University *of*
Leicester

DNA methylation, genomic imprinting and polyphenism in the bumblebee, *Bombus terrestris*

Thesis submitted for the degree of

Doctor of Philosophy

at the University of Leicester

by

Harindra Amarasinghe BSc. (Hons)

Department of Biology

University of Leicester

March 2015

DNA methylation, genomic imprinting and polyphenism in the bumblebee, *Bombus terrestris*

Harindra Amarasinghe

Abstract

Genomic imprinting, the parent-of-origin specific expression of alleles is an important area of research in evolutionary biology and human health (cancers and developmental syndromes). Haig's kinship theory suggests that the maternally and paternally derived alleles of offspring resource allocation genes have evolved under different selectional pressures. Thus within different kin related individuals they are expressed unequally, each allele favouring their own inclusive fitness. Social insects provide the best independent model system to study the evolution of imprinting. However, imprinting has not been discovered in any social insect.

My PhD lays the groundwork for a social insect model of genomic imprinting. Methylation is a common epigenetic tag of genomic imprinting in mammals and flowering plants. I found that a functional methylation system which is involved in the reproductive caste formation, development and social behaviour is present in the bumblebee, *Bombus terrestris*. Under queenless conditions, reproducing and non-reproducing worker castes show different brain methylome patterns. Alteration of methylation can cause a sterile worker to turn into a reproductive worker with increased aggressive behaviour and ovary development. Next I found monoallelic methylation associated with monoallelic expression in genes predicted to be imprinted by Haig's theory. Also, differential allele specific expression that are apparently due to parent-of-origin effects is present in reproduction loci of *B. terrestris*. Reciprocal crosses at these loci is recommended as further work, to check whether these expression patterns are due to genomic imprinting.

I assess the effects of maternally and paternally contributed sociobiological factors on worker male production and found that the paternity or the queen mating frequency has a significant influence on worker male production in eusocial Hymenoptera.

Finally, I also studied the polyphenism involved in phase dependent behavioural plasticity of locusts. I found that the transition of solitarious to fully gregarious behaviour in the desert locust, *Schistocerca gregaria* begins without significant changes in the DNA methylation landscape of the CNS but subjected to the pronounced differences at a later stage.

Acknowledgements

First and foremost I owe my deepest gratitude to my supervisor Dr. Eamonn Mallon for his guidance and endless support provided throughout this study. I am ever indebted to you for everything you have taught me. Thank you for believing in me and giving me this amazing chance in life.

I am very grateful to Dr. Robert Hammond for his advice at various stages of this PhD. Also, I would like to offer my sincere thanks to Dr. Swidbert Ott for providing locust specimens and guidance for chapter seven.

In particular, I would like to thank Kate Lee for her support with bioinformatic analysis and to Ceinwen for her technical assistance. I would also like to thank past and present members in my lab: Cris, Akram, Sony, Bobby, Bradley, Zoë and to all my friends at the University of Leicester for their invaluable advice and encouragement.

I must also acknowledge and thank the University of Leicester for their financial support of my study.

The most important thanks goes to my family. First, to my father. If there is any father who sacrificed every second of his life for his children, then that's you, my dad. Whatever success I achieved in life, it's all because of your courage, protection and limitless love. Second, to my mother, the inspiration of my education. As much as twenty years ago, you said that you wouldn't be around to see me completing a PhD. I can't believe that day has arrived and such a lot of time has passed. I deeply miss your presence, not only in this moment but throughout all my life. Third, to my husband for always being there for me at difficult times. Thank you for everything you have ever done for me and I love you endlessly. Fourthly, to my brother for his constant encouragement and belief that I could do this. Next, to the cutest thing in the world, to my little son, for being patient when I use the time that belonged to him for this thesis. Last but certainly not least, I would like to thank everyone else who supported and encouraged me throughout this PhD.

Finally, I would like to mention one thing which is always in my thoughts and gives me courage to walk ahead.

"*When battered by the realities of the world,*
If one's mind never tends to tremble,
And it stays secure, griefless and stainless,
That is their highest achievement of success."

-Maha mangala Sutta

Contents

List of Figures

List of Tables

Abbreviations

A	**A**denine
AIMS	**A**mplification of **I**nter-**M**ethylated **S**ite
bp	**B**ase **p**airs (of nucleic acid)
BSA	**B**ovine **S**erum **A**lbumin
CNS	**C**entral nervous system
cDNA	**C**omplementary **DNA**
C	**C**ytosine
^{o}C	**D**egrees **C**entigrade
DNA	**D**eoxyribbonucleic acid
dNTP	**D**eoxynucleotide triphosphate
DMSO	**D**imethylsulphoxide
Dnmt1	**D**NA (5'-cytosine) methyltransferase 1
Dnmt2	**D**NA (5'-cytosine) methyltransferase
Dnmt3	**D**NA (5'-cytosine) methyltransferase 3 subfamily
dd$H_2$0	**D**ouble distilled water
EtOH	**E**thanol
EDTA	**E**thylenediaminetetraacetic acid
ESTs	**E**xpressed **s**equence **t**ags
GMA	**G**ene **M**utaions **A**nalysis gels
Gb	**G**iga base pairs (1 000 000 000 bp)
g	**G**ram
G	**G**uanine
IPTG	**I**sopropyl-β-D-thio-galactoside
kb	**K**ilobase
LB	**L**uria-**B**ertani broth
$MgCl_2$	**M**agnesium chloride

mRNA	**M**essenger RNA
MS-AFLP	**M**ethylation **S**ensitive**A**mplified **F**ragment **L**ength **P**olymorphism
MSRE	**M**ethylation **S**ensitive **R**estriction **E**nzymes
µg	**M**icrogram
µl	**M**icrolitre
mm	**M**illimetre
M	**M**olar concentration (mol per litre)
ng	**N**anogram
NTE	**N**aCl **T**ris **E**DTA buffer
NEB	**N**ew **E**ngland **B**iolabs
PBS	**P**hosphate **B**uffered **S**aline
pmol	**P**icomoles
PCR	**P**olymerase **C**hain **R**eaction
PolyNAT	**P**oly- **N**- **A**cryloyl- **T**ris gels
r	**R**elatedness
RT-qPCR	**R**everse **T**ranscription **q**uantitative **PCR**
rpm	**R**evolutions per minute
RNA	**R**ibonucleic acid
RNase	**R**ibonuclease
RNAi	**R**NA **i**nterference
s	**S**econds
SNP	**S**ingle **N**ucleotide **P**olymorphism
SDS	**S**odium **D**odecyl **S**ulphate
ρ	**S**pearman's rank correlation (rho)
SOC	**S**uper **O**ptimal broth with **C**atabolite repression
T	**T**hymine
TAE	**T**ris **A**cetate **E**DTA buffer
g	**U**nit of force equal to that exerted by gravity
UTR	**U**ntranslated region
U	**U**racil
V	**V**olt
v/v	**V**olume to volume ratio
X-gal	5-bromo-4-chloro-3-indolyl-beta-D-galacto-pyranoside
YB	**Y**orkshire **B**ioscience

For my beloved parents,

everything I have and will accomplish is due to their love and encouragement.

