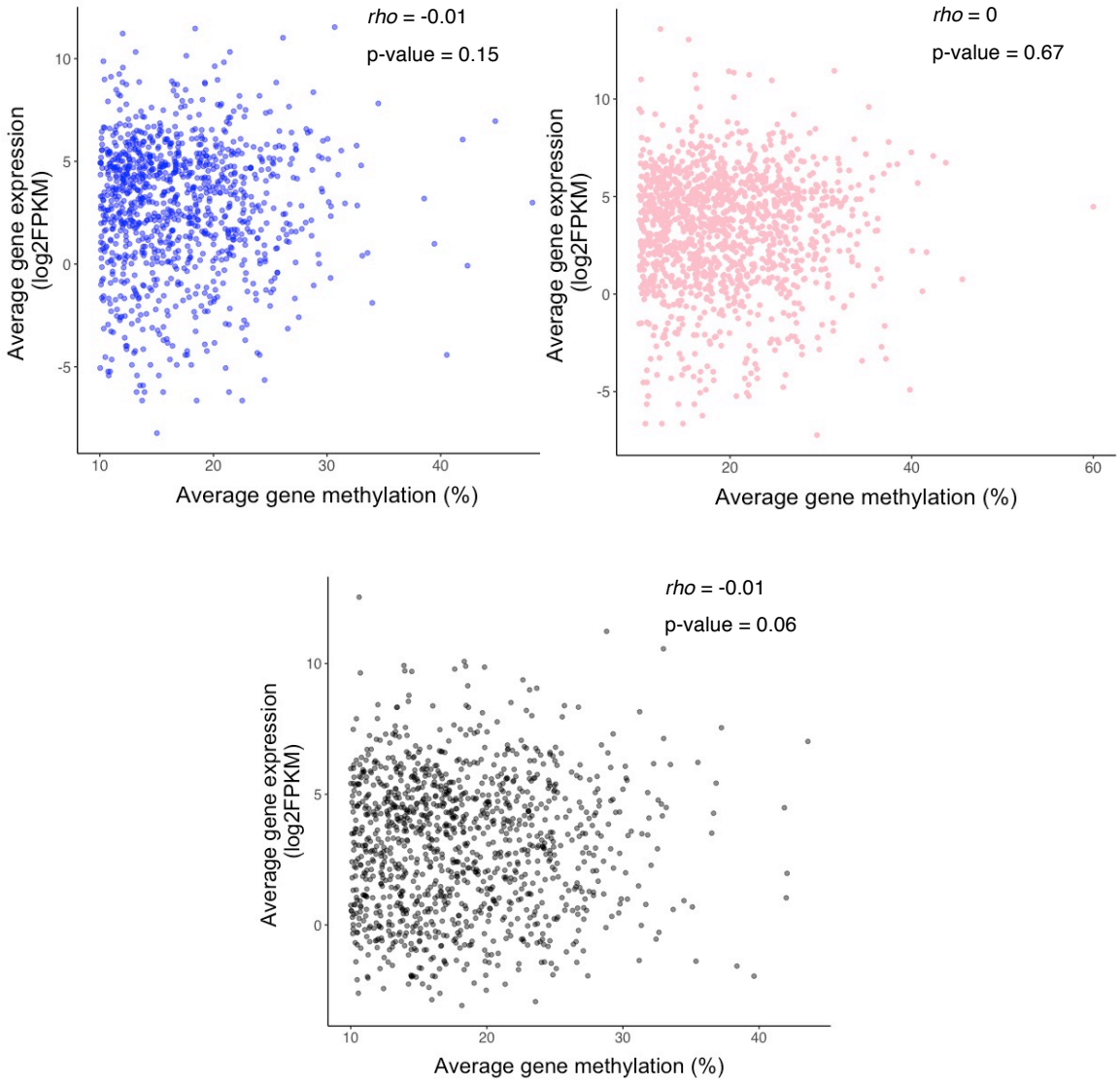
GO:0008528	G-protein coupled peptide receptor activity	0.003886169
GO:0004198	calcium-dependent cysteine-type endopeptidase activity	0.003939454
GO:0015171	amino acid transmembrane transporter activity	0.004712245
GO:0004869	cysteine-type endopeptidase inhibitor activity	0.007795034
GO:0016714	oxidoreductase activity	0.008570663
GO:0008121	ubiquinol-cytochrome-c reductase activity	0.008570663
GO:0005248	voltage-gated sodium channel activity	0.008570663
GO:0003735	structural constituent of ribosome	0.008896594
GO:0003887	DNA-directed DNA polymerase activity	0.009475679
GO:0004129	cytochrome-c oxidase activity	0.010120493
GO:0008241	peptidyl-dipeptidase activity	0.010120493
GO:0008408	3'-5' exonuclease activity	0.011757395
GO:0008138	protein tyrosine/serine/threonine phosphatase activity	0.016471582
GO:0033897	ribonuclease T2 activity	0.020072683
GO:0019901	protein kinase binding	0.020072683
GO:0004435	phosphatidylinositol phospholipase C activity	0.020072683
GO:0019001	guanyl nucleotide binding	0.020072683
GO:0031683	G-protein beta/gamma-subunit complex binding	0.020072683
GO:0005158	insulin receptor binding	0.020072683
GO:0003995	acyl-CoA dehydrogenase activity	0.022346383
GO:0009055	electron transfer activity	0.022346383
GO:0004649	poly(ADP-ribose) glycohydrolase activity	0.026581498
GO:0003951	NAD+ kinase activity	0.026581498
GO:0004185	serine-type carboxypeptidase activity	0.02815012
GO:0005102	signaling receptor binding	0.034584723
GO:0003676	nucleic acid binding	0.042398671

Table S10: Molecular function GO terms for female biased genes

GO number	Molecular function	Corrected *p*-value
GO:0005515	protein binding	3.02E-107
GO:0003735	structural constituent of ribosome	2.28E-40
GO:0005524	ATP binding	1.06E-27
GO:0003723	RNA binding	6.26E-22
GO:0003676	nucleic acid binding	4.64E-20
GO:0001733	galactosylceramide sulfotransferase activity	1.27E-17
GO:0008237	metallopeptidase activity	8.50E-15
GO:0008270	zinc ion binding	2.86E-14
GO:0003677	DNA binding	1.06E-11
GO:0004672	protein kinase activity	2.37E-09
GO:0008234	cysteine-type peptidase activity	1.87E-08
GO:0008168	methyltransferase activity	1.87E-08
GO:0016705	oxidoreductase activity	2.42E-08
GO:0004812	aminoacyl-tRNA ligase activity	3.62E-08
GO:0020037	heme binding	3.87E-08
GO:0022857	transmembrane transporter activity	4.92E-08
GO:0005506	iron ion binding	4.93E-08
GO:0016787	hydrolase activity	9.38E-07
GO:0030983	mismatched DNA binding	2.05E-06
GO:0046872	metal ion binding	2.49E-06
GO:0003824	catalytic activity	2.86E-06
GO:0016758	transferase activity, transferring hexosyl groups	3.46E-05
GO:0004518	nuclease activity	6.35E-05
GO:0004984	olfactory receptor activity	0.000204266
GO:0030170	pyridoxal phosphate binding	0.000385714
GO:0008236	serine-type peptidase activity	0.000396609
GO:0005201	extracellular matrix structural constituent	0.000427737
GO:0000049	tRNA binding	0.000434109
GO:0000166	nucleotide binding	0.000586956
GO:0003899	DNA-directed 5'-3' RNA polymerase activity	0.000956959
GO:0003755	peptidyl-prolyl cis-trans isomerase activity	0.000956959
GO:0003950	NAD+ ADP-ribosyltransferase activity	0.001030725
GO:0004386	helicase activity	0.001644868
GO:0004003	ATP-dependent DNA helicase activity	0.002706424

GO:0016810	hydrolase activity, acting on carbon-nitrogen (but not peptide) bonds	0.004265841
GO:0050660	flavin adenine dinucleotide binding	0.004745511
GO:0009982	pseudouridine synthase activity	0.006496391
GO:0008173	RNA methyltransferase activity	0.006496391
GO:0004540	ribonuclease activity	0.006496391
GO:0003746	translation elongation factor activity	0.006496391
GO:0046923	ER retention sequence binding	0.009715767
GO:0004553	hydrolase activity, hydrolyzing O-glycosyl compounds	0.012850341
GO:0030151	molybdenum ion binding	0.012850341
GO:0004560	alpha-L-fucosidase activity	0.012850341
GO:0016763	transferase activity, transferring pentosyl groups	0.012850341
GO:0008484	sulfuric ester hydrolase activity	0.013161943
GO:0016491	oxidoreductase activity	0.014080055
GO:0004222	metalloendopeptidase activity	0.014839734
GO:0019843	rRNA binding	0.015033586
GO:0003743	translation initiation factor activity	0.018231378
GO:0004298	threonine-type endopeptidase activity	0.02378681
GO:0008080	N-acetyltransferase activity	0.024324424
GO:0018024	histone-lysine N-methyltransferase activity	0.028395069
GO:0005525	GTP binding	0.028395069
GO:0004175	endopeptidase activity	0.028395069
GO:0008026	ATP-dependent helicase activity	0.038637487
GO:0004499	N,N-dimethylaniline monooxygenase activity	0.042156842
GO:0016614	oxidoreductase activity, acting on CH-OH group of donors	0.043852848
GO:0008528	G-protein coupled peptide receptor activity	0.046082757
GO:0016702	oxidoreductase activity	0.048020188
GO:0008641	ubiquitin-like modifier activating enzyme activity	0.048020188

Figure S3: Spearman's correlation analysis of gene methylation and expression in genes with > 10% methylation (blue = male, pink = female, black = average in both sexes).



Chapter 3: Role of parent-of-origin specific DNA methylation in Paternal Genome Elimination

3.1 Chapter Summary

In this chapter, I investigate the role of parent-of-origin specific DNA methylation in Paternal Genome Elimination, a genomic imprinting phenomenon found in several insect taxa including the scale insect genus, *Planococcus*. I perform whole genome bisulfite sequencing (WGBS) on the hybrid offspring of two closely related PGE species, *Planococcus citri* and its sister species, *Planococcus ficus*. I identify allele-specific DNA methylation levels and patterns that could possibly act as the molecular identifier of an allele's parental origin, allowing for the specific silencing and elimination of paternally inherited alleles in males.

3.2 Introduction

According to the principles of Mendelian inheritance, the two copies of a gene in a diploid organism – one maternally inherited and the other paternally inherited – are functionally equivalent in determining phenotype (Mendel, 1865). However, exceptions to this rule can occur when the genetic interests of the mother and father diverge (Haig, 2000; Normark, 2006). Genomic imprinting is an epigenetic process in which genes, chromosomes, or entire haploid complements exhibit different functional behaviour dependent upon their parental origin (Moore and Haig, 1991). This phenomenon is thought to have arisen 150 million years ago (Murphy and Jirtle, 2003) and was generally believed to be exclusive to mammals and flowering plants, where it has been extensively studied over the past decades. However, the term *'imprinting'* was first used to describe parent-of-origin specific chromosome behaviour discovered in the Dipteran insect family, *Sciaridae* (Crouse, 1960).

The imprinting behaviour described in *Sciara coprophila* is an example of one of the most striking cases of genomic imprinting, a phenomenon known as Paternal Genome Elimination (PGE). In PGE species, both sexes develop from fertilised eggs, however, males eliminate paternally inherited alleles from their germline (Metz, 1938; Brown and Nelson-Rees, 1961; Bongiorni *et al.*, 2004). As a result, similar to haplodiploid males, they only transmit maternally inherited alleles to offspring. Additionally, in some PGE taxa such as the mealybug genus *Planococcus* (Pseudococcidae, Hemiptera), the paternally inherited alleles in males are transcriptionally silenced during early embryogenesis (Bongiorni and Prantera, 2001). Thus, in these insects, male gene expression is haploid and maternal, whilst females are diploid. As the name suggests, it is exclusively paternally inherited chromosomes that are subject to the process of PGE (Brown and Nelson-Rees, 1961). However, it is unclear how these are recognised and specifically targeted for silencing and elimination.

One possible mechanism for parent-of-origin recognition of chromosomes in species with PGE is DNA methylation. DNA methylation is a vital epigenetic modification that is associated with a number of regulatory processes, including parent-of-origin specific imprinting (Li, Beard and Jaenisch, 1993). In mammals, differential methylation of imprinting control regions (ICRs) regulates allelic repression of imprinted genes (Kota and Feil, 2010). The establishment of these imprints occurs in the germline and in most cases, ICR methylation originates from the egg (Delaval and Feil, 2004). Whilst the majority of methylation marks are stripped from the genome of the embryo during the early stages of embryogenesis (Richards, 2006), ICRs retain parental imprints (Hajkova *et al.*, 2002), which mediate imprinted expression of genes (Delaval and Feil, 2004). In insects, the mechanisms and patterns of DNA methylation maintenance and reprogramming are not fully understood. Allele-specific methylation associated with allele-specific gene

expression is reported in the bumblebee *Bombus terrestris* (Lonsdale *et al.*, 2017) and the ants *Camponotus floridanus* and *Harpegnathos saltator,* (Bonasio *et al.*, 2010). This suggests the possibility of DNA methylation-mediated parent-specific expression in these insects. However, patterns of parent-of-origin specific DNA methylation may be confounded by *cis*-mediated allele-specific methylation that is associated with haplotype rather than parent-of-origin (Remnant *et al.*, 2016; Wang, Werren and Clark, 2016; Wedd, Kucharski and Maleszka, 2016). As such, in insects, the role of DNA methylation in imprinting remains unclear.

Parallels can, however, be drawn between specific silencing of the paternal chromosomes in PGE males and the imprinted inactivation of the paternal X chromosome that occurs in female marsupial mammals. Genetic sex determination in mammals generates females with two copies of an X chromosome and males with one copy. To correct the X-linked gene expression imbalance between males and females, females transcriptionally silence one of their X chromosomes (Lyon, 1961). The inactivated X chromosome condenses into a compact structure called a Barr body and is maintained in a silenced state (Boumil and Lee, 2001). This is similar to the process by which the paternally inherited alleles in mealybug males are transcriptionally silenced through heterochromatinization and form a heterochromatic body, which is visible in somatic tissue (Hughes-Schrader, 1948). In placental mammals, this X chromosome inactivation in the embryo is 'random' in the sense that either the maternal or paternal X chromosome can be silenced in any given cell. However, marsupial females specifically silence the paternally inherited X chromosome (Huynh and Lee, 2005). Interestingly, studies in marsupial and placental mammals have identified differences in levels and patterns of DNA methylation between the inactive X chromosomes and the active X chromosomes (Bernardino *et al.*, 2000; Hellman and Chess, 2007; Rens *et al.*, 2010). The recognition and silencing of paternally inherited alleles under PGE appears to be regulated by the

same epigenetic machinery, suggesting an evolutionarily conserved mechanism for the recognition of parental origin (Bongiorni, Cintio and Prantera, 1999). However, the molecular details are poorly understood and the extent to which DNA methylation plays a direct role in the recognition, transcriptional suppression and germline elimination of paternally inherited alleles is unknown.

Planococcus provides an excellent invertebrate model system in which to study DNA methylation and its role in genomic imprinting. These insects are genetically tractable with a short generation time, ease of laboratory rearing, availability of highly inbred lines, and closely related cross-fertile species. The parent-of-origin specific behaviours of their chromosomes provide a unique opportunity to investigate the molecular mechanisms of this form of imprinting and its consequences at a chromosomal level.

Planococcus citri shows patterns of sex-specific DNA methylation and levels of global DNA methylation that are high relative to other insect species (see Chapter 2). In-situ nick translation studies in *Planococcus* species also suggest that there may be differences in levels of DNA methylation between paternally and maternally inherited alleles. However, these studies produced conflicting results, with one suggesting that paternal alleles are hypomethylated relative to maternal alleles in both sexes (Bongiorni, Cintio and Prantera, 1999), a second suggesting no significant difference in DNA methylation between parental alleles in either sex (Buglia, Predazzi and Ferraro, 1999) and a third, using methylation-specific PCR, suggesting that methylation of paternal DNA is greater in males than in females (Mohan and Chandra, 2005). Therefore, the presence of allele-specific DNA methylation and its role(s) in genomic imprinting remains inconclusive.

Here, I used allele-specific whole genome bisulfite sequencing (WGBS) on the hybrid offspring of two closely related PGE species, *Planococcus citri* and

Planococcus ficus, to identify parent-of-origin specific DNA methylation patterns. There are 4.5 million species-specific SNPs between *P. citri* and *P. ficus* (de la Filia, Thesis 2018) allowing assignment of parental alleles in offspring. I analysed differences in levels and patterns of DNA methylation between maternally and paternally inherited alleles within and between sexes to determine whether DNA methylation levels and/or patterns could act as the molecular identifier of an allele's parental origin, allowing specific silencing and elimination of only those paternally inherited in males.

3.3 Methods

3.3.1 Insect husbandry and experimental populations

I used laboratory lines of the citrus mealybug *Planococcus citri* (Risso) and the closely related vine mealybug *Planococcus ficus* (Signoret). Both *Planococcus citri* and *Planococcus ficus* were cultured on sprouting potatoes in sealed plastic bottles. Insect cultures were kept at 25°C and ~70% relative humidity. Under these conditions, the species have a generation time (time from oviposition until sexual maturity) of approximately 30 days (*P. citri*) and 40 days (*P. ficus*).

All the experimental crosses in this study were conducted between females from *P. citri* (WYE3-2) and males from *P. ficus* (PF1-1). These isofemale lines were reared in the laboratory under a sib-mating regime. In each generation, one mated female is taken from culture and transferred to a new container to give rise to the next generation. The *P. citri* line had undergone 22 generations of sib-mating prior to this experiment. The *P. ficus* line is derived from an Israeli population and had undergone 8 generations of sib-mating.

3.3.2 Interspecific crosses (P. citri x P. ficus)

For the experimental crosses, *P. citri* virgin females were isolated after becoming sexually distinguishable from males (3^{rd}-4^{th} instar) and kept in separate containers until sexual maturity (35-days old). *P. ficus* males were isolated at pupal stage and kept in separate containers until sexually mature (~21 days). Hybrid matings took place in small glass Petri dishes with the aid of synthetic pheromones from *P. ficus*, the paternal species, and the occurrence of mating was visually monitored. The synthetic pheromone used in this experiment was the racemic ester (S)-lavandulyl senecioate (Hinkens, McElfresh and Millar, 2001) and was provided by Professor Jocelyn Miller (University of California, Riverside). The pheromone was diluted in 100% ethanol to a concentration of 10ng/µl. Reciprocal crosses could not be carried out in this experiment (i.e. *P. ficus* mother and *P. citri* father) as very few male offspring survive (Rotundo and Trembley, 1982). After mating, the mating pair was kept in a glass vial for 3-5 days until egg-laying was observed, after which males were removed and females were left to lay eggs for 7 days or until death. The eggs were then transferred to a sprouting potato in a new container. Three biological replicates (i.e. mating pairs) were set up. F1 virgin females and adult males were reared and collected under the same conditions as their parents. 35-day old virgin females and adult males were collected and stored at -80°C for DNA extraction.

3.3.3 DNA extraction

Genomic DNA was extracted from pools of ~60 whole adult males and 15 whole virgin adult females using DNeasy Blood and Tissue kit (Qiagen, CA) and Promega DNA Clean and Prep Kit (Promega) in a custom DNA extraction protocol. DNA samples were cleaned and concentrated using Zymo DNA Clean and Concentrator Kit (Zymo Research) according to manufacturer's instructions. DNA A260/A280 absorption ratios were measured with a NanoDrop ND-1000 Spectrophotometer (Thermo Scientific,

USA) and concentrations were measured with a Qubit Fluorometer (Life Technologies, CA).

3.3.4 Microsatellite analysis

A microsatellite genotyping approach was used to confirm that offspring from interspecific crosses were true hybrids: Total genomic DNA from F0 (parents) and F1(offspring) individuals from experimental crosses was extracted using prepGEM Insect Kit (ZyGEM) according to manufacturer's instructions. F0 parental lines were genotyped to determine the alleles present in their gene pools and a subset of F1 offspring were genotyped to confirm expected hybrid genotypes. Microsatellite primers for PCR amplification were obtained from Martins *et al.* (2014) (for full details, see Appendix A).

3.3.5 Library preparation and sequencing

The number of hybrid males produced from *P. citri* x *P. ficus* crosses is low (even 0 in some cases) and it is difficult to sample the adequate amount of DNA required for bisulfite sequencing. Due to the nature of this analysis, each sample could only contain full siblings, as half-siblings (males with different fathers or mothers) would affect SNP identification and allele-specific methylation analysis. Therefore, a low input bisulfite library preparation kit Pico Methyl-Seq Library Prep Kit (Zymo) was used. Bisulfite conversion and library preparation was carried out on male and female DNA samples (50ng input gDNA) using according to manufacturer's instructions. Non-methylated *Escherichia coli* DNA (Zymo D5016, 0.5ng/μL) was added to the hybrid DNA samples and simultaneously bisulfite-treated and sequenced to allow calculation of the bisulfite conversion efficiency. Libraries were sequenced on HiSeq 4000 by Edinburgh Genomics to generate 150b paired-end reads. Number of reads generated for each sample can be found in Table S1.

3.3.6 Analysis of DNA methylation from WGBS data

3.3.6.1 Trimming of reads and quality control

Initial quality control of Illumina reads was performed using FastQC v0.11.7 (Andrews, 2010). Reads were trimmed using Trimmomatic v0.33 (Bolger, Lohse and Usadel, 2014) with additional quality trimming of 10bp from the 3-prime and 5-prime ends of reads in accordance with library preparation analysis guidelines (Krueger and Andrews, 2011). Average length of reads after trimming was 112bp. After trimming, Illumina reads were first aligned to the converted non-methylated *Escherichia coli* K-12 strain reference genome (Anton *et al.*, 2015) using Bismark v0.19.0 (Krueger and Andrews, 2011) to estimate the conversion rate of the C to T conversion. A total of 2% of reads were uniquely mapped to the *E. coli* genome and the bisulfite conversion efficiency was calculated to be 99.00%. Illumina reads for each sample which did not map to the *E. coli* reference genome were then aligned to a synthetic genome composed of the reference genomes of *P. citri* and *P.ficus* (*Planococcus citri* version v0 and *Planococcus ficus* version v0, both publicly available on mealybug.org) using Bismark v0.19.0 with Bowtie2 (Krueger and Andrews, 2016) at default alignment mismatch settings. Reads derived from PCR duplicates and those that map to multiple locations in the genome were removed from downstream analysis (see Table S1 for full mapping details).

Merged coverage for each sex was calculated using CGmapTools v0.1.0 (Guo *et al.*, 2017). Coverage was calculated to be 19.0X for males and 15.0X for females. Coverage at cytosine sites was calculated as 11.0X in males and 9.0X in females. A breakdown of coverage by sample can be found in Table S2.

3.3.6.2 Reasons for lower than expected coverage

The random priming that occurs in post-bisulfite methods such as the Pico Methyl-Seq Library Prep Kit (Zymo) introduce errors, indels and methylation

biases that can detrimentally affect mapping efficiencies and methylation calls. This low input bisulfite sequencing kit generated a considerable level of bias at 3' and 5' ends during the 'random' priming stage and so 15% of the read length was lost in the quality trimming process, which required additional 10bp trimming from 5' and 3' ends after adapter trimming. This resulted in lower coverage than expected across the genome.

3.3.7 Assigning parental origin to DM sites, regions and genes

3.3.7.1 SNP calling

To identify sites and regions in the hybrid genomes with allele-specific methylation, I first identified heterozygous SNPs in each sample using CGmapTools' BayesWC strategy. This strategy is designed specifically for bisulfite data and particularly useful in low coverage contexts. ATCGmap files produced in CGmapTools provide read counts on the Watson and Crick strands at each base, from which an ATCG table is produced. From this table, a genotype is generated from the bisulfite data. Due to the nature of bisulfite data, cytosines (C) may be converted to uracils, therefore the presence of a thymine (T) in a BS-seq read may indicate a T or a C in the unconverted genome. For example, if the ATCG table only has a T on the Watson strand, the site could arise from genotypes such as TT, CC or TC genotypes. CGmapTools assigns wildcard genotypes to denote ambiguity in the predicted genotypes (Y = T or C; R = A or G). A Bayesian model is then used to resolve ambiguity and allow reliable variant calling: The genotype with the highest posterior probability from the exact genotype set and wildcard genotype set is selected as the predicted genotype (see Guo *et al.,* 2017 for full description). This strategy retains more data than SNP-calling strategies designed for non-bisulfite data in which ambiguous genotypes are simply removed. The SNPs identified in this analysis were then verified using the *P. citri* and *P. ficus* reference genomes.

3.3.7.2 Allele-specific methylation

Using predicted SNPs (common to each sample and at a minimum read-depth of 5), I used the asm function in CGmapTools v0.1.0 to identify allele-specific methylation at sites and regions within the hybrid genomes. An allele-specific site or region was considered to be differentially methylated if the following criteria were satisfied: [1] minimum read depth of 5 at C site in order to call methylation level; [2] at least 2 CpG sites in the region; [3] the methylation level on the hypomethylated allele should be < 0.2 and on the hypermethylated allele should be > 0.8 and [4] corrected *p*-value from multiple t-tests should be < 0.05, FDR <5% (Benjamini and Hochberg, 1995). Again, only DM sites and regions common to all samples were considered for analysis.

3.3.7.3 Sex–specific methylation analyses

Overall levels of methylation in various cytosine (C) contexts were calculated for each sample using CGmapTools v0.1.0 and average CpG methylation levels of genes were identified using BedTools v1.6 (Quinlan and Hall, 2010). Methylation differences between males and females were assessed using a principal component analysis (PCA) and by identifying differentially methylated (DM) sites, regions and genes. PCA was carried out using MethylKit v1.7.0 (Akalin *et al.*, 2012). DM sites and regions were identified using CGmapTools v0.1.0 and only CpG sites and regions with a minimum coverage of 5 reads per sample and common to all replicates were considered for analyses. To be considered differentially methylated, a site had to have at least a 15% methylation difference at a 5% FDR (Q < 0.05) (Benjamini and Hochberg, 1995). Differentially methylated regions were calculated using the Dynamic Fragment Strategy in CGmapTools v0.1.0. Regions were calculated using the following criteria: [1] maximum fragment size is 1000bp, [2] fragment must have at least 5 cytosines, and [3] the maximum distance between two adjacent common cytosines is 100bp. I

considered regions to be differentially methylated if they had at least a 10% methylation difference at a 5% FDR (Q < 0.05).

3.3.7.4 Genomic context of differentially methylated sites and GO term enrichment analysis

GO term enrichment analyses of differentially methylated gene sets were performed using GOAtools version v0.6.10 (Klopfenstein *et al.*, 2018). Redundant terms were then removed using REVIGO (Supek *et al.*, 2011).

3.4 Results

3.4.1 Global DNA methylation analysis

As with sex-specific analyses in *Planococcus citri* (Chapter 2), global methylation levels in *Planococcus* hybrids were assessed in following C contexts: C, CG, CHG, CHH, CA, CC, CT, CH and CW (see Chapter 2 for details). In all samples, methylation predominantly occurs in a CG context, indicating that CpG methylation is the predominant form of DNA methylation in these species (Figure 1).

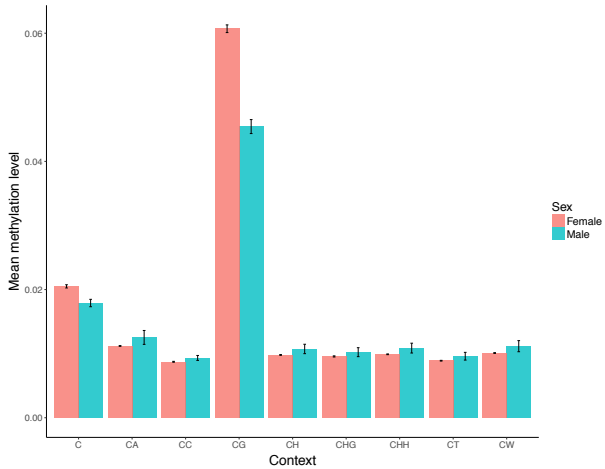


Figure 1: Average level of global methylation in all C contexts (C, CG, CHG, CHH, CA, CC, CT, CH, CW). In both males and females, methylation is predominantly present in the context of CpG.

3.4.2 Allele-specific differential methylation in hybrids

The differences in behaviours of the two parental chromosome sets in PGE males provide a striking example of genomic imprinting. Here I use the hybrid offspring of two *Planococcus* species, *P. citri (PC)* and *P. ficus (PF)*, to investigate the role of DNA methylation in mediating these parent-of-origin specific behaviours. Firstly, I identified 217,836 species-specific SNPs in the *citri* x *ficus* hybrid bisulfite genomes (Table 1) (minimum read depth = 5 in each sample). These SNPs were confirmed by cross-referencing with PC and PF reference genomes (*P. citri* version v0 and *P. ficus* version v0, both publicly available on mealybug.org) and then used as markers of parental origin (PC = maternal, PF = paternal) for differential methylation (DM) analyses. I conducted DM analyses at site, region and gene levels in order to reveal allele-specific patterns of DNA methylation in the male and female hybrid genomes. Out of 217,836 SNPs identified between the PC and PF parental genomes, less than 1% of these are differentially methylated (1026 SNPs). At site-level, I find 444 differentially methylated (DM) SNPs in female hybrids and 582 DM SNPs in male hybrids (common to all replicates). In both sexes, the majority of these DM SNPs show significantly higher levels of DNA methylation at the maternal site (PC specific SNPs) relative to the paternal site (PF specific SNPs) (p-value = 0.0001, Figure 2). In females, 316 (76%) of DM SNPs are hypermethylated at the maternally inherited site and 91 (24%) are hypermethylated at the paternally inherited site. In males, 410 (71%) of DM SNPs are hypermethylated at the maternally inherited site and 144 (29%) are hypermethylated at the paternally inherited site (Figure 3).

Table 1: Number of allele-specific DM sites in different regions of the hybrid genomes (only those common to all replicates are included). The majority of DM sites are located within gene bodies in both sexes, predominantly in exons.

		DM SNPs in males		DM SNPs in females	
Region	Total number of SNPs identified	Paternally hypermethylated	Maternally hypermethylated	Paternally hypermethylated	Maternally hypermethylated
Exon	48,780	104	357	65	264
Intron	59,611	5	7	2	15
Gene body (intron + exon)	108,391	109	364	67	279
Promoter	32,136	11	50	8	57
Intergenic	127,513	25	23	18	15

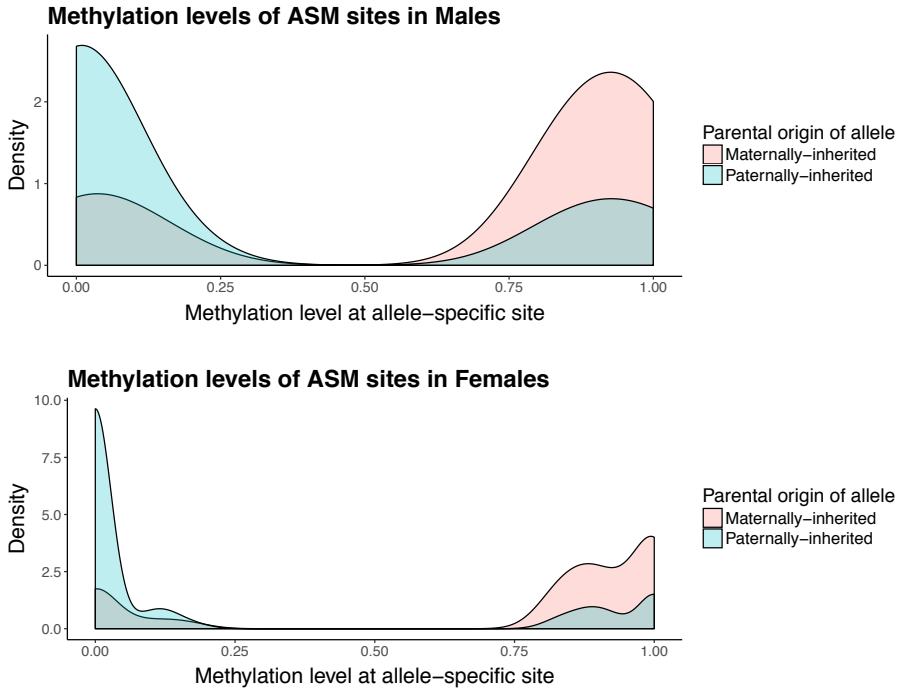


Figure 2: Methylation levels of paternally and maternally inherited sites in males and females. In both sexes, maternally inherited sites have higher levels of DNA methylation than paternally inherited sites (*p*-value = 0.0001 in both sexes, Welch Two-sample T-test).

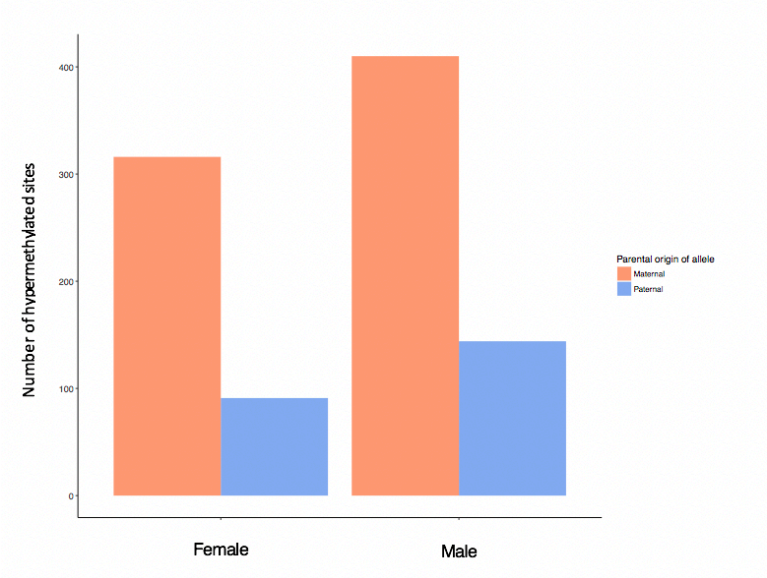


Figure 3: Number of sites at which a parental allele is hypermethylated at allele-specific methylation sites in males and females. Only sites common to all replicates are included in analysis.

Allele-specific methylation analysis of regions in the hybrid genome show a similar pattern to site analysis. Here, regions are defined using a differential fragmentation strategy (Guo *et al.*, 2017). Regions are between 5-1000bp in length, must contain at least 5 cytosine residues and the distance between adjacent cytosines must be no greater than 100bp. Overall, in both sexes, DNA methylation levels in DM regions are higher in females than in males (*p*-value = 0.00013, Figure 4). In females, the number of hypermethylated maternally inherited regions (n=8) is higher than the number of hypermethylated paternally inherited regions (n=1). The same pattern is found in males with 11 regions hypermethylated in the maternally inherited region and 3 hypermethylated in the paternally inherited region (Figure 5). Due to low coverage, the ability to reliably detect differential methylation of SNPs is impaired. However, the overall pattern is consistent between biological replicates, with paternally inherited alleles showing lower levels of methylation than maternally inherited alleles in both regions and sites. These analyses reveal consistent patterns of allele-specific methylation that could

be involved in imprinting and parent-of-origin specific expression in PGE males.

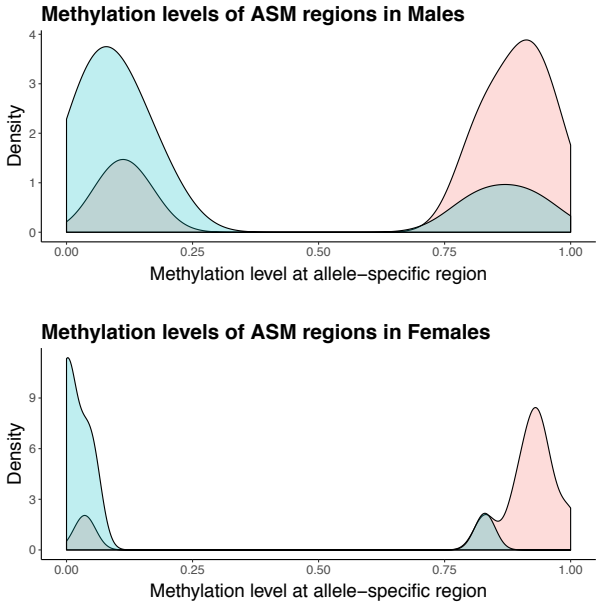


Figure 4: Methylation levels of paternally and maternally inherited regions in males and females. Blue = paternally inherited alleles; pink = maternally inherited alleles. In females, maternally inherited regions have higher levels of methylation relative to paternally inherited regions (p = 0.00013 in both sexes, Welch Two-sample T-test).

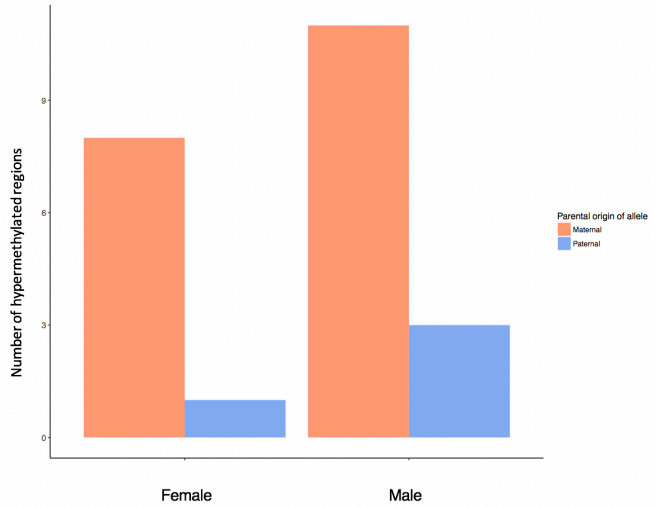


Figure 5: Number of regions at which a parental allele is hypermethylated in males and females. Only regions common to all replicates are included in analysis.

3.4.3 Genomic context of allele-specific differential methylation

To investigate possible effects of this allele-specific DNA methylation, I examined the genomic context of DM CpG sites between paternal and

maternal alleles in both sexes. Previous studies in *P. citri* show that the majority of methylated CpG sites are found within gene bodies (exons and introns) (Chapter 2). Similarly, I find that the majority of all DM allele-specific sites in males and females are located in gene bodies, predominantly in exons (total number in gene bodies: males = 473 (461 in exons), females = 346 (329 in exons), see Table 1). In both sexes, the number of maternally hypermethylated sites in gene bodies is higher than the number paternally hypermethylated (males: paternally hypermethylated = 104, maternally hypermethylated = 357; females: paternally hypermethylated = 65, maternally hypermethylated = 264; see Table 1). This pattern also holds true at intron and exon levels (Table 1). DM allele-specific sites are also located in the promoter regions of both sexes and again I find that the number of maternally hypermethylated sites is higher than the number of paternally methylated sites (males: paternally hypermethylated = 11, maternally hypermethylated = 50; females: paternally hypermethylated = 8, maternally hypermethylated = 57; see Table 2). Intergenic regions host a number of DM allele-specific sites in both sexes. Here, however, there is no significant difference between the number of paternally hypermethylated sites and the number of maternally hypermethylated sites (males: paternally hypermethylated = 25, maternally hypermethylated = 23; females: paternally hypermethylated = 18, maternally hypermethylated = 15; see Table 1). At all regions, the number of DM SNPs identified is a small proportion of the overall number of SNPs detected (Table 1). Therefore, although these results highlight potentially important differences in allele-specific methylation patterns at sites associated with regulation of gene expression in insects, the low number may suggest a limited or highly specific role in parent-of-origin recognition and gene expression.

To further investigate the role of parent-of-origin methylation in parent-of-origin expression, I conducted a gene level analysis of DM SNPs. In males, there are 28 unique genes hypermethylated at paternal sites and 100 unique genes hypermethylated at maternal sites. In females, 21 unique genes are hypermethylated at paternal sites relative to the maternal site and 97 unique genes are hypermethylated at maternal sites. Out of 142 maternally hypermethylated genes, 55 genes (39%) are found in both sexes, whilst 42 genes and 45 genes are found specifically in females and males, respectively (Figure 6a). Out of a total of 49 paternally hypermethylated genes, 8 (16%) are common to both sexes whilst 13 and 20 of these are found only in females and males, respectively (Figure 6b). DM genes found in only one of the sexes may encode proteins that have sex-specific roles mediated by DNA methylation. To investigate this hypothesis, GO term enrichment analyses were run on the following six DM gene categories:

- paternal hypermethylated in both sexes
- maternal hypermethylated in both sexes
- maternal hypermethylated in females
- maternal hypermethylated in males
- paternal hypermethylated in females
- paternal hypermethylated in males

GO terms associated with fructose metabolism were found to be enriched in paternal and maternal hypermethylated genes common to both sexes. In other categories, no significant GO term enrichment was identified, thus sex-specific genes were not identified.

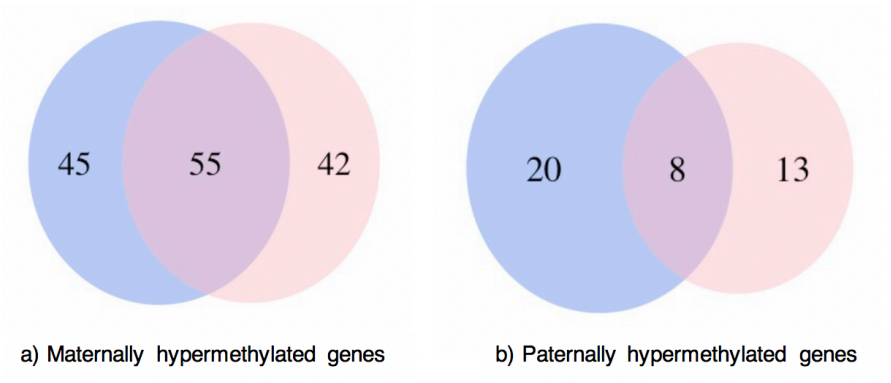


Figure 6: Venn diagrams of the number of a) maternally and b) paternally hypermethylated genes found in each sex and common to both sexes (pink = female; blue = male).

3.4.4 Relationship between parent-of-origin specific methylation and parent-of-origin specific expression

Overall, the number of maternally and paternally hypermethylated genes in each sex is very similar, even though parent-of-origin specific gene is expression is only known to occur in males. As studies suggest an association between increased levels of DNA methylation and elevated gene expression in insects (Wang *et al.*, 2013; Glastad, Hunt and Goodisman, 2014; Mathers *et al.*, 2018), I compared expression levels of parent-of-origin specific DM genes in hybrid males with parent-of-origin specific expression patterns using hybrid male transcriptome data (de la Filia, Thesis 2018). However, I find no pattern of parent-of-origin specific gene expression associated with hyper- or hypomethylation of parental alleles.

3.4.5 Sex-specific DNA methylation in Planococcus hybrids

In order to test if *citri* x *ficus* hybrid patterns of methylation are consistent with pure *P. citri* patterns, I conducted a sex-specific analysis of DNA methylation patterns in hybrid males and females. A clustering analysis based on methylation at CpG sites shows distinct clustering of male and female

samples indicating reproducible differences in global patterns of CpG methylation between the sexes (Figure 7).

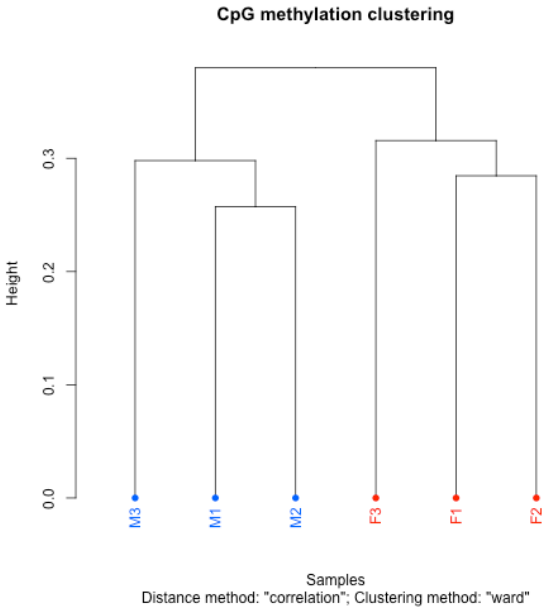


Figure 7: Clustering analysis of all samples based on methylated CpG sites. Male samples are indicated by blue text and female samples are indicated by red text.

I also find that, similar to *P. citri*, *citri* x *ficus* hybrids have high global methylation levels relative to other insect taxa and sex-specific methylation patterns are apparent. Hybrid females and *P. citri* females have similar levels of global CpG methylation (hybrid = 6%, se = +/- 0.06%; *P. citri* = 6%, se = +/- 0.01%; Welch Two-sample T-test *p*-value = 0.78) (Figure 8). However, hybrid males have significantly lower global methylation than *P. citri* males (hybrid = 4.5%, se = +/- 0.1%; *P. citri* = 7.9%, se = +/- 0.1%; Welch Two-sample T-test *p*-value = 0.00006) (Figure 8b). Thus, in *citri* x *ficus* hybrids, males have significantly lower global DNA methylation levels than females (Welch Two-sample T-test, *p*-value = 0.0009764), which is opposite to the pattern found in pure *P. citri* (Figure 9). Interestingly, patterns of methylation at DM sites and regions are also contrary to those found in *P. citri*, with the majority hypermethylated DM sites and regions found in males (Table 2). At the gene level, most of the DM genes detected show higher methylation levels in males relative to females (Table 3), again, showing the opposite pattern to *P. citri* in which the majority of DM genes show higher levels of

methylation in females (see Chapter 2). This suggests that there may be species-specific differences in DNA methylation between *P. citri* and *P. ficus.*

Table 2: Number of differentially methylated sites and regions (methylation difference > 15%, FDR < 5%) between hybrid males and females. 97.5% of DM sites are hypermethylated in males and only 2.5% of DM sites are hypermethylated in females. A similar pattern is found in regional analysis with 92.7% of DM regions hypermethylated in males and only 7.3% of DM regions hypermethylated in females.

	Total number	Number of hypermethylated in males	Number hypermethylated in females
Sites	3621	3532 (97.5%)	89 (2.5%)
Regions	17017	15778 (92.7%)	1239 (7.3%)

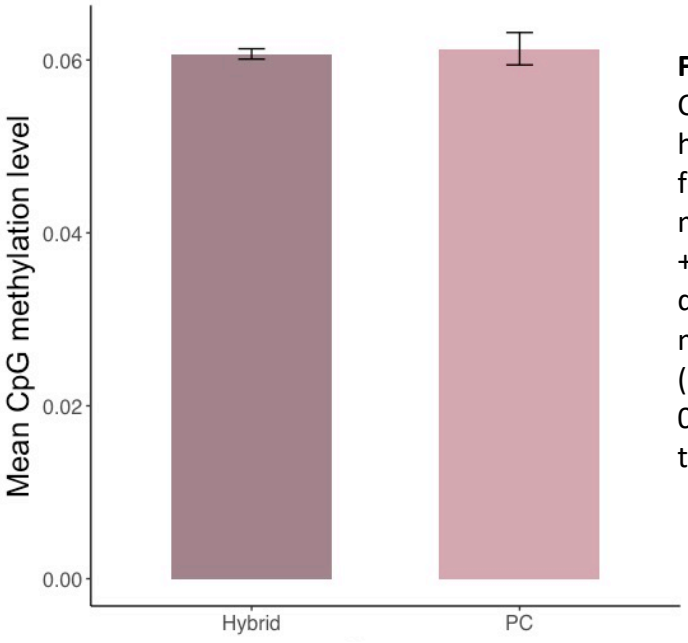


Figure 8a: Average global CpG methylation levels of hybrid and *P. citri* (PC) females. Hybrid female DNA methylation level (6%, se = +/- 0.06%) is not significantly different to levels of DNA methylation in PC females (6%, se = +/- 0.1%), p-value = 0.78 (Welch Two-sample T-test).

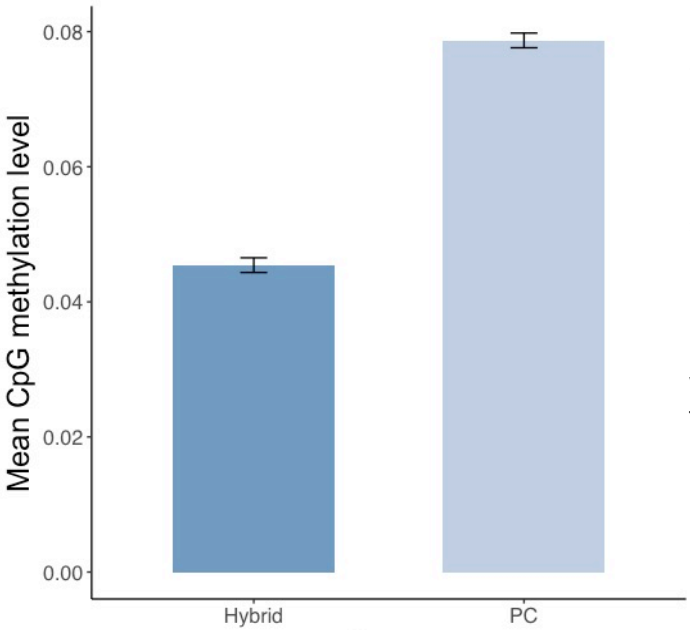


Figure 8b: Average global CpG methylation levels of hybrid and *P. citri* (PC) males. Hybrid male DNA methylation levels (4.5%, se = +/- 0.01%) are significantly lower than levels of DNA methylation in PC males (7.5%, se = +/- 0.01%), p-value = 0.000058 (Welch Two-sample T-test).

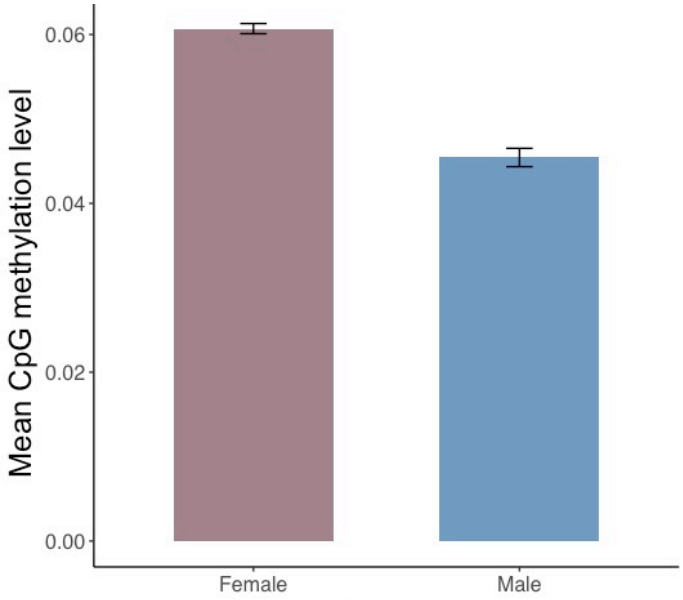


Figure 9: Average global CpG methylation levels of hybrid males and females. Hybrid female DNA methylation level (6%, se = +/- 0.06%) is significantly higher than hybrid male methylation level (4.5%, se = +/- 0.1%), *p*-value = 0.0009764 (Welch Two-sample T-test).

Table 3: Number of unique genes at differentially methylated regions genes that are hypermethylated in males and females.

Total number of unique genes	Number hypermethylated in males	Number hypermethylated in females
1411	1369	42

3.5 Discussion

The relationship between DNA methylation and genomic imprinting has been well studied in plants and mammals but remains elusive in insects. This is largely due to the fact that in species for which there is DNA methylation data, there is no clear evidence of parent-of-origin specific imprinting. Paternal genome elimination is an extreme form of genomic imprinting in which the paternally inherited haploid complement in males is subject to somatic silencing and germline elimination. This parent-of-origin specific chromosome behaviour is found in a number of insect taxa, including the mealybug genus, *Planococcus*, which provide a unique opportunity to investigate the epigenetic mechanisms underlying genomic imprinting in insects. Here, I use base-pair resolution analysis of allele-specific DNA methylation to investigate the role of this epigenetic modification in two key but distinct processes of PGE: i) recognition of the parental origin of alleles and ii) silencing of the paternal genome in males.

The intention of this study was to investigate potential patterns of parent-of-origin specific methylation, however, due to male hybrid inviability from crosses with *P. ficus* mothers and *P. citri* fathers, this analysis lacks

reciprocal crosses. Therefore, it cannot be ruled out that the allele-specific differences in DNA methylation described here may in fact relate to species-specific DNA methylation differences between *P. citri* and *P. ficus*. For clarity, I will discuss potential species-specific and parent-of-origin specific interpretation of results and move on to describe useful future studies to elucidate these findings.

In *Planococcus* hybrids, I detect the presence of allele-specific DNA methylation in both sexes. In males and females, paternally inherited (*P. ficus*) SNP sites and regions have overall lower levels of methylation relative to those inherited maternally. An enzymatic study of DNA methylation levels of parental alleles in *Planococcus citri* (Bongiorni, Cintio and Prantera, 1999) also showed lower levels of DNA methylation on paternally inherited chromosomes relative to those maternally inherited in both sexes. Furthermore, this methylation pattern is also similar to that found on the X chromosomes in female mammals, where the inactive X chromosome has overall lower DNA methylation levels than the active X chromosome (Bernardino *et al.*, 2000; Weber *et al.*, 2005; Rens *et al.*, 2010). Therefore, it is possible that these methylation differences are parent-of-origin specific and that in *Planococcus* DNA methylation acts as a molecular marker distinguishing the parental origin of alleles after fertilization. Thus, allowing specifically paternally inherited alleles to be silenced and, in the germline, eliminated through PGE.

However, parent-of-origin specific methylation differences cannot be disentangled from species-specific differences or indeed haplotype-specific differences. To confirm that these methylation differences are parent-of-origin specific – as opposed to species-specific - this experiment should be repeated using intraspecific reciprocal crosses of genetically distinct *P. citri* lines. Furthermore, allele-specific methylation and expression results should be interpreted with caution as allele-specific methylation can be determined

by DNA sequence in *cis*, confounding parent-of-origin analyses (Meaburn, Schalkwyk and Mill, 2010; Remnant *et al.*, 2016). Thus, understanding the relationship between DNA methylation and genomic imprinting across taxa is challenging.

Interestingly, this study also revealed a remarkable difference between sex-specific DNA methylation levels found in pure *P. citri* males and *citri* x *ficus* hybrid males. Pure *P. citri* and hybrid females have no significant difference in global methylation levels, however, *P. citri* males have significantly higher methylation levels than hybrid males. This suggests that there are species-specific differences in DNA methylation or that hybridisation somehow affects the epigenetic landscape of these organisms. It is particularly curious that hybridisation only appears to affect global methylation levels in males, as hybrid male mortality rates are high (Rotundo and Trembley, 1982). There is growing evidence to support that mis-regulation of gene silencing plays a role in hybrid incompatibility (Bomblies, 2006). Although there are few studies addressing species-specific epigenetic landscapes and hybrid incompatibility in insects, a study in the plant genus, *Arabidopsis*, shows that epigenetic variation contributes to hybrid genome incompatibility (Blevins *et al.*, 2017). Methylation analysis of pure *P. ficus* males and females is required to further understand this result and investigate whether DNA methylation has a role in high hybrid male mortality.

As well as functioning as a marker of parental origin, in mammals and flowering plants, DNA methylation can serve as a mechanism to regulate expression of alleles in a sex-specific manner. In PGE males, the paternally inherited alleles are transcriptionally silenced and only those maternally inherited are expressed. Therefore, as well as a mechanism to allow recognition of the parental origin of alleles, there must also be a mechanism by which paternal alleles are silenced. It is known that paternal alleles are hyper-condensed into a heterochromatic state in early embryogenesis

(Bongiorni and Prantera, 2001) but the role of DNA methylation in this process remains unclear. My results show that although there are clear allele-specific methylation differences in *Planococcus* hybrids, the proportion of methylated SNPs is very small (<1%). This may indicate a limited role for DNA methylation in the silencing of an entire chromosome. Furthermore, in *P. citri,* levels of gene methylation do not correlate with levels of gene expression, and the role of DNA methylation in gene expression is unclear (Chapter 2). Further investigation into the role of DNA methylation in the regulation of gene expression is crucial to our understanding of how this epigenetic mechanism may be involved in the complex process of genomic imprinting.

3.6 SUPPLEMENTARY TABLES AND FIGURES

Table S1: Hybrid and *E. coli* DNA reads

Hybrid reads

Sample	Total number uniquely aligned reads (million)	Single-end mapping efficiency	Number of Cs analysed	Total CpGs
Male 1	28.1	41.40%	439,037,404	97,629,607
Male 2	52.5	43.80%	917,209,427	204,098,648
Male 3	37.1	46.70%	632,328,790	132,805,140
Female 1	41.2	38.50%	590,927,140	128,691,516
Female 2	33.5	38.00%	480,471,220	107,051,621
Female 3	16.2	35.90%	358,071,266	79,685,514

E. coli reads

Sample	Total number uniquely aligned reads (million)	Single-end mapping efficiency	Number of Cs analysed	Total CpGs
Male 1	0.4	2.8%	11,483,633	3,586,057
Male 2	1	3.7%	29,877,414	9,279,114
Male 3	1	5.6%	30,636,786	9,335,944
Female 1	0.9	4.3%	25,630,552	7,958,175
Female 2	0.6	3.3%	16,248,329	5,082,874
Female 3	0.4	2.8%	11,020,768	3,463,275

Libraries were sequenced over two lanes. The above combines information for both lanes.

Table S2: Per sample coverage

Sample	Overall global coverage	Methylation effective coverage
Female 1	6.8x	4.6x
Female 2	5.9x	4.2x
Female 3	4.8x	3.5x
Male 1	5.4x	3.9x
Male 2	9.6x	6.2x
Male 3	6.8x	4.5x
Female Average	5.8x	4.1x
Male Average	7.3x	4.9x

Table S3: GO term enrichment analysis

GO Number	Name	Corrected *p*-value (BH, FDR)
GO:0008152	metabolic process	3.13E-08
GO:0015074	DNA integration	3.52E-05
GO:0055114	oxidation-reduction process	5.23E-05
GO:0006468	protein phosphorylation	0.00177005
GO:0006508	proteolysis	0.00387038
GO:0006741	NADP biosynthetic process	0.0093269
GO:0006471	protein ADP-ribosylation	0.03470807
GO:0006030	chitin metabolic process	0.04886199
GO:0007608	sensory perception of smell	0.07565018
GO:0006779	porphyrin-containing compound biosynthetic process	0.08095038
GO:0018149	peptide cross-linking	0.08095038
GO:0006000	fructose metabolic process	0.08095038
GO:0016021	integral component of membrane	0.00189506
GO:0016020	membrane	0.00387038
GO:0005634	nucleus	0.04886199
GO:0005576	extracellular region	0.07565018
GO:0005622	intracellular	0.07565018
GO:0005737	cytoplasm	0.11330395
GO:0019898	extrinsic component of membrane	0.12402419
GO:0005739	mitochondrion	0.18036842
GO:0005515	protein binding	2.56E-08
GO:0003676	nucleic acid binding	3.13E-08
GO:0005524	ATP binding	1.03E-06
GO:0003677	DNA binding	5.47E-06
GO:0046872	metal ion binding	0.00090839
GO:0004672	protein kinase activity	0.00177005
GO:0008236	serine-type peptidase activity	0.00177005
GO:0016491	oxidoreductase activity	0.00486791
GO:0003723	RNA binding	0.00905909
GO:0003951	NAD+ kinase activity	0.0093269
GO:0005506	iron ion binding	0.01948975
GO:0003824	catalytic activity	0.02087307
GO:0016705	oxidoreductase activity	0.03004809

GO:0016624	oxidoreductase activity, acting on the aldehyde or oxo group of donors	0.03470807
GO:0008270	zinc ion binding	0.03791025
GO:0003899	DNA-directed 5'-3' RNA polymerase activity	0.03791025
GO:0003810	protein-glutamine gamma-glutamyltransferase activity	0.04803397
GO:0008410	CoA-transferase activity	0.04803397
GO:0000166	nucleotide binding	0.04886199
GO:0008061	chitin binding	0.04886199
GO:0003700	DNA binding transcription factor activity	0.04998994

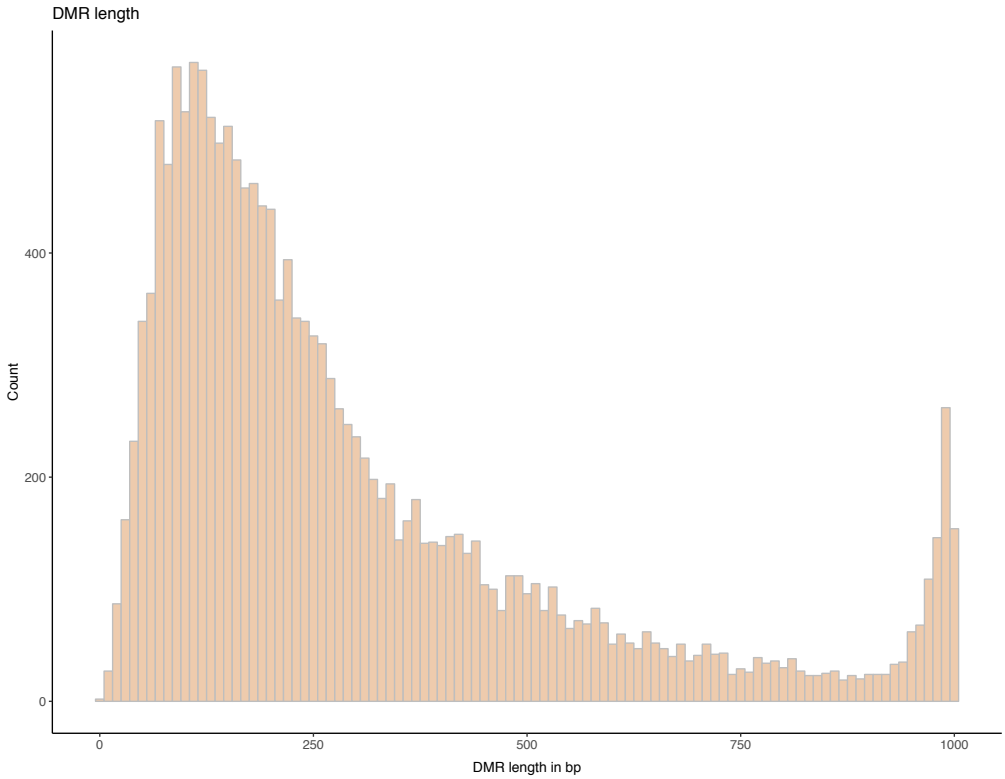


Figure S1: Variation in length of DMRs generated by Dynamic Fragment Strategy (Guo *et al.*, 2017) . Minimum length is 5 base pairs and maximum length is 1000 base pairs.

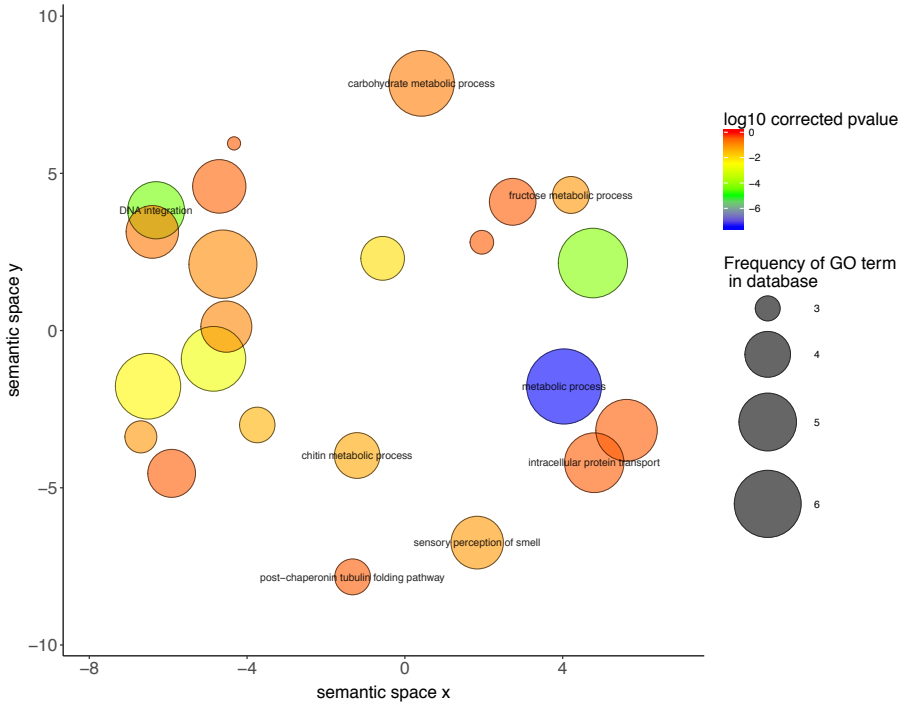


Figure S2: Enriched GO terms related to biological processes plotted in semantic space for differentially methylated regions in which males are hypermethylated relative to females.

Chapter 4: Role of two key heterochromatin pathways in Paternal Genome Elimination

4.1 Chapter Summary

This study investigates the role of two evolutionarily conserved heterochromatin pathways, H3K9me3-HP1 and H3K27me3-PRC2, in Paternal Genome Elimination (PGE). I use an immunocytological approach to detect the presence of these histone modifications in the two key processes of PGE: 1) silencing of the paternal genome in somatic tissues and 2) recognition, elimination and imprinting of paternal chromosomes during spermatogenesis.

4.2 Introduction

For most sexually reproducing organisms, the two parentally inherited copies of a gene are equivalent in transmission and expression. However, there are exceptions to this rule. Genomic imprinting is an epigenetic process in which expression of one gene copy is favoured depending on its parental origin (Moore and Haig, 1991). This phenomenon is extensively studied mammals and flowering plants, however, imprinting was first discovered in insects decades before its discovery elsewhere (Crouse, 1960). One of the most striking cases of genomic imprinting in insects is Paternal Genome Elimination (PGE), a process that involves the elimination of an entire haploid genome in a parent-of-origin specific manner. Under PGE, both sexes develop from fertilized eggs and initially possess a diploid euchromatic chromosome complement. However, males subsequently eliminate paternally inherited chromosomes from their germline. Different PGE species vary in the timing of the elimination of the paternal genome, and in whether it

becomes transcriptionally silenced or not. As a result, male gene expression varies from haploid to diploid with various intermediates. PGE is found in several insect taxa, including the citrus mealybug, *Planococcus citri* (Pseudococcidae, Hemiptera). In *P. citri,* in addition to the elimination of paternal chromosomes from the germline, the entire haploid set of paternally inherited chromosomes becomes condensed and transcriptionally silenced in somatic cells at mid-cleavage embryogenesis (Brown and Nur, 1964).

PGE in *P. citri* is an example of two important genetic phenomena: differential regulation of homologous chromosomes and genomic imprinting. The silencing of the paternal genome in male somatic tissue is an example of differential expression of homologous chromosomes in a parent-of-origin specific manner. In PGE, this regulation is achieved through facultative heterochromatinization, where the paternally inherited chromosomes are condensed and merge to form a transcriptionally silenced heterochromatic body (Hughes-Schrader, 1948; Brown and Nelson-Rees, 1961). During spermatogenesis, complete segregation of the maternal and paternal genomes occurs and the nuclei containing the paternal genome disintegrate, with only the maternal genome packaged into mature sperm (Brown and Nelson-Rees, 1961). This process essentially renders males functionally haploid.

The recognition and silencing of paternally derived chromosomes under PGE remain poorly understood. However, striking parallels can be drawn between PGE and the process of paternal X chromosome inactivation that occurs in marsupial mammals and in early development in mice. In these mammals, the paternally inherited X chromosomes are recognised and specifically silenced in all cells (Lyon, 1961; Sharman, 1971). Indeed, PGE in *P. citri* and X chromosome inactivation appear to be regulated by similar epigenetic mechanisms, namely histone modifications associated with chromatin condition (Bongiorni and Prantera, 2001; Heard, 2005). Heterochromatin has

a critical role in a number of cellular and evolutionary processes, including gene silencing and chromosome segregation (Brown, 1966). There are two pathways critical to the formation of heterochromatin, both of which involve methylation of histone H3. The H3K9me3-HP1 pathway involves tri-methylation of histone H3 at lysine 9 (H3K9me3) by the histone methyltransferase Suppressor of Variegation 3-9 (SU(VAR)3-9); in turn, H3K9me3 binds to Heterochromatin Protein 1 (HP1), which mediates methylation of H3K9me3, generating a positive feedback loop that creates and maintains heterochromatin (Schotta *et al.*, 2002; Schotta, Ebert and Reuter, 2003) (Figure 1a). Although traditionally associated with constitutive heterochromatin, the H3K9me3-HP1 pathway is now known to have a more diverse function in the epigenetic silencing of chromosomal regions through facultative heterochromatinization. In *P. citri,* HP1 and H3K9me3 precede the onset of heterochromatinization of the paternal chromosomes in PGE males and, furthermore, HP1 knockouts show a loss of heterochromatinization and associated histone modifications in male embryos (Bongiorni *et al.*, 2007). This suggests a causative role in the silencing of the paternal genome in PGE.

The establishment of sex-specific epigenetic marks during gametogenesis is a key feature of genomic imprinting and studies suggest that H3K9me3 may also have a role in the genomic imprinting that occurs during meiosis in *P. citri* males (Buglia and Ferraro, 2004; Bongiorni *et al.*, 2009). Firstly, H3K9 methylation is intimated as the molecular marker that distinguishes the parental origin of chromosomes in the male germline allowing for the non-independent assortment of parental chromosomes during meiosis (Bongiorni *et al.*, 2009). This process is crucial for PGE as it produces nuclei that contain only the maternal genome, which then elongate into mature sperm. The nuclei containing only the paternal genomes hyper-condense into pyknotic nuclei. Secondly, it is proposed that H3K9me3 is carried into the ooplasm on the nuclei of mature sperm and acts as the imprint that identifies

the paternally inherited genome in male embryos (Bongiorni *et al.*, 2007). Furthermore, different levels of H3K9me3 found on the two sperm derived from the same meiotic division are believed to have a role in sex determination (Buglia and Ferraro, 2004). However, support for these hypotheses in *P. citri* is, at best, limited; therefore, further investigation is required in order to explain the role that H3K9me3 may have in the recognition of the paternally inherited genome after fertilisation and the imprinting that occurs during spermatogenesis in PGE.