Chapter 1

Introduction

Most organisms are diploid. That is they inherit two alleles for every autosomal gene, one copy from the mother and one copy from the father. Both copies of these genes are functional or expressed for the majority of these genes. However, in certain genes only one copy is expressed, according to which parent it was inherited from. That is certain genes are expressed only when they are inherited from a person's mother while others are expressed only when they are inherited from a person's father. This biological process of activating or silencing gene expression depending on it's parent-of-origin is referred to as genomic imprinting (Reik and Walter, 2001).

The fundamental concept of Mendelian principles is that parent-of-origin of a gene doesn't influence its dominance or recessiveness in phenotype determination and both the parental copies contribute equally to the outcome. Thus imprinting shows a clear deviation from classical Mendelian inheritance (Pardo-Manuel de Villena et al., 2000).

Since the discovery of the first naturally occurring imprinted gene (*insulin-like growth factor type 2 receptor* or *Igf2r*) in mice (Barlow et al., 1991), imprinted genes have been found to be involved in many important aspects of biology. Appropriate expression of imprinted genes is essential for a normal development in mammals and flowering plants. Improper imprinting can cause developmental and neurological disorders when they occur during early development and can cause cancer when altered later in life (Barlow and Bartolomei, 2014).

In this introduction, first I will summarize different theories proposed to explain the evolution of genomic imprinting. Then I will describe Haig's kinship theory (the most widely accepted) and applications of its predictions for a social insect model system such as *Bombus terrestris* emphasizing four different social contexts: monandrous, polyandrous, queen-right and queenless. Thirdly I will discuss monoallelic gene expression and parent-of-origin effects as evidence to support genomic imprinting being functional in social Hymenoptera. Fourthly I will explain the mechanism of genomic imprinting where I will discuss DNA methylation as an epigenetic marker of parental imprints and the different methylation systems that exist in mammals and insects. Next I will discuss phenotypic plasticity, a consequence of epigenetic modifications which are central to the success of insects. Lastly the overall aim of my PhD and specific research aims in each chapter will be summarized.

1.1 Theories of Genomic imprinting

Imprinted genes are under greater selective pressure than normal genes because only one copy is transcriptionally active at a time. Any variation present in that copy will be expressed without any back-up copy to mask its effects. As a consequence, imprinted genes evolve more rapidly than other genes (Wilkins and Haig, 2003).

One of the major advantages of diploidy is it's masking effect against deleterious recessive mutations. Imprinting diminishes this ability making an organism vulnerable to these mutations (Wilkins and Haig, 2003). Despite this, evolution maintains imprinting across generations. It has been found that imprinting has evolved roughly at 100 mammalian loci (Henckel and Arnaud, 2010, Williamson, 2012) and maintained throughout the mammalian radiation over 150 million years. For evolutionary biologists the origin of genomic imprinting is an intriguing topic. Several hypotheses have been proposed to explain how the apparent disadvantage of imprinting has been overcome.

The Ovarian time bomb hypothesis proposed by Varmuza and Mann 1994 states that imprinting occurs as an adaptation against trophoblastic disease in female mammals. According to this, inactivation of the maternal copy of growth-enhancing genes in

oocytes has evolved, to prevent the development of unfertilized oocytes into ovarian cancers. Sex-specific selection, another hypothesis which describes imprinting suggests it has evolved due to different selection pressures on the two sexes (Iwasa and Pomiankowski, 1999). However, these theories, arguably, do not explain the full extent of genomic imprinting. The most complete theory proposed to explain the evolution of genomic imprinting is Haig's kinship theory of genomic imprinting (Haig, 2000).

1.2 Kinship theory of genomic imprinting

Kin selection is a form of natural selection that promotes the reproductive success of an individual's relatives even at a cost to the individual's survival and reproduction. This is because an individual can improve its overall genetic success (fitness) when a gene is transmitted to the future generation through many of its relatives than only through its own offspring. Hamilton 1964 described this as the 'inclusive fitness theory' which states that altruistic social behavior has evolved as a combined effect of relatedness, benefit and cost. This means, through the survival of kin, benefits of indirect fitness exceeds the cost occurred due to direct fitness (Bourke, 2014). According to Hamilton, kin selection is mainly dependent on the degree of relatedness. i.e. a higher relatedness increases the chance of a specific allele being shared. Therefore understanding the kin structure, which determines relatedness among relatives, is central to the study of reproductive conflicts.

Kinship theory predicts that the genes inherited from the mother and the father are equally expressed in the offspring. However, Parent-offspring conflict theory states that the genetic interests of parent and offspring are different (Trivers, 1974). This is because parents are equally related to all offspring. Thus, they prefer an equal investment of resources among them. In contrast, offspring are fully related only to themselves and are only half or less related to their siblings. Thus their genes seek more parental investment than the parents had intended to provide. This leads to a conflict between the parents and offspring. The kinship theory of genomic imprinting was postulated as a combination of these two theories, in the sense that it is a gene centered view of kinship and parent-offspring conflict. For example, if

any gene acts selfishly and obtain additional resources from parents, that will be at cost to its siblings. This in turn would be a cost to its own fitness since passing on of that gene to the next generation via relatives would be affected due to this selfish behaviour.

Where Trivers saw the genes of an offspring as a coherent actor in conflict with the parents for resources, Haig extended this such that each locus in a diploid offspring allele could be under different selectional pressure, depending on its relatedness to each parent (Haig, 2000). Therefore maternally (matrigene) and paternally (patrigene) derived gene loci can be silenced or expressed unequally. Thereby each parental locus tries to manipulate the growth and survival of its offspring to be beneficial to the fitness of itself. According to Haig's theory two prerequisites are required for genomic imprinting to be functional: first, epigenetic marks to differentiate matrigenes from patrigenes; second, a difference in relatedness of matrigenes and patrigenes among the social group members. Then kin selection will act independently on genes of maternal and paternal origin and the evolutionary stable strategy for each parental gene would be selected.

Due to the capacity to explain many empirical aspects of the biology of imprinted genes, Haig's kinship theory has become the most widely accepted among all theories that address genomic imprinting: It explains, why imprinting is common in mammals and angiosperms, taxa with prolonged provisioning of resources to the offspring. It also provides answers to why many imprinted genes are expressed in tissues (placenta and endosperm) that are involved in mediating resource transfer from the mother to embryo. Furthermore it accounts for why paternally expressed genes tend to increase the offspring size while a maternally expressed gene does the opposite (Burt and Trives, 2006, Haig, 2000). Haig's theory addresses above questions as follows:

The kinship theory hypothesizes that imprinting grew out of a competition between males for maternal resources. In a polyandrous mating system, many fathers may contribute to the offspring of the same mother. Therefore paternally derived alleles (patrigenes) which are inherited from different fathers are unrelated among the siblings (Figure 1.1). As a result, different patrigenes evolved to be more competitive and aggressive, in extracting maternal resources. So that, it would be beneficial

to their own offspring, although it is a reproductive cost to the mother. In contrast, the maternally derived allele (matrigene) is equally related among all siblings. Hence in matrigene's point of view, it cannot favour any offspring during resource provisioning, since it would be a cost to the offspring who carry the same allele as those she favoured. In this scenario the matrigene tends to react altruistically and becomes silenced/imprinted, so that it could minimize or control the resource flow. The patrigenes act selfishly thus tend to be expressed.