The H3K9me3-HP1 pathway is suggested to interact with a second pathway involved in facultative heterochromatinization in which histone H3K27 is tri-methylated by the Polycomb Repressive Complex 2 (PRC2) (Boros *et al.*, 2014; Jamieson *et al.*, 2016) (Figure 1b). The H3K27me3-PRC2 pathway is a hallmark of facultative heterochromatin (Schwartz and Pirrotta, 2007) and plays a key role in X chromosome inactivation in mammals (Plath *et al.*, 2002). However, its role in the recognition, silencing and elimination of the paternal genome in *P. citri* males has not yet been tested. Interestingly, a study in *Maconellicoccus hirsutus,* another mealybug species with PGE, shows higher expression of H3K27me3 in male nuclei relative to female nuclei (Mathur *et al.*, 2010) suggesting a potential role in the formation of facultative heterochromatin of the paternally inherited genome in males.

In this study, I use *P. citri* to investigate the role of these evolutionarily conserved heterochromatin pathways, H3K9me3-HP1 and H3K27me3-PRC2, in the distinct processes of Paternal Genome Elimination: 1) the silencing of the paternal genome in male somatic nuclei and 2) the recognition, elimination and imprinting of the paternal genome during spermatogenesis. To elucidate the role(s) of these pathways in genomic silencing, I use an immunocytological approach to test whether histone modifications H3K9me3 and H3K27me3 are associated with the condensed paternal genome in male somatic nuclei. I also examine the presence of

these modifications on the paternal and maternal genomes throughout spermatogenesis. I focus on comparing histone modifications on the heterochromatic paternal chromosomes and the euchromatic maternal chromosomes that may act as a molecular identifier of the chromosomes parental origin. In pupal sperm cysts, I examine whether the presence of H3K9me3 and H3K27me3 supports the hypothesis that these histone modifications act as the imprint, which is transmitted to the oocyte at fertilization.

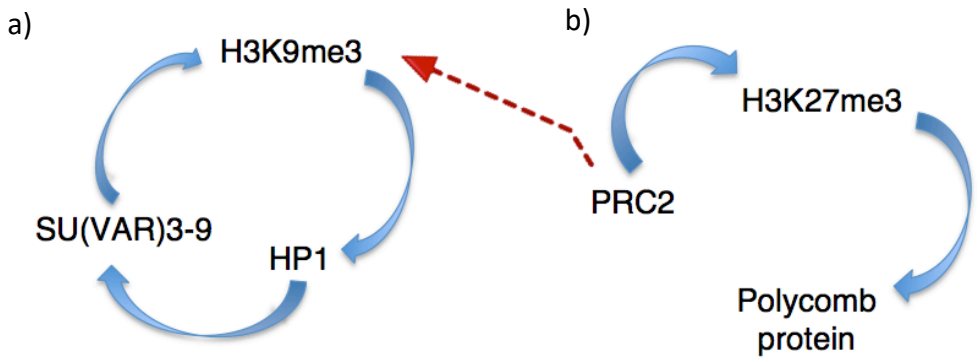


Figure 1: Two histone H3 methylation pathways crucial for the formation of heterochromatin. a) H3K9me3-HP1 pathway: Histone methyltransferase, Suppressor of Variegation 3-9 (SU(VAR)3-9), selectively methylates histone H3 at Lysine 9 (H3K9) generating a binding site for the heterochromatin protein 1 family (HP1). HP1 is, in turn, associated with SU(VAR)3-9, promoting methylation of H3K9 thus generating a heterochromatin feedback loop (Schotta *et al.*, 2002); b) H3K27me3-PRC2 pathway: Polycomb Repressive Complex 2 (PRC) methylates Histone H3 at Lysine 27. Tri-methylated H3K27 specifically binds to the chromodomain of Polycomb protein, which is involved in gene silencing through heterochromatinization (Schwartz and Pirrotta, 2007). Red arrow indicates potential function of PRC2 in methylating H3K9 (Boros *et al.*, 2014). These schematics are based on information from studies in *Drosophila* and *Neurospora*.

4.3 Methods

4.3.1 Insect husbandry

Planococcus citri are genetically tractable with a short generation time and ease of laboratory rearing. Insects were cultured on sprouting potatoes in sealed plastic bottles at 25°C and ~70% relative humidity. Under these conditions, *P. citri* has a generation time (time from oviposition until sexual maturity) of approximately 30 days. Experimental isofemale lines (CP1-2) originate from natural populations in Israel and are reared in the laboratory under a sib-mating regime. In each generation, one mated female is taken from culture and transferred to a new container to give rise to the next generation.

4.3.2 Tissue collection and fixation

I conduct immunocytological analyses of histone modifications H3K9me3 and H3K27me3 throughout spermatogenesis in *P. citri*. The life cycle of males is approximately 37 days. Males go through three larval stages before pupating and emerging as winged adults at around day 35. Male germinal tissues divide actively toward the end of the second instar. Testes were dissected from 2^{nd} – 3^{rd} instar males and pupal males. Whole males were fixed in Bradley-Carnoy solution (4:3:1 chloroform, 90% ethanol, acetic acid) overnight at 4°C. Fixed males were then dissected on a siliconized coverslip in a drop of PFA:acetic acid fixative and then squashed using a microscope slide. Slides were then immediately submerged in liquid nitrogen and coverslips were popped off using a razorblade. Slides were then rehydrated and washed in 3 x 10min with 1X PBT (1X phosphate-buffered saline with 0.1% Triton-X 100).

4.3.3 Immunostaining

Slides were then submerged in 5% Normal Donkey Serum (NDS) in PBT blocking solution for 30 minutes. Primary antibodies were added to the fixed tissue on the microscope slide and mixed by placing a coverslip on the slide and gently moving up and down several times. The coverslip was left on the

slide and the tissue was incubated with primary antibody in a humidity chamber at 4°C overnight. Primary antibodies used in this study and their dilutions in 1X PBT with 5% NDS were: 1:50 rabbit anti-H3K9me3 (Active Motif) and 1:50 rabbit anti-H3K27me3 (Active Motif). After incubation in primary antibody, the coverslip was removed, and slides were washed 3 times at 10 minutes each with 1X PBT. The tissue was then stained with fluorescently conjugated secondary antibodies and Hoechst solution for 1hr at room temperature as described above. The secondary antibody used in this study was anti-rabbit Alexa Fluor 488 (at 1:100 dilution in 1X PBT with 5% NDS; Abcam). Hoechst solution was diluted at 1:1000 with 1X PBT. Tissues were washed as above and then left to air-dry for 10 minutes in the dark. A drop of 80% glycerol was then added to the slide and coverslip was positioned and sealed onto the slide with nail polish. Slides were stored at 4°C in the dark until imaging.

4.3.4 Confocal microscopy and image processing

Fluorescent microscopic imaging was conducted with a Leica TCS SPE-5 confocal microscope. Images were collected as Z-series for each laser channel and subsequently merged for visual capture of all features within the same nucleus that were not in the same focal plane. Merged images were exported as Lif files and then remerged and processed in ImageJ.

4.4 Results

4.4.1 The role of two key heterochromatin pathways in silencing the paternally inherited genome in PGE males

At the 7th cleavage division of embryogenesis, the paternally inherited chromosomes of PGE males become transcriptionally silenced through facultative heterochromatinization (Bongiorni and Prantera, 2001). Previous research shows that the H3K9me3-HP1 pathway plays a critical role in this silencing process (Bongiorni *et al.*, 2007). Immunocytological analyses of

male somatic tissue in my study further support these findings as H3K9me3 staining co-localises with the heterochromatic bodies found in male nuclei (Figure 2a). The H3K9me3-HP1 pathway is also suggested to interact with the H3K27me3-PRC2 pathway to regulate transcriptional silencing of genic and non-genic regions (Boros *et al.*, 2014). Therefore, I tested for the presence of H3K27me3 in the heterochromatic bodies of male nuclei and find that H3K27me3 staining also co-localises with this region (Figure 2b). The pattern of H3K27me3 staining is consistent with the pattern of H3K9me3 staining found in these cells, suggesting that both pathways may have a role in silencing the paternal genome through the formation of heterochromatin. Additionally, this supports previous suggestions of crosstalk between these pathways to regulate transcriptional silencing of chromosomal regions (Boros *et al.*, 2014; Jamieson *et al.*, 2016).

4.4.2 The role(s) of histone modifications H3K9me3 and H3K27me3 in recognition, elimination and imprinting of paternal chromosomes during spermatogenesis

Spermatogenesis in Planococcus citri: an overview

Male meiosis in *P. citri* is characterised by two specialised events: i) an inversion of the meiotic divisions in which the first division is equational and the second is reductional and ii) striking non-independent assortment of chromosomes in the second division where a monopolar spindle separates the maternally inherited chromosomes from those that are paternally-inherited (Bongiorni *et al.*, 2004). Each of the sperm cysts in male testes contains 16 primary spermatocytes, each of which undergoes the meiotic divisions described above. Following these meiotic divisions, each of the four haploid products of a single meiotic division contains either exclusively maternal or paternal alleles. These nuclei then form a quadrinucleate spermatid in which the two paternal nuclei are in a heterochromatic state and the two maternal nuclei are in a euchromatic state. The final product is a cyst containing 64 nuclei, 32 of which contain the maternal genome and elongate

into mature sperm. The other 32 nuclei contain the paternal genome and degenerate in situ (Brown and Nelson-Rees, 1961; Nur, 1962). Hoechst stained images of nuclei throughout these stages of spermatogenesis are shown in Figure 3.

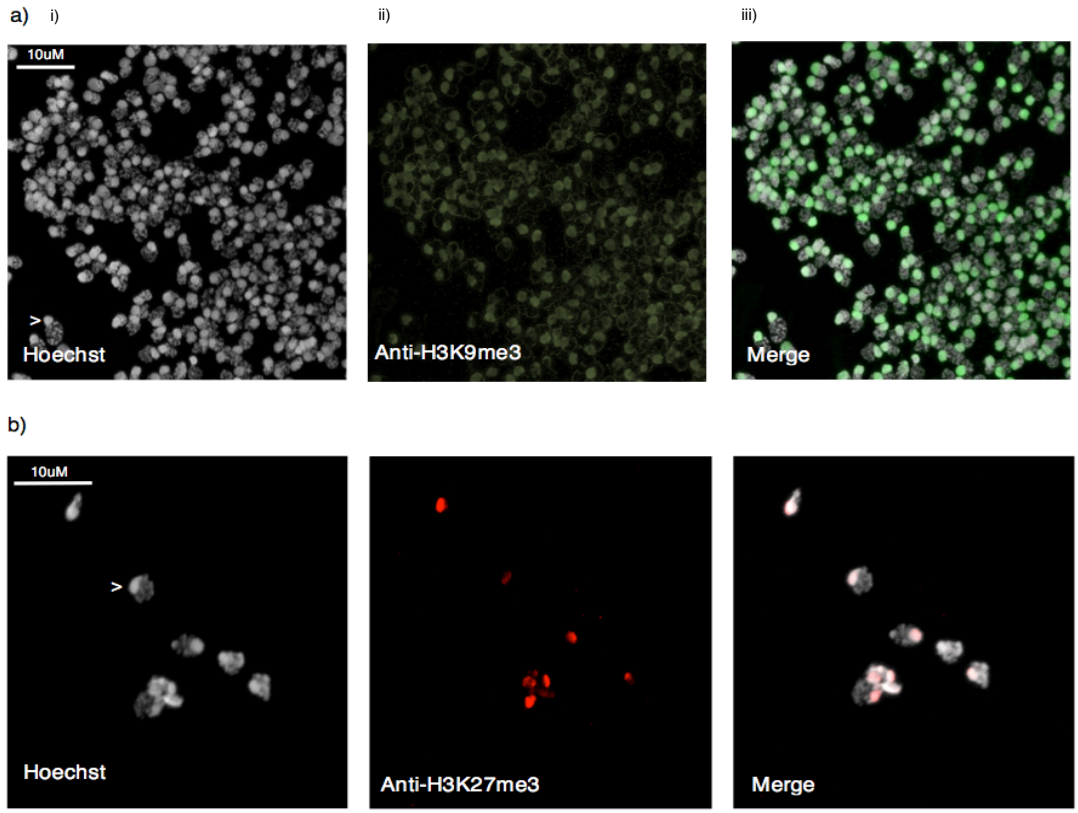


Figure 2: Histone modifications in male somatic nuclei. These images show somatic nuclei from male body tissue. The condensed paternally inherited chromosomes are visible in the nuclei as heterochromatic bodies; examples are labelled with arrowheads (>). Row a) i) shows nuclei stained with Hoechst (grey) and ii) anti-H3K9me3 (green) iii) merged image showing H3K9me3 staining co-localising at the heterochromatic bodies within male somatic nuclei. Row b) is as described for row a) but shows anti-H3K27me3 staining (red) in ii) and iii). H3K27me3 is also found to co-localise with heterochromatic bodies in male somatic nuclei.

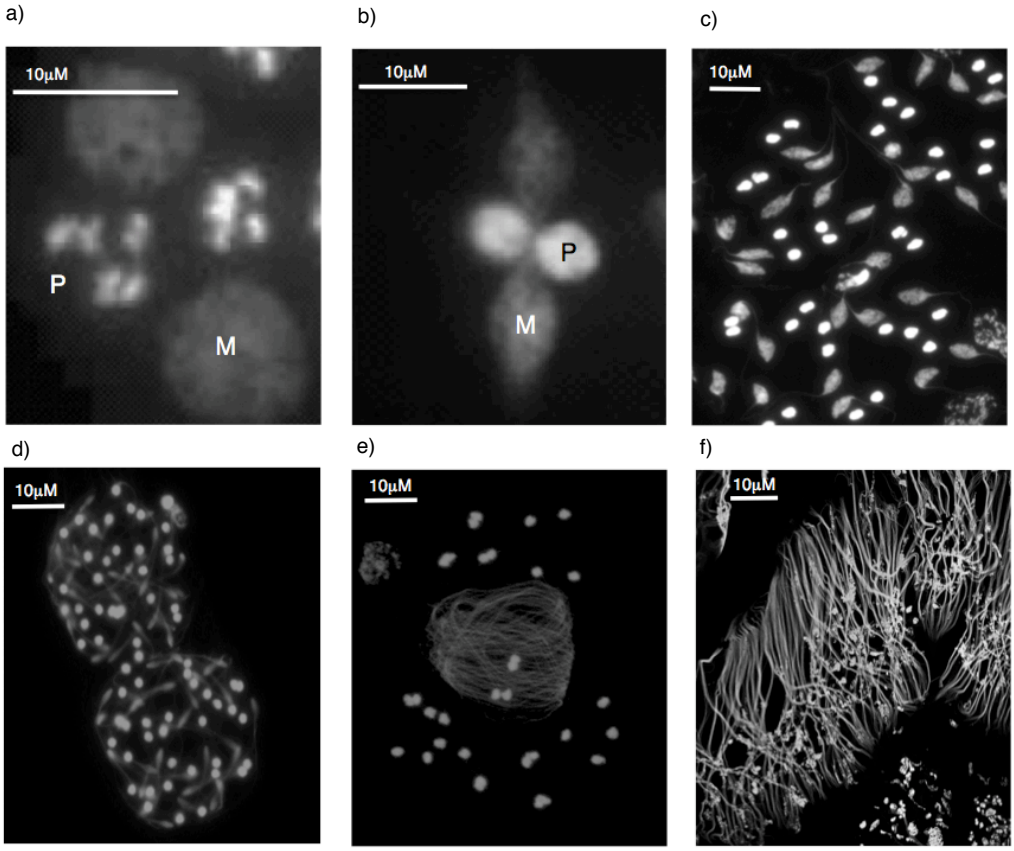


Figure 3: Stages of spermatogenesis in *Planococcus citri.* Hoechst stained images of nuclei during spermatogenesis in *P. citri*. P = paternal chromosomes, M = maternal chromosomes. Image a) shows the four haploid products of male meiosis. Paternal chromosomes are distinguishable from maternal chromosomes due to their high level of condensation, in which each individual chromosome can be seen. Maternal chromosomes form a diffuse haploid nucleus. Image b) shows the four haploid nuclei in quadrinucleate spermatid formation, where all nuclei produced from a single meiotic division share a cytoplasm. Nuclei containing paternal chromosomes are highly condensed compared to those containing maternal chromosomes. Image c) shows a squash of many quadrinucleate spermatids. Maternal nuclei are beginning to elongate into spermatids whilst paternal nuclei remain in a condensed chromatin state and do not undergo elongation. Image d) shows maternal spermatids continuing to elongate within the cyst, forming a 'ball of yarn' structure. Paternal nuclei remain suspended in a heterochromatic state. Image e) shows maternal spermatids forming long, thin sperm bundle structures whilst condensed male nuclei are pushed outwards of this structure. Image f) shows an open sperm cyst containing many sperm bundles. Individual sperm nuclei can be seen at the tips of bundle structures (marked by >).

4.4.3 Recognition and elimination of parental chromosomes during meiosis

Firstly, to investigate the role of H3K9me3 and H3K27me3 in the recognition of parental origin of chromosome, I analyse the germline of 2nd instar males in which the four haploid products of a single meiotic division are distinguishable. At telophase, the five condensed paternal chromosomes are visible and beginning to separate from the maternal nucleus, in which individual chromosomes can no longer be seen (Figure 3a). At this key stage, H3K9me3 is absent in both maternal and paternal chromosomes (Figure 4a). H3K9me3 staining in the surrounding somatic nuclei provide a control. In contrast, H3K27me3 is found to mark both the paternal and maternal chromosome sets (Figure 4b). These results suggest that H3K9me3 and H3K27me3 have different roles during this stage of meiosis. However, since in each case both parental sets of chromosomes have equivalent markings, these marks do not appear to function as a distinguishing feature of parental origin at this critical stage in meiosis.

In late 2nd instar males, the quadrinucleate stage of spermatogenesis is visible (Figure 3b). Here, the four haploid nuclei are found together within a shared cytoplasm and the different levels of chromatin condensation between the maternal and paternal nuclei is apparent. The maternal nuclei are in a euchromatic state and have a teardrop shape as they progress into the elongation process. The paternal nuclei are heterochromatic and remain circular, gradually getting smaller as degradation occurs. Due to the nature of tissue preparations, the quadrinucleate shape of these nuclei is disturbed but individual nuclei can be seen clearly (Figure 5). At this stage, H3K9me3 and H3K27me3 are found on both the euchromatic and the heterochromatic nuclei (Figure 5). This suggests that between telophase and formation of the quadrinucleate spermatid, H3K9me3 is acquired on both sets of nuclei. Once again, there is no difference in presence of histone modifications between maternal and paternal nuclei providing further evidence that these marks are

not involved in recognition of a nucleus' or chromosome's parental origin during spermatogenesis.

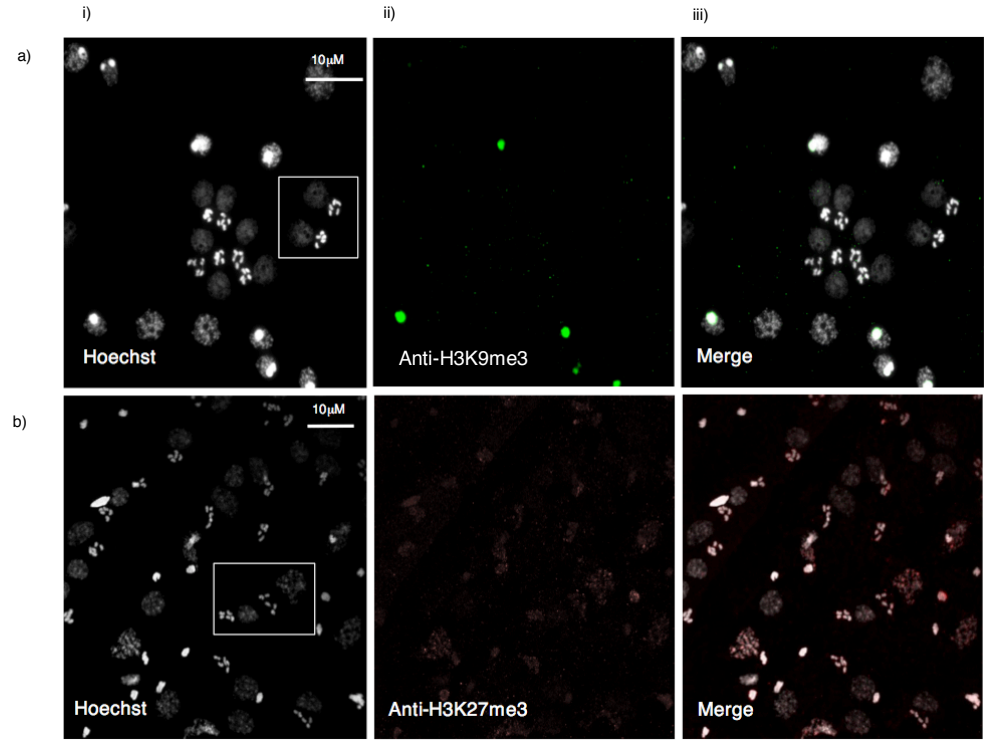


Figure 4: Histone modifications in sperm cyst of 2nd-instar male. These images show the four haploid products of meiosis present in sperm cysts dissected from 2nd-instar males. P = paternal chromosomes and M = maternal chromosomes. The four haploid products are clearly visible at this stage (examples in white boxes). Paternal chromosomes are distinguishable from female chromosomes due to their high level of condensation. Row a) shows chromosomes stained with i) Hoechst (grey) ii) anti-H3K9me3 (green) iii) shows a merged image, H3K9me3 staining (green) co-localises with heterochromatic bodies in the somatic tissues present on the slide but does not co-localise with maternal haploid nuclei or paternal chromosomes. Row b) is as described for row a) but shows anti-H3K27me3 staining (red) in ii) and iii). H3K27me3 co-localises with both maternal haploid nuclei and paternal chromosomes.

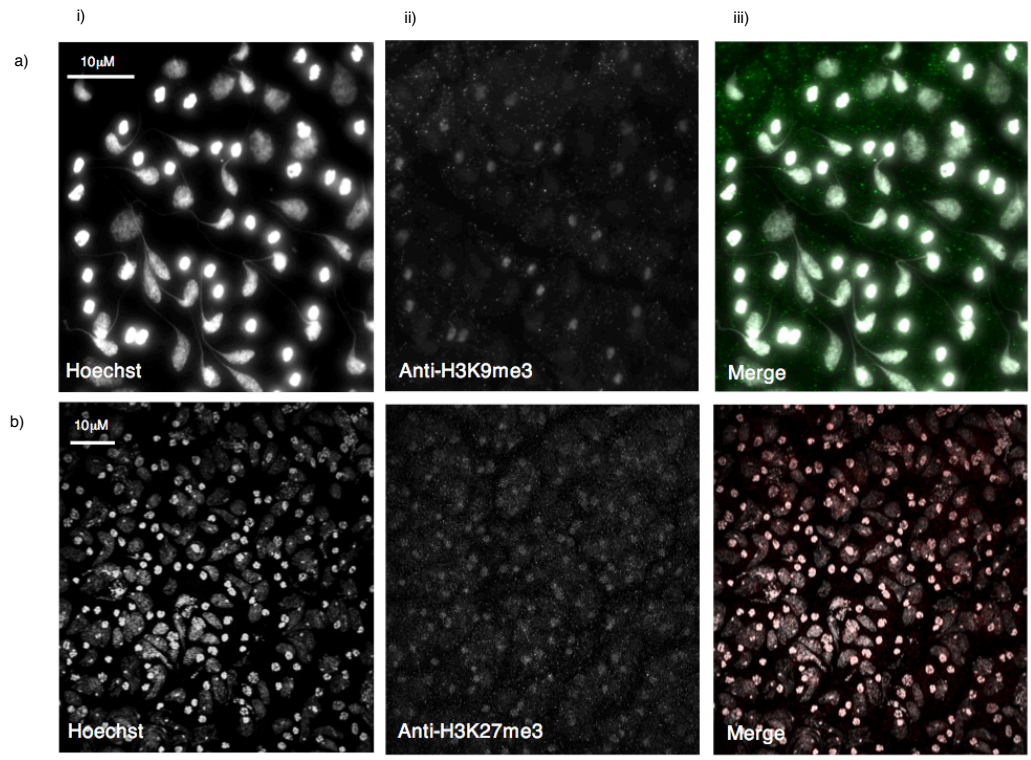


Figure 5: **Histone modifications in sperm cysts of late 2nd-instar male.** The above images show the quadrinucleate spermatid stage of spermatogenesis. Nuclei are dispersed across the slide due to squash fixation technique. Row a) shows i) Hoechst stained haploid nuclei, paternal nuclei are highly condensed and maternal nuclei are less condensed and beginning to elongate into spermatids; ii) shows H3K9me3 staining of nuclei (green); iii) shows a merged image, Hoechst in grey and H3K9me3 staining in green. H3K9me3 staining co-localises with all haploid nuclei. Row b) as described for a) but image ii) shows H3K27me3 (red) staining of nuclei; iii) shows a merged image, Hoechst in grey and H3K27me3 staining in red. H3K27me3 staining co-localises with all haploid nuclei.

Crucially, beyond this stage of spermatogenesis, H3K9me3 and H3K27me3 are no longer found on the euchromatic maternal nuclei but consistently mark the heterochromatic paternal nuclei (Figures 6). Here, whilst the nuclei containing the maternally inherited genome elongate, the paternal nuclei remain in a heterochromatic condition. The presence of these histone modifications on the paternal nuclei suggest that the same pathways involved in silencing the paternal genome in male somatic tissue, H3K9me3-HP1 and H3K27me3-PRC2, may also be involved in the elimination of the

paternal genome during spermatogenesis. The heterochromatic condition of these nuclei likely prevents their elongation into mature sperm cells and thus, they degrade in situ. However, further investigation is required to confirm this role. The loss of H3K9me3 and H3K27me3 from the maternal nuclei as they elongate might be a consequence of histone-to-protamine exchange, where histones are removed from sperm nuclei and replaced with smaller protamine molecules to allow efficient packaging of the nuclei for transfer into female (Braun, 2001). However, testing for absence of all histones (not just their specific modifications) and presence of protamines throughout spermatogenesis would be required to confirm this.

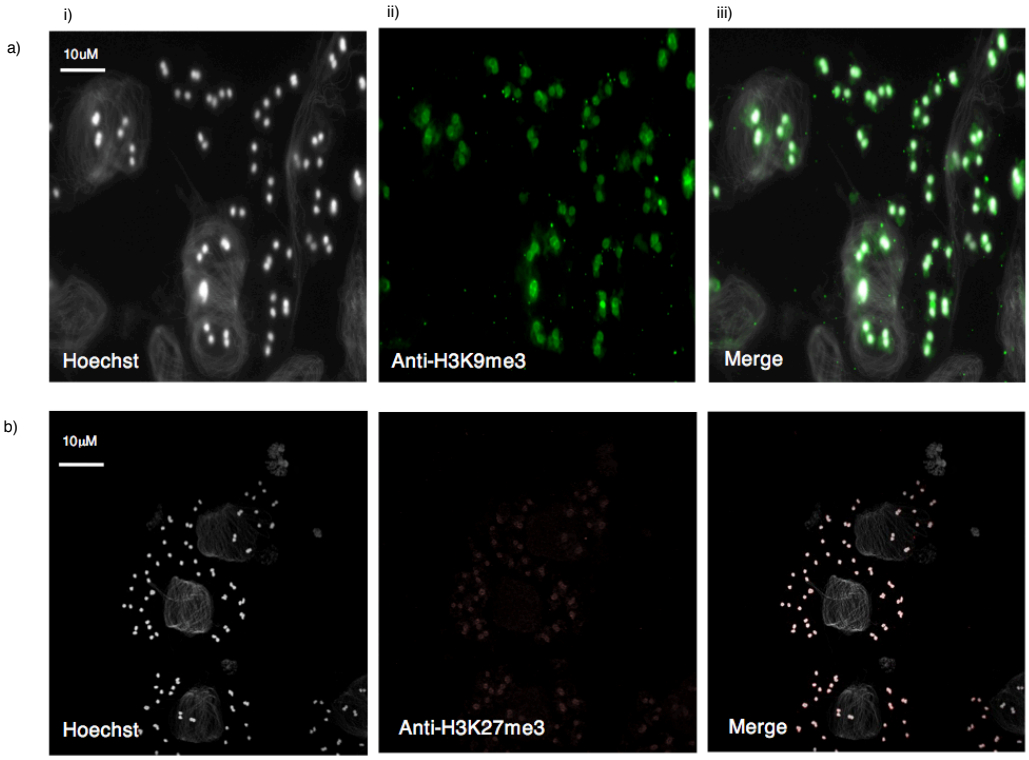


Figure 6: **Histone modifications in the sperm cysts of 3rd-instar males.** The above images show sperm cysts dissected from 3rd instar males. Elongating sperm nuclei containing maternal chromosomes are found in bundles in the maturing male. Heterochromatic nuclei containing paternal chromosomes do not elongate and form sperm bundle structures. Row a) i) shows Hoechst stained sperm nuclei and the heterochromatic nuclei containing the paternally inherited chromosomes; ii) shows anti-H3K9me3 (green) of nuclei; iii) shows the merged image, Hoechst in grey and H3K9me3 staining in green. H3K9me3 staining co-localises with heterochromatic male nuclei but is not found on the elongating maternal nuclei. Row b) is as described in a) but ii) shows anti-H3K27me3 staining (red) of nuclei; iii) shows merged image, Hoechst in grey and anti-H3K27me3 in red. H3K27me3 staining co-localises with heterochromatic male nuclei but is not found on the elongating maternal nuclei.

4.4.4 Role of histone modifications in imprinting and transgenerational epigenetic inheritance

Although H3K9me3 and H3K27me3 appear to not be involved in the mechanism that recognizes the parental origin of chromosomes during spermatogenesis, further analyses were conducted to investigate their role in imprinting. Contrary to previous results, I find that neither H3K9me3 nor H3K27me3 are present on the sperm nuclei in the cysts of male pupae (Figures 7 & 8). The long, thin nuclei that extend through almost the entire length of the sperm cells lack any staining associated with these histone modifications. Instead, H3K9me3 and H3K27me3 staining are found in the cytoplasm surrounding these nuclei. These histone modifications appear to be located at the 'tips' of the elongating sperm; however, this is most likely due to the fact that these are squashed tissue preparations. Sperm cysts form a skein like shape as they elongate through the testes of the developing male and the heterochromatic nuclei (containing the paternally inherited chromosomes) are pushed to the outer surface (Nur, 1962). As the sample is squashed, the heterochromatic nuclei are pushed to the edges of the sperm bundle, resulting in the staining pattern observed. In this study, confocal microscopy allows a clear 3-dimensional image of the sperm nuclei within the cyst therefore antibody staining in the cytoplasm can be distinguished from staining that co-localises with the nuclei. Although histones are not generally associated with cytoplasm, in this case it can be concluded that the histone modifications present in the cytoplasm of Figures 7 and 8 were once associated with the now degraded heterochromatic paternal nuclei. Due to the suppression of cytokinesis at the end of meiosis II, the four haploid products of meiosis (2 euchromatic nuclei containing maternal chromosomes and 2 heterochromatic nuclei containing paternal chromosomes) share a cytoplasm. Thus, these histone modifications are still present in the cyst cytoplasm as the euchromatic nuclei elongate and mature.

Overall, these findings show that although H3K9me3 and H3K27me3 appear to be involved in the silencing that occurs in male somatic nuclei and the elimination of the paternal genome during spermatogenesis, they do not act as the paternal imprint. The presence of these histone modifications throughout spermatogenesis is summarised in Figure 9.

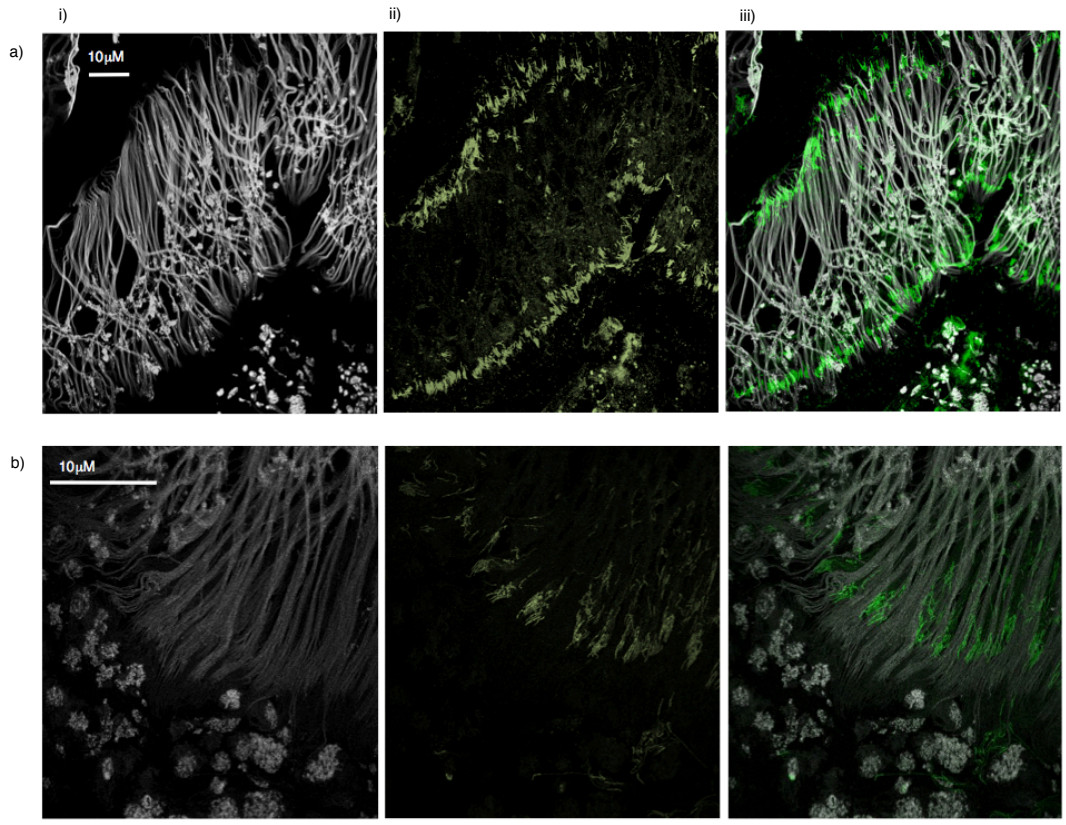


Figure 7: **Histone modification H3K9me3 in sperm cysts of male pupa.** The above images show H3K9me3 staining in sperm cysts dissected from male pupa. Row a) i) shows Hoechst stained sperm nuclei in the sperm cyst; ii) shows H3K9me3 staining (green) localised towards the end of the sperm bundles within the cyst; iii) shows the merged image, Hoechst in grey and H3K9me3 staining in green. H3K9me3 staining does not appear to co-localise with the sperm nuclei and is found in surrounding cytoplasmic regions lacking nuclear staining. Row b) is as described for row a) but at a higher magnification in which individual sperm nuclei can be visualised towards the tips of the bundles.

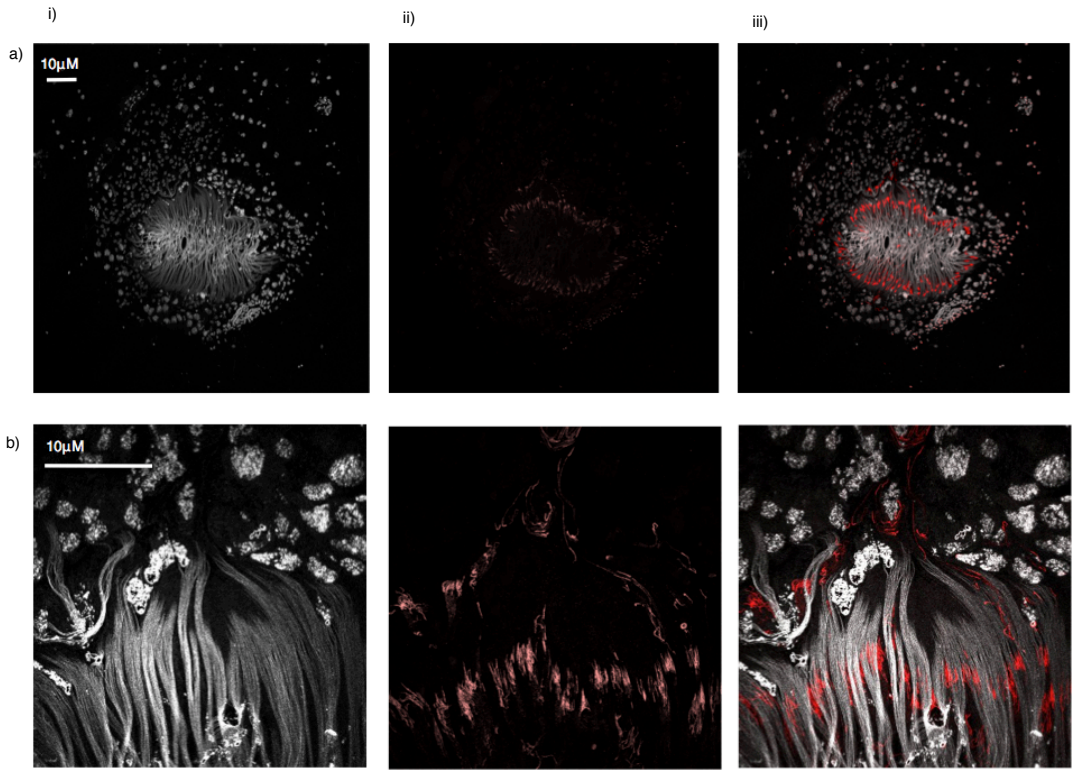


Figure 8: **Histone modification H3K27me3 in sperm cysts of male pupa.** The above images show H3K27me3 staining in sperm cysts dissected from male pupa. Row a) i) shows Hoechst stained sperm nuclei in the sperm cyst; ii) shows H3K27me3 staining (red) localised towards the end of the sperm bundles within the cyst; iii) shows the merged image, Hoechst in grey and H3K27me3 staining in red. H3K27me3 staining does not appear to co-localise with the sperm nuclei and is found in surrounding cytoplasmic regions lacking nuclear staining. Row b) is as described for row a) but at a higher magnification in which individual sperm nuclei can be visualised towards the tips of the bundles.

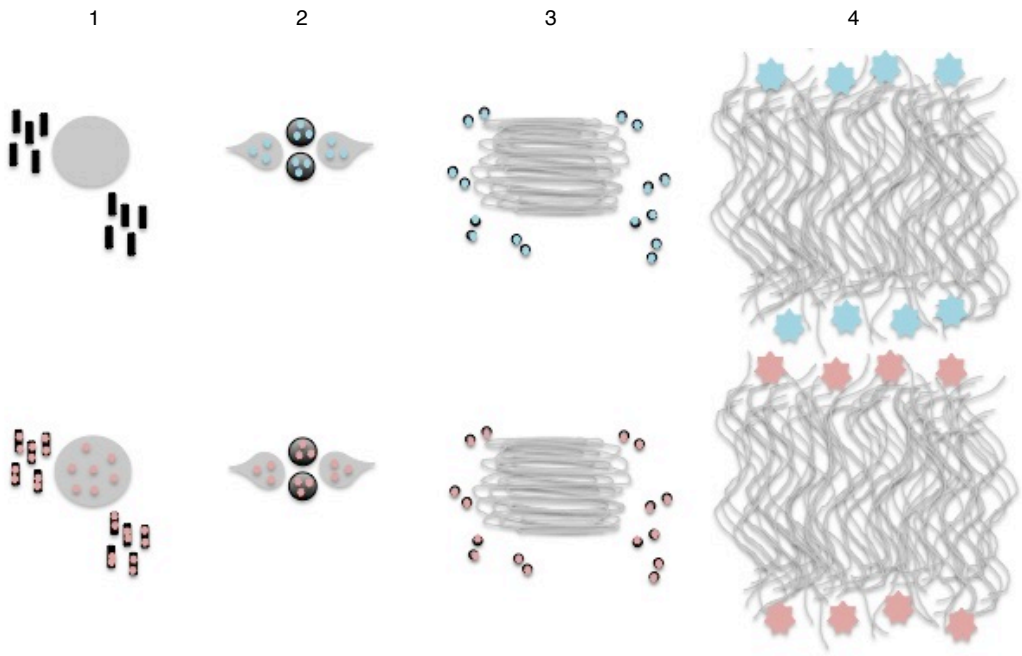


Figure 9: Summary of histone modifications, H3K9me3 and H3K27me3, throughout *P. citri* spermatogenesis. Blue stars indicate H3K9me3 staining; pink stars indicate H3K27me3 staining. Stars are for visualisation only and do not represent levels of histone modifications found in nuclei. At stage 1, where the four haploid products of meiosis are visible (nuclei containing maternal chromosomes and the condensed paternal chromosomes), H3K9me3 staining is absent from all nuclei and chromosomes. H3K27me3, on the other hand, is present on both maternal nuclei and individual paternal chromosomes. At stage 2, the quadrinucleate spermatid stage, both histone modifications are present on paternal and maternal haploid nuclei. At stage 3, both histone modifications are lost from the maternal nuclei, which are now elongating to form mature sperm cells. Both H3K9 and H3K27 tri-methylation remain on the heterochromatic paternal nuclei. At stage 4, where sperm bundles are clearly visible within the testes of pupa, both histone modifications remain absent from the maternal nuclei, which are now elongating and forming bundles. H3K9me3 and H3K27me3 are found towards the edges of the sperm bundles, but do not co-localise with the sperm nuclei. These may be the remnants of the heterochromatic nuclei, which are no longer visible but did share a cytoplasm with maternal nuclei.

4.5 Discussion

Paternal Genome Elimination is a striking example of genomic imprinting and parent-of-origin specific gene expression and behaviour. Whilst these phenomena and their molecular mechanisms are well studied in mammals and flowering plants, little is known about how this process is regulated in insects. Facultative heterochromatin plays a critical role in the silencing of the paternal genome that occurs in PGE males (Bongiorni *et al.*, 2007). Furthermore, histone modifications involved in heterochromatin formation are also suggested to play a role in the elimination and imprinting of the paternal genome during spermatogenesis (Buglia and Ferraro, 2004; Bongiorni *et al.*, 2009). In this study, I perform immunocytological analyses of two key heterochromatin pathways, H3K9me3-HP1 and H3K27me3-PRC2, in *Planococcus citri* to elucidate their involvement in PGE. These pathways are crucial for the formation of heterochromatin and evidence supports their co-localisation and co-operation in gene silencing (Boros *et al.*, 2014).

My findings support earlier work showing evidence for the involvement of the H3K9me3-HP1 pathway in the silencing of the paternal genome in males (Bongiorni *et al.*, 2007). However, the role of the H3K27me3-PRC2 pathway – a key regulator of facultative heterochromatin – has never been tested until now. I show that H3K27me3 does indeed co-localise with the heterochromatinized paternal genome in male somatic nuclei, suggesting involvement of the H3K27me3-PRC2 pathway in the silencing of the paternal genome during PGE, although further testing is required to confirm a causative role. This result does, however, lend support to the hypothesis that H3K9me3-HP1 and H3K27me3-PRC2 heterochromatin pathways interact with one other to suppress transcription of large chromosomal regions (Boros *et al.*, 2014; Jamieson *et al.*, 2016). These findings also highlight striking parallels between the role of histone modifications in silencing the paternal genome during PGE and their involvement in X chromosome inactivation in

mammals. The inactive X chromosome is associated with methylation of histones H3K27 and H3K9 and evidence suggests that the H3K27me3-PRC2 pathway may be involved in initiation and/or maintenance of this inactive state (Heard, 2004). Additionally, both the inactive X chromosome in mammals and the paternal genome in PGE males are capable of developmental reversibility, a hallmark feature of epigenetic phenomena (Nur, 1967). These parallels strongly suggest that similar epigenetic mechanisms, namely histone modifications, underlie the regulation of chromosome condensation and silencing in both cases.

Although involved in the silencing of the paternal genome in somatic tissue, the role of histone modifications in the recognition and elimination of paternal chromosomes during spermatogenesis remains inconclusive. The suggestion that H3K9me3 is carried on the sperm nuclei and acts as molecular marker to distinguish parental origin of chromosomes is contentious (Buglia and Ferraro, 2004; Bongiorni *et al.*, 2009). This is due to limited evidence and to the fact that a process of histone to protamine exchange occurs during spermatogenesis, where the majority of histones are removed from sperm nuclei and replaced with smaller molecules called protamines to allow compaction of the nuclei (Hecht, 1989). However, not all histones are removed and therefore the presence of histones on *P. citri* sperm cannot be ruled out.

Contrary to previous studies, I find that H3K9me3 is not present on sperm nuclei found in *P. citri* male pupae. Furthermore, H3K27me3 is also absent from these sperm nuclei. Confocal microscopy of sperm cysts dissected from pupae shows clearly that although these modifications are present within the cyst cytoplasm, they do not co-localise with the elongating sperm nuclei and are likely the remnants of heterochromatic paternal nuclei that fall to form mature sperms after meiotic divisions. At the pupal stage, spermatogenesis is complete although the elongation process continues until pupae develop

into adult males (Nur, 1962). Since only a single mature sperm cell enters the ooplasm at fertilisation, these histone modifications would not be transferred to the oocyte. Therefore, these specific histones do not act as the imprint that identifies the paternal genome within the embryo. Indeed, this does not rule out the role of histones and their modifications completely and further testing is required to fully elucidate the histone and protamine complement in sperm nuclei.

The role of H3K9me3 and H3K27me3 in the recognition of chromosome parental origin during spermatogenesis is also ruled out. I find that H3K9me3 is absent from both the paternal and maternal chromosomes at telophase whilst H3K27me3 is present on both. At this stage of meiosis, it appears that these two histone modifications differ in their functions; however, it is clear that neither is involved in distinguishing the parental origin of chromosomes as in each case both parental sets are marked equivalently. At the quadrinucleate stage, both H3K9me3 and H3K27me3 are present on the euchromatic and heterochromatic nuclei suggesting a role in regulation of gene expression within the maternal nuclei and a role in facultative heterochromatinization of the paternal nuclei. Interestingly, as spermatogenesis progresses and the euchromatic maternal nuclei are elongating into spermatids, both H3K9me3 and H3K27me3 are associated with the heterochromatic paternal nuclei but are absent from the euchromatic maternal nuclei. This strongly suggests a role for these histone modifications in the heterochromatinization of the paternally inherited nuclei in late spermatogenesis.

Therefore, I propose that an alternative role of H3K9me3 and H3K27me3 during spermatogenesis in *P. citri* is to prevent the nuclei carrying paternally inherited chromosomes from completing the elongation process required to develop into mature sperm. Thus, potentially playing a crucial role in the process of PGE. Indeed, studies of supernumerary B chromosomes in other

PGE species (Nur, 1962; Nur and Brett, 1988) suggest that chromatin state plays a crucial role in the elimination of the paternal genome. B chromosomes can escape elimination during spermatogenesis by decondensing and segregating with maternal alleles. However, in certain genetic backgrounds, B chromosomes are unable to decondense and remain heterochromatic. In this state, they are eliminated along with the paternal alleles (Nur and Brett, 1988). Thus, being heterochromatic appears to be necessary to ensure elimination. Indeed, heterochromatinization of the paternal genome also appears to play a mechanistic role in diaspidid elimination in which paternal chromosomes are eliminated from somatic tissues (Brown, 1965). A similar example of histone modification involvement in chromosome elimination was described recently in the haplodiploid insect, *Nasonia vitripennis* (Aldrich *et al.*, 2017). The *N. vitripennis* male genome possesses a non-essential, selfish B chromosome called Paternal Sex Ratio (PSR) which distorts sex ratio by eliminating paternally derived inherited chromosomes during early embryogenesis (Nur *et al.*, 1988). PSR alters patterns of histone modifications found in paternal chromosomes and disrupts the chromatin remodelling process during early embryonic development. Consequentially, paternal chromosomes are eliminated from the embryo whilst PSR remains unaffected (Aldrich *et al.*, 2017).

My study shows that histone modifications H3K9me3 and H3K27me3 are likely involved in the heterochromatinization of the paternal genome during embryonic development and spermatogenesis. Further testing is required to confirm their functional role in the elimination of the paternal complement during the latter process. I also show that these histone modifications are not carried on the mature sperm nuclei, ruling out their role as the molecular identifier of paternally-inherited chromosomes after fertilisation. Thus, the mechanism by which the parental origin of chromosomes is distinguished in PGE species remains elusive.

Chapter 5: Heterochromatin genes in species with Paternal Genome Elimination

5.1 Chapter summary

Heterochromatin is known to play a critical role in the silencing of the paternally inherited chromosomes during the process of Paternal Genome Elimination in the citrus mealybug, *Planococcus citri*. In this chapter, I identify and characterise key genes in the evolutionarily conserved H3K9me3-HP1 heterochromatin pathway: The Heterochromatin Protein 1 (*HP1*) gene family and histone methyltransferase *SU(VAR)3-9* genes. I study expression profiles of these genes throughout development in both sexes and evaluate their similarities to extensively studied *HP1* family and *SU(VAR)3-9* genes in *Drosophila*.

5.2 Introduction

Heterochromatin is a highly conserved component of the eukaryotic genomic architecture and comprises a large proportion of the genome in many metazoans (Ho *et al.*, 2014). It is characterised by a highly condensed chromatin structure, which plays an important role in the regulation of gene expression and chromosome segregation (Vermaak and Malik, 2009). There are two classes of heterochromatin, both of which are associated with these functions: i) constitutive heterochromatin, which remains condensed throughout the cell cycle and an organism's development, and is often characterised by highly repetitive sequence structure (Dorer and Henikoff, 1994) and ii) facultative heterochromatin, which is developmentally regulated, allowing for specific regulation of chromatin condensation throughout development and between cell types (Heard and Disteche, 2006). A striking

example of gene regulation through facultative heterochromatin is the inactivation of the X chromosome in mammals (Lyon, 1961; Heard and Disteche, 2006). Female mammals possess twice as many X chromosomes as male mammals. To compensate for differences in X-linked dosage, one of the X chromosomes in females is inactivated through increased compaction of DNA and formation of facultative heterochromatin (Barr and Carr, 1962; Heard and Disteche, 2006).

However, decades before its discovery in mammals, the silencing of chromosomes through facultative heterochromatinization was first described in insects in a process known as Paternal Genome Elimination (PGE) (Brown and Nelson-Rees, 1961). PGE is a genomic imprinting phenomenon found in thousands of insect species, including the citrus mealybug, *Planococcus citri*. In this species, PGE involves the silencing and elimination of an entire haploid genome in a parent-of-origin specific manner (Hughes-Schrader, 1948). Both sexes develop from fertilized eggs and initially possess a diploid euchromatic chromosome complement. However, in males the entire paternally inherited haploid chromosome set becomes heterochromatic in early embryos, whilst the maternally inherited complement remains euchromatic. The heterochromatic state of the paternal genome is maintained throughout the lifetime of the male and can be visualised as heterochromatic bodies within the somatic nuclei. Furthermore, the paternal chromosomes remain in this heterochromatic state throughout male meiosis and do not give rise to mature sperm (Nur, 1962; Chapter 3) Thus, *P. citri* males only express and transmit maternally inherited chromosomes. Females, on the other hand, do not undergo the process of PGE and both maternally and paternally-derived chromosomes remain euchromatic throughout development.

The facultative heterochromatinization of the paternal genome in *P. citri* is mediated by an evolutionarily conserved heterochromatin pathway involving

the Heterochromatin Protein 1 (*HP1*) gene family and the histone modification H3K9me3 (Bongiorni *et al.*, 2007; Chapter 3). Across all studied taxa, *HP1* genes contain two functionally important domains. The first is an N-terminal chromodomain (CD), which binds specifically to methylated Histone H3 at Lysine 9 (Bannister *et al.*, 2001). The second domain is a C-terminal chromoshadow domain (CSD), which interacts with other non-histone proteins (Li *et al.*, 2003) and facilitates HP1 interactions with SU(VAR)3-9, a histone methyltransferase that specifically methylates histone H3 at lysine 9 (Schotta *et al.*, 2002). A less conserved 'hinge' region, also involved in protein targeting, links these two domains (Smothers and Henikoff, 2001). The H3K9me3-HP1 pathway described above generates a positive feedback loop that maintains the higher order state of heterochromatin (Schotta *et al.*, 2002; Ebert *et al.*, 2004). This pathway was shown to have a causative role in the silencing of the paternal genome during PGE when dsRNA knockdowns of an HP1 homolog in *P. citri* embryos resulted in a loss of heterochromatinization of the paternal genome (Bongiorni *et al.*, 2007). Furthermore, it is hypothesised that the heterochromatic state of the paternally inherited chromosomes prevents them from successfully completing spermatogenesis (Chapter 3). This implicates heterochromatin and *HP1* in the elimination of the paternal genome. However, despite the crucial role of heterochromatin in the silencing, and potentially the elimination, of the paternal genome in PGE males, very little is known about *HP1* genes in *P. citri* and their involvement in the establishment and maintenance of heterochromatin.

The *HP1* gene family is phylogenetically and functionally diverse both within species and across taxa (Vermaak and Malik, 2009). The first HP1 protein to be described was found in *Drosophila melanogaster* (James and Elgin, 1986) and since then the majority of information regarding the *HP1* gene family and their functions comes from studies conducted on the *Drosophila* genus. The *HP1* gene family is rapidly evolving with a dynamic pattern of gene losses

and gains of lineage specific paralogs (Levine *et al.*, 2012). A number of new *HP1* paralogs were discovered in the *Drosophila* genus increasing the total number from five to ten (Levine *et al.*, 2012). Additionally, sixteen partial *HP1* genes, containing either a CD or a CSD, have been described (Vermaak and Malik, 2009; Levine *et al.*, 2012). The first five paralogs of *HP1* to be discovered in *Drosophila* have been extensively studied. All contain both chromodomains and chromoshadow domains but differ in their localisation and function. HP1A is strongly associated with heterochromatic regions (James *et al.*, 1989), but also has a role in telomere capping and euchromatic gene regulation (Fanti *et al.*, 2003); HP1C is associated with transcriptionally active regions of euchromatin (Greil *et al.*, 2003) and HP1B localises to both heterochromatic and euchromatic regions (Font-Burgada 2008; Smothers and Henikoff, 2001). Finally, HP1D and HP1E are expressed preferentially in the male and female germline, respectively, of *Drosophila melanogaster* (Vermaak, Henikoff and Malik, 2005; Levine and Malik, 2015). An *HP1* homolog isolated from *P. citri, PCHET2,* is 57% identical to the *D. melanogaster HP1A* chromodomain at the amino acid level (Epstein, James and Singh, 1992) and is predominantly localised to heterochromatic bodies in male somatic tissues (Bongiorni and Prantera, 2001). Although one *HP1* homolog is identified, the *HP1* gene family in *P. citri* remains uncharacterised. An understanding of the diversity of this gene family is the first step towards elucidating its role in PGE.

Analyses of *Drosophila* suggest that *HP1* genes both within a *Drosophila* genome and across the genus are playing multiple chromatin-based roles (Levine *et al.*, 2012; Levine and Malik, 2013). Functions and expression patterns of *HP1* genes can vary between species and, interestingly, some have key roles in meiotic drive and sex determination: In *Drosophila simulans*, *HP1D2*, an *HP1* paralog that has been lost in *D. melanogaster,* highlights a central role for chromatin and chromatin modifiers in sexual conflict. *HP1D2* gene expression is testes restricted and involved in the

distortion of offspring sex ratio through the Paris sex ratio system (Helleu *et al.*, 2016). The sex-ratio allele of *HP1D2* is located on the X chromosome and produces a protein that localises to the Y chromosome, disrupting its segregation during meiosis II (Cazemajor, Joly and Montchamp-Moreau, 2000; Helleu *et al.*, 2016). Consequentially, males cannot produce sperm containing Y chromosomes and so the sex ratio of a male's progeny is female-biased. This finding confirms a role for members of the *HP1* gene family in genetic conflict and meiotic drive.

The process of PGE is a striking example of genetic conflict and meiotic drive in which chromatin modifications mediated by *HP1* genes appear to play a role. Understanding the diversity of the *HP1* gene family and elucidating the role of the H3K9me3-HP1 pathway in *P. citri* is key to understanding the epigenetic mechanisms which underlie the following processes in PGE: i) the establishment and maintenance of paternal genome heterochromatinization in males and ii) the role of heterochromatin in preventing the paternal genome from completing spermatogenesis, resulting in its elimination from the germline.

In order to elucidate the role of the *HP1* gene family in PGE, I begin by characterising the *HP1* genes present in the *P. citri* genome. As PGE is a sex-specific process in which the paternal chromosomes of males are specifically targeted, I study the expression profile of putative *HP1A* homolog, *PCHET2*, in both sexes to identify patterns of male-specific expression. In addition to sex-specificity, the two key events in PGE occur at distinct stages of development: 1) the silencing of the paternal genome begins during embryogenesis and is maintained throughout a male's life and 2) the elimination of the paternal genome occurs during spermatogenesis in juvenile males. In order to elucidate the role of *PCHET2* in these separate processes, I analyse expression levels of this gene at various key stages of development. Additionally, to further investigate the involvement of the

H3K9me3-HP1 heterochromatin pathway in PGE, I identify the histone methyltransferase gene *SU(VAR)3-9* in the *P. citri* genome and analyse its expression throughout development in males and females to identify sex-specific and development-specific patterns.

5.3 Methods

5.3.1 Insect Husbandry

The citrus mealybug (*Planococcus citri)* is a tractable laboratory model with a short generation time and ease of rearing. Insects were cultured on sprouting potatoes in sealed plastic bottles at 25°C and ~70% relative humidity. Under these conditions, *P. citri* has a generation time (time from oviposition until sexual maturity) of approximately 30 days. Experimental isofemale lines (CP1-2) originate from natural populations in Israel and were reared in the laboratory under a sib-mating regime. In each generation, one mated female is taken from culture and transferred to a new container to give rise to the next generation.

5.3.2 Heterochromatin Protein 1 family in Planococcus citri

I use a similar approach to that used in a recent genomic study of *HP1* family genes across Hymenopteran species (Fang, Schmitz and Ferree, 2015). *HP1* orthologs were identified by BLAST-searching concatenated amino acid sequences of the chromodomains (CD) and chromoshadow domains (CSD) of five *HP1* gene family members from *Drosophila melanogaster* against the *P. citri* genome (mealybug.org, version v0). The default tBLASTn settings on Geneious 8.1.5 were used for these searches. The domains of each gene with an E-value of 0.1 or lower were then BLAST-searched against the *D. melanogaster* reference genome (NCBI). A *P. citri* gene was considered to belong to the *HP1* family if it fulfilled the following criteria:

(i) Best matched a *D. melanogaster HP1* gene

(ii) Retrieved the same *P. citri* gene used for the original query when re-BLAST-searched against the *P. citri* genome

The CDs and the CSDs from the newly identified *P. citri* sequences were then added to the *D. melanogaster* concatenated sequences to further BLAST-search against the *P. citri* genome. This increased chances of discovering all likely matches. This iterative BLAST searching approach was repeated until no new *HP1* genes were found in the *P. citri* genome. My search did not include querying with the less conserved hinge and tail regions of *HP1* genes, which lie outside the CDs and CSDs. Thus, it should be acknowledged that this study could have missed some *HP1* family genes in the *P. citri* genome.

In line with previous *HP1* gene family studies (Levine *et al.*, 2012; Fang, Schmitz and Ferree, 2015), any gene containing both a CD and a CSD was considered to be a full *HP1* gene. Genes with only a CSD were considered partial *HP1* genes because this domain is unique to *HP1* genes (Aasland and Stewart, 1995). Since CDs are present in a number of non-HP1 genes that perform chromatin-related functions (Messmer *et al.,* 1992; Woodage *et al.,* 1997), genes possessing only a CD domain were only considered to be partial *HP1* genes if they best matched CDs from known *HP1* genes and not CDs from other chromatin genes such as polycomb and chd-1.

5.3.3 Phylogenetic analyses of HP1 genes

HP1 gene sequences from a variety of taxa were collected from the NCBI database in order to assess phylogenetic relationships with newly identified *P. citri HP1* genes (Table S1). Amino acid sequences for chromodomains and chromoshadow domains of *HP1* homologs in *P. citri*, *Drosophila melanogaster*, *Acrythosiphon pisum, Caenorhabditis elegans, Drosophila yakuba, Drosophila simulans, Ceratitus capitata, Folsomia candida and Mus musculus* (outgroup) were extracted using NCBI Conserved Domain website

(https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi; Table S2). Amino acid sequences were aligned using MAFFT with default parameters (Katoh, Rozewicki and Yamada, 2017). These sequences were used as the dataset for tree building. Phylogenetic trees were generated using MrBayes v3.2.5 (Ronquist and Huelsenbeck, 2003). The amino acid model parameter was set to 'mixed', where the Markov chain samples each model according to its probability. The MCMC chains were run for 1,000,000 generations. Trees were annotated using Figtree (http://tree.bio.ed.ac.uk/software/figtree/).

5.3.4 Identification of SU(VAR)3-9 genes in P. citri genome

Amino acid sequences of *SU(VAR)3-9* genes from *Acrythosiphon pisum* (NP_001119634.2), *Apis mellifera* (NP_001035367.1), *Drosophila melanogaster* (NP_524357) and *Lepeophtheirus salmonis* (CDW38775.1) were collected using the NCBI database and used to BLAST search against the *P. citri* genome (mealybug.org, version v0). tBLASTn searches were carried out in Geneious R8.1.5 (Kearse *et al.*, 2012) using default settings. Hits were discarded if they did not meet the following threshold values: E-value ≤ 1e-10 and query coverage ≥ 50%. *P. citri* sequences meeting threshold criteria were then used as queries in BLAST-searches against the NCBI database (NCBI Resource Coordinators, 2016) to identify the presence of the conserved domains, SET (IPR001214) and Pre-SET (IPR007728), and relevant orthologs from other species. A *P. citri* gene was considered to be a full, functional *SU(VAR)3-9* homolog if it fulfilled the following criteria: (1) top NCBI blast hit was the relevant gene in another species and (2) contained SET and Pre-SET conserved domains required for functionality. Protein sequence alignments of these conserved domains were produced using MAFFT (Katoh, Rozewicki and Yamada, 2017) and visualised in Jalview (Waterhouse *et al.*, 2009).

5.3.5 RNA extraction

Total RNA was extracted from genetically inbred male and female *P. citri* at key developmental stages: male- and female-biased embryos, 3rd instar males and females, adult males, virgin and mated females (Table 1). Females' first broods are known to be male-biased while embryos laid on day 3 are female-biased (Ross *et al.*, 2010, 2011). As there is no way to identify the sex of an embryo without destroying it, we used first broods as a proxy for male embryos and third-day broods as a proxy for female embryos. These stages were chosen as they represent key developmental stages in *P. citri*: in male embryos silencing of the paternal genome has occurred but meiosis does not begin until the 2nd-instar; in female embryos there is no silencing of the paternal genome and the first stage of meiosis is underway; in 3rd instar males late spermatogenesis is occurring, which is completed by the adult stage. In virgin females, meiosis remains incomplete as sperm entry is required to trigger meiosis II and in mated females embryos are developing (Nur, 1962). To avoid pot-effects, each sample included insects from at least 3 breeding bottles. 10 biological replicates were prepared for each sample where 1 biological replicate equals: 10 females (for mated females, virgin females, 3rd instar females); 20 males (for adult males and 3rd instar males); 3 egg masses (for embryo groups). RNA extraction was performed using TRIzol® reagent (Thermo Fisher Scientific, USA) according to the manufacturer's instructions. RNA was treated with DNase to remove any DNA (Thermo Scientific DNase I, RNase-free kit) according to manufacturer instructions. Quantity and quality of extracted genetic material was assessed using NanoDrop ND-1000 Spectrophotometer (Thermo Scientific, USA) and Qubit (Thermo Fisher Scientific, USA) assays. A260/A280 and A260/A230 ratios were calculated for all samples and only samples with A260/A280 of 1.7 > 2.0 and A260/A230 of >1.0 were processed.

5.3.6 Reverse Transcriptase and cDNA preparation

RNA was reverse transcribed using M-MLV RT and random hexamers according to manufacturer's (Promega, USA) instructions. Negative RT controls for use in qPCR were also prepared. In total, 10 cDNA samples for each group were prepared.

5.3.7 PCR validation of HP1 genes in Planococcus citri

RT-PCR was used to confirm expression of *HP1-like* transcripts in adult males and females. For primer details see Table S2. RT-PCR was performed on cDNA using two biological replicates for each group. Negative controls were used to identify contamination and primer dimer. Products were amplified in 25uL reactions using MyTaqTM Red PCR mix (Bioline, UK) under the following cycling conditions: [1] 1 min at 95°C for initial denaturation, [2] 35 cycles of 15 seconds at 95°C, 15 seconds at 65°C and 10 seconds at 72°C and then [3] 5 mins at 72°C for final extension.

5.3.8 Identification of RP49 and RP17 housekeeping genes in Planococcus citri

Commonly used reference housekeeping genes, Ribosomal Protein 49 (*RP49*) and *RP17*, were identified in the *P. citri* genome using the methods described in Chapter 2. *RP49* has been used in qPCR analysis on the Japanese mealybug, *P. kraunhiae* and other Hemiptera (Sugahara *et al.*, 2017) and *RP17* has been used to normalize gene expression data in *P. citri* (Duncan *et al.*, 2014). The expression stability of both housekeeping genes between *P. citri* groups was validated.

5.3.9 qRT-PCR

Quantitative real-time PCR was carried out to investigate expression levels of *HP1* and *SU(VAR)3-9* genes in both sexes throughout development. The *HP1* sequence *PCHET2* was used for qPCR as knocking down this gene in

P. citri embryos coincides with the loss of heterochromatic bodies and associated heterochromatin marker, H3K9me3 (Bongiorni *et al.*, 2007).

qRT-PCR reactions were performed on an Applied Biosystems StepOnePlus system using Fast SYBR® Green master mix (Applied Biosystems, Thermo Fisher Scientific, USA). Primers for reference sequences and target sequence were designed such that amplicons produced for reference and target genes were similar in length (142-231 bp) in order to minimize differential effects of RNA degradation or PCR inhibition (Table S3). PCR cycle using StepOne Real-Time PCR systems (Thermofisher) as follows: [1] Holding stage: 2 mins at 60°C, [2] Cycling stage: 40 cycles of 10 secs at 95°C and 30 secs at 60°C, [3] melt curve stage (step and hold): 15 secs at 95°C, 1 min at 60°C and 15 secs at 95°C. All other settings were left as default. Applied Biosystems StepOnePlus system was used to validate amplification efficiency and specificity. 10-fold serial dilution standard curves were run with each primer pair on representatives from each sample group to ensure reaction efficiencies in the range of 90-100% and R^2 values of >0.9. Melt curves were also visualized to ensure specificity of reactions. All cDNA samples were diluted 1:7 and run in triplicate to account for technical variation. For each sample, all target and reference genes were assayed on a single plate.

5.3.10 Analysis of gene expression

The relative expression levels of each gene were calculated using the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001). One-way ANOVAs and Tukey pairwise comparisons were used to identify significant differences between groups. These statistical tests were performed in R Studio v3.5.0.

Table 1: Expression levels of *HP1* and *SU(VAR)3-9* genes were analysed in 7 *Planococcus citri* groups, which represent both sexes throughout development: embryos, 3rd-instar juveniles and adults. Illustrations are provided to detail morphology at different stages.

Group	Morphology	Description
Adult male		1-2 days after pupation
Mated female		~35 days old, immediately after first egg deposit
Virgin female		35 days old
3rd instar female		Sexually immature
3rd instar male		Sexually immature
Embryos		Male-biased embryos – 1st brood Female-biased embryos – 3rd brood

5.4 Results

5.4.1 Identification of six full HP1-like genes in Planococcus citri

The *HP1* gene family is highly conserved from yeast to humans (Singh *et al.*, 1991; Eissenberg and Elgin, 2000; Lomberk, Wallrath and Urrutia, 2006), although the chromatin functions of *HP1* paralogs are dynamic between species (Levine *et al.*, 2012). A putative *HP1A* homolog in *P. citri*, *PCHET2*, has a critical role in the silencing the paternal chromosomes in males through facultative heterochromatinization (Bongiorni *et al.*, 2007). However, the diversity of the *HP1* gene family in this species, and indeed in Hemiptera, remains unknown. In this study, I characterise the family of *HP1* genes present in *P. citri* using a reciprocal BLAST search with five *HP1* genes found

in *Drosophila melanogaster*. I identify six full *HP1* gene sequences within the *P. citri* genome, including the previously isolated homolog, *PCHET2* (Epstein, James and Singh, 1992). All *P. citri HP1* genes identified contain both the chromodomain (CD) and chromoshadow domain (CSD) that are characteristic of this gene family (Table 2). Expression of these genes is confirmed using RT-PCR on *P. citri* whole adult male and female cDNA. Five out of six, *HP1* genes in *P. citri* are expressed in both sexes (Figure 1). One of the *HP1* genes, *PC_HP1_5,* appears to only be expressed in females (Figure 1). This expression pattern is similar to that of *D. melanogaster* HP1D, in which expression is predominantly restricted to the female germline (Vermaak, Henikoff and Malik, 2005). However, germline-specific RT-PCR of *PC_HP1_5* is required to confirm germline specific expression patterns in *P. citri*. None of the *HP1* genes identified in *P. citri* are exclusively expressed in males, as is the case with testes-restricted *HP1E* homolog in *D. melanogaster* and the testes-restricted *HP1D2* in *D. simulans*. *PCHET2*, a putative homolog of *D. melanogaster HP1A*, is expressed in both sexes, which is in accordance with *HP1A* homolog expression patterns identified in *Drosophila* and other taxa (Mandrioli and Borsatti, 2007; Eissenberg and Elgin, 2014).