Therefore the kinship theory of genomic imprinting predicts that, at any locus that enhances offspring growth, the matrigene would be imprinted whereas at a locus that suppress offspring growth, the matrigene would be expressed. Similarly, at growth enhancing loci the selfish patrigene would be expressed whilst at growth suppressing loci it will be imprinted (Haig, 2000). The expression patterns of some mammalian embryonic genes are consistent with these predictions.

For instance, maternal imprinting of *insulin-like growth factor 2/Igf2* (fetal growth enhancer) locus in mice decreases placental size and thereby minimizes maternal resource transfer to the fetus. Imprinting of *Igf2* from the paternal side increases the fetal size via more resource diversion from the mother to the offspring (Haig and Graham, 1991). Likewise predictions of Haig's kinship theory are mainly based on mother-embryo relationship, the best place at which parents could manipulate offspring, to ensure the survival of their traits in future generations. Since offspring that grow up with better parental resources will survive, reproduce and propagate parental traits. This would maximize parental fitness.

Kin conflict is the foundation of Haig's theory. However, it relies on evidence from mammals and flowering plants which have very few kin related interactions compared to those that exist within an eusocial insect colony (Queller, 2003).

1.3 Eusocial Hymenoptera: a model system for Haig's kinship theory

Eusocial Hymenoptera provide an independent model system, matching the prerequisites of Haig's kinship theory. The two main requirements for genomic imprinting are found in social insects: a) Presence of haplodiploidy which creates a network of individuals with different degrees of kin relationships (Queller, 2003). b) Presence of DNA methylation, one of a key regulatory mechanisms of genomic imprinting in mammals (Wang et al., 2006).

Hymenopterans such as ants, bees and wasps are haplodiploid. i.e haploid males develop from unfertilized eggs and diploid females (queens and workers) develop from fertilized eggs. Thus the distribution of maternally and paternally derived alleles among colony members varies (Figure 1.1). As a result, the relatedness of matrigenes and patrigenes to different offspring varies and a conflict arises between parental alleles over the treatment of offspring to which they are differently related. In addition, different kinds of kin relationships arise between members of the colony. Therefore resource allocation within the colony is not just between the mother and her daughters but also among brothers, sisters, nephews, sons etc (Figure 1.1). Haig's kinship theory predicts that there is a conflict between maternally and paternally derived alleles when provisioning resources among different kin related offspring during reproduction. Genomic imprinting has evolved during this resource allocation as a means of manipulating the reproductive potential of their offspring to maximize the fitness of each parental allele (Queller, 2003). Hence, reproductive loci in eusocial hymenopteran workers are a potential place to check for genomic imprinting. Presence of many kin related interactions within a single colony and the existence of different social contexts (e.g. monandrous, polyandrous, queen-right, queenless etc.) make eusocial Hymenoptera the most satisfying model system available to test many independent theoretical predictions about the evolution of imprinting.

Apart from the conflicts arising in direct resource allocation to offspring, there are many other conflicts in eusocial Hymenoptera: sex allocation; reproductive division of labour (brood rearing, policing, male production by workers, defense); caste fate of female larvae etc. (Queller, 2003, Queller and Strassmann, 2002). This makes

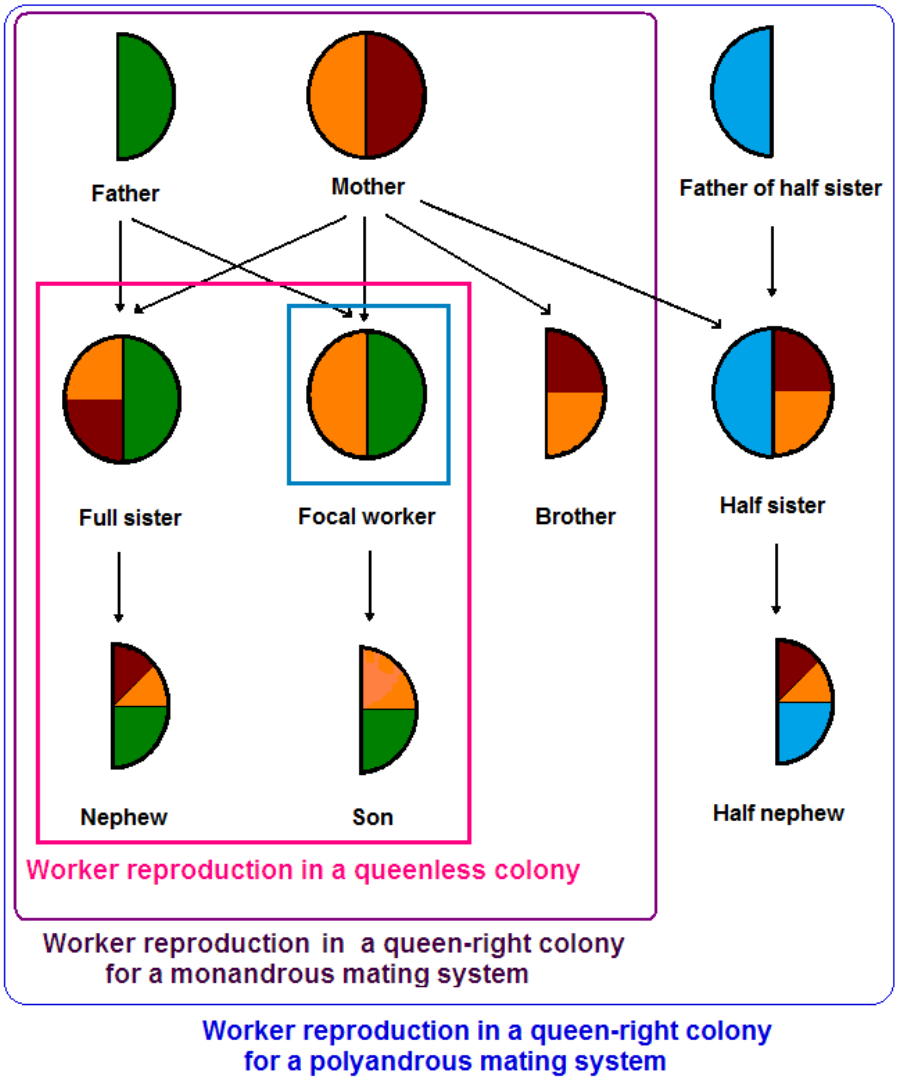

Figure 1.1: Distribution of parental genes among a focal worker and her relations. Females are diploid and represented by full circles. Haploid males are represented by half circles. Matrigene and patrigene of the focal worker are depicted in orange and green respectively. Matrigene of the queen that are not inherited by the focal worker are shown in brown. Patrigene of an unrelated father to the focal worker is shown in blue. The proportion filled by a certain colour is the probability of that individual possessing that allele (adapted from Drewell et al. 2012).

an even stronger case for eusocial Hymenoptera as a model system to test Haig's theory.

However, imprinting in social insects has not been discovered yet. My research is focused on the bumblebee, *B.terrestris* as a model organism, to check for the presence of imprinting in social insects.

1.4 The bumblebee as a model organism

As stated in the kinship theory, there is a conflict between matrigenes and patrigenes when provisioning resources among offspring during reproduction. Therefore a reproductive trait which shows variation among these genes would be an ideal model to test the theory. In addition, these variations should be able to be quantified either phenotypically or genotypically (gene expression). Worker reproduction in the buff-tailed bumblebee, *Bombus terrestris* is such a trait which fulfills all above requirements.

B.terrestris belongs to an exceedingly successful insect order, the Hymenoptera. It is considered to have an intermediate level of eusociality. This means that excepting the size difference, the queen and workers are morphologically similar (Figure 1.2). Hence in reproductive division of labour they are much more plastic than workers in a higher eusocial taxa (Cardinal and Danforth, 2011), which are morphologically specialized and fixed for distinct colony tasks (e.g. mouth parts in honeybee workers are turned into spoon-shaped for feeding and wax moulding tasks; certain ant species have highly specialized soldier castes with enlarged heads and mandibles for defensive tasks. Thus these anatomical features are less likely to be reversible). In contrast, workers belonging to an intermediate eusocial taxa can switch into different roles easily. Thus they even can compete with the queen for reproduction, leading to numerous conflicts within the colony.

B.terrestris colonies are composed of a singly mated queen and non-reproducing workers (van Honk and Hogeweg, 1981). Towards the end of the colony life, the queen starts to produce male and queen destined eggs (the switching point), after which the workers begin to lay male destined eggs (the competition point). Conflicts between the queen and workers and among workers arise at this point in the colony (Duchateau and Velthuis, 1988). This is because workers have the potential to lay male destined eggs, even though they are unmated.

Compared to a polyandrous mating system, worker reproduction is common in a monandrous mating system (Ratnieks, 1993, Wenseleers and Ratnieks, 2006a). Therefore when making predictions for a monandrous mating system, the distribution of maternal and paternal alleles in worker produced offspring will also have to be

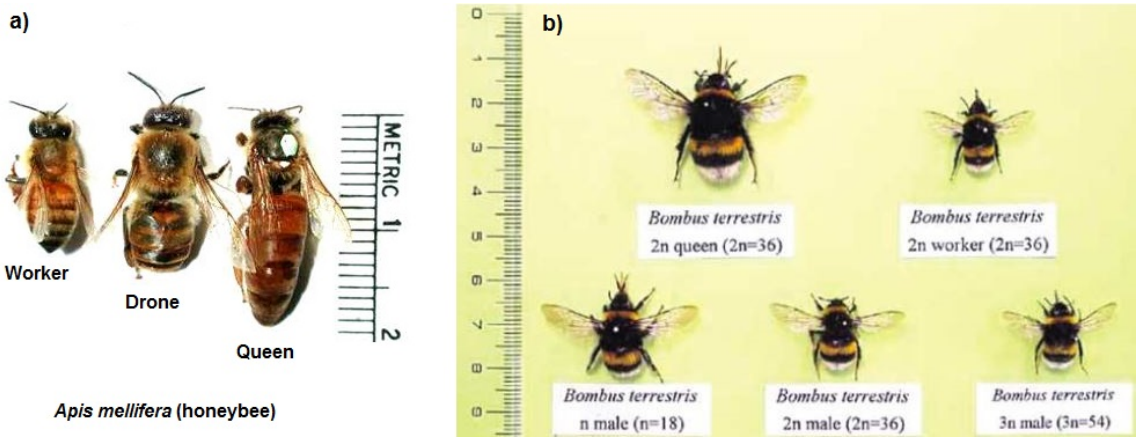

Figure 1.2: Comparison of colony member morphology between highly eusocial and intermediately eusocial bees. a) In the highly eusocial honeybee *Apis mellifera*, queen, workers and drones differ in size, wing length, mouthparts etc. (Adapted from http://www.uni.illinois.edu/~stone2/bee_life_stages.html), b) In *B.terrestris* which has an intermediate eusociality, queen, workers and drones are morphologically similar and only differ in size (Adapted from Ayabe et al. 2004).

taken into account. Predictions of Haig's kinship theory for a monandrous mating system such as in *B.terrestris* and their applications for two selected social contexts (queen-right and queenless) are described below.

As predicted for a polyandrous species, from Figure 1.1, we could also predict the selectional pressures imposed on a locus in a monandrous species. In a queen-right (colonies where a queen is present), monandrous colony, the patrigene in a worker is unrelated to the queen's sons but has a 50% chance of being in a worker's son (This is because the queen uses unfertilized eggs/only the queen's alleles to produce her sons while workers use either the queen's allele or father's allele to produce their sons). For the patrigene, its fitness will increase only if the workers reproduce and propagate it to the future generations. Therefore it is predicted that patrigene tends to promote worker reproduction. However, the matrigene in a worker has an equal (50%) chance of being in a queen's son as well as in a worker's son. Therefore the matrigene tries to ensure an equal distribution of her resources among both her sons and daughters. As a result, at any locus which enhances worker reproduction, the matrigene should react altruistically and be imprinted while the patrigene react selfishly and thus should be expressed. Conversely at any locus that suppresses

worker reproduction, the patrigene is expected to be imprinted while matrigene is expressed.

This situation is reversed when the colony becomes queenless. This is represented in Figure 1.1 in the inner red box. The patrigene in a given worker has a 50% chance of being in any worker reproduced male. In other words, it is equally likely to be in that worker's son or a different worker's son. The matrigene has a 50% chance of being in the focal bee's sons but only a 25% chance of being in a nephew. For a patrigene, it is of equal value if the worker it is in reproduces or the worker's siblings reproduce. Thus the patrigene behaves altruistically to promote worker reproduction. For a matrigene it is twice as beneficial if the worker herself reproduces. Therefore the matrigene behaves selfishly and tries to inhibit worker reproduction. Hence the kinship theory predicts that for a locus that enhances worker reproduction in a monandrous queenless colony, the matrigene should be expressed and the patrigene should be imprinted. Similarly for a locus that inhibits worker reproduction, the matrigene should be imprinted and patrigene should be expressed.

1.5 Monoallelic expression and parent-of-origin effects in the Hymenoptera

Although imprinting in social insects is yet to be discovered, in non-social Hymenoptera it is central to the sex determination. For example normal expression of maternally provided transformer mRNA (Nvtra) in *Nasonia vitripennis* produces females while imprinting of maternal Nvtra is essential for male development (Zwier et al., 2012).

Monoallelic expression is defined as a stable expression of one allele over the other. It could occur due to various reasons such as dominant-recessive effects, parent-of-origin effects, X chromosome inactivation etc (Eckersley-Maslin and Spector, 2014). Therefore genomic imprinting is only one such reason that genes should display monoallelic expression. Although imprinting has not been discovered yet, there is

evidence for the existence of monoallelic expression and phenotypic effects of parent-of-origin in social Hymenoptera. For example caste-specific monoallelic methylation and monoallelic expression has been found in ants, *Camponotus floridanus* and *Harpegnathos saltator* (Bonasio et al., 2012). Queen caste determination and caste allocation in Argentine ants (*Linepithema humile*) are controlled by the paternal lineage whereas their maternal lineage decides the number of offspring produced for the colony (Libbrecht et al., 2011). *Apis mellifera scutellata*, a mixed species of both Africanized and European honeybees, displays different parental effects on offspring behaviour. A cross between European queen x Africanized drones or Africanized queen x European drones produce genotypically identical progeny. However, when the father is an Africanized drone, the resulted progeny is highly defensive while they are less aggressive if the father is an European drone (Guzman-Novoa et al., 2005). In another similar experiment on reciprocal crosses between *Apis mellifera scutellata* and *A. m. capensis*, worker ovary size has been found to increase when the father is *A. m. scutellata* drone, despite the progenies of the two crosses are genotypically identical (Oldroyd et al., 2014). In both of the above reciprocal cross experiments, imprinting of the patrigene is possible, since it enhances the male's reproductive success via increased fertility of daughter workers. Therefore all this evidence suggests the likely existence of imprinting in eusocial Hymenoptera.