Table 2: Putative *HP1* genes identified in the *Planococcus citri* genome. Genes were found in v0 of *P. citri* genome, which is publicly available on mealybug.org.

Name	Gene ID (mealybug.org, *P. citri* v0)	Location in genome	Exons	Size (amino acids)
PC_HP1_1	g11218.t1	PCITRI_01187: 85,626-86,156	1	176
PC_HP1_2	g5977.t1	PCITRI_00720: 58,125-58,676	1	183
PC_HP1_3	g5978.t1	PCITRI_00720: 59,538-60,122	1	194
PC_HP1_4	g17698.t1	PCITRI_01895: 33,033-33,596	1	187
PC_HP1_5	g3685.t1	PCITRI_00548: 101,285-102,027	2	215
PCHET2	g5976.t1	PCITRI_00720: 55,372-55,956	1	194

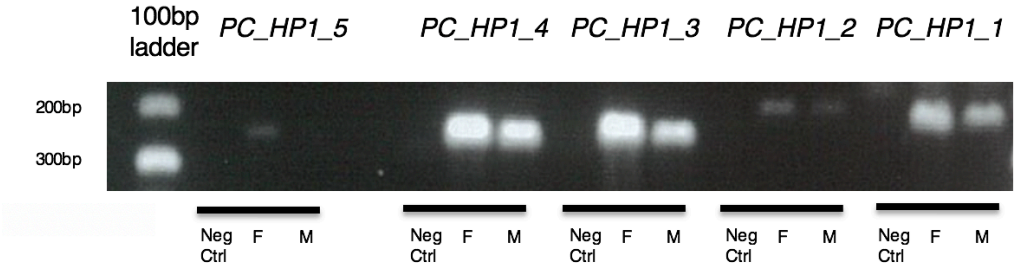


Figure 1: RT-PCR of *HP1* genes identified in the *P. citri* genome. cDNA of whole adult male and female *P. citri* was used to analyse expression patterns of different *PC_HP1* genes. All genes are expressed in both sexes with the exception of *PC_HP1_5*, which appears to only be expressed in females.

5.4.2 Phylogenetic analysis of newly identified HP1 genes in P. citri

In order to establish the evolutionary relationships of *HP1* genes in *P. citri* and those in other taxa, I performed a phylogenetic analysis using conserved chromodomain and chromoshadow domain sequences (Figure 2). The *HP1* gene family is understudied in species outside of the *Drosophila* genus and so consequently diversity of species within this phylogenetic analysis is limited. However, from these results I was able to assign homology to three out of six *P. citri HP1* genes based on amino acid sequence similarity to *HP1* genes in *Drosophila* species. Both *PC_HP1_1* and *PC_HP1_4* appear to be homologous to the *Drosophila HP1B* gene, suggesting that there are two copies of this *HP1* paralog in *P. citri*. In *D. melanogaster, HP1B* co-localises with both euchromatic and heterochromatic regions, and has key functions in transcriptional activation and development (Zhang, Wang and Sun, 2010). However, localisation and knockdown studies of these putative *P. citri HP1B* genes is essential to confirm their role in *P. citri* and clarify how they may be involved in the processes of PGE. This analysis also shows that *PC_HP1_5* is homologous to *HP1C* genes found in *Drosophila* species. In *P. citri*, *PC_HP1_5* is only *HP1* gene to be exclusively expressed in females; however, *HP1C* does not show sex-specific expression in other species. Indeed, as previously discussed, the roles and expression patterns of *HP1* paralogs between even closely related species is diverse (Vermaak and Malik, 2009), therefore *HP1C* in *P. citri* may have evolved a specialised

function. As with the putative *HP1B* homologs in *P. citri*, further studies are required to test the role of *HP1C* in this species.

Thus far, all analyses of *HP1* gene and protein function in *P. citri* have focused on *PCHET2. PCHET2* was cloned using *Drosophila melanogaster HP1A* chromodomain sequence and shares 57% amino acid sequence similarity with this protein (Epstein, James and Singh, 1992). Unexpectedly, my analysis does not confirm that *PCHET2* is homologous to *Drosophila HP1A* genes but does highlight that *PC_HP1_3* is an additional copy of *PCHET2*. Knockdown studies of *PC_HP1_3* in *P. citri* males are required to test for expected functional similarities between this gene and *PCHET2*. Overall, *PCHET2*, *PC_HP1_3* and *PC_HP1_2* could not be assigned homologous to any of the *Drosophila HP1* paralogs used in this study. Hemiptera diverged from Diptera around 350 million years ago (Kazemian *et al.*, 2014) and thus a lack of species diversity in this study could explain why homology could not be assigned. Alternatively, the fact that three of the six *HP1* genes could not be assigned as homologous to any of the *Drosophila HP1* genes in this study may suggest a dynamic pattern of diversification within *P. citri* or in the Hemiptera. Indeed, *HP1* gene family evolution is dynamic and studies continue to discover *HP1* paralogs that are unique to certain species (Levine *et al.*, 2012), such as *HP1D2* in *D. simulans* which evolved 25 million years ago and has been lost at least twice in the *Drosophila* genus (Meiklejohn, 2016). Therefore, *P. citri* may have evolved *HP1* genes unique to the species or to the Hemiptera. However, this cannot be confirmed without further analyses of *HP1* genes in the Hemiptera. Additionally, this phylogeny only contained the most well studied *HP1* genes present in *Drosophila*. Therefore, it could be that the three *P. citri* HP1 genes for which homology could not be assigned may be homologous to other *Drosophila HP1* homologs not present in this analysis.

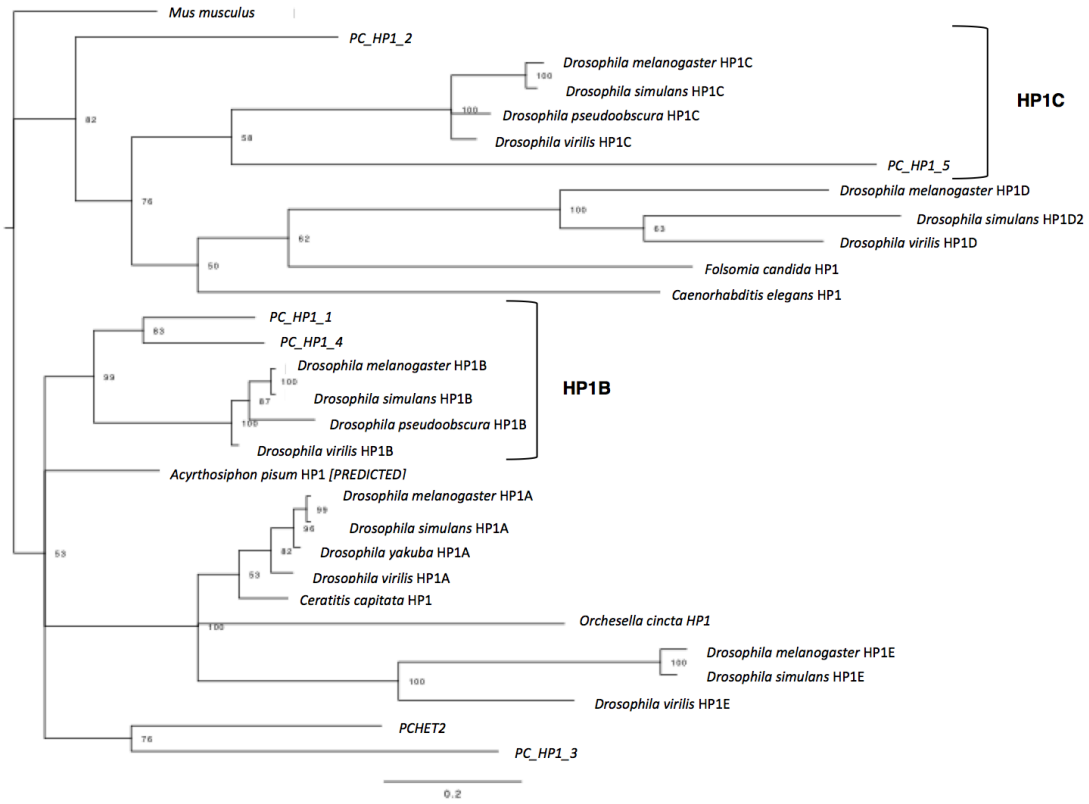


Figure 2: Phylogenetic analysis of *HP1* genes in *Planococcus citri* and other species. Phylogenetic tree of CD and CSDs in *HP1* genes found in *P. citri, Drososphila* species, *Acrythosiphon pisum, Ceratitus capitata, Folsomia candida,* and *Mus musculus* (outgroup). Tree was constructed using MrBayes to assess the evolutionary relationships of *HP1* genes in *P. citri* and those identified in other species.

5.4.3 Expression analysis of an HP1 homolog in P. citri

Expression analyses of *P. citri HP1* homolog, *PCHET2,* across different developmental stages in both sexes of *P. citri* confirm that expression is found in all groups. However, adult males have highest levels of *PCHET2* expression, significantly higher than expression levels than virgin adult females (Figure 3; *p*-value = 0.03, ANOVA Tukey HSD). These findings are consistent with *HP1's* role in paternal genome silencing, as approximately half of the male genome is in a heterochromatic state, therefore, it is likely that expression of *HP1* genes is higher in males than in virgin females. However, silencing of the male genome begins early in embryogenesis therefore it is unclear why levels of *PCHET2* expression are only significantly higher in adult males and not in all male stages. In embryos, the reason for

this could be that I use day of laying as a proxy for embryo sex, as a female's first brood is known to be male-biased while broods laid on day 3 are female-biased (Nelson-Rees, 1960). Therefore, the 'male-biased embryos' group in this study will contain a number of female embryos and the 'female-biased embryos' group will contain a number of male embryos. With regards to 3rd-instars, it is unclear why expression is not significantly different between the sexes. It is also a possibility that *PCHET2* functions in the maintenance of heterochromatin in adult tissues. Expression analyses of other *P. citri HP1* genes may highlight potential roles associated with sex and developmental stage. Mated females do not have significantly different expression to other female groups or male groups (Figure 3). This is possibly due to the fact that mated females are carrying male embryos, which are undergoing facultative heterochromatinization of their paternally inherited genomes. Hence, *HP1* expression is at an intermediate level between males and females. However, these results should be interpreted with caution, as expression of this gene is low in all groups (on average 0.1-fold of HK gene expression) and variability between biological replicates is high. Therefore, it is difficult to confidently distinguish true signal from background noise. To reduce noise and allow a more precise investigation of *PCHET2* expression, tissue-specific expression analysis could be conducted. Testes-specific analysis, in particular, may highlight expression patterns relevant to the elimination events that occur during spermatogenesis.

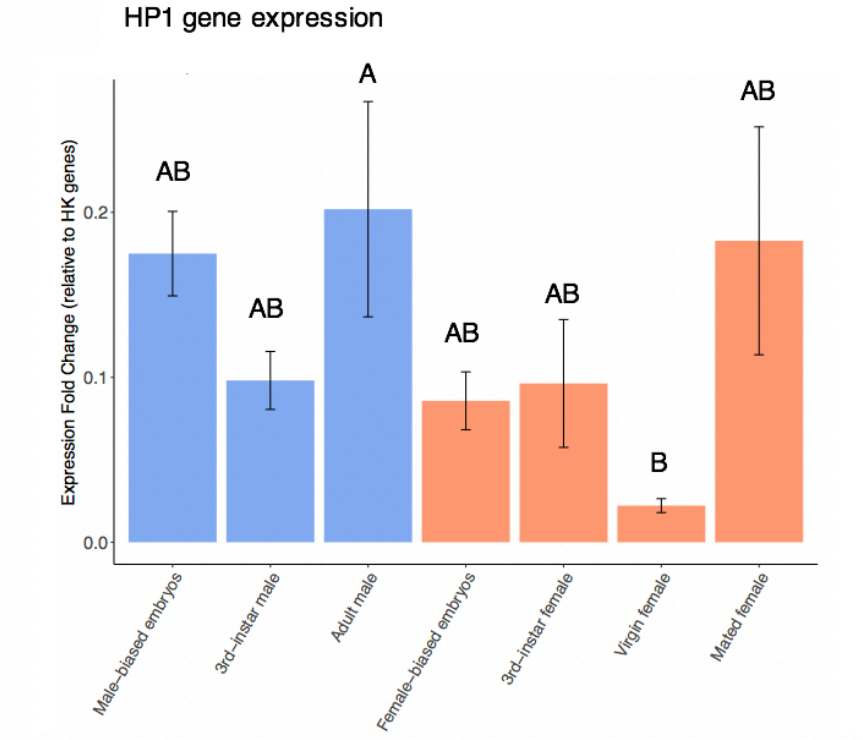


Figure 3: *HP1 (PCHET2)* gene expression throughout development in *Planococcus citri* males and females relative to housekeeping genes (*RP49* and *RP17*). Male groups are coloured blue and female groups are coloured orange. Bar represents average expression across 10 biological replicates per group with standard error of the mean. Samples marked with the same letter are not significantly different from one another (ANOVA, Tukey Pairwise analysis). Adult males have the highest expression of *HP1*, significantly higher than expression in adult virgin females. Table 3 lists *p*-values for each comparison.

Table 3: *HP1* (*PCHET2*) gene expression level comparisons between males and females at key developmental stages. ANOVA and Tukey HSD pairwise comparison tests were carried out and corrected *p*-values are shown.

Comparison	**Tukey HSD corrected *p*-value**
Adult male x Virgin female	0.032
Adult male x Mated female	0.99
Male-biased embryo x Female-biased embryo	0.62
3rd-instar male x 3rd-instar female	1.00

5.4.4 Identification of SU(VAR)3-9 genes in P. citri

SU(VAR)3-9 has a key role in the evolutionarily conserved facultative heterochromatin pathway that silences the paternal genome in PGE, as it produces a protein that tri-methylates histone H3 at lysine 9. This, in turn, generates a binding site for HP1 family proteins and creates a positive feedback loop for the formation of facultative heterochromatin (Eskeland, Eberharter and Imhof, 2007). I identify one *SU(VAR)3-9* gene in the *P. citri* genome that contains both the SET (Su(Var)3-9, enhancer-of-zeste, trithorax) (IPR001214) and Pre-SET (IPR007728) conserved domains that are crucial to its histone methyltransferase function (Figure 4).

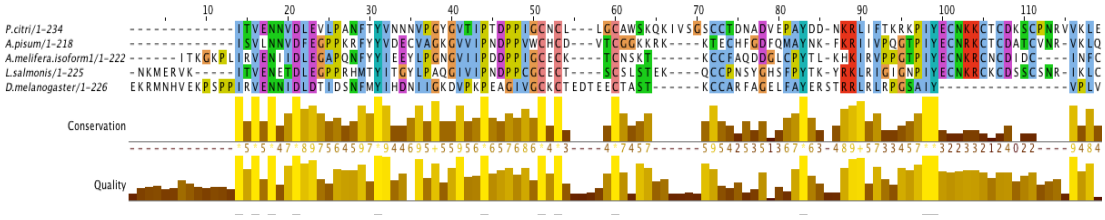


Figure 4: Alignment of SU(VAR)3-9 protein conserved domains (Pre-SET and SET) found in *P. citri* and other insects.

5.4.4 Expression analysis of P. citri SU(VAR)3-9 gene

To further investigate the role of *SU(VAR)3-9* and the H3K9me3-HP1 pathway in PGE, I studied the expression of *SU(VAR)3-9* throughout development in both sexes. Similar to *PCHET2* expression, I find that *SU(VAR)3-9* is expressed in both sexes at all developmental stages. However, in this case, there is no clear sex-specific pattern of expression (Figure 5). On average, adult males have the highest *SU(VAR)3-9* gene expression levels but not significantly higher than expression in adult female groups (compared to virgin females *p*-value = 0.98, compared to mated females *p*-value = 0.47; ANOVA Tukey HSD). This lack of sex-specific

expression may be due to the fact that tri-methylated H3K9 plays an important role in transcriptional regulation of both paternally and maternally inherited genes in both sexes and is not exclusively associated with facultative heterochromatin in males. However, high variability between biological replicates makes these results difficult to interpret. 3rd-instar groups of both sexes have significantly lower expression of *SU(VAR)3-9* than adult males and virgin adult females. It is unclear why SU(VAR)3-9 expression is significantly lower in 3rd-instars compared to adult males and virgin females. However, low overall expression levels of *SU(VAR)3-9* (on average 0.25-fold of HK gene expression) and high variability between biological replicates in these groups, render these results inconclusive.

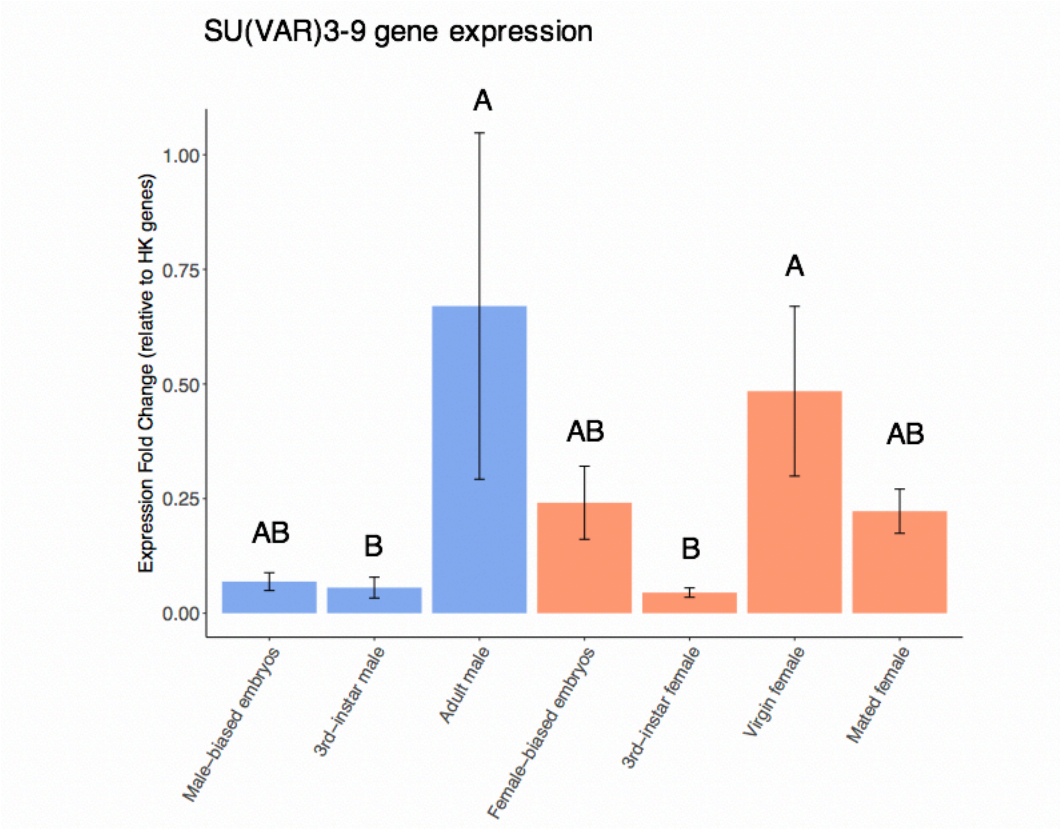


Figure 5: *SU(VAR)3-9* gene expression throughout development in *Planococcus citri* males and females relative to housekeeping genes (*RP49* and *RP17*). Male groups are coloured blue and female groups are coloured orange. Bar represents average expression across 10 biological replicates per group with standard error of the mean. Samples marked with the same letter are not significantly different from one another (ANOVA, Tukey Pairwise analysis).

5.5 Discussion

Heterochromatin has a key role in the regulation of gene expression, chromosome segregation and, potentially, in meiotic drive (Vermaak and Malik, 2009; Helleu *et al.*, 2016). The highly conserved *HP1* gene family is crucial for the formation and maintenance of heterochromatin across taxa from yeast to humans (Singh *et al.*, 1991; Eissenberg and Elgin, 2000). Extensive studies in the *Drosophila* genus have revealed the complexity of this gene family in which heterogeneity of paralogs is matched by the heterogeneity of their functions (Levine *et al.*, 2012). The diverse functions of *HP1* paralogs are a direct result of their abilities to interact with a variety of proteins (Vermaak and Malik, 2009). Many eukaryotes possess only one *HP1* gene but several express two or more paralogs. In *P. citri*, an *HP1* gene and its involvement in heterochromatin formation is known to have a key role in the genomic imprinting phenomenon, Paternal Genome Elimination (PGE). The paternally inherited genome in *P. citri* males is transcriptionally silenced during early embryogenesis through the formation of facultative heterochromatin (Hughes-Schrader, 1948) and subsequently lost from the germline during spermatogenesis. An *HP1* gene, *PCHET2*, has a causative role in the silencing of the paternal genome through its involvement in the evolutionarily conserved H3K9me3-HP1 heterochromatin pathway (Bongiorni *et al.*, 2007). The heterochromatic state of the paternal genome is maintained in the germline and it is hypothesised that this chromatin condition prevents the paternal chromosomes from successfully completing spermatogenesis (Chapter 3). However, the mechanisms involved in establishing and maintaining heterochromatin in this species remain unclear and, despite its important role in PGE, very little was known about the diversity of the *HP1* gene family in *P. citri* or in other Hemipteran insects.

Using a computational approach, I identify six full *HP1* genes in the *P. citri* genome, including the *PCHET2* gene. Although smaller than the *HP1* gene family in *D. melanogaster*, which has ten full *HP1* paralogs, the *P. citri HP1* gene family is larger than *HP1* gene families in mice and humans, which each have three paralogs (Saunders *et al.*, 1993; Furuta *et al.*, 1997; Jones, Cowell and Singh, 2000) and in *Nasonia vitripennis*, which has 1 full and 2 partial *HP1* genes (Fang, Schmitz and Ferree, 2015). The *HP1* gene family has a vital role in the formation of heterochromatin, which in turn, has a vital role in regulation of gene expression and chromosome condensation during PGE. Therefore, a larger repertoire of *HP1* genes in *P. citri* may be representative of these complex chromatin-based processes.

Phylogenetic analyses reveal that two of the identified *HP1* genes in *P. citri* are homologous to the *Drosophila HP1B* paralog and one is homologous to the *Drosophila HP1C* paralog. *HP1B* is required for transcription of both heterochromatic and euchromatic genes and plays a key role in developmental processes (Zhang, Wang and Sun, 2010). *HP1C* functions least like the canonical heterochromatin protein *HP1A* and localises exclusively to euchromatin where is associated with active transcription (Kwon and Workman, 2011). Interestingly, the *P. citri HP1C* homolog is the only *HP1* gene found to be exclusively expressed in females. Sex-specific expression of *HP1* genes in *Drosophila* occurs in the germline. Therefore, germline-specific RT-PCR of the *HP1C* homolog in *P. citri* would be of particular interest and allow for further speculation of its functional role. In *D. melanogaster*, the *HP1D* gene is predominantly expressed in heterochromatic regions in the ovaries (Vermaak, Henikoff and Malik, 2005) where it is associated with transposon silencing (Klattenhoff et al., 2009). Phylogenetic studies in *Drosophila* species reveal that although *HP1D* is highly conserved, it is rapidly evolving and notable differences in amino acid sequence occurs between *Drosophila* species (Vermaak, Henikoff and Malik, 2005). Both conserved domains in *HP1D* have evolved far more rapidly than

their counterparts in other *HP1* genes and thus, it is difficult to identify orthologs in other organisms (Vermaak, Henikoff and Malik, 2005). In *D. simulans*, expression of HP1D2 is testes-restricted and codes for a protein that inhibits the successful segregation of the Y chromosome during meiosis, meaning males can only transmit an X chromosome in their sperm, thus distorting the sex ratio of offspring (Helleu *et al.*, 2016). The *P. citri HP1C* homolog is an obvious candidate for future studies investigating the role of chromatin-associated proteins in sexual conflict. Furthermore, investigating the evolutionary rates of *HP1* genes in PGE species would also highlight those genes potentially involved in such an arms race. With genomes now becoming available for multiple mealybug species this would be a sensible next step.

Expression analyses of *PCHET2* show higher expression in adult males than in adult virgin females. This provides further evidence for the role of *HP1* genes and heterochromatin in silencing the paternal genome, as it is expected that male *HP1* gene expression levels will be higher due to the fact that half of their genome is in a heterochromatic state. However, this pattern is only found in adult stages, with embryos and 3rd-instars showing no sex-specific expression patterns. Therefore, it would be of interest to conduct expression analyses of other *P. citri HP1* homologs, particularly the *PCHET2* paralog, *PC_HP1_3* to compare expression profiles. High expression levels of different *HP1* paralogs could be indicative of increased functional importance in a particular sex at different stages of development. For example, high expression in early embryonic stages may suggest a role in the establishment of heterochromatinization of the paternal genome. Additionally, high expression in immature males – or specifically in the testes of immature males – could be indicative of a role in elimination of the paternal genome during spermatogenesis. However, these results would not confirm the roles of *HP1* paralogs and therefore additional analyses such as

RNAi knockdown studies and cytological investigations would be required to support these hypotheses.

This study provides a preliminary overview of the *HP1* gene family present within *P. citri*. These results show that this species has a higher number of *HP1* genes than found in Hymenopteran species and in mammals, which may be associated with their role(s) in PGE. However, further functional and phylogenetic analyses within the Hemiptera are required to assess the evolution of the *HP1* gene family in this order, elucidate the functions of *HP1* paralogs in different species and identify their potential role in the silencing and loss of paternal chromosomes under PGE.

5.6 SUPPLEMENTARY TABLES AND FIGURES

Table S1: HP1 genes used to construct phylogenetic analysis. All sequences were collected from the NCBI database (https://www.ncbi.nlm.nih.gov). Conserved domains were then extracted using NCBI conserved domain search (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi)

Organism name	Genbank/NCBI accession number	Description
Mus musculus	NP_001345879	chromobox protein homolog 5
Drosophila simulans	XP_016024109	heterochromatin protein 1A
Drosophila virilis	XP_002051948	heterochromatin protein 1A
Drosophila yakuba	XP_002088757	heterochromatin protein 1A
Drosophila melanogaster	AAF52618	heterochromatin protein 1A
Drosophila melanogaster	AFH07300	heterochromatin protein 1B
Drosophila melanogaster	AAF56059	heterochromatin protein 1C
Drosophila melanogaster	NP_536794.1	heterochromatin protein 1D
Drosophila melanogaster	AAF54354.2	heterochromatin protein 1E
Drosophila simulans	XP_002106478	heterochromatin protein 1D2
Drosophila pseudoobscura	XP_001354996.3	heterochromatin protein 1B
Drosophila virilis	XP_002055065.2	heterochromatin protein 1B
Drosophila simulans	XP_002106513.1	heterochromatin protein 1B
Drosophila pseudoobscura	XP_001358118.1	heterochromatin protein 1C
Drosophila virilis	XP_002054276.1	heterochromatin protein 1C
Drosophila simulans	XP_002104480.1	heterochromatin protein 1C
Drosophila virilis	XP_002049857.1	heterochromatin protein 1D,

		isoform A
Drosophila simulans	XP_002104147.1	heterochromatin protein 1E
Drosophila virilis	XP_002052989.1	heterochromatin protein 1E
Ceratitis capitata	ODM96524.1	heterochromatin protein 1
Folsomia candida	OXA38520.1	heterochromatin protein 1
Caenorhabditis elegans	NP_001022653.1	HP1 Like (heterochromatin protein)
Acyrthosiphon pisum	XP_008184084	heterochromatin protein 1-like

Table S2: *Planococcus citri HP1* primers used for RT-PCR

Gene	Amplicon size (bp)	Primer-F	Primer-R
PC_HP1_1	234	AAAAACGACGAATCGAGCAG	ATTCGCTTCTTTAGCCACGA
PC_HP1_2	183	CGCTCCAATAATGATCCTCAA	CGCATCCTTTGAACATACCA
PC_HP1_3	240	AGGAGTTTGAGTCGCGAAAA	GTCAGACCTCGTGCAAATCC
PC_HP1_4	246	CGGTAGAGAACCTCGATTGC	TAGCGCCGATTATCTTCTCG
PC_HP1_5	229	TTGGTGAGGAACTCCGAATC	GCCATTCAACGAGCAATTTT

Table S3: *Planococcus citri* qPCR primers used in this study

Gene	Amplicon size (bp)	Primer-F	Primer-R
HP1	142	AATGGAAGGGCTATGGCGAC	CCGTCGTCATCAGGCTCAAT
SU(VAR)3-9	179	ATCCGCATTATCCGTACAGC	CCGTCGAAAATAACGTCGAT
RP49	165	AAGAAGGTTCAAGGGCCAGT	TGGGCAACTTCAGCACAATA
RP17	231	CTGCGAACCCTACATCACCT	TTGAAAGGCCAGAAGAATCG

Chapter 6: Epilogue

6.1 Thesis overview

Paternal Genome Elimination (PGE) is a genomic imprinting phenomenon found in thousands of insect species that involves the elimination of an entire haploid genome in a parent-of-origin specific manner. The repeated evolution of PGE suggests the existence of an evolutionarily conserved mechanism for parent-of-origin recognition, expression and transmission of genes across arthropods. The aim of this thesis was to investigate the epigenetic mechanisms underlying the key processes of PGE using the citrus mealybug, *Planococcus citri* as a study organism.

6.2 Key findings and future studies

Chapter 2

In this chapter, I used sex-specific whole genome bisulfite sequencing (WGBS) data and transcriptome (RNA-seq) data to describe the methylation landscape of *P. citri* and address the following questions: 1) does *P. citri* exhibit patterns of sex-specific DNA methylation; 2) is there evidence of sex-biased gene expression 3) what is the relationship between gene methylation and gene expression and 4) what do these patterns tell us about the regulation of sexual dimorphism and PGE?

I found that *P. citri* has high levels of DNA methylation compared to other Hemiptera – despite the apparent loss of the *de novo* DNA methyltransferase gene, *DNMT3*. This suggests that *DNMT1* can perform both maintenance and *de novo* DNA methylation functions or that an as yet unclassified gene may be involved in the process. Similar to patterns described in other insects, DNA methylation in *P. citri* is predominantly located in gene bodies and particularly enriched in exons. However, whilst gene body methylation in

other species is associated with elevated gene expression, evidence for this relationship in *P. citri* is not convincing. This indicates that the functional activity of DNA methylation is not entirely conserved across different insect species. Indeed, this highlights the importance of expanding DNA methylation studies across a variety of invertebrate taxa, as strong focus on Hymenopteran species has led to a major gap in our knowledge of how DNA methylation functions in insects.

Approximately 26% of all genes in the *P. citri* genome exhibit sex-biased expression. This is considerably less than the 75% of genes that exhibit sex-biased expression in *Drosophila melanogaster* and *Nasonia vitripennis* (Assis, Zhou and Bachtrog, 2012; Wang, Werren and Clark, 2015), despite the extreme sexual dimorphism in *P. citri.* However, sex-biased gene expression is highly variable throughout development and can be specific to particular developmental stages and then lost in adults (Grath and Parsch, 2016). Therefore, investigating expression throughout development will provide greater insight into sex-biased expression patterns and their role(s) in sexual dimorphism. Future studies should also analyse tissue-specific expression patterns, particularly in the germline as sex-biased expression tends to be highest in the gonads (Grath and Parsch, 2016).

Sex-specific patterns of DNA methylation and gene expression are found in *P. citri*; however, there is no correlation between sex-specific differences in gene methylation and sex-specific differences in gene expression. This finding is particularly striking as *P. citri* males and females are genetically identical and so it is likely that sexual dimorphism is a consequence of sex-specific expression of genes that are present in both sexes. Therefore, I hypothesise that alternative mechanisms are involved in the regulation of sex-biased expression. There are very few studies in insects that directly investigate the relationship between sex-specific methylation and sex-specific expression, therefore it is difficult to place my results in the context of

other work. However, one study, in the parasitoid wasp *Nasonia*, finds that despite extensive sex-specific differences in gene expression, there are no clear differences in DNA methylation between adult males and females (Wang, Werren and Clark, 2015). In this species, both sexes also have the same gene complement and there are no differences in gene absence/presence between males and females. This supports the suggestion that mechanisms other than DNA methylation may underlie the process of sex-specific gene expression in insects. One alternative mechanism may be the presence of sex-specific transcription factor binding sites as is the case in the fig wasp, *Ceratosolen solmsi* (Sun *et al.*, 2015).

However, although a clear correlation between methylation and expression on a gene-by-gene basis is not identified, unlike in *N. vitripennis, P. citri* shows evident patterns of sex-specific methylation, and these may regulate gene expression differences through *trans* rather than *cis* effects. Conducting a similar WGBS-seq and RNA-seq analysis with intraspecific hybrid crosses of two genetically distinct lines of *P. citri* would allow the identification of *cis* versus *trans* regulatory effects. If DNA methylation differences are due only to changes in *cis*-regulatory sequences, then allele-specific methylation in offspring will resemble parental methylation status. If changes are exclusively a result of *trans* factors (e.g. methylation status is remodelled in every generation) then offspring allele-specific methylation will be ~50% on both parental alleles with no intraspecific differences.

Chapter 2 provides a preliminary analysis of DNA methylation and gene expression in an insect with PGE. There is much scope for future analyses, particularly as sequencing methods develop. Both DNA methylation and expression are dynamic biological processes and the data in this study only provides a snapshot of levels and patterns at this specific stage. These analyses were carried out on pooled whole adult samples and thus cannot detect variation between individuals, tissues or developmental stages.

Therefore, a relationship between DNA methylation and expression cannot be ruled out in specific tissues or at specific periods in development. Furthermore, a recent study has highlighted the occurrence of individual variation in DNA methylation patterns and the implications this may have in assigning functions to this epigenetic modification (Libbrecht *et al.*, 2016). Future studies should consider both germline-specific and stage-specific analyses. Germline studies are of particular interest as this is where elimination of the paternal genome occurs in PGE males. Analyses of early stages of development where key sex-specific gene expression occurs could also provide information about the role of DNA methylation in sex determination.