In the next section I will discuss the molecular mechanism of genomic imprinting.

1.6 Methylation, an epigenetic tag

Every somatic cell in a multicellular organism habours the same DNA code. However, the functions they perform for the body are highly diverse. Cells receive this phenotypic flexibility mainly through epigenetic modifications, the heritable changes that occur in the genome without any alteration of the DNA sequence. Specific epigenetic tags determine which genes are required to be expressed or silenced among different tissues. There are many forms of epigenetic modifications (e.g DNA methylation, chromatin modifications, noncoding RNA etc.) which may play equally important roles on gene regulation (Bbosa et al., 2013). During this

study, I particularly focused on DNA methylation, a common regulatory mechanism of genomic imprinting in mammals (Gehring, 2013).

DNA methylation is the addition of a methyl group to the 5^{th} position of the cytosine pyrimidine ring or the number 6 nitrogen of the adenine purine ring. This reaction is catalyzed by a group of enzymes called DNA methyltransferases (Dnmts). Cytosine methylation primarily occurs at CpG dinucleotides (i.e. the cytosine residues followed by a guanine). In response to physiological, behavioural, environmental and maturational influences, *de novo* methyltransferase (Dnmt3) methylates new CpG sites. Then Dnmt1 copies these methylation patterns to daughter strands (hemi-methylated DNA) during DNA replication and ensures the transmission of previously established methylation/imprinting patterns across generations (Jurkowska et al., 2011). Dnmt2 is not considered a *bona fide* DNA methyltransferase as although it is structurally and catalytically similar to Dnmt1 and Dnmt3 it functions *in vivo* as a tRNA methyltransferase (Goll et al., 2006).

DNA methylation is closely associated with histone modifications. Together they are involved in the regulation of chromatin structure and gene expression. During histone modifications, various molecules (e.g CH_3 groups) are attached to histone tails through processes such as acetylation, phosphorylation and methylation. Binding of such molecules alter the activity of DNA around histone molecules. For example when DNA wrapped around a histone molecule is unmethylated, such chromatin is less condensed. Thus genes are open to transcription factors allowing them to be expressed. Conversely if DNA is methylated, they are tightly wrapped around the histone molecule (nucleosome) making a heterochromatin environment around a gene. This makes the gene less exposed to transcription factors leading it to be silenced. Hence methylation can directly inhibit the binding of transcription activating factors and thereby determines if a gene is to be activated or repressed (Li, 2002). In addition, DNA methylation can induce gene silencing through the recruitment of transcription suppression factors such as histone deacetylases (HDAC) which leads to the production of heterochromatin. It has been discovered that the recruitment of HDAC to chromatin is controlled by methyl-binding domain proteins (e.g MeCP2) which are specifically bound at methylated cytosines (Prokhortchouk and Hendrich, 2002). Cancers which arise due to the silencing of tumour suppressor genes are mediated by transcription suppression factors.

Likewise, DNA methylation is vital for normal development. It is directly linked with various biological processes such as X chromosome inactivation, suppression of repetitive elements, reprogramming of germline during embryonic development, formation of chromatin structure and prevention of cancers (Varriale, 2014). Relevant to this study it plays an important role in establishing and maintaining allele-specific expression of imprinted genes in mammals and flowering plants.

The parent-of-origin-specific expression of genes is regulated by differential DNA methylation at imprinting control loci of maternal and paternal alleles (Ferguson-Smith, 2011). During this process first, the existing methylation pattern in primordial germ cells in the parental generation is erased. Then during gametogenesis, new methylation tags are established on different alleles in a parent-of-origin-specific manner. i.e if any imprint is established in a developing sperm it is a paternal imprint and vice versa. Finally, following fertilization, during the pre-implantation stage these methylation tags are removed and tissue and gene specific *de novo* methylation patterns are established in the developing embryo. Imprinting must be reprogrammed in the germ line, because a maternal allele in one generation may be a paternal allele in the next. However, during the pre-implantation stage parental methylation/imprinting signatures are sometimes preserved as a trait in the current generation (Messerschmidt et al., 2014). Such imprinting patterns could affect normal development of the embryo and lead to various genetic disorders.

For example both Prader-Willi and Angelman syndromes in humans are associated with the chromosome 15q(11-13) region which harbours both maternally and paternally expressed genes. If genes in this region in the paternally inherited copy are silenced (usually the maternal copy is silenced in a normal human and the paternal copy is expressed. In the disease the paternal copy also becomes non-functional, leading to a lack of expression in both the parental copies) it will lead to Prader-Willi syndrome while maternal inheritance of imprinted genes in this region causes Angelman syndrome (Horsthemke and Wagstaff, 2008). However, 25% of patients with these syndromes receive them due to uniparental disomy. i.e. both the copies of a gene are inherited from a single parent and lack of expression of the other parent causes the syndrome (`https://www.uic.edu/classes/bms/bms655/lesson8.html`). Supporting the predictions made by Haig's theory for the mother-embryo relationship, infants with Prader-Willi syndrome (both chromosome 15 are derived from the

mother) display poor suckling and intake less resources from their mother whereas infants with Angelman syndrome (both chromosomes 15 are derived from the father) have more paternal contribution thus are more active and display strong sucking.

NOEY2 is a maternally imprinted, paternally expressed tumor suppressor gene in human. If the paternal copy of this gene becomes silent, lack of expression in this gene will lead to ovarian cancer (Fu et al., 2014). Therefore DNA demethylation is essential to maintain the right balance of expression levels in a genome. It regulates important processes such as reactivation of silenced genes, epigenetic reprogramming of embryonic stem cells (pluripotency of cells allow them to differentiate into any required cell type in the body) etc.

Loss of global DNA methylation and loss of imprinting in oncogenes have been reported in many cancer types (Bbosa et al., 2013). On the other hand, CpG islands, clusters of CpG sites located at the 5' ends/promoters of genes (Bird, 1980), which remain unmethylated in normal cells are methylated in certain cancer cells. Hence both hypermethylation of growth-suppressing genes (e.g tumor-suppressor genes) and hypomethylation of specific growth-promoting genes (e.g proto-oncogenes enhance uncontrolled proliferation of cells) play an important roles in carcinogenesis. As a result, DNA methylation profiles of body tissues are extensively used in recent cancer diagnosis (Das and Singal, 2004).

DNA methylation is a conserved phenomenon across taxa. However, the proportion of methylcytosine in the genome varies among taxa: 0-3% in insects; 4-6% in mammals and birds; 10% in fish and amphibians; 30% in some plants (Field et al., 2004). Here I focus on insect methylation systems and their modifications from the mammalian system for diverse functional roles.