Chapter 3

This study aimed to identify patterns of parent-of-origin specific DNA methylation using the hybrid offspring of two closely related PGE species, *Planococcus citri* and *Planococcus ficus*. It has been previously suggested that differences in levels of DNA methylation on the paternal and maternal chromosomes in *P. citri* may have a role in their silencing and elimination during PGE (Bongiorni, Cintio and Prantera, 1999). However, the nature of this study – *in-situ* nick translation using methylation sensitive enzymes – provides no indication of the levels or genomic context of methylation on the chromosomes. Furthermore, a subsequent study in the same species found no significant DNA methylation differences between maternal and paternal chromosomes (Buglia, Predazzi and Ferraro, 1999). Therefore, I used a next-generation whole genome bisulfite sequencing approach to analyse methylation levels at a base-pair resolution and identify parent-of-origin specific patterns in both sexes. Interestingly, although unfortunately for this study, the mortality rate of hybrid males is high making it difficult to obtain the sufficient quantity of DNA required for bisulfite sequencing. Due to the nature of the analysis each sample could only contain full siblings, as half-siblings (males with different fathers or mothers) would affect SNP identification and

allele-specific methylation analysis. Additionally, hybrid crosses can only produce male offspring in one direction (only with *P. citri* mothers and *P. ficus* fathers). As a consequence, the low coverage and lack of reciprocal crosses in this study makes it difficult to draw any solid conclusions about parent-of-origin specific DNA methylation and its role in PGE. Overall, my results do support previous findings that paternally inherited alleles have lower levels of DNA methylation than maternally inherited alleles. However, the number of differentially methylated SNPs is small proportion of the total number of species-specific SNPs identified. Furthermore, the lack of a clear correlation between gene methylation and gene expression found in Chapter 2 suggests that differential methylation is not necessarily indicative of a function in their transcriptional silencing but may have a role in marking their paternal origin. A key issue in this study is that parent-of-origin specific methylation differences cannot be disentangled from species-specific differences or indeed haplotype-specific differences. To confirm that these methylation differences are parent-of-origin specific – as opposed to species-specific - this experiment should be repeated using intraspecific reciprocal crosses of genetically distinct *P. citri* lines. The higher survival rate of male offspring would allow for better coverage (no need to use a low input bisulfite kit) and, thus, a more reliable dataset. Any lineage specific effects can also be identified and accounted for. Additionally, immunostaining of chromosomes for 5-methylcytosine would allow a visual analysis of methylation differences across the length of individual chromosomes. As previously discussed, DNA methylation analyses in this thesis are conducted using pooled adult whole-body samples and therefore do not give insight into tissue-specific patterns of this epigenetic modification. Interestingly, silencing of the paternal genome is reversed in some male tissues including the Malpighian tubules and the testes (Nur, 1967). Analysis of the specific tissues in which heterochromatinization of the paternal genome has been reversed – compared to tissue in which the paternal genome remains silenced - could reveal a role for DNA methylation in the transcriptional regulation of the

paternally inherited alleles. This data should be combined with RNA-seq data to elucidate the relationship between allele-specific methylation and expression.

My analysis also revealed a remarkable difference between sex-specific DNA methylation levels found in pure *P. citri* adults and *citri* x *ficus* hybrid adults. Pure *P. citri* and hybrid females have no significant difference in global methylation levels, however, *P. citri* males have significantly higher methylation levels than hybrid males. This suggests that there are species-specific differences in DNA methylation or that hybridisation somehow affects the epigenetic landscape of these organisms. It is particularly interesting that hybridisation only appears to affect global methylation levels in males, as hybrid male mortality rates are high (Rotundo and Trembley, 1982). There is growing evidence to support that mis-regulation of gene silencing plays a role in hybrid incompatibility (Bomblies, 2006). Although there are few studies addressing species-specific epigenetic landscapes and hybrid incompatibility in insects, a study in the plant genus, *Arabidopsis*, shows that epigenetic variation contributes to hybrid genome incompatibility (Blevins *et al.*, 2017). Methylation analysis of pure *P. ficus* males and females is required to further understand this result and investigate whether DNA methylation has a role in high hybrid male mortality.

Chapter 4

In this chapter, I described the role of two evolutionarily conserved heterochromatin pathways, H3K9me3-HP1 and H3K27me3-PRC2, in Paternal Genome Elimination (PGE). I used an immunocytological approach to detect the presence of histone modifications, H3K9me3 and H3K27me3, in the two key processes of PGE: 1) silencing of the paternal genome in somatic tissues and 2) recognition, elimination and imprinting of paternal chromosomes during spermatogenesis. I found that both of these histone modifications are present on the heterochromatinized paternal genome

during male embryonic development and spermiogenesis. This suggests a role for these pathways in both the silencing and elimination of the paternal genome during PGE. I propose that the heterochromatinization of the nuclei carrying paternally inherited chromosomes via the H3K9me3-HP1 and H3K27me3-PRC2 pathways prevents their successful completion of meiosis and results in their elimination from mature sperm. Studies of B chromosomes in *Pseudococcus affinis* and Paternal Sex Ratio (PSR) in *Nasonia vitripennis* support that the chromatin structure of selfish genetic elements may allow them to escape elimination (Nur and Brett, 1988; Aldrich *et al.*, 2017). However, an RNAi approach to target genes involved in the condensation of paternal chromosomes in the male germline is required to test whether or not uncondensed paternal chromosomes can escape elimination during spermatogenesis.

Contrary to previous studies (Buglia and Ferraro, 2004; Bongiorni *et al.*, 2009), I found that H3K9me3 and H3K27me3 are not present on pupal sperm nuclei and therefore are unlikely to act as the imprint identifying paternally inherited chromosomes in the embryo. Therefore, the mechanism by which the parental origin of a chromosome is distinguished in PGE species still remains unclear, although findings in Chapter 4 support a role for differential DNA methylation. These results, of course, do not completely rule out histone involvement in imprinting and one should perform further immunocytological analyses of all core histones. Additionally, analyses of protamines throughout spermatogenesis would provide further insight into the occurrence (or absence) of histone to protamine exchange in this highly specialised type of meiosis. Genome-wide analyses of these histone modifications on maternal and paternal chromosomes in males and females can also be conducted using ChipSeq. Combining this data with an allele-specific RNA-seq dataset will allow an assessment of how different histone modifications may regulate parent-of-origin specific gene expression. Additionally, this would allow a direct comparison between patterns of DNA

methylation and histone modifications at a base-pair resolution, which is important as mechanisms for epigenetic regulation of gene expression are not mutually exclusive. In the ant, *Camponotus floridanus*, DNA methylation patterns in genes are strongly predicted by the presence of histone modifications. Furthermore, these epigenetic modifications are more predictive of gene expression when considered together than when considered independently (Glastad, Hunt and Goodisman, 2015).

Chapter 5

In this chapter, I identified and characterised key genes in the evolutionarily conserved H3K9me3-HP1 heterochromatin pathway: The Heterochromatin Protein 1 (*HP1*) gene family and histone methyltransferase *SU(VAR)3-9* genes. I identify six full *HP1* genes in the *P. citri* genome, two of which are homologous to *Drosophila HP1B* and one that is homologous to *Drosophila HP1C*. This study reveals that *P. citri* has a higher number of *HP1* genes than Hymenopteran species and mammals, which may be associated with the complex chromatin-based processes that occur during PGE. The *HP1C* homolog in *P. citri* is expressed exclusively in females suggesting a sex-specific function for this gene. Studies in *Drosophila* have shown that sex-specific expression of *HP1* genes is a result of germline-specific expression, therefore germline-specific RT-PCR of the *HP1C* homolog in *P. citri* would be of particular interest and allow for further speculation of its functional role.

I also analysed the expression profiles of a putative *HP1A* homolog, *PCHET2,* and histone methyltransferase gene, *SU(VAR)3-9* throughout development in both sexes. Whilst results for *SU(VAR)3-9* expression were inconclusive, *PCHET2* expression is significantly higher in adult males than in adult virgin females. This implicates this *HP1* gene in the silencing of the paternal genome. However, as discussed in Chapter 5, there is considerable variability in expression between biological replicates and tissue-specific expression analysis may reduce this noise.

Here I have provided a preliminary overview of the *HP1* gene family present within *P. citri*. With genomes now available for multiple Hemipteran species (mealybug.org), a phylogenetic analysis of this gene family in the Hemiptera would be a sensible next step. This would provide insight into evolutionary rates of *HP1* gene family members in this group and highlight those that may be involved in sexual conflict. Gene losses and gains in different Hemipteran species would also be revealed and a comparison between PGE and non-PGE species may indicate *HP1* homologs involved in genomic imprinting.

6.3 Implications of this work

Altogether, the findings of this thesis provide evidence for the role of both DNA methylation and histone modifications in key processes that occur during PGE. However, these results are preliminary and there is scope for more specialised functional analyses. As well as providing insight into the mechanisms of genomic imprinting, an understanding of the epigenetic tools involved in PGE will provide insight into its evolution and its involvement in sex determination.

Sex determination in PGE species is hypothesised to be under maternal control as a number of maternal factors influence the sex ratio of offspring (Ross *et al.*, 2010, 2011). However, it is unknown what triggers the process of PGE in males and whether or not heterochromatinization of the paternal genome is the consequence of being male or whether it causes maleness. To test this, RNAi knockdowns of genes involved in the heterochromatinization to the paternal genome can be constructed and given to mothers or embryos in an artificial growth environment. This is similar to a test performed by Bongiorni *et al.* (2007) in which heterochromatinization of the paternal genome in male embryos was reversed by dsRNA knockdown of the *HP1* gene, *PCHET2*. Unfortunately, in this experiment embryos were

removed from their mothers and soaked in dsRNA solution so therefore did not survive. If an approach can be developed that allows the survival of embryos, one can investigate how losing the ability to condense the paternal genome affects the sex-ratio of offspring. For example, if all RNAi-treated embryos develop as females, this would suggest heterochromatin may have a causative role in sex determination. If the sex of an embryo is unaffected by a loss of heterochromatinization, this suggests that silencing of the paternal genome is a consequence of another sex determining mechanism but is not essential for male development. I suggest a likely scenario is that haploid gene expression, resulting from maternally-induced paternal genome silencing, is what causes male development. The apparent ability of mothers to adjust the sex-ratio of their offspring supports the hypothesis that sex determination is under maternal control, however empirical support is lacking. If mothers do indeed control offspring sex, they may do so through maternal effects deposited by the female into her egg cytoplasm as is the case in *N. vitripennis* (Verhulst, Beukeboom and van de Zande, 2010). Likely candidates for these maternal effects are small RNAs, long non-coding RNAs, or, as with *N. vitripennis*, messenger RNAs. In order to elucidate the role of maternal RNAs in sex determination and PGE, the RNA profiles of eggs destined to become males and females should be analysed.

6.4 Concluding remarks

This research provides insight into the putative roles of sex-specific and parent-of-origin specific epigenetic modifications in the recognition, silencing and elimination of the paternal genome during PGE. Understanding the mechanisms involved in this striking form of genomic imprinting will provide general insights into the role of epigenetic regulation and allow the development of an invertebrate model system for studying parent-of-origin effects. Additionally, analyses of DNA methylation and histone modifications

in a non-social, non-hymenopteran insect broadens understanding of the function(s) and evolution of epigenetic modifications within arthropods.

References

Aasland, R. and Stewart, A. F. (1995) 'The chromo shadow domain, a second chromodomain in heterochromatin-binding protein1, HP1.', *Nucleic Acids Res.*, 23(16), pp. 3168–3173.

Achwal, C. W., Iyer, C. A. and Chandra, H. S. (1983) 'Immunochemical evidence for the presence of 5mC, 6mA and 7mG in human, Drosophila and mealybug DNA', *FEBS Letters*, 158(2), pp. 353–358. doi: 10.1016/0014-5793(83)80612-7.

Akalin, A. *et al.* (2012) 'MethylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles', *Genome Biology*. BioMed Central Ltd, 13(10), p. R87. doi: 10.1186/gb-2012-13-10-R87.

Aldrich, J. C. *et al.* (2017) 'A "selfish" B chromosome induces genome elimination by disrupting the histone code in the jewel wasp Nasonia vitripennis', *Scientific Reports*. Nature Publishing Group, pp. 1–14.

Andrews, S. (2010) 'FastQC: A quality control tool for high throughput sequence data.', *Available: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/*.

Anton, B. P. *et al.* (2015) 'Complete genome sequence of ER2796, a DNA methyltransferase-deficient strain of Escherichia coli K-12', *PLoS ONE*, 10(5), pp. 1–22. doi: 10.1371/journal.pone.0127446.

Assis, R., Zhou, Q. and Bachtrog, D. (2012) 'Sex-biased transcriptome evolution in drosophila', *Genome Biology and Evolution*, 4(11), pp. 1189–1200. doi: 10.1093/gbe/evs093.

Bachtrog, D. *et al.* (2014) 'Sex Determination: Why So Many Ways of Doing It?', *PLoS Biology*, 12(7), pp. e1001899-13.

Bannister, A. J. *et al.* (2001) 'Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain', *Nature*. Nature Publishing Group, 410(6824), pp. 120–124.

Barr, M. L. and Carr, D. (1962) 'Correlations between sex chromatin and sex chromosomes', *Acta Cytologica*, 6, pp. 34–45.

Benjamini, Y. and Hochberg, Y. (1995) 'Controlling the False Discovery Rate : A Practical and Powerful Approach to Multiple Testing Author (s): Yoav Benjamini and Yosef Hochberg Source : Journal of the Royal Statistical Society . Series B (Methodological), Vol . 57 , No . 1 Published by ':, *Journal of the Royal Statistical Society Series B*, 57(1), pp. 289–300.

Bernardino, J. *et al.* (2000) 'Common methylation characteristics of sex chromosomes in somatic and germ cells from mouse, lemur and human', *Chromosome Research*, 8(6), pp. 513–525. doi: 10.1023/A:1009271706488.

Bewick, A. J. *et al.* (2016) 'Evolution of DNA methylation across insects.', *Molecular Biology and Evolution*. Department of Genetics, University of Georgia, Athens, GA 30602, USA

bewickaj@uga.edu schmitz@uga.edu.: Oxford University Press, pp. msw264-12.

Bewick, A. J. *et al.* (2018) 'Gene-regulatory independent functions for insect DNA methylation', *bioRxiv*.

Bierl-Leonhardt, B. A. . *et al.* (1981) 'Isolation, identification, and synthesis of the sex pheromone of the citrus mealybug, Planococcus citri (Risso)', *Tetrahedron Letters*, 22, p. 389.

Bird, A. (2002) 'DNA methylation patterns and epigenetic memory', *Genes Dev*, 16, pp. 6–21. doi: 10.1101/gad.947102.6.

Bird, A. P. (1980) 'DNA methylation and the frequency of CpG in animal DNA', *Nucleic Acids Research*, 8(7).

Bird, A. P. and Wolffe, A. P. (1999) 'Methylation-Induced Repression— Belts, Braces, and Chromatin', *Cell*, 99(5), pp. 451–454. doi: 10.1016/S0092-8674(00)81532-9.

Blevins, T. *et al.* (2017) 'Hybrid incompatibility caused by an epiallele', *Proceedings of the National Academy of Sciences*, 114(14), pp. 3702–3707. doi: 10.1073/pnas.1700368114.

Bolger, A. M., Lohse, M. and Usadel, B. (2014) 'Trimmomatic: A flexible trimmer for Illumina sequence data', *Bioinformatics*, 30(15), pp. 2114–2120. doi: 10.1093/bioinformatics/btu170.

Bomblies, K. (2006) 'Hybrid Incompatibility: When Opposites Attract with a Fatal Outcome', *Current Biology*, 16(14), pp. 542–544. doi: 10.1016/j.cub.2006.06.038.

Bonasio, R. *et al.* (2010) 'Genomic Comparison of the Ants Camponotus floridanus and Harpegnathos saltator', *Science*, 329(5995), pp. 1068–1071. doi: 10.1126/science.1192428.

Bonasio, R. *et al.* (2012) 'Genome-wide and caste-specific DNA methylomes of the ants camponotus floridanus and harpegnathos saltator', *Current Biology*, 22(19), pp. 1755–1764. doi: 10.1016/j.cub.2012.07.042.

Bongiorni, S. *et al.* (2004) 'Inverted meiosis and meiotic drive in mealybugs', *Chromosoma*, 112(7), pp. 331–341.

Bongiorni, S. *et al.* (2007) 'Epigenetic regulation of facultative heterochromatinisation in Planococcus citri via the Me(3)K9H3-HP1-Me(3)K20H4 pathway.', *Journal of Cell Science*. Department of Agrobiologia e Agrochimica, University of Tuscia, 01100 Viterbo, Italy.: The Company of Biologists Ltd, 120(Pt 6), pp. 1072–1080.

Bongiorni, S. *et al.* (2009) 'Epigenetic marks for chromosome imprinting during spermatogenesis in coccids', *Chromosoma*. Springer-Verlag, 118(4), pp. 501–512.

Bongiorni, S., Cintio, O. and Prantera, G. (1999) 'The Relationship Between DNA Methylation and Chromosome Imprinting in the Coccid Planococcus citri', *Genetics Society of America*, pp. 1–8.

Bongiorni, S. and Prantera, G. (2001) 'Facultative heterochromatization in parahaploid male mealybugs: involvement of a heterochromatin-associated protein', *Development*, pp. 1–9.

Boros, J. *et al.* (2014) 'Polycomb repressive complex 2 and H3K27me3 cooperate with H3K9

methylation to maintain heterochromatin protein 1α at chromatin.', *Molecular and cellular biology*, 34(19), pp. 3662–74. doi: 10.1128/MCB.00205-14.

Borsa, P. and Kjellberg, F. (1996) 'Experimental evidence for pseudo-arrhenotoky in Hypothenemus hampei (Coleoptera: Scolytidae)', *Heredity*, 76(2), pp. 130–135. doi: 10.1038/hdy.1996.20.

Boulton, R. A., Collins, L. A. and Shuker, D. M. (2014) 'Beyond sex allocation: the role of mating systems in sexual selection in parasitoid wasps', *Biological Reviews*, p. n/a-n/a.

Boumil, R. M. and Lee, J. T. (2001) 'Forty years of decoding the silence in X-chromosome inactivation', *Human Molecular Genetics*, 10(20), pp. 2225–2232. doi: 10.1093/hmg/10.20.2225.

Boyko, A. and Kovalchuk, I. (2011) 'Genome instability and epigenetic modification-heritable responses to environmental stress?', *Current Opinion in Plant Biology*. Elsevier Ltd, 14(3), pp. 260–266. doi: 10.1016/j.pbi.2011.03.003.

Braun, R. E. (2001) 'Packaging paternal chromosomes with protamine', *Nature Genetics*, 28(1), pp. 10–12. doi: 10.1038/88194.

Brown, S. W. (1964) 'Automatic frequency response in the evolution of male haploidy and other coccid chromosomal systems.', *Genetics*, 49(May), pp. 797–817. doi: papers2://publication/uuid/6EE1DB3E-CC7C-4652-86C0-D00E7157A83B.

Brown, S. W. (1965) 'Chromosomal survey of the armored and palm scale insects (Coccoidea: Diaspididae and Phoenicoccoccidae)', *Hilgardia*, 36(5), pp. 189–294.

Brown, S. W. (1966) 'Heterochromatin', *Science*, 151(3709), pp. 417–425.

Brown, S. W. and Nelson-Rees, W. A. (1961) 'Radiation analysis of a lecanoid genetic system', *Genetics*, 46(August), pp. 983–1007.

Brown, S. W. and Nelson-Rees, W. A. (1961) 'Radiation Analysis of a Lecanoid Genetic System', *Genetics*. Department of Genetics, University of California, Berkeley, California.: Genetics Society of America, 46(8), pp. 983–1007.

Brown, S. W. and Nur, U. (1964) 'Heterochromatic Chromosomes in the Coccids', *Science*, 145.

Buglia, G. L. and Ferraro, M. (2004) 'Germline cyst development and imprinting in male mealybug Planococcus citri', *Chromosoma*. Springer-Verlag, 113(6), pp. 284–294.

Buglia, G. L., Predazzi, V. and Ferraro, M. (1999) 'Cytosine methylation is not involved in the heterochromatization of the paternal genome of mealybug Planococcus citri', *Chromosome Research*, 7, pp. 71–73.

Bull, J. J. (1979) 'Advantage for the evolution of male haploidy and systems with similar genetic transmission', *Heredity*, 43, pp. 361–381.

Bull, J. J. (1983) *Evolution of sex determining mechanisms*. Menlo Park, California: The Benjamin/Cummings Publishing Company, Inc.

Bullard, B., Dabrowska, R. and Winkelman, L. (1973) 'The contractile and regulatory proteins of insect flight muscle', *Biochem.J.*, 135(2), pp. 277–286.

Burggren, W. W. (2017) 'Chapter One - Epigenetics in Insects: Mechanisms, Phenotypes and Ecological and Evolutionary Implications', in Verlinden, H. (ed.) *Advances in insect physiology*. Elsevier, pp. 1–30.

Burt, A. and Trivers, R. (2008) *Genes in conflict: the biology of selfish genetic elements*. Belknap.

Cazemajor, M., Joly, D. and Montchamp-Moreau, C. (2000) 'Sex-ratio meiotic drive in Drosophila simulans is related to equational nondisjunction of the Y chromosome', *Genetics*, 154(1), pp. 229–236.

Chen, S. *et al.* (2018) 'fastp: an ultra-fast all-in-one FASTQ preprocessor', *bioRxiv*, p. 274100. doi: 10.1101/274100.

Cooper, D. W. *et al.* (1993) 'X-inactivation in marsupials and monotremes', *Seminars in developmental biology.* London :, pp. 117–128. doi: 10.1006/sedb.1993.1014.

Crouse, H. V (1960) 'The controlling element in sex chromosome behavior in Sciara', *Science*, 166(3911), pp. 1428–1443. doi: 10.1126/science.166.3911.1422.

Daane, K. M. *et al.* (2012) *Biology and Management of Mealybugs in Vineyards. In: Bostanian N., Vincent C., Isaacs R. (eds) Arthropod Management in Vineyards:. Springer, Dordrecht*.

Deans, C. and Maggert, K. A. (2015) 'What do you mean, "Epigenetic"?', *Genetics*, 199(4), pp. 887–896. doi: 10.1534/genetics.114.173492.

Delaval, K. and Feil, R. (2004) 'Epigenetic regulation of mammalian genomic imprinting', *Current Opinion in Genetics and Development*, 14(2), pp. 188–195. doi: 10.1016/j.gde.2004.01.005.

Dhayalan, A. *et al.* (2010) 'The Dnmt3a PWWP domain reads histone 3 lysine 36 trimethylation and guides DNA methylation', *Journal of Biological Chemistry*, 285(34), pp. 26114–26120. doi: 10.1074/jbc.M109.089433.

Dobin, A. *et al.* (2013) 'STAR: Ultrafast universal RNA-seq aligner', *Bioinformatics*, 29(1), pp. 15–21. doi: 10.1093/bioinformatics/bts635.

von Dohlen, C. D. *et al.* (2001) 'Mealybug beta-proteobacterial endosymbionts contain gamma-proteobacterial symbionts.', *Nature*, 412(26), pp. 433–436. doi: 10.1038/35086563.

Dong, X. and Weng, Z. (2013) 'The correlation between histone modifications and gene expression', *Epigenomics*, 5(2), pp. 113–116. doi: 10.2217/epi.13.13.

Dorer, D. R. and Henikoff, S. (1994) 'Expansions of transgene repeats cause heterochromatin formation and gene silencing in Drosophila', *Cell*, 77(7), pp. 993–1002. doi: 10.1016/0092-8674(94)90439-1.

Drewell, R. A. *et al.* (2014) 'The dynamic DNA methylation cycle from egg to sperm in the

honey bee Apis mellifera', *Development*, 141(13), pp. 2702–2711. doi: 10.1242/dev.110163.

Duncan, R. P. *et al.* (2014) 'Dynamic recruitment of amino acid transporters to the insect/symbiont interface', *Molecular Ecology*, 23(6), pp. 1608–1623.

Ebert, A. *et al.* (2004) 'Su(var) genes regulate the balance between euchromatin and heterochromatin in Drosophila', *Genes and Development*, 18(23), pp. 2973–2983. doi: 10.1101/gad.323004.

Ehrich, M. *et al.* (2007) 'A new method for accurate assessment of DNA quality after bisulfite treatment', *Nucleic Acids Research*, 35(5), pp. 1–8. doi: 10.1093/nar/gkl1134.

Eissenberg, J. C. and Elgin, S. C. (2000) 'The HP1 protein family: Getting a grip on chromatin', *Current Opinion in Genetics and Development*, 10(2), pp. 204–210. doi: 10.1016/S0959-437X(00)00058-7.

Eissenberg, J. C. and Elgin, S. C. R. (2014) 'HP1a: A structural chromosomal protein regulating transcription', *Trends in Genetics*. Elsevier Ltd, 30(3), pp. 103–110. doi: 10.1016/j.tig.2014.01.002.

Elango, N. *et al.* (2009) 'Evolutionary rate variation in Old World monkeys', *Biology Letters*, 5(3), pp. 405–408. doi: 10.1098/rsbl.2008.0712.

Ellegren, H. and Parsch, J. (2007) 'The evolution of sex-biased genes and sex-biased gene expression', *Nature Reviews Genetics*, 8(9), pp. 689–698. doi: 10.1038/nrg2167.

Epstein, H., James, T. C. and Singh, P. B. (1992) 'Cloning and expression of Drosophila HP1 homologs from a mealybug , Planococcus citri', *Journal of Cell Science*, 101.

Escribá, M. C. and Goday, C. (2013) 'Histone H3 phosphorylation and elimination of paternal X chromosomes at early cleavages in sciarid flies.', *Journal of Cell Science*. Centro de Investigaciones Biológicas, Departamento de Biología Celular y Molecular, CSIC, Ramiro de Maeztu 9, 28040 Madrid, Spain.: The Company of Biologists Ltd, 126(Pt 14), pp. 3214–3222.

Eskeland, R., Eberharter, A. and Imhof, A. (2007) 'HP1 Binding to Chromatin Methylated at H3K9 Is Enhanced by Auxiliary Factors', *Molecular and Cellular Biology*, 27(2), pp. 453–465. doi: 10.1128/MCB.01576-06.

Falckenhayn, C. *et al.* (2013) 'Characterization of genome methylation patterns in the desert locust Schistocerca gregaria', *Journal of Experimental Biology*. Division of Epigenetics, DKFZ-ZMBH Alliance, German Cancer Research Center, Heidelberg 69120, Germany.: The Company of Biologists Ltd, 216(8), pp. 1423–1429.

Fang, C., Schmitz, L. and Ferree, P. M. (2015) 'An Unusually Simple HP1 Gene Set in Hymenopteran Insects', *Biochemistry and Cell Biology*, 93, pp. 596–603.

Fanti, L. *et al.* (2003) 'Chromosomal distribution of Heterochromatin Protein 1 (HP1) in Drosophila: A cytological map of euchromatic HP1 binding sites', *Genetica*, 117(2–3), pp. 135–147. doi: 10.1023/A:1022971407290.

Feng, S. *et al.* (2010) 'Conservation and divergence of methylation patterning in plants and animals', *Proceedings of the National Academy of Sciences*, 107(19), pp. 8689–8694. doi: 10.1073/pnas.1002720107.

Ferguson-Smith, A. C. (2011) 'Genomic imprinting: the emergence of an epigenetic paradigm', *Nature Reviews Genetics*. Nature Publishing Group, 12(8), pp. 565–575.

Fisher, R. A. (1930) *Genetical Theory of Natural Selection*. Oxford University Press.

Flores, K. *et al.* (2012) 'Genome-wide association between DNA methylation and alternative splicing in an invertebrate', *BMC Genomics*, 13(1), p. 480. doi: 10.1186/1471-2164-13-480.

Foret, S. *et al.* (2009) 'Epigenetic regulation of the honey bee transcriptome: unravelling the nature of methylated genes', *BMC Genomics*, 10(1), p. 472. doi: 10.1186/1471-2164-10-472.

Foret, S. *et al.* (2012) 'DNA methylation dynamics, metabolic fluxes, gene splicing, and alternative phenotypes in honey bees', *Proceedings of the National Academy of Sciences*, 109(13), pp. 4968–4973. doi: 10.1073/pnas.1202392109.

Furuta, K. *et al.* (1997) 'Heterochromatin protein HP1Hsbeta (p25beta) and its localization with centromeres in mitosis.', *Chromosoma*, 106, pp. 11–19. doi: 10.1007/s004120050219.

Gardner, A. (2014) 'Dynamics of sex ratio and female unmatedness under haplodiploidy', *Ecology and Evolution*, 4(9), pp. 1623–1628. doi: 10.1002/ece3.1045.

Gardner, A. and Ross, L. (2014) 'Mating ecology explains patterns of genome elimination', *Ecology Letters*. Edited by M. van Baalen, 17(12), pp. 1602–1612.

Gerbi, S. A. (1986) 'Unusual chromosome movements in sciarid flies', in *Germline - Soma Differentiation*, pp. 71–104.

Glastad, K. M. *et al.* (2011) 'DNA methylation in insects: on the brink of the epigenomic era', *Insect Molecular Biology*, 20(5), pp. 553–565.

Glastad, K. M. *et al.* (2014) 'Epigenetic inheritance and genome regulation : is DNA methylation linked to ploidy in haplodiploid insects ?', *Proceedings of the Royal Society B: Biological Sciences*, (April). doi: 10.1098/rspb.2014.0411.

Glastad, K. M. *et al.* (2014) 'Epigenetic inheritance and genome regulation: is DNA methylation linked to ploidy in haplodiploid insects?', *Proceedings of the Royal Society B: Biological Sciences*, 281(1785), p. 20140411.

Glastad, K. M. *et al.* (2016) 'The caste- and sex-specific DNA methylome of the termite Zootermopsis nevadensis', *Scientific Reports*. Nature Publishing Group, 6(October), pp. 1–14. doi: 10.1038/srep37110.

Glastad, K. M., Hunt, B. G. and Goodisman, M. A. D. (2012) 'Evidence of a conserved functional role for DNA methylation in termites', *Insect Molecular Biology*, 22(2), pp. 143–154.

Glastad, K. M., Hunt, B. G. and Goodisman, M. A. D. (2014) 'ScienceDirectEvolutionary insights into DNA methylation in insects', *Current Opinion in Insect Science*. Elsevier Inc, 1,

pp. 25–30.

Glastad, K. M., Hunt, B. G. and Goodisman, M. A. D. (2015) 'DNA methylation and chromatin organization in insects: Insights from the ant Camponotus floridanus', *Genome Biology and Evolution*, 7(4), pp. 931–942. doi: 10.1093/gbe/evv039.

Goll, M. G. *et al.* (2006) 'Methylation of tRNA Asp by the DNA Methyltransferase Homolog Dnmt2', *Science (New York, N.Y.)*, 311(January), pp. 395–398. doi: 10.1126/science.1120976.

Goll, M. G. and Bestor, T. H. (2005) 'Eukaryotic Cytosine Methyltransferases', *Annual Review of Biochemistry*, 74(1), pp. 481–514. doi: 10.1146/annurev.biochem.74.010904.153721.

Gonzalgo, M. L. and Jones, P. A. (1997) 'Mutagenic and epigenetic effects of DNA methylation', *Mutation Research - Reviews in Mutation Research*, 386(2), pp. 107–118. doi: 10.1016/S1383-5742(96)00047-6.

Grath, S. and Parsch, J. (2016) 'Sex-Biased Gene Expression', *Annual Review of Genetics*, 50(1), pp. 29–44. doi: 10.1146/annurev-genet-120215-035429.

Greil, F. *et al.* (2003) 'Distinct HP1 and Su(var)3-9 complexes bind to sets of developmentally coexpressed genes depending on chromosomal location', *Genes and Development*, 17(22), pp. 2825–2838. doi: 10.1101/gad.281503.

Grunau, C., Clark, S. J. and Rosenthal, A. (2001) 'Bisulfite genomic sequencing: systematic investigation of critical experimental parameters.', *Nucleic acids research*, 29(13), pp. E65-5. doi: 10.1093/nar/29.13.e65.

Guo, W. *et al.* (2017) 'CGmapTools improves the precision of heterozygous SNV calls and supports allele-specific methylation detection and visualization in bisulfite-sequencing data', *Bioinformatics*, (January 2018), pp. 1–7. doi: 10.1093/bioinformatics/btx595.

Haig, D. (1993) 'The evolution of unusual chromosomal systems in coccoids: extraordinary sex ratios revisited', *Journal of Evolutionary Biology*, 6(1), pp. 69–77. doi: 10.1046/j.1420-9101.1993.6010069.x.

Haig, D. (2000) 'The kinship theory of Genomic Imprinting', (34), pp. 9–32.

Hajkova, P. *et al.* (2002) 'Epigenetic reprogramming in mouse primordial germ cells', *Mechanisms of Development*, 117(1–2), pp. 15–23. doi: 10.1016/S0925-4773(02)00181-8.

Hall, E. *et al.* (2014) 'Sex differences in the genome-wide DNA methylation pattern and impact on gene expression, microRNA levels and insulin secretion in human pancreatic islets', *Genome Biol*, 15(12), p. 522. doi: 10.1186/s13059-014-0522-z.

Hamilton, W. D. (1964) 'The genetical theory of social behavior', *Journal of Theoretical Biology*, 7(1), pp. 17–52. doi: 10.1016/0022-5193(64)90039-6.

Heard, E. (2004) 'Recent advances in X-chromosome inactivation', *Current Opinion in Cell Biology*, 16(3), pp. 247–255. doi: 10.1016/j.ceb.2004.03.005.

Heard, E. (2005) 'Delving into the diversity of facultative heterochromatin: The epigenetics of the inactive X chromosome', *Current Opinion in Genetics and Development*, 15(5 SPEC. ISS.), pp. 482–489. doi: 10.1016/j.gde.2005.08.009.

Heard, E. and Disteche, C. M. (2006) 'Dosage compensation of the X chromosome in mammals', *Genes & Development*, 38(1), pp. 47–53. doi: 10.1016/j.ics.2006.06.012.

Hecht, N. B. (1989) 'Mammalian protamines and their expression', in Hnilica, L., Stein, G., and Stein, J. (eds) *Histones and other basic nuclear proteins*. CRC Press, pp. 347–373.

Helleu, Q. *et al.* (2016) 'Rapid evolution of a Y-chromosome heterochromatin protein underlies sex chromosome meiotic drive', *Proceedings of the National Academy of Sciences*, 113(15), pp. 4110–4115. doi: 10.1073/pnas.1519332113.

Hellman, A. and Chess, A. (2007) 'Gene Body – Specific Methylation on the Active X Chromosome', *Science*, 315(5815), pp. 1141–1143. doi: 10.1126/science.1136352.

Hendrich, B. and Tweedie, S. (2003) 'The methyl-CpG binding domain and the evolving role of DNA methylation in animals.', *Trends in genetics : TIG*. Institute for Stem Cell Research, The University of Edinburgh, Roger Land Building, The King's Buildings, Edinburgh EH9 3JQ, UK. Brian.Hendrich@ed.ac.uk: Elsevier Ltd, 19(5), pp. 269–277.

Henter, H. J. (2003) 'Inbreeding Depression and Haplodiploidy: Experimental Measures in a Parasitoid and Comparisons Across Diploid and Haplodiploid Insect Taxa', *Evolution*, 57(8), pp. 1793–1803. doi: 10.1111/j.0014-3820.2003.tb00587.x.

Herrick, G. and Seger, J. (1999) 'Imprinting and paternal genome elimination in insects', *Results and Problems in Cell Differentiation*, 25, pp. 41–71.

Hinkens, D. M., McElfresh, J. S. and Millar, J. G. (2001) 'Identification and synthesis of the sex pheromone of the vine mealybug, Planococcus ficus', *Tetrahedron Letters*, 42(9), pp. 1619–1621. doi: 10.1016/S0040-4039(00)02347-9.

Ho, J. W. K. *et al.* (2014) 'Comparative analysis of metazoan chromatin organisation', *Nature*, 512(7515), pp. 449–452. doi: 10.1021/nn300902w.Release.