1.7 Methylation systems

The basic methylation system has evolved differently among mammals and insects. For example mammals harbour one paralog of *Dnmt1* and three paralogs of *Dnmt3*: 3a, 3b and 3L. Conversely in insects, *Dnmt3* subgroup is limited to a single paralog but show an expansion in *Dnmt1* subgroup: 1a, 1b and 1c (Lyko and Maleszka,

2011). Hence different taxa may have used the same methylation system for contrasting functions: as a defense against transposable elements, for differential expression of parental alleles (genomic imprinting), as a developmental response to changing environment (phenotypic plasticity) etc.

For example, *Dnmt3* in honeybees is involved in queen-worker caste differentiation (Kucharski et al., 2008) and alternative splicing in fat bodies (Li-Byarlay et al., 2013). However, in complex genomes such as in mammals, *Dnmt3* is responsible for multiple tasks other than its main role of establishing new methylation marks on DNA: regulation of methylation-independent gene expression (3a and 3b) and establishment of maternal imprints during gametogenesis (3L). *Dnmt1* in *Nasonia vitripennis* is comprised of 3 paralogs (Beeler et al., 2013) and inhibition of maternal *Dnmt1a* is lethal for embryonic development (Zwier et al., 2012). Paternally inherited chromosomes in *Planococcus citri* are hypomethylated and silenced in male mealybugs, whereas maternally inherited chromosomes remain hypermethylated and active (Bongiorni et al., 1999).

Through evolution, multiple versions of the basic *Dnmt* system have arisen among insect species (Figure 1.3). The complete methylation system comprising of all three *Dnmts* is well conserved in both social (ants and some species of bees) and non-social (*Nasonia vitripennis*) Hymenoptera. In addition, the pea aphid *Acyrthosiphon pisum* is also reported to have all 3 *Dnmts* required for DNA methylation (Glastad et al., 2011). Conversely in *Drosophila* both *de novo* and maintenance methyltransferases are absent (Zemach et al., 2010) but the presence of Dnmt2 and small quantities of 5-methylcytosine (in non CpG context) only in embryos have been detected (Boffelli et al., 2014). *Dnmt3* is lost in the flour beetle (*Tribolium castaneum*) and moths (*Bombyx mori*, *Mamestra brassicae*; Glastad et al. 2011). However, *M. brassicae* has been reported to have the highest level of methylation reported in any insect. Approximately 10% of all cytosines in its genome are methylated thus resembling a vertebrate-like methylation level. This is because 5-methylcytosines present in its genome are not only restricted to CpG dinucleotides but could also present as the external cytosine in a CCGG site (Mandrioli and Volpi, 2003). Orthoptera, which shares, a common ancestor with Hymenoptera (diverged about 350 million years ago) shows substantially higher methylation levels (i.e 1.3-1.9% of cytosines in the genome are methylated compared to 0.3% in *A. mellifera*). However, *Dnmt3*

have been found lost in both the desert locust (*Schistocerca gregaria*) and migratory locust (*Locust migratoria*) species (Falckenhayn et al., 2013).

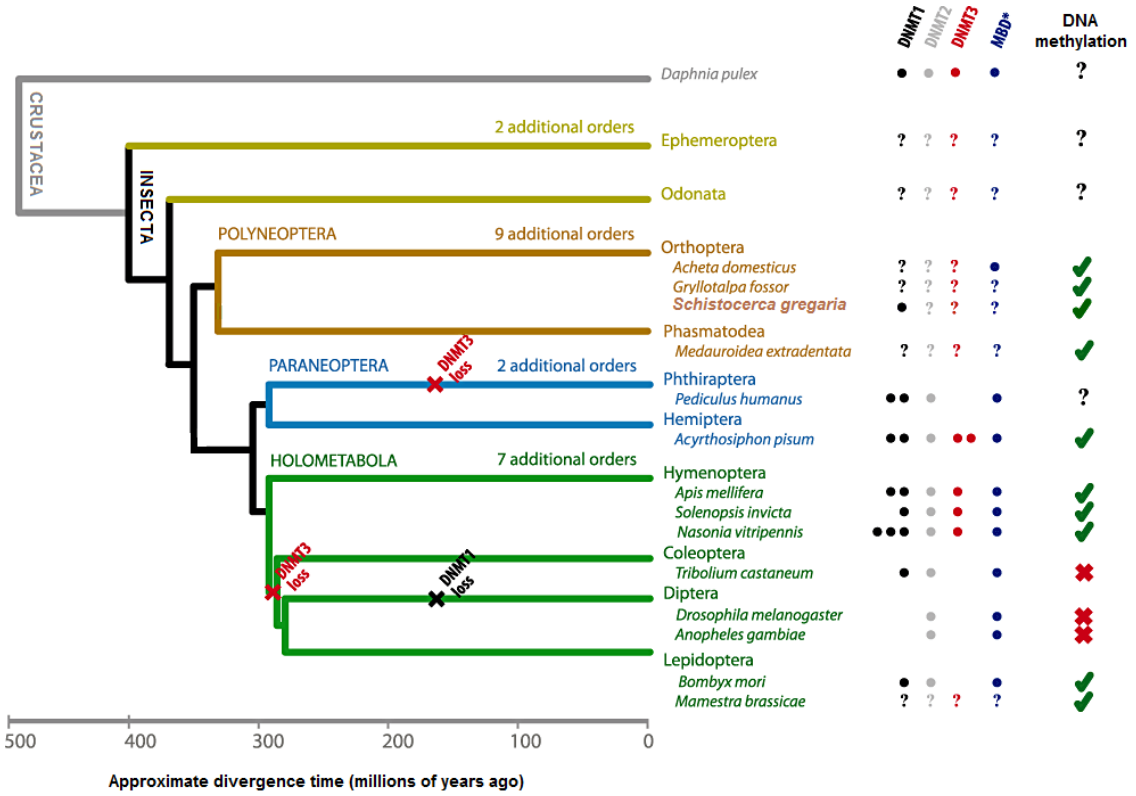

Figure 1.3: Evolution of DNA methylation systems in insects. The presence of methyl CpG binding domain proteins (MBD) and the number of paralogs present in each *Dnmt* type are represented by dots. Absence of clear sequenced genomic data is denoted by a question mark. The presence and absence of an active methylation system is indicated by a tick and a cross respectively. Adapted from (Glastad et al., 2011).

DNA methylation is most active during embryogenesis where cells undergo rapid cell division. However, methylation machinery remains active throughout the lifetime and can be influenced by external environmental agents (Lillycrop et al., 2014). For example the methylation pattern in monozygotic twins is indistinguishable early in life but changes substantially with age (Fraga et al., 2005). Maternal care which helps to establish epigenetic marks on rat pups plays an important role in fear response later in their life (Szyf et al., 2008). Likewise epigenetic modifications play a vital role in regulating phenotypic plasticity of organisms.

1.8 Epigenetics, phenotypic plasticity and polyphenism

Phenotypic plasticity is the ability of an organism to change its phenotype in response to different environmental conditions. This change could be a suite of behavioural, physiological or morphological changes that are accommodated by a single genome. Also it could be a gradual difference which produce intermediate phenotypes or the production of discrete intraspecific variations. Polyphenism belongs to the latter and is defined as an extreme form of phenotypic plasticity where a single genome expresses two or more distinct morphs (Capinera, 2008).