Hodson, C. N. *et al.* (2017) 'Paternal genome elimination in Liposcelis booklice (Insecta: Psocodea)', *Genetics*, 206(2), pp. 1091–1100. doi: 10.1534/genetics.117.199786.

Hughes-Schrader, S. (1948) 'Cytology of Coccids (Coccoïdea-Homoptera)', *Advances in Genetics*. Academic Press, 2, pp. 127–203. doi: 10.1016/S0065-2660(08)60468-X.

Hunt, B. G. *et al.* (2010) 'Functional Conservation of DNA Methylation in the Pea Aphid and the Honeybee', *Genome Biology and Evolution*, pp. 1–11.

Hunt, B. G. *et al.* (2013) 'Patterning and Regulatory Associations of DNA Methylation Are Mirrored by Histone Modifications in Insects', *Genome Biology and Evolution*, 5(3), pp. 591–598.

Hunt, B. G. *et al.* (2013) 'The function of intragenic DNA methylation: Insights from insect epigenomes', *Integrative and Comparative Biology*, 53(2), pp. 319–328. doi:

10.1093/icb/ict003.

Huynh, K. D. and Lee, J. T. (2005) 'X-chromosome inactivation: A hypothesis linking ontogeny and phylogeny', *Nature Reviews Genetics*, 6(5), pp. 410–418. doi: 10.1038/nrg1604.

Immler, S. (2008) 'Sperm competition and sperm cooperation: The potential role of diploid and haploid expression', *Reproduction*, 135(3), pp. 275–283. doi: 10.1530/REP-07-0482.

James, T. C. *et al.* (1989) 'Distribution patterns of HP1, a heterochromatin-associated nonhistone chromosomal protein of Drosophila', *European journal of cell biology*, 50(1), p. 170—180.

James, T. C. and Elgin, S. C. (1986) 'Identification of a nonhistone chromosomal protein associated with heterochromatin in Drosophila melanogaster and its gene.', *Molecular and Cellular Biology*, 6(11), pp. 3862–3872. doi: 10.1128/MCB.6.11.3862.

Jamieson, B. (1987) *The Ultrastructure and Phylogeny of Insect Spermatozoa*. Cambridge University Press.

Jamieson, K. *et al.* (2016) 'Loss of HP1 causes depletion of H3K27me3 from facultative heterochromatin and gain of H3K27me2 at constitutive heterochromatin', *Genome Research*, 26(1), pp. 97–107. doi: 10.1101/gr.194555.115.

Jones, D. O., Cowell, I. G. and Singh, P. B. (2000) 'Mammalian chromodomain proteins: Their role in genome organisation and expression', *BioEssays*, 22(2), pp. 124–137. doi: 10.1002/(SICI)1521-1878(200002)22:2<124::AID-BIES4>3.0.CO;2-E.

Kagiwada, S. *et al.* (2013) 'Replication-coupled passive DNA demethylation for the erasure of genome imprints in mice', *EMBO Journal*. Nature Publishing Group, 32(3), pp. 340–353. doi: 10.1038/emboj.2012.331.

Katoh, K., Rozewicki, J. and Yamada, K. D. (2017) 'MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization', *Briefings in Bioinformatics*, (August), pp. 1–7. doi: 10.1093/bib/bbx108.

Kay, S., Skowronski, D. and Hunt, B. G. (2017) 'Developmental DNA methyltransferase expression in the fire ant Solenopsis invicta', *Insect Science*, 6, pp. 6513–6519.

Kazemian, M. *et al.* (2014) 'Evidence for deep regulatory similarities in early developmental programs across highly diverged insects', *Genome Biology and Evolution*, 6(9), pp. 2301–2320. doi: 10.1093/gbe/evu184.

Kearse, M. *et al.* (2012) 'Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data', *Bioinformatics*, 28(12), pp. 1647–1649. doi: 10.1093/bioinformatics/bts199.

Kirkpatrick, M. and Hall, D. W. (2004) 'Sexual Selection and Sex Linkage', *Evolution*, 58(1998), pp. 683–691.

Klopfenstein, D. V. *et al.* (2018) 'GOATOOLS: A Python library for Gene Ontology analyses',

Scientific Reports, 8(1), pp. 1–17. doi: 10.1038/s41598-018-28948-z.

Klose, R. J. and Bird, A. P. (2006) 'Genomic DNA methylation: The mark and its mediators', *Trends in Biochemical Sciences*, 31(2), pp. 89–97. doi: 10.1016/j.tibs.2005.12.008.

Kota, S. K. and Feil, R. (2010) 'Epigenetic Transitions in Germ Cell Development and Meiosis', *Developmental Cell*. Elsevier Inc., 19(5), pp. 675–686. doi: 10.1016/j.devcel.2010.10.009.

Kouzarides, T. (2007) 'Chromatin Modifications and Their Function', *Cell*, 128(4), pp. 693–705. doi: 10.1016/j.cell.2007.02.005.

Kraaijeveld, K. *et al.* (2009) 'Effects of parthenogenesis and geographic isolation on female sexual traits in a parasitoid WASP', *Evolution*, 63(12), pp. 3085–3096. doi: 10.1111/j.1558-5646.2009.00798.x.

Krueger, F. and Andrews, S. R. (2011) 'Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications', *Bioinformatics*, 27(11), pp. 1571–1572.

Krueger, F. and Andrews, S. R. (2016) 'SNPsplit: Allele-specific splitting of alignments between genomes with known SNP genotypes', *F1000Research*, 5, pp. 1415–1479.

Kwon, S. H. and Workman, J. L. (2011) 'The changing faces of HP1: From heterochromatin formation and gene silencing to euchromatic gene expression: HP1 acts as a positive regulator of transcription', *BioEssays*. doi: 10.1002/bies.201000138.

de la Filia, A. G., Bain, S. A. and Ross, L. (2015) 'Haplodiploidy and the reproductive ecology of Arthropods', *Current Opinion in Insect Science*. Elsevier Inc, 9(0), pp. 36–43.

Lawrence, M., Daujat, S. and Schneider, R. (2016) 'Lateral Thinking: How Histone Modifications Regulate Gene Expression', *Trends in Genetics*. Elsevier Ltd, 32(1), pp. 42–56. doi: 10.1016/j.tig.2015.10.007.

Leng, N. *et al.* (2013) 'EBSeq: An empirical Bayes hierarchical model for inference in RNA-seq experiments', *Bioinformatics*, 29(8), pp. 1035–1043. doi: 10.1093/bioinformatics/btt087.

Levine, M. T. *et al.* (2012) 'Phylogenomic Analysis Reveals Dynamic Evolutionary History of the Drosophila Heterochromatin Protein 1 (HP1) Gene Family', *PLoS Genetics*. Edited by C. D. Jones. Public Library of Science, 8(6), pp. e1002729-12.

Levine, M. T. and Malik, H. S. (2013) 'A rapidly evolving genomic toolkit for Drosophila heterochromatin', *Fly*, 7(3). doi: 10.4161/fly.24335.

Levine, M. T. and Malik, H. S. (2015) 'Mitotic fidelity requires transgenerational action of a testis-restricted HP1', *eLife*, pp. 1–20.

Li, B. and Dewey, C. N. (2011) 'RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome', *BMC Bioinformatics*, 12. doi: 10.1186/1471-2105-12-323.

Li, E., Beard, C. and Jaenisch, R. (1993) 'Role for DNA methylation in genomic imprinting', *Nature*, 366(6453), pp. 362–365. doi: 10.1038/366362a0.

Li, E. and Zhang, Y. (2014) 'DNA methylation in mammals', *Cold Spring Harbor Perspectives in Biology*, 6(5). doi: 10.1101/cshperspect.a019133.

Li, N. *et al.* (2010) 'Whole genome DNA methylation analysis based on high throughput sequencing technology', *Methods*. Elsevier Inc., 52(3), pp. 203–212. doi: 10.1016/j.ymeth.2010.04.009.

Li, Y. *et al.* (2003) 'Effects of tethering HP1 to euchromatic regions of the Drosophila genome', *Development*, 130(9), pp. 1817–1824. doi: 10.1242/dev.00405.

Libbrecht, R. *et al.* (2016) 'Robust DNA methylation in the clonal raider ant brain', *Current Biology*. Elsevier Ltd, 26(3), pp. 391–395. doi: 10.1016/j.cub.2015.12.040.

Lister, R. *et al.* (2009) 'Human DNA methylomes at base resolution show widespread epigenomic differences', *Nature*. Nature Publishing Group, 462(7271), pp. 315–322. doi: 10.1038/nature08514.

Livak, K. J. and Schmittgen, T. D. (2001) 'Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2−ΔΔCT Method', *Methods*, 25(4), pp. 402–408.

Lomberk, G., Wallrath, L. L. and Urrutia, R. (2006) 'The Heterochromatin Protein 1 family', *Genome Biology*, 7(7), pp. 1–8. doi: 10.1186/gb-2006-7-7-228.

Lonsdale, Z. *et al.* (2017) 'Allele specific expression and methylation in the bumblebee, *Bombus terrestris*', *PeerJ*, 5, p. e3798. doi: 10.7717/peerj.3798.

Lyko, F. *et al.* (2010) 'The honey bee epigenomes: Differential methylation of brain DNA in queens and workers', *PLoS Biology*, 8(11). doi: 10.1371/journal.pbio.1000506.

Lyko, F. (2017) 'The DNA methyltransferase family: a versatile toolkit for epigenetic regulation', *Nature Reviews Genetics*. Nature Publishing Group. doi: 10.1038/nrg.2017.80.

Lyon, M. F. (1961) 'Gene action in the X-chromosome of the mouse (mus musculus L.)', *Nature*, 190(4773), pp. 372–373. doi: 10.1038/190372a0.

Macke, E., Olivieri, I. and Magalhães, S. (2014) 'Local Mate Competition Mediates Sexual Conflict over Sex Ratio in a Haplodiploid Spider Mite', *Current Biology*. Elsevier Ltd, pp. 1–5.

Mandrioli, M. and Borsatti, F. (2007) 'Analysis of heterochromatic epigenetic markers in the holocentric chromosomes of the aphid Acyrthosiphon pisum', *Chromosome Research*, 15(8), pp. 1015–1022.

Martins, R. F. *et al.* (2014) 'Isolation and characterization of fifteen polymorphic microsatellite loci for the citrus mealybug, Planococcus citri (Hemiptera: Pseudococcidae), and cross-amplification in two other mealybug species', *Journal of genetics*. Springer, 93(1), pp. 75–78.

Maschietto, M. *et al.* (2017) 'Sex differences in DNA methylation of the cord blood are related to sex-bias psychiatric diseases', *Scientific Reports*. Nature Publishing Group, 7, pp. 1–11. doi: 10.1038/srep44547.

Mathers, T. C. *et al.* (2018) 'Sex-specific changes in the aphid DNA methylation landscape', *bioRxiv*.

Mathur, V. *et al.* (2010) 'An analysis of histone modifications in relation to sex-specific chromatin organization in the mealybug maconellicoccus hirsutus', *Cytogenetic and Genome Research*, 129(4), pp. 323–331. doi: 10.1159/000315899.

Meaburn, E. L., Schalkwyk, L. C. and Mill, J. (2010) 'Allele-specific methylation in the human genome: Implications for genetic studies of complex disease', *Epigenetics*, 5(7), pp. 578–582. doi: 10.4161/epi.5.7.12960.

Meiklejohn, C. D. (2016) 'Heterochromatin and genetic conflict', *Proceedings of the National Academy of Sciences*, 113(15), pp. 3915–3917. doi: 10.1073/pnas.1603015113.

Mendel, G. (1865) *Experiments in Plant Hybridisation*, *Verhand- lungen des naturforschenden Vereines in Brünn, Bd. IV für das Jahr*. doi: 10.2307/2528198.

Metz, C. W. (1928) 'Genetic evidence of a selective segregation of chromosomes in a second species of Sciara (Diptera)', *Proceedings of the National Academy of Sciences*, pp. 140–141.

Metz, C. W. (1938) 'Chromosome Behavior, Inheritance and Sex Determination in Sciara', *The American Naturalist*, 72(743), pp. 485–520. doi: 10.1086/280803.

Misof, B. *et al.* (2014) 'Phylogenomics resolves the timing and pattern of insect evolution', *Science*, 346(6210), pp. 763–768. doi: 10.1126/science.1257570.

Mohan, K. N. and Chandra, H. S. (2005) 'Isolation and analysis of sequences showing sex-specific cytosine methylation in the mealybug Planococcus lilacinus', *Molecular Genetics and Genomics*. Springer-Verlag, 274(6), pp. 557–568.

Moore, T. and Haig, D. (1991) 'Genomic imprinting in mammalian development: a parental tug-of-war', *Trends in Genetics*, 7(2), pp. 45–49. doi: 10.1016/0168-9525(91)90230-N.

Murphy, S. K. and Jirtle, R. L. (2003) 'Imprinting evolution and the price of silence', *BioEssays*, 25(6), pp. 577–588. doi: 10.1002/bies.10277.

Niemitz, E. (2013) 'TET function in primordial germ cells', *Nature Genetics*. Nature Publishing Group, 45(3), p. 233. doi: 10.1038/ng.2577.

Normark, B. B. (2003) 'THE EVOLUTION OF ALTERNATIVE GENETIC SYSTEMS IN INSECTS', *Annual Review of Entomology*, 48(1), pp. 397–423. doi: 10.1146/annurev.ento.48.091801.112703.

Normark, B. B. (2004) 'Haplodiploidy as an outcome of coevolution between male-killing cytoplasmic elements and their hosts', *Evolution*, 58(4), pp. 790–798. doi: 10.1111/j.0014-3820.2004.tb00412.x.

Normark, B. B. (2006) 'Perspective: Maternal Kin Groups and the Origins of Asymmetric Genetic Systems—Genomic Imprinting, Haplodiploidy, and Parthenogenesis', *Evolution*, 60(4), p. 631. doi: 10.1554/05-546.1.

Nur, U. (1962) 'Sperms, Sperm Bundles and Fertilization', *Journal of Mor*, pp. 173–199.

Nur, U. (1967) 'Reversal of heterochromatization and the activity of the paternal

chromosome set in the male mealy bug.', *Genetics*, 56(3), pp. 375–389.

Nur, U. (1980) 'Evolution of unusual chromosome systems in scale insects (Coccoidea : Homoptera)', *Insect Cytogenetics*. Blackwell Scientific Publications, pp. 97–117.

Nur, U. *et al.* (1988) 'A "selfish" B chromosome that enhances its transmission by eliminating the paternal genome.', *Science (New York, N.Y.)*, 240(4851), pp. 512–514. doi: 10.1126/science.3358129.

Nur, U. (1990) 'Heterochromatization and euchromatization of whole genomes in scale insects (Coccoidea: Homoptera).', *Development (Cambridge, England). Supplement*, pp. 29–34.

Nur, U. and Brett, B. L. H. (1988) 'Genotypes affecting the condensation and transmission of heterochromtic B chromosomes in the mealybug', *Chromosoma*, pp. 1–8.

Oldroyd, B. P. and Fewell, J. H. (2007) 'Genetic diversity promotes homeostasis in insect colonies', *Trends in Ecology and Evolution*, 22(8), pp. 408–413. doi: 10.1016/j.tree.2007.06.001.

Parisi, M. *et al.* (2003) 'Paucity of genes on the Drosophila X chromosome showing male-biased expression', *Science*, 299(5607), pp. 697–700. doi: 10.1126/science.1079190.

Peters, J. (2014) 'The role of genomic imprinting in biology and disease: An expanding view', *Nature Reviews Genetics*. Nature Publishing Group, 15(8), pp. 517–530. doi: 10.1038/nrg3766.

Plath, K. *et al.* (2002) 'Role of Histone H3 Lysine 27 Methylation in X Inactivation', *Science*, 298(5616), pp. 1039–1043. doi: 10.1126/science.1084274.

Quinlan, A. R. and Hall, I. M. (2010) 'BEDTools: A flexible suite of utilities for comparing genomic features', *Bioinformatics*, 26(6), pp. 841–842. doi: 10.1093/bioinformatics/btq033.

Rae, P. M. M. and Steele, R. E. (1979) 'Absence of cytosine methylation at C-C-G-G and G-C-G-C sites in the rDNA coding regions and intervening sequences of Drosophila and the rDNA of other higher insects', *Nucleic Acids Research*, 6(9), pp. 2987–2995. doi: 10.1093/nar/6.9.2987.

Reeve, H. K. and Pfennig, D. W. (2003) 'Genetic biases for showy males: Are some genetic systems especially conducive to sexual selection?', *Proceedings of the National Academy of Sciences*, pp. 1–6.

Regev, A., Lamb, M. J. and Jablonka, E. (1998) 'The Role of DNA Methylation in Invertebrates: Developmental Regulation or Genome Defense?', *Molecular Biology and Evolution*, 15(7), pp. 880–891. doi: 10.1093/oxfordjournals.molbev.a025992.

Remnant, E. J. *et al.* (2016) 'Parent-of-origin effects on genome-wide DNA methylation in the Cape honey bee (Apis mellifera capensis) may be confounded by allele-specific methylation', *BMC Genomics*. BMC Genomics, 17(1), p. 226. doi: 10.1186/s12864-016-2506-8.

Rens, W. *et al.* (2010) 'Epigenetic modifications on X chromosomes in marsupial and monotreme mammals and implications for evolution of dosage compensation', *Proceedings of the National Academy of Sciences*, 107(41), pp. 17657–17662. doi: 10.1073/pnas.0910322107.

Richards, E. J. (2006) 'Inherited epigenetic variation - Revisiting soft inheritance', *Nature Reviews Genetics*, 7(5), pp. 395–401. doi: 10.1038/nrg1834.

Ronquist, F. and Huelsenbeck, J. P. (2003) 'MrBayes 3: Bayesian phylogenetic inference under mixed models', *Bioinformatics*, 19(12), pp. 1572–1574. doi: 10.1093/bioinformatics/btg180.

Ross, L. *et al.* (2010) 'Sex allocation in a species with paternal genome elimination: the roles of crowding and female age in the mealybug', *Evolutionary Ecology Research*, 12, pp. 89–104.

Ross, L. *et al.* (2011) 'Temperature, age of mating and starvation determine the role of maternal effects on sex allocation in the mealybug Planococcus citri', *Behavioral Ecology and Sociobiology*, 65(5), pp. 909–919. doi: 10.1007/s00265-010-1091-0.

Ross, L. *et al.* (2012) 'Temporal variation in sex allocation in the mealybug Planococcus citri: adaptation, constraint, or both?', *Evolutionary Ecology*, 26(6), pp. 1481–1496.

Ross, L., Pen, I. and Shuker, D. M. (2010) 'Genomic conflict in scale insects: the causes and consequences of bizarre genetic systems', *Biological Reviews*, p. no-no.

Ross, L. and Shuker, D. M. (2009) 'Scale insects', *Current Biology*, 19(5), pp. R184-6. doi: 10.1016/j.cub.2008.12.023.

Rotundo, G. and Trembley, E. (1982) 'Hybridization and sex pheromone responses in two closely related mealybug species (Homoptera: Pseudococcidae)', *Systematic Entomology*. Wiley/Blackwell (10.1111), 7(4), pp. 475–478. doi: 10.1111/j.1365-3113.1982.tb00459.x.

Sacktor, B. (1970) 'Regulation of Intermediary Metabolism, with Special Reference to the Control Mechanisms in Insect Flight Muscle', *Advances in insect physiology*. London, pp. 267–347. doi: 10.1016/S0065-2806(08)60243-X.

Sarda, S. *et al.* (2012) 'The Evolution of Invertebrate Gene Body Methylation', *Molecular Biology and Evolution*, 29(8), pp. 1907–1916.

Saunders, W. S. *et al.* (1993) 'Molecular cloning of a human homologue of Drosophila heterochromatin protein HP1 using anti-centromere autoantibodies with anti-chromo specificity.', *Journal of cell science*, 104 (Pt 2, pp. 573–582.

Schotta, G. *et al.* (2002) 'Central role of Drosophila SU(VAR)3±9 in histone H3-K9 methylation and heterochromatic gene silencing', *The EMBO Journal*, 21(5), pp. 1121–1131.

Schotta, G., Ebert, A. and Reuter, G. (2003) 'SU(VAR)3-9 is a conserved key function in heterochromatic gene silencing', *Genetica*, 117, pp. 149–158.

Schrey, A. W. *et al.* (2016) 'Epigenetic response to environmental change: DNA methylation

varies with invasion status', *Environmental Epigenetics*, 2(2), p. dvw008. doi: 10.1093/eep/dvw008.

Schwartz, Y. B. and Pirrotta, V. (2007) 'Polycomb silencing mechanisms and the management of genomic programmes', *Nature Reviews Genetics*, 8(1), pp. 9–22. doi: 10.1038/nrg1981.

Selker, E. U. *et al.* (2003) 'The methylated component of the *Neurospora crassa* genome', *Nature*, 422(6934), pp. 893–897. doi: 10.1038/nature01564.

Sharman, G. B. (1971) 'Late DNA replication in the paternally derived X chromosome of female kangaroos', *Nature*, 230(5291), pp. 231–232. doi: 10.1038/230231a0.

Shuker, D. M. *et al.* (2005) 'Male influence on sex allocation in the parasitoid wasp Nasonia vitripennis', *Behavioral Ecology and Sociobiology*. Springer-Verlag, 59(6), pp. 829–835.

Shuker, D. M., Moynihan, A. M. and Ross, L. (2009) 'Sexual conflict, sex allocation and the genetic system', *Biology Letters*, 5(5), pp. 682–685. doi: 10.1098/rsbl.2009.0427.

Simpson, V. J., Johnson, T. E. and Hammen, R. F. (1986) 'Caenorhabditis elegans DNA does not contain 5-methylcytosine at any time during development or aging', *Nucleic Acids Research*, 14(16), pp. 6711–6719. doi: 10.1093/nar/14.16.6711.

Singh, P. B. *et al.* (1991) 'A sequence motif found in a Drosophila heterochromatin protein is conserved in animals and plants.', *Nucleic Acids Research*. Department of Molecular Embryology, Institute of Animal Physiology and Genetics Research, Babraham, Cambridge, UK.: Oxford University Press, 19(4), pp. 789–794.

Slater, L. M., Allen, M. D. and Bycroft, M. (2003) 'Structural variation in PWWP domains', *Journal of Molecular Biology*, 330(3), pp. 571–576. doi: 10.1016/S0022-2836(03)00470-4.

Smothers, J. F. and Henikoff, S. (2001) 'The hinge and chromo shadow domain impart distinct targeting of HP1-like proteins.', *Molecular and cellular biology*, 21(7), pp. 2555–69. doi: 10.1128/MCB.21.7.2555-2569.2001.

Sugahara, R. *et al.* (2017) 'Tissue-specific expression and silencing phenotypes of mitochondrial phosphate carrier paralogues in several insect species', *Insect Molecular Biology*, 26(3), pp. 332–342. doi: 10.1111/imb.12297.

Sun, B.-F. *et al.* (2015) 'Regulation of transcription factors on sexual dimorphism of fig wasps', *Scientific Reports*. Nature Publishing Group, 5(1), p. 10696. doi: 10.1038/srep10696.

Supek, F. *et al.* (2011) 'Revigo summarizes and visualizes long lists of gene ontology terms', *PLoS ONE*, 6(7). doi: 10.1371/journal.pone.0021800.

Suzuki, M. M. and Bird, A. (2008) 'DNA methylation landscapes: Provocative insights from epigenomics', *Nature Reviews Genetics*, 9(6), pp. 465–476. doi: 10.1038/nrg2341.

Takagi, N. and Sasaki, M. (1975) 'Preferential inactivation of the paternally derived X chromosome in the extraembryonic membranes of the mouse', *Nature*, 256(5519), pp. 640–642. doi: 10.1038/256640a0.

Úbeda, F. and Normark, B. B. (2006) 'Male killers and the origins of paternal genome elimination', *Theoretical Population Biology*, 70(4), pp. 511–526. doi: 10.1016/j.tpb.2006.06.011.

Vea, I. M. *et al.* (2016) 'Differential juvenile hormone variations in scale insect extreme sexual dimorphism', *PLoS ONE*, 11(2), pp. 1–18. doi: 10.1371/journal.pone.0149459.

Verhulst, E. C., Beukeboom, L. W. and van de Zande, L. (2010) 'Maternal control of haplodiploid sex', *Science*, 328(April), pp. 620–623. doi: 10.1126/science.1185805.

Vermaak, D., Henikoff, S. and Malik, H. S. (2005) 'Positive selection drives the evolution of rhino, a member of the heterochromatin protein 1 family in drosophila', *PLoS Genetics*, 1(1), pp. 0096-0108. doi: 10.1371/journal.pgen.0010009.

Vermaak, D. and Malik, H. S. (2009) 'Multiple Roles for Heterochromatin Protein 1 Genes in *Drosophila*', *Annual Review of Genetics*, 43(1), pp. 467–492. doi: 10.1146/annurev-genet-102108-134802.

Waddington, C. H. (1942) 'The epigenotype', *International journal of epidemiology*, 41(1), pp. 10–13. doi: 10.1093/ije/dyr184.

Wang, X. *et al.* (2013) 'Function and Evolution of DNA Methylation in Nasonia vitripennis', *PLoS Genetics*. Edited by C. Desplan, 9(10), p. e1003872.

Wang, X., Werren, J. H. and Clark, A. G. (2015) 'Genetic and epigenetic architecture of sex-biased expression in the jewel wasps Nasonia vitripennisand giraulti', *Proceedings of the National Academy of Sciences*, 112(27), pp. E3545–E3554.

Wang, X., Werren, J. H. and Clark, A. G. (2016) 'Allele-Specific Transcriptome and Methylome Analysis Reveals Stable Inheritance and Cis-Regulation of DNA Methylation in Nasonia', *PLoS Biology*. Edited by N. H. Barton. Public Library of Science, 14(7), pp. e1002500-21.

Wang, Y. *et al.* (2006) 'Functional CpG Methylation System in a Social Insect', *Science*, 1(13), pp. 645–648.

Waterhouse, A. M. *et al.* (2009) 'Jalview Version 2-A multiple sequence alignment editor and analysis workbench', *Bioinformatics*, 25(9), pp. 1189–1191. doi: 10.1093/bioinformatics/btp033.

Weaver, J. and Bartolomei, M. S. (2014) 'Chromatin regulators of Genomic Imprinting', *Biochimica et Biophysica Acta*, 3, pp. 169–177.

Weber, M. *et al.* (2005) 'Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells', *Nature Genetics*, 37(8), pp. 853–862. doi: 10.1038/ng1598.

Wedd, L., Kucharski, R. and Maleszka, R. (2016) 'Differentially methylated obligatory epialleles modulate context-dependent LAM gene expression in the honeybee Apis mellifera', *Epigenetics*. Taylor & Francis, 11(1), pp. 1–10. doi:

10.1080/15592294.2015.1107695.

Werren, J. H. (1993) 'The Evolution of Inbreeding in Haplodiploid Organisms', in Thornhill, N. W. (ed.) *The Natural History of Inbreeding and Outbreeding: Theoretical and Empirical Perspectives*. University of Chicago Press, pp. 42–59.

Wutz, A., Smrzka, O. W. and Schweifer, N. (1997) 'Imprinted expression of the Igf2r gene depends on an intronic CpG island', 389(October), pp. 745–749.

Xiang, H. *et al.* (2010) 'Single base–resolution methylome of the silkworm reveals a sparse epigenomic map', *Nature Biotechnology*. Nature Publishing Group, 28(5), pp. 516–520.

Yi, S. V and Goodisman, M. A. D. (2009) 'Computational approaches for understanding evolution of DNA methylation in animals', *Epigenetics*, pp. 551–556.

Yoshino, Y. *et al.* (2017) 'DNA methylation changes in intron 1 of Triggering receptor expressed on myeloid cell 2 in Japanese schizophrenia subjects', *Frontiers in Neuroscience*, 11(MAY), pp. 1–7. doi: 10.3389/fnins.2017.00275.

Zahavi, A. (1975) 'Mate selection-A selection for a handicap', *Journal of Theoretical Biology*, 53(1), pp. 205–214. doi: 10.1016/0022-5193(75)90111-3.

Zemach, A. *et al.* (2010) 'Genome-wide evolutionary analysis of eukaryotic DNA methylation', *Science*, 11928(MAY), pp. 916–920.

Zhang, D., Wang, D. and Sun, F. (2010) 'Drosophila melanogaster heterochromatin protein HP1b plays important roles in transcriptional activation and development', *Chromosoma*. Springer-Verlag, 120(1), pp. 97–108.

Zhang, X. *et al.* (2010) 'Methylation of a Single Intronic CpG Mediates Expression Silencing of the PMP24 Gene in Prostate Cancer', *Prostate*, 70(7), pp. 765–776. doi: 10.1021/nn300902w.Release.

APPENDIX A: Microsatellite information

Microsatellite primers for PCR amplification were obtained from (Martins *et al.*, 2012). A panel of 6 multiplexed loci (Pci-7, Pci-16, Pci-17, Pci-21, Pci-22 and Pci-24) was used to genotype hybrid offspring. PCR amplification of microsatellite loci was performed using Type-it Microsatellite PCR kit (QIAgen, The Netherlands) in a 10µl reaction volume containing 1µL of prepGEM reaction product, 5µL of 2x Master Mix, 0.25 µM of the reverse primer and 0.25 µM of each 5' fluorescently-tagged primer. PCR reactions were performed under the following conditions: initial denaturation at 95°C for 5 min, 32 cycles of denaturation at 95°C for 30s, annealing at 55°C for 90s and extension at 72°C for 30s and a final extension step at 60°C for 30min. 1µl of PCR product was sent to Edinburgh Genomics for microsatellite genotyping on the ABI 3730 DNA Analyzer system (ThermoFisher Scientific, United Stated of America) with LIZ 500 as size standard. Microsatellite peaks were scored using the Microsatellite Plugin implemented in Geneious 8.1.3 (Biomatters Ltd., New Zealand) and corrected manually.

Appendix A Table 1. Microsatellite loci (Martins *et al.,* 2012) used in this study: allele richness (N), allele size range and genomic location in *P. citri* and *P. ficus* experimental populations. Table prepared by A. G de la Filia.

Locus	*Planococcus citri*			*Planococcus ficus*		
	N	*Alleles*	*Location**	*N*	*Alleles*	*Location**
Pci-7	3	137, 140, 143	00125 (F 0.01 / R 0.004)	4	137, 140, 143, 146	09272 (F 0.01 / R 0.004)
Pci-16	2	191, 194	01444 (F 0.004 / R 0.01)	1	200	09756 (F 0.004 / R 0.01)
Pci-17	2	197, 200	00790 (F 0.0004 / R 0.02)	1	194	08745 (F 0.02 / R 0.02)
Pci-21	1	288	00133 (F 0.01 / R 0.004)	1	279	01154 (F 0.01 / R 0.004)
Pci-22	1	295	00083 (F 0.01 / R 0.0001)	2	289, 292	00369 (F 0.6 / R 0.0001)
Pci-24	1	168	00250 (F 0.00 / R 0.004)	1	172	03585 (F 0.001 / R 0.004)

* For genomic locations of each locus, scaffold numbers corresponding to best BLAST hits of primer sequences to assemblies PCITRI.V1 and PFICUS.V0 are given. All forward (F) and reverse (R) pairs had best hits to the same scaffold in all loci; E-values are indicated in superscript.

APPENDIX B: Publication related to thesis

de la Filia, A. G*., **Bain, S. A***. and Ross, L. (2015) 'Haplodiploidy and the reproductive ecology of Arthropods', Current Opinion in Insect Science. 9(0), pp. 36–43.

*joint first authors

Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Insect Science

Haplodiploidy and the reproductive ecology of Arthropods

Andrés G de la Filia[1], Stevie A Bain[1] and Laura Ross

Approximately 15% of all arthropods reproduce through haplodiploidy. Yet it is unclear how this mode of reproduction affects other aspects of reproductive ecology. In this review we outline predictions on how haplodiploidy might affect mating system evolution, the evolution of traits under sexual or sexual antagonistic selection, sex allocation decisions and the evolution of parental care. We also give an overview of the phylogenetic distribution of haplodiploidy. Finally, we discuss how comparisons between different types of haplodiploidy (arrhenotoky, PGE with haploid vs *somatically* diploid males) might help to discriminate between the effects of virgin birth, haploid gene expression and those of haploid gene transmission.

Address
School of Biological Sciences, Institute of Evolutionary Biology, University of Edinburgh, Edinburgh EH9 3JT, UK

Corresponding author: Ross, Laura (laura.ross@ed.ac.uk)

[1] Authors contributed equally.

Current Opinion in Insect Science 2015, **9**:36–43

This review comes from a themed issue on **Behavioural ecology**

Edited by **Cécile Lelann** and **Jacintha Ellers**

doi:10.1016/j.cois.2015.04.018

Introduction

Behavioural ecology focuses on understanding how natural selection shapes the way organisms behave. Insects have featured prominently as model systems [1]. Despite providing important general insights, these studies fail to include the full diversity of reproductive systems in arthropods. Most assume that each parent is contributing an equal share of their genes to their offspring. Yet as many as 15% of arthropods are haplodiploids [2–4], where mothers monopolize the production of male offspring, either by the asexual production of sons (arrhenotoky) or by producing sons that eliminate their father's genome from their germline (paternal genome elimination, PGE) [4]. Haplodiploidy has received attention in the context of eusociality (though its importance has increasingly fallen out of favour [5]), yet how it affects other aspects of species' ecology has barely been addressed. Here we consider its role in reproductive behaviour and mating system evolution. We summarize available theory (main text and Table 1) and empirical data (supplementary Table S1), provide verbal models when formal ones are lacking, and identify areas that need addressing in the future.

Most biologists are familiar with haplodiploidy in the Hymenoptera. The vast majority of hymenopterans reproduce through arrhenotoky [4], and most behavioural ecology studies on haplodiploid species involve members of this order. Yet, it constitutes just one of two-dozen independent origins of haplodiploidy [4]: arrhenotoky is also found among thrips, some hemipterans and several clades of beetles and mites. PGE, where males develop from fertilized eggs but subsequently eliminate the paternal chromosomes, is found in most scale insects (Hemiptera), some beetles, flies, springtails, lice and mites (in total about 20 000 species) [6] (Figures 1 and 2). Different PGE species vary in the timing of the elimination of the paternal genome, and in whether it becomes transcriptionally silenced or not [4,6]. As a result, male gene expression varies from haploid to diploid (Figure 3) with various intermediates. This variability is important as it might allow differentiation of the effects of haploid gene transmission and those of haploid gene expression, while comparisons between arrhenotokous and PGE taxa could provide insights into the importance of virgin birth (see Table 1).

Evolutionary genetics under haplodiploidy

Haplodiploidy affects the evolutionary genetics of species in a number of ways. Under arrhenotoky and some types of PGE, gene expression in males is haploid and maternal. Therefore, recessive mutations are exposed to selection in males, firstly, reducing genetic load, due to a lower effective mutation rate and the exposure of deleterious recessive alleles in haploid males [7] and secondly, increasing the rate at which rare recessive beneficial mutations can spread. As a result, these species are expected to adapt faster to changing environments. This is true only for non-sex specific traits. The evolution of male-limited traits is complex, as sons do not inherit them from their fathers (Figure 3). In addition, selection among females has a relatively greater impact on evolutionary change as each gene finds itself more frequently in females than males [8].