Polyphenism can be seen in every stage of life. One of the best examples is that the origin of phenotypically diverse cell types in multicellular organisms occurred from a zygote with a single genetic identity. This diversification occurs through a variety of molecular pathways that activate gene expression only in the relevant part of the genome, in different cell types (Bonasio, 2014).

Once an individual perceives an external environmental stimulus (via visual, olfactory receptors etc.), it is transduced into an internalized signal through intracellular signal transduction pathways such as the neuro-endocrine system. If that signal exceeds the required threshold level, it will induce epigenetic modifications which lead to over expression, silencing or alternatively splicing of genes to produce appropriate behavioural and morphological effects. Optimal threshold value required to produce epigenetic changes, may vary with age; for example in *Harpegnathos saltator* young workers perform in-nest tasks while older workers specialize in foraging (Yan et al., 2014).

Due to the remarkable ability to evolve well into environmental plasticities, insects have become a highly successful group of animals on earth (Simpson et al., 2011). In other words, polyphenism ensures their survival in many ways: it allows them a) to divide the life history into distinct developmental stages (in holometabolic insects, larval stages are adapted to feed and grow while adult stages are dedicated to disperse and reproduce) thus specialized to utilize different niches, feeding modes and habitats (Moczek, 2010); b) to change into different phenotypes that are well suited to the changing environment (seasonal morphs of Lepidoptera); c) to survive

in temporally heterogeneous environments (dispersal morphs - wing polyphenism in aphids, migratory phenotypes of locusts); d) to divide labour within a social group (caste system of eusocial insects). Certain polyphenic traits are mediated by dietary cues. For example, catkin and twig like morphs in caterpillars of the moth *Nemoria arizonaria* depend on whether they feed on the catkin or leaves of the plant (Erick Greene et al., 2009).

The most remarkable examples of polyphenism are found among eusocial insects which lead them to be the most successful among insects: caste polyphenism of the termite *Reticulitermes speratus* is comprised of a king, founder queen, neotenic (secondary reproductive) queens, workers and soldiers (Simpson et al., 2011); sexual and asexual generations in aphids are determined by the season/temperature (Dixon, 1977), whereas winged and wingless morphs are determined by both genotype and environmental factors (Ogawa and Miura, 2014). However only certain species of aphids are eusocial. In the eusocial aphid *Ceratovacuna japonica* a sterile soldier caste with smaller weapon is produced in the season when predators are not abundant (Hattori et al., 2013); The majority of bee, wasp and ant species show queen-worker castes with clear differences in morphology, reproductive physiology, behavior and even lifespan (Keller and Genoud, 1997); In certain ant and termite species worker caste is further specialized morphologically and behaviourally into subcastes (e.g. major and minor subcastes of ants - *Pheidologeton* spp., *Pheidole* spp. with markedly different size polyphenism).

In some cases, colony members show polyethism, i.e. castes differ only in behaviour but not in morphology. For example, the non-reproductive worker caste in *Apis mellifera* is specialized into nursing and foraging subcastes. They are capable of switching into either of these roles depending on the colony requirement (Herb et al., 2012).

Polyphenism plays an important role in the social organization and division of labour in eusocial insects. Reproductive caste formation in *Bombus terrestris* and phase polyphenism in locusts are discussed in detail, in chapter 3 and 7 respectively.

1.9 Thesis Aims

The fundamental aim of my PhD was to establish a social insect model to study genomic imprinting. Reproduction is a potential trait where genomic imprinting could evolve. Using eusocial Hymenoptera, *Bombus terrestris* worker reproduction as a model system, I studied the above topic in three main steps.

The first step was to check for the presence of molecular mechanism involved in genomic imprinting. By comparing methylation differences between different worker reproductive castes, I aimed to show a functional DNA methylation system in *Bombus terrestris*. Second, through an experimental alteration of methylation levels in *B. terrestris* workers, I expected to find out the role of DNA methylation in *B. terrestris* worker reproduction (**Chapter three**).

Work presented in Chapter 3 is published in Amarasinghe et al. 2014.

The third step was to find genes that could potentially be imprinted in the *Bombus terrestris* genome. This aim was achieved with two different approaches: (i) Candidate gene approach - My expectation was to find parent-of-origin allele specific expression, in targeted genes that are associated with *B. terrestris* worker reproduction (**Chapter four**). (ii) Next generation sequencing approach - during this monoallelically methylated and monoallelically expressed genes were searched for in *B. terrestris* at a genome-wide scale (**Chapter five**).

Next, I aimed to focus on a different aspect of worker reproduction in eusocial Hymenoptera. Similar to the reproductive division of labour that exists among different castes, worker reproduction and cooperative brood rearing (i.e. all the offspring in the colony; produced by the mother queen, by a worker on its own and by other workers) are also important aspects of eusociality. To address how worker male production is affected by sociobiological factors such as the relatedness differences derived due to parental influence (maternal; queen/s derived factors and paternal; drone derived factors) and colony size, I conducted a meta-analysis of data across 90 eusocial hymenopteran species in **Chapter six**.

Phenotypic plasticity, the key to the ecological success of eusocial Hymenoptera can also be seen in non-social insect lineages. As with caste specific morphs (reproductive polyphenism) present in *Bombus terrestris*, locusts show extreme phenotypic morphs (phase polyphenism) that are different in behaviour, colour, metabolism and development. The functional role of DNA methylation in phase polyphenism of desert locust, *Schistocerca gregaria* was studied in **Chapter seven**.

Chapter 2

General methods

2.1 Bumblebee colony rearing

All experiments were carried out on commercially sourced *B. terrestris* colonies (Koppert Biological Systems, Haverhill, UK). They were reared in wooden nest boxes (Inner dimensions - 24cm x 16cm x 13.5cm) under red light at 26°C and 60% humidity (Alaux et al., 2007). They were fed *ad libitum* with pollen (Percie du sert, France) and 50% volume/volume apiary solution (Meliose - Roquette, France) as a source of protein and energy, respectively. Apiary syrup was provided through a mouse feeding bottle which was attached to a separate Perspex box (18.5 cm x 12.5 cm x 6.5 cm) with cat litter. The main nest and the feeding area were interconnected by a small plastic tube that serves as a passage for the bees to walk freely between the two areas.

2.2 DNA extraction

Several methods and commercial kits were used to extract DNA during this thesis.

2.2.1 QIAGEN QlAaMP DNA Mini Kit

Proteinase K solution (20mg/ml) was prepared by dissolving 5mg of proteinase K (Sigma-Aldrich) in 250µl of dH_20.

First, the tissue (e.g. bee head, leg, etc.) was crushed inside an eppendorf tube and incubated at 56°C overnight with 20µl of freshly prepared Proteinase K and 180µl of Buffer ATL. Next, 200µl of Buffer AL was mixed with the sample followed by incubation at 70°C for 10 minutes. After adding 200µl of absolute ethanol, the sample was vortexed and transferred in to a QlAamp Mini spin column placed in a 2 ml collection tube. The mini column assembly was centrifuged at 6000g for 1 minute. The column was transferred on to a new collection tube and the previous collection tube with the flow through was discarded. The column was then washed with 500µl of Buffer AW1 by centrifuging at 6000g for 1 minute followed by another wash with 500µl of Buffer AW2 for 3 minutes at 20,000g. The column was centrifuged for another 1 minutes at 20,000g to remove all residual ethanol and transferred into a clean eppendorf tube. 200µl of nuclease free water was added to the membrane and left for 5 minutes at room temperature before it was centrifuged at 8000g for 1minute. DNA samples were stored at -20°C.