Sexual selection

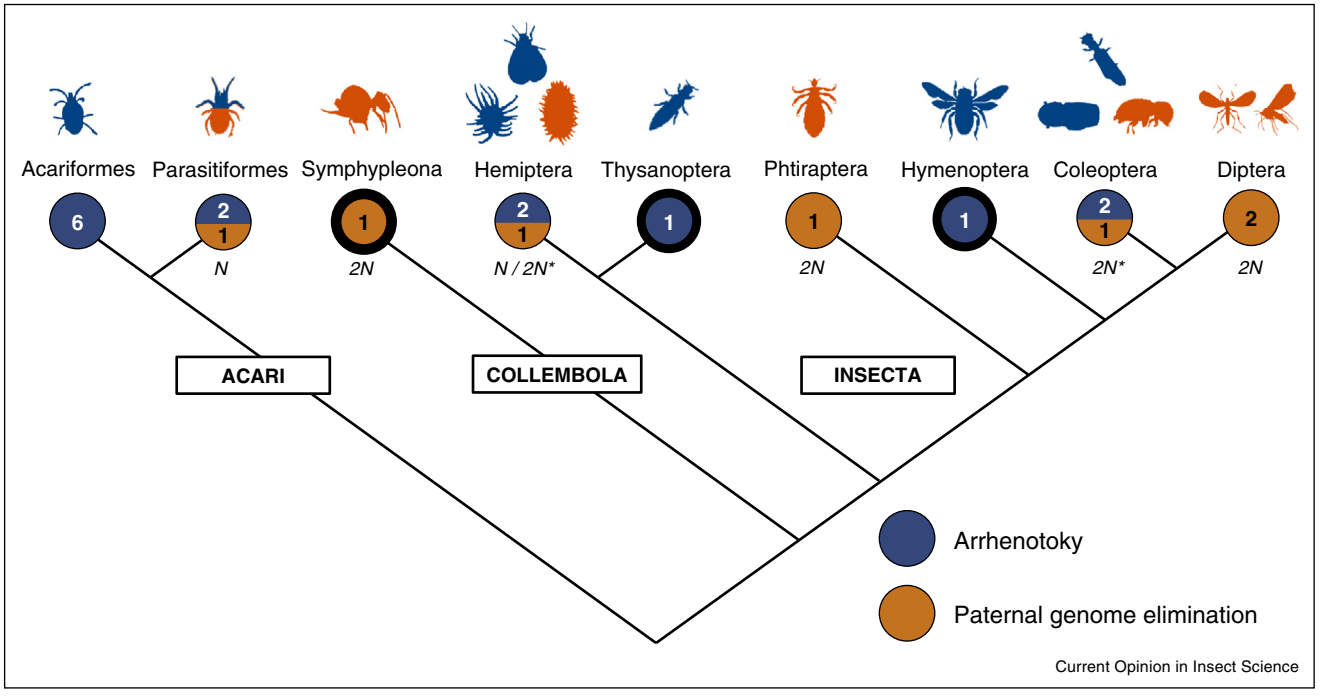
Sexual selection arises through competition within a sex (usually males) for access to mates (and their gametes) [9]

Table 1

An overview of the prediction for each of the reproductive ecology traits discussed in the manuscript. We summarize how we expect the three different types of haplodiploid organisms to differ with respect to diplodiploid taxa. + indicates that trait is promoted relative to diplodiploidy, − that the type of haplodiploidy inhibits the evolution of the trait, while = indicates that there is no expected difference between haplodiploids and diplodiploids. Please note that most of these predictions, especially differences between the different types of haplodiploids, are based on verbal models and will need to be corroborated by formal theory in the future.

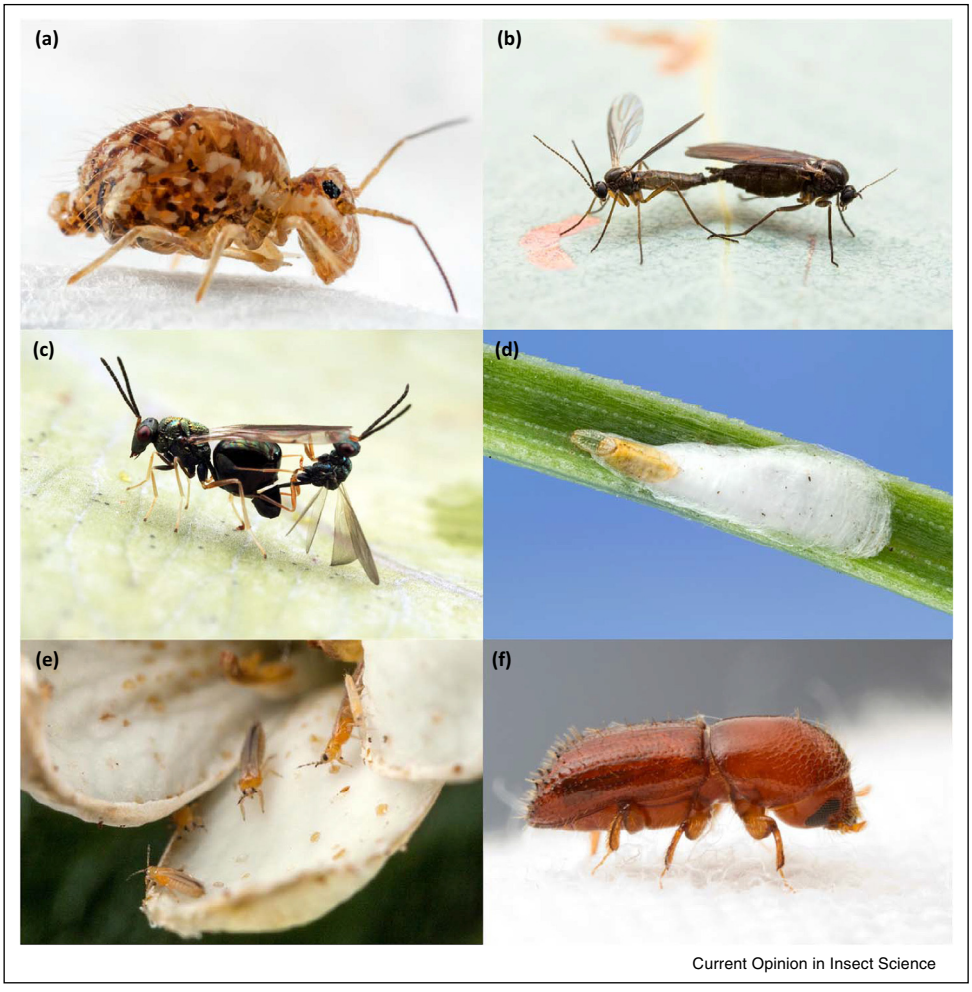
	Arrhenotoky	PGE (haploid soma)	PGE (diploid soma)
Inbreeding depression	− [24]	−	=
Exaggerated sexual selected traits (under Fisherian runaway selection)	− [12]	−	−/= Sons might express their fathers' ornament, increasing their reproductive success, yet are unable to pass it on to their offspring
Exaggerated sexual selected traits (handicap principle)	+ [13]	+	+
Intra-locus conflicts	Resolved in favour of females (dominant traits), polymorphism (recessive traits) [16]	Resolved in favour of females (dominant traits), polymorphism (recessive traits)	Resolved in favour of females (both dominant and recessive traits).
Inter-locus conflicts	Females are more likely to evolve resistance to male trait [16]	Females are more likely to evolve resistance to male trait	= (?) Mothers equally likely to evolve resistance as sons might express their fathers' trait, benefitting the mother through their increased reproductive success
Sperm cooperation	+ [52]	+ [52]	+ [52]
Fertility assurance	+	− Virgin females unable to produce offspring	− Virgin females unable to produce offspring
Facultative sex ratio control	+ [33,53]	+/= [54,55]	+/= [56••]
Polyandry	−/= [11••]	+	+
Maternal care	= [49]	=	=
Paternal care	+ [50••]	+	+

Figure 1



Schematic cladogram of arrhenotokous (blue) and PGE (orange) groups in Arthropoda. The number of independent origins of haplodiploidy is indicated within the circles. Clades in which all members are haplodiploid are indicated with a black ring around the circle. The type of PGE is indicated below the circle with *2N* for germline PGE, *2N** for germline PGE, where the paternal genome is transcriptionally silenced in somatic cells and *N* for embryonic PGE. Origins outside the Arthropoda (rotifers and nematodes) are not shown.

Figure 2

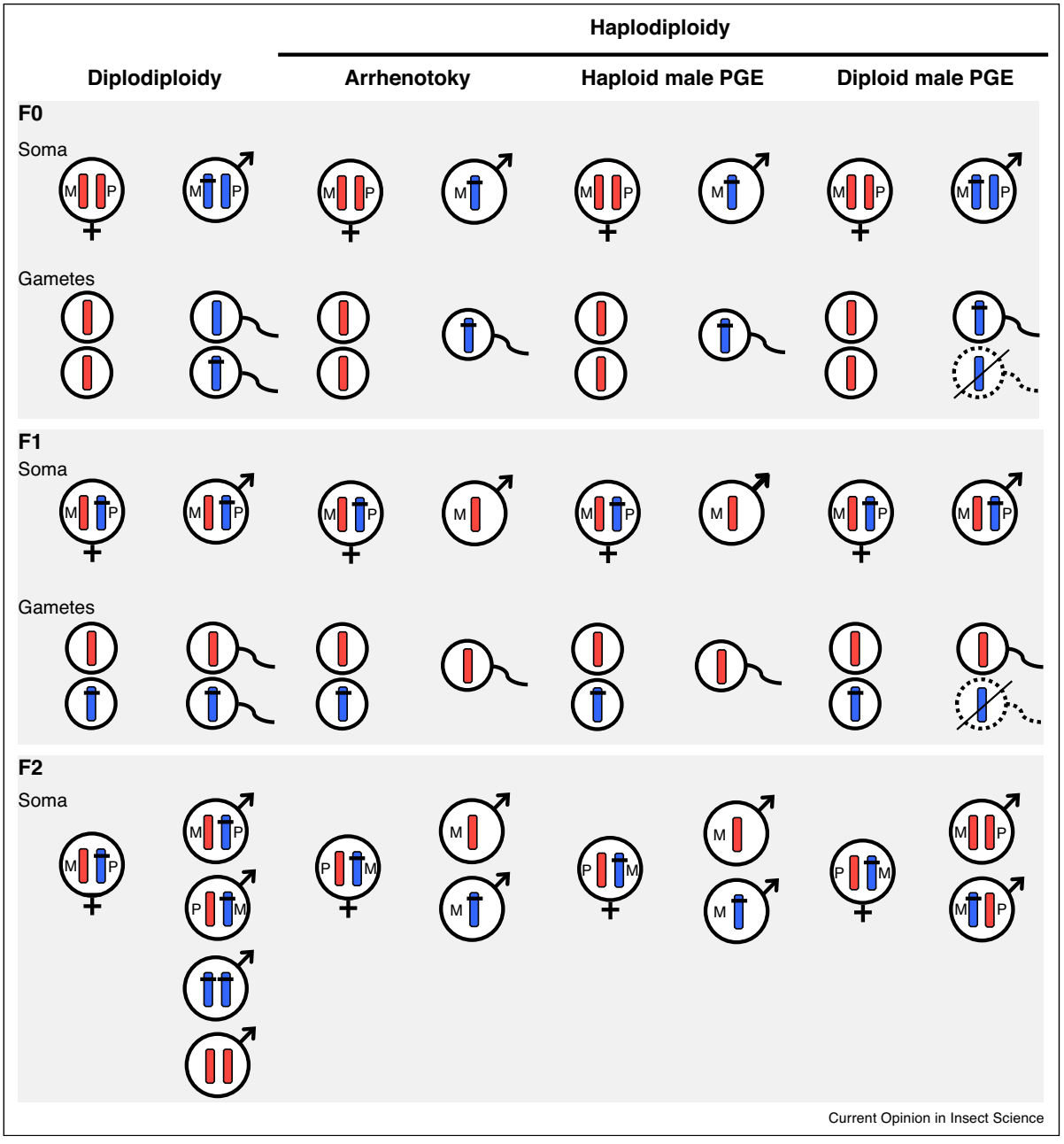


A number of examples of PGE species: **(a)** a globular springtail (Symphypleona), **(b)** a pair of mating fungus gnats (Sciaridae), **(d)** the armoured scale insect *Chionaspis pinifoliae.* And a number of arrhenotokous species: **(c)** Eucharitid wasps mating, **(e)** flower thrips, **(f)** *Xyleborus* sp. ambrosia beetle.
Source: Images b-f © Alex Wild and image a © Gil Wizen, used with permission.

and can result in the evolution of exaggerated traits. Such traits evolve if females chose to mate with males carrying them, either because the trait signals genetic quality (the handicap principle) [10], or because their sons will inherit it and therefore be attractive to other females (Fisherian runaway selection) [9].

As haplodiploid sons do not inherit traits from their fathers their maternal grandfathers are their closest male progenitors, so that selection on male traits skips generations (Figure 3) [11••]. A simulation study [12] showed that, due to this delay, rare alleles encoding male ornaments are particularly likely to be lost through genetic drift. The same might be true for alleles underlying traits that increase a male's reproductive success without being a direct target of female choice, such as combat ability. Subsequent deterministic models showed that haplodiploid transmission genetics also affects the genetic correlation between male traits and female preference, thereby promoting sexual selection through the handicap principle, but impeding Fisherian runaway selection [13]. Together, these models suggest that haplodiploidy should affect the evolution of exaggerated male traits. Comparative efforts to identify the prevalence of such traits and the degree of sexual dimorphism between haplodiploid and diplodiploid species might therefore, in principle, provide insight into the relative importance of runaway versus handicap selection. However these predictions are based on a number of simplifying assumptions, and there is an urgent need for more formal theory considering finite population sizes, costs of female preference, sex-specific mutation rates and allelic dominance.

Figure 3



Genetic inheritance of a paternal allele under diplodiploidy and the three different types of haplodiploidy. Somatic genotypes are represented for three generations (F0, F1 and F2) and gamete genotypes for F0 and F1. (For simplicity, assume that there is no meiotic recombination and that offspring mate to produce the next generation.) Maternal chromosomes in F0 are shown in pink and paternal chromosomes in F0 are shown in blue. Black lines in the upper half of some chromosomes represent a given male trait (e.g., an advantageous trait in inter-locus conflict or sexual selection), whose inheritance we follow across three generations. M, inherited from the mother P, inherited from the father. Colours refer to maternal (red) and paternal (blue) genomes in F0. In F1 and F2, only one out of four possible female soma genotypes are shown, while all possible male soma genotypes carrying the original paternal allele are indicated for both generations. The figure shows how transmission of the male trait is affected by the different genetic systems. Paternal line inheritance is possible under diplodiploidy only. Due to the fact that arrhenotokous males develop from unfertilized eggs, they do not inherit paternal chromosomes. In haploid male PGE, the situation is similar because paternal chromosomes are eliminated in the early developmental stages of the zygote. Also, in certain groups with diploid male PGE, such as in Neococcidae, the paternal genome is transcriptionally repressed and F1 males will not express the male trait. In these three scenarios, selection does not act upon this trait in F1 males.

Sexual conflict

Sexual conflicts result from the discordance of genetic interests between the sexes; males and females might differ in their optimal allele state or expression level at the same gene (intra-locus conflict) [14] or different loci (inter-locus conflict) [15••]. Specific theory on the role of haplodiploidy is scarce, but analogies to X-chromosome inheritance allow us to utilize theory of sexual conflict under sex linkage [16]. As haplodiploid males obtain reproductive success only through daughters, male-beneficial traits that reduce female fitness are particularly unlikely to spread [16]. Intra-locus conflicts, in particular, will tend to be resolved in favour of females. Only if the trait is recessive and its effects are masked in females, could a polymorphism with a female advantageous allele arise [14], though not under PGE with somatically diploid males. By contrast, inter-locus conflict, especially over sex-limited traits, will not necessarily be resolved in favour of females. For example, a trait that reduces female fecundity but increases sperm competitive ability is as likely to spread under haplodiploidy as under diplodiploidy [15••], yet, while diplodiploid females could benefit indirectly, through sons inheriting it, haplodiploid mothers cannot. As a result haplodiploid females are more likely to evolve resistance mechanisms [16]. Unfortunately there are few empirical studies on either intra-locus or inter-locus sexual conflicts and the predictions outlined here remain to be corroborated.

Sperm cooperation

In diploids, each individual sperm carries a unique haploid genome, different from the diploid genome of the male [17, 18••,19]. Under haplodiploidy sperm are produced mitotically, so individual sperm are genetically identical (barring mutations). As a result, there might be more scope for sperm cooperation, especially under post-copulatory sexual selection [18••]. Empirical data on sperm behaviour under haplodiploidy are limited. However, a peculiar type of sperm cooperation has been found among scale insects with PGE. In this group, individual sperm cells have lost their motility, which they regain by assembling into motile sperm bundles, consisting of tens or even hundreds of sperm cells [20].

Mating systems and inbreeding

There is a strong empirical association between the occurrence of haplodiploidy and certain mating systems, especially those in which inbreeding is systematic (Table S1). Examples include arrhenotokous species such as many parasitoid wasps or bark and ambrosia beetles [21,22], and PGE species such as the coffee-borer beetle [23]. Haplodiploids are more resistant to inbreeding depression due to their reduced genetic load [24,25••,26,27]. However, inbreeding can be detrimental under some conditions: hymenopterans with complementary sex determination (CSD) are greatly affected, as inbreeding produces sterile diploid homozygous males [28]. In PGE species in which the paternal genome is transcriptionally active [29] (diploid male PGE, Figure 3) deleterious recessive alleles are not exposed to selection, so such species are expected to suffer from substantial inbreeding depression. Some of them might have evolved monogeny (where all offspring of each individual female are either exclusively male or exclusively female) as an elaborate mechanisms to avoid inbreeding [30]. Finally we expect substantial inbreeding depression in females under all types of haplodiploidy when inbreeding depression is caused by genes with female-limited expression [27].

Another aspect that could have strong implications on mating systems of arrhenotokous, but not PGE, species is that unmated females can still reproduce by producing all-son broods, which could result in relaxed selection for mate-finding traits compared to diploid/PGE females [31], or allow females to be more choosy. The capacity for virgin birth might also make arrhenotokous females good colonizers: a single arrhenotokous female could theoretically establish a population by producing sons and mating with them. Sex ratio control under arrhenotoky allows for the female-biased sex ratios favoured under such conditions [32,33]. Empirical support comes from ambrosia beetles, where incestuous arrhenotokous species are predominant over diploid outbreeding species with similar ecology on remote islands [34].

Finally, haplodiploidy might affect female mating rates. Monogamy has received considerable attention in the Hymenoptera as an important pre-requisite for the evolution of eusociality. Yet, although there is a huge literature on the link between haplodiploidy and eusociality, few authors have discussed whether monogamy is more or less prevalent among haplodiploids (although see [11••,31,35]). Females are thought to mate multiply to obtain either direct (nuptial gifts, replenishment of sperm supplies) or indirect benefits (promote genetic diversity, increase probability of genetic compatibility) [11••]. In theory, haplodiploidy could affect both. Arrhenotokous females use sperm only to fertilize their female eggs and are able to produce sons without sperm. As a result they might both be less likely to become sperm depleted, and to suffer low reproductive success [22,31,36]. In addition, because the cost of remaining unmated is less severe, females can afford to be choosier about whom to mate with. This is not expected under PGE, as females require sperm to fertilize zygotes of both sexes. In terms of indirect genetic effects, both PGE and arrhenotokous females produce broods that are less genetically diverse on average than diplodiploid females do. They might, therefore, be selected to compensate for this by multiple mating. This is supported by various studies on haplodiploid obligately eusocial species [37,38]. To summarize, female remating rates might vary substantially among haplodiploids, but the relative balance between direct and indirect benefits suggests that, on average, they

would be lowest for arrhenotokous female and highest for those with PGE.

Sex allocation

Sex allocation is perhaps the only aspect of insect reproductive behaviour where studies on haplodiploid species are over-represented. The ability of haplodiploids to precisely alter the sex ratio of their offspring is well documented [39]. Increased control over sex allocation is obvious in haplodiploid taxa, where, unlike under genetic sex determination in diploids, there is no default sex ratio of 50:50. Sex ratio control might have allowed haplodiploid species to evolve a wide range of mating systems and promote alloparental brood care, as mothers are able to bias their sex ratio towards the more helpful sex [40]. However, this flexibility might come at the cost of increased conflicts over sex allocation: First of all, sexual conflict arises between parents over the sex ratio of their offspring [41,42]. Haplodiploid mothers generally favour an equal investment into each sex [43]. Yet fathers, who are not related to male offspring, favour a strongly female biased sex ratio and may evolve ways to persuade their partner to increase fertilization rates (under arrhenotoky) or manipulate the sex determining mechanism (under PGE). Support for the possibility that arrhenotokous males can, under some conditions, manipulate sex allocation decisions of their partners comes from parasitoid wasps [44,45], and spider mites [46••]. Although no studies have yet considered male influence on sex allocation under PGE, it might be more likely to occur as fathers' genes are present in sons [41].

Haplodiploidy might also lead to conflicts among siblings over sex allocation, and between parents and offspring, in those species where siblings interact. Under haplodiploidy, a female is more closely related to her sisters than to her brothers, and should favour a more female-biased sex ratio. The occurrence of these conflicts and how they are resolved has been studied extensively in the eusocial Hymenoptera [47] but have received less attention in other taxa where they are expected to occur, such as social thrips and mites.

Parental care

Another aspect of insect reproduction that varies dramatically between species is the presence of parental care and the relative energy expenditure males and females devote to caring for their young. A number of studies have investigated how haplodiploidy affects the evolution of paternal versus maternal care. A population genetic model by Wade [48] suggested that haplodiploidy facilitates the evolution of maternal care, but assumed that the cost of maternal care rests on both parents, not just on the mother. A subsequent model [49] included the latter possibility as well as effects of inbreeding and alternative assumptions about the genetic underpinning of the maternal care. The result of this model suggests that haplodiploidy does not generally promote maternal care.

What about paternal care? Intuitively, haplodiploidy might be expected to inhibit paternal care as males are selected to care only for their female offspring. However, although fathers value their sons less than under diplodiploidy, they value their daughters more, and these two effects exactly cancel [50••]. So under outbreeding, haplodiploidy neither promotes nor inhibits paternal care. By contrast, under inbreeding haplodiploidy may promote paternal care, as it inflates a male's relatedness to his offspring more than under diploidy [50••]. This suggests that paternal care might be overrepresented in haplodiploids, although empirical support is ambiguous (Table S1). Exclusively paternal care is rare among insects (probably for reasons unrelated to ploidy). The only clear examples are found in three (sub)families of diploid Hemipterans and one family of haplodiploid and strong inbreeding thrips [51], which seem to fit the model well.

One important assumption of these models [48–50••] is that parents are unable to preferentially care for the offspring to which sex they are most related. Under outbreeding, haplodiploid females are equally related to both offspring sexes, but fathers are related only to daughters. Therefore, if males are able to preferentially care for their daughters, paternal care might be promoted. The same might be true for maternal care under inbreeding, as mothers become more related to their daughters than to their sons.

Conclusion

Haplodiploid reproduction is widespread among arthropods. Males either do not inherit any genes from their fathers or, if they do, they fail to pass them on to their offspring. Here we discussed how haplodiploidy can profoundly alter mating system evolution, sex allocation and the evolution of traits under sexual selection or sexual antagonism. These predictions are not just significant to understanding the evolution of haplodiploid taxa, but, in comparison, with diplodiploid taxa, could provide more general insights into these phenomena. Unfortunately, predictions are overwhelmingly based on verbal or very simplistic models, and much of the formal theory that is available was developed for other purposes: either extrapolated from models of X-linkage or motivated by a presumed link with eusociality and therefore tailored specifically to hymenopterans. It is clear that more formal theoretical effort is needed. A particular challenge will be to address how the different types of haplodiploidy (arrhenotoky and PGE with haploid or somatically diploid males) could help dissect the relative importance of the ability of virgin birth and effects of haploid gene expression or transmission. In Table 1, we present how we expect them to affect each of the traits discussed in

the manuscript. As each of the three types of haplodiploidy has evolved repeatedly, these predictions lend themselves well to a formal phylogenetic comparative approach. Data presented in Table S1 could serve as an excellent starting point. Finally, there is scope for a multitude of empirical tests to test predictions.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.cois.2015.04.018.

Acknowledgements

We would like to thank Deborah Charlesworth, Koos Boomsma, David Shuker, Andy Gardner, Ben Normark and two anonymous reviewers for their valuable comments and encouragement. SAB is supported by a BBSRC Eastbio DTP, AGF by the Darwin Trust and LR by a NERC IRF.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Shuker D, Simmons L: *The Evolution of Insect Mating Systems*. Oxford University Press; 2014.

2. Bachtrog D, Mank JE, Peichel CL, Kirkpatrick M, Otto SP, Ashman T-L, Hahn MW, Kitano J, Mayrose I, Ming R *et al.*: **Sex determination: why so many ways of doing it?** *PLoS Biol* 2014, **12**:e1001899 Extensive review on the diversity and evolution sex determining systems across different organisms.

3. Normark BB: **Modes of reproduction**. *The Evolution of Insect Mating Systems*. Oxford University Press; 2014:: 1-19.

4. Normark BB: **The evolution of alternative genetic systems in insects**. *Annu Rev Entomol* 2003, **48**:397-423 The most comprehensive review of the genetic/reproductive systems found in insects.

5. Ross L, Gardner A, Hardy N, West SA: **Ecology, not the genetics of sex determination, determines who helps in eusocial populations**. *Curr Biol* 2013, **23**:2383-2387.

6. Gardner A, Ross L: **Mating ecology explains patterns of genome elimination**. *Ecol Lett* 2014, **17**:1602-1612 Theoretical model of the evolution of paternal genome elimination under inbreeding.

7. Werren JH: **The evolution of inbreeding in haplodiploid organisms**. *The Natural History of Inbreeding and Outbreeding: Theoretical and Empirical Perspectives*. University of Chicago Press; 1993.

8. Price GR: **Selection and covariance**. *Nature* 1970, **227**:520-521.

9. Fisher RA: *The Genetical Theory of Natural Selection*. Clarendon Press; 1930.

10. Zahavi A: **Mate selection — a selection for a handicap**. *J Theor Biol* 1975, **53**:205-214.

11. •• Boulton RA, Collins LA, Shuker DM: **Beyond sex allocation: the role of mating systems in sexual selection in parasitoid wasps**. *Biol Rev Cambridge Phil Soc* 2014 http://dx.doi.org/10.1111/brv.12126.
A comprehensive review of research on sexual selection in parasitoid wasps.

12. Reeve HK, Pfennig DW: **Genetic biases for showy males: are some genetic systems especially conducive to sexual selection?** *Proc Natl Acad Sci* 2003, **100**:1089-1094.

13. Kirkpatrick M, Hall DW: **Sexual selection and sex linkage**. *Evolution* 2004, **58**:683-691.

14. Albert AYK: **Sexual selection can resolve sex-linked sexual antagonism**. *Science* 2005, **310**:119-121.

15. •• Andrés JA, Morrow EH: **The origin of interlocus sexual conflict: is sex-linkage important?** *J Evol Biol* 2003, **16**:219-223.
A theoretical study on sex-linked interlocus sexual conflict that is applicable to haplodiploidy.

16. Kraaijeveld K: **Male genes with nowhere to hide; sexual conflict in haplodiploids**. *Anim Biol* 2009, **59**:403-415 Comprehensive review and theory paper on sexual conflict under haplodiploidy.

17. Parker GA, Begon ME: **Sperm competition games: sperm size and number under gametic control**. *Proc R Soc Lond B Biol* 1993, **253**:255-262.

18. •• Immler S: **Sperm competition and sperm cooperation: the potential role of diploid and haploid expression**. *Reproduction* 2008, **135**:275-283.
Review paper that discusses the selective pressures that determine sperm cooperation and conflict under diplodiploidy as well as haplodiploidy.

19. Pizzari T, Foster KR: **Sperm sociality: cooperation, altruism, and spite**. *PLoS Biol* 2008, **6**:e130.

20. Jamieson BGM: *The Ultrastructure and Phylogeny of Insect Spermatozoa*. CUP Archive; 1987.

21. Keller L, Peer K, Bernasconi C, Taborsky M, Shuker DM: **Inbreeding and selection on sex ratio in the bark beetle *Xylosandrus germanus***. *BMC Evol Biol* 2011, **11**:359.

22. Gottlieb D, Lubin Y, Harari AR: **The effect of female mating status on male offspring traits**. *Behav Ecol Sociobiol* 2014, **68**:701-710.

23. Borsa P, Kjellberg F: **Experimental evidence for pseudo-arrhenotoky in *Hypothenemus hampei* (Coleoptera: Scolytidae)**. *Heredity* 1996, **76**:130-135.

24. Werren JH: **The evolution of inbreeding in haplodiploid organisms**. In *The Natural History of Inbreeding and Outbreeding: Theoretical and Empirical Perspectives.* Edited by Thornhill NW. University of Chicago Press; 1993.

25. •• Henter H: **Inbreeding depression and haplodiploidy: experimental measures in a parasitoid and comparisons across diploid and haplodiploid insect taxa**. *Evolution* 2003, **57**:1793-1803.
A comprehensive experimental and comparative analysis of inbreeding depression across insect taxa.

26. Antolin MF: **A genetic perspective on mating systems and sex ratios of parasitoid wasps**. *Res Popul Ecol* 1999, **41**:29-37.

27. Tien NSH, Sabelis MW, Egas M: **Inbreeding depression and purging in a haplodiploid: gender-related effects**. *Heredity* 2015, **114**:327-332.

28. Heimpel GE, de Boer JG: **Sex determination in the hymenoptera**. *Annu Rev Entomol* 2008, **53**:209-230.

29. Sánchez L: **Sciara as an experimental model for studies on the evolutionary relationships between the zygotic, maternal and environmental primary signals for sexual development**. *J Genet* 2010, **89**:325-331.

30. Tabadkani SM, Khansefid M, Ashouri A: **Monogeny, a neglected mechanism of inbreeding avoidance in small populations of gall midges**. *Entomol Exp Appl* 2011, **140**:77-84.

31. Macke E, Magalhães S, Khanh HD-T, Frantz A, Facon B, Olivieri I: **Mating modifies female life history in a haplodiploid spider mite**. *Am Nat* 2012, **179**:E147-E162 http://dx.doi.org/10.1086/665002.

32. Adamson M, Ludwig D: **Oedipal mating as a factor in sex allocation in haplodiploids**. *Phil Trans R Soc Lond B: Biol Sci* 1993, **341**:195-202.

33. Hamilton WD: **Extraordinary sex ratios**. *Science* 1967, **156**:477-488.

34. Jordal BH, Beaver RA, Kirkendall LR: **Breaking taboos in the tropics: incest promotes colonization by wood-boring beetles**. *Global Ecol Biogeogr* 2001, **10**:345-357.

35. Ridley M: **Clutch size and mating frequency in parasitic Hymenoptera**. *Am Nat* 1993, **142**:893-910.

36. Rautiala P, Helanterä H, Puurtinen M: **Unmatedness promotes the evolution of helping more in diplodiploids than in haplodiploids**. *Am Nat* 2014, **184**:318-325.

37. Baer B, Schmid-Hempel P: **Experimental variation in polyandry affects parasite loads and fitness in a bumble-bee**. *Nature* 1999 http://dx.doi.org/10.1038/16451.

38. Oldroyd BP, Fewell JH: **Genetic diversity promotes homeostasis in insect colonies**. *Trends Ecol Evol* 2007, **22**:408-413.

39. Bull JJ: *The Evolution of Sex Determining Mechanisms*. Benjamin Cummings; 1983.

40. Gardner A, Ross L: **Haplodiploidy, sex-ratio adjustment, and eusociality**. *Am Nat* 2013, **181**:E60-E67.

41. Shuker D, Moynihan A, Ross L: **Sexual conflict, sex allocation and the genetic system**. *Biol Lett* 2009, **5**:682.

42. Shuker DM, Cook N: **Evolution: conflict by the sexes, for the sexes**. *Curr Biol* 2014, **24**:R1135-R1137.

43. Gardner A: **Total reproductive value of juvenile females is twice that of juvenile males under X-linkage and haplodiploidy**. *J Theor Biol* 2014, **359**:236-237.

44. Shuker D, Sykes EM, Browning LE, Beukeboom L, West S: **Male influence on sex allocation in the parasitoid wasp *Nasonia vitripennis***. *Behav Ecol Sociobiol* 2006, **59**:829-835.

45. King BH, Kuban KA: **Should he stay or should he go: male influence on offspring sex ratio via postcopulatory attendance**. *Behav Ecol Sociobiol* 2012, **66**:1165-1173.

46. •• Macke E, Olivieri I, Magalhães S: **Local mate competition mediates sexual conflict over sex ratio in a haplodiploid spider mite**. *Curr Biol* 2014, **24**:2850-2854.
Experimental evolution study that provides support that haplodiploid males can alter the sex ratio his mate produces.

47. Ratnieks FLW, Boomsma JJ: **Facultative sex allocation by workers and the evolution of polyandry by queens in social Hymenoptera on JSTOR**. *Am Nat* 1995, **145**:969-993.

48. Wade MJ: **Maternal effect genes and the evolution of sociality in haplo-diploid organisms**. *Evolution* 2001, **55**:453-458.

49. Gardner A: **Evolution of maternal care in diploid and haplodiploid populations**. *J Evol Biol* 2012, **25**:1479-1486
Theoretical model exploring how haplodiploidy can affect the evolution of maternal care.

50. •• Davies NG, Gardner A: **Evolution of paternal care in diploid and haplodiploid populations**. *J Evol Biol* 2014, **27**:1012-1019.
Theoretical model exploring how haplodiploidy can affect the evolution of paternal care.

51. Tallamy DW: **Evolution of exclusive paternal care in arthopods**. *Annu Rev Entomol* 2001, **46**:139-165.

52. Normark BB: **Unusual gametic and genetic systems**. In *Sperm Biology: An Evolutionary Perspective.* Edited by Hosken DJ, Birkhead T. Academic Press; 2009.

53. Wells HG, Huxley JS, Wells GP: *The Science of Life*. Amalgamated Press; 1929.

54. Varndell N, Godfray H: **Facultative adjustment of the sex ratio in an insect (*Planococcus citri*, Pseudococcidae) with paternal genome loss**. *Evolution* 1996, **50**:2100-2105.

55. Ross L, Langenhof M, Pen I, Beukeboom L, West S, Shuker D: **Sex allocation in a species with paternal genome elimination: the roles of crowding and female age in the mealybug *Planococcus citri***. *Evol Ecol Res* 2010, **12**:89-104.

56. •• Featherston R, Jones TM, Elgar MA: **Female resistance behaviour and progeny sex ratio in two Bradysia species (Diptera: Sciaridae) with paternal genome elimination**. *J Evol Biol* 2013, **26**:919-928.
One of only a few experimental studies on the reproductive ecology of a species with PGE.