2.2.2 QIAGEN QlAaMP DNA Micro Kit

First, 10 mg of tissue was incubated with 180µl of Buffer ATL at room temperature for 1 minute. Then, 20µl of freshly prepared Proteinase K was added to the sample, mixed by pulse vortexing and incubated overnight at 56°C on a heat block. Following the 15 hours of incubation, 200µl of Buffer AL was added to the sample and mixed by pulse vortexing for 15 seconds to ensure an efficient lysis. Then, the entire lysate was transferred into a QlAaMP mini-elute column placed on a collection tube, followed by a 5 minutes incubation at room temperature with 200µl of absolute ethanol. The entire mini-column assembly was centrifuged at 6000g for 1 minute. Then the column was transferred in to a clean collection tube and the previous collection tube containing the flow-through was discarded. Next the above centrifugation step was repeated with 500µl of Buffer AW1 and 500µl of Buffer AW2 respectively. After the final wash, the mini-column was placed on a new collection tube the entire assembly

was centrifuged at full speed (20,000g) for 3 minutes. Finally, the mini-column was placed in a clean eppendorf and 200µl of nuclease free water was added onto the center of the membrane. Followed by incubation at room temperature for 5 minutes the mini-column assembly was centrifuged at 20,000g for 1 minute. Eluted DNA samples were subsequently stored at -20°C.

2.2.3 Phenol:chloroform method

NTE was prepared according to Wang et al. 2006 ; 25ml of NaCl (100mM), 12.5ml of Tris (50mM), 2.5ml of EDTA (10mM), 12.5ml of 1% SDS, 25µl of 0.01% Triton X-100 and 197.5ml of dH_20 to make 250ml of NTE.

First, the tissue was crushed with liquid nitrogen inside a 1.5 ml eppendorf tube. Then the samples were incubated on a heat block at 50°C for 3 hours with 700µl of NTE and 17.5µl of 20mg/ml Proteinase K (Sigma-Aldrich). One volume of phenol:chloroform (Fisher Scientific) was added to that and vortex vigorously to mix the phases. It was then spun in a centrifuge (Progen) for 10 minutes at 10,000g to separate the phases. The upper aqueous phase with DNA was transferred in to a new epi tube without any contamination of the interphase proteins. Another phenol:chloroform extraction step was performed with the above collected aqueous phase to confirm that there were no more denatured protein collected at the interface. It was then treated with 1µl RNase (10mg/ml) for 10 minutes. The samples were precipitated with two volumes of absolute ethanol, incubated at room temperature for 20 minutes and then centrifuged at the maximum speed (14,000g) for 15 minutes. All of the supernatant was discarded and the tubes were dried inside the fume hood. The sides of the tube were washed with 50µl of nuclease free water to elute the DNA. DNA samples were stored at -80°C.

2.3 PCR amplifications

All oligonucleotides were purchased from Sigma-Aldrich, UK. The stock solutions were made to 100µM/µl, while working solutions were 10µM/µl and stored at -20°C. YB-Taq 2x reaction mix (York Biosciences, UK) was used for many PCR

amplifications in this thesis with optimized amounts of $MgCl_2$, primers, distilled water and DNA template. It is composed of Taq DNA Polymerase (0.1 units/µl) and 200µM of each dNTP. PCR amplifications were performed in a Biometra T1 thermocycler and a Biometra T Professional Gradient thermocycler.

2.4 Quantification and visualization of DNA

The quantity of DNA was measured using the NanoDrop^{TM} 1000 Spectrophotometer (Thermo Scientific, UK) and the results were analysed using the software, NanoDrop 1000 v3.7. Prior to use, the samples were defrosted and vortexed for 5 minutes.

The quality of DNA was assessed by 1 - 3% agarose gel electrophoresis. 1% agarose gels were used for separation of larger DNA fragments (>1kb) whereas 3% gels were used for smaller fragment sizes. 1% agarose gels were made by heating 1g of 1% agarose (Melford, UK) with 100 ml of 1x TAE (EDTA, acetic acid and ddH_20, prepared according to Sambrook and Russell 2001) in a microwave. One microlitre of ethidium bromide (Fisher Scientific, UK) was mixed with the melted agarose and poured in to gel trays with combs and left until solidified. Five microlitres of PCR amplified sample was mixed with 1µl loading buffer and loaded in to the gels. Q Step 1, 100bp or 50 bp ladders (New England Biolabs, UK) was used as standard markers. Gels were then run in 1x TAE in electrophoresis tank under 65 - 100 V using a Standard Biometra power pack for 30 - 90 minutes. GeneFlash (Syngene, UK) transilluminator system was used to visualize gels under the UV light.

2.5 Purification of PCR products

Purification of DNA from gel bands was carried out with the Promega Wizard SV Gel and PCR Clean-Up System as follows. First, the correct DNA band was excised using a clean razor blade with a minimal exposure to UV and also including a minimal volume of agarose. The gel slice was then placed in a pre-weighed 1.5ml eppendorf. The weight of the band was calculated by subtracting the weight of the empty eppendorf from the weight after introducing the gel slice.

Membrane binding solution was added at a ratio of 10µl of solution per 10mg of gel slice. Then the mixture was vortexed and incubated at 65°C on a heating block until the gel slice was completely dissolved. The eppendorf was vortexed every 2 minutes to increase the rate of agarose melting. The dissolved gel mixture was transferred in to a SV minicolumn placed in a collection tube and incubated at room temperature for 1 minute before being centrifuged for 1 minute at 16,000*g*. The minicolumn was transferred in to another clean collection tube and the liquid accumulated in the previous collection tube was discarded. The column was washed by adding 700µl of membrane wash solution and centrifuged for 1 minute at 16,000*g*. The collection tube was emptied as before. Then, 500µl of membrane wash solution was added to the minicolumn before being centrifuged for 5 minutes at 16,000*g*. The flow through was discarded and the minicolumn assembly was centrifuged for 3 minutes to evaporate residual ethanol. The minicolumn was then transferred to a new 1.5ml eppendorf and 50µl of nuclease free water was applied directly to the center of the column. Following an incubation for 1 minute at room temperature, the assembly was centrifuged for 1 minute at 16,000*g*. The eluted DNA was stored at -20°C.

2.6 Methylation sensitive amplified fragment length polymorphism (MS-AFLP)

MS-AFLP has been used as a genome-wide methylation screening technique in many epigenetic studies (Schrey et al., 2013). It detects variation in methylation status of CCGG recognition sites, the CpG enriched regions in a genome.

During this technique DNA is cut with methylation sensitive (HpaII and MspI, frequent cutters) and insensitive (EcoRI, rare cutter) restriction enzymes and double stranded adapters are ligated to the ends of the digested fragments, to produce template DNA for subsequent amplifications (Vos et al., 1995). Depending on the genome size, restriction-ligation generates thousands of adapter-ligated fragments. For visualization after electrophoresis, only a subset of these fragments is amplified using selective primers, which are extended into the unknown part of the fragments usually 1 - 3 arbitrarily chosen bases beyond the restriction site (Figure 2.1). Because of the high selectivity, primers differing by only a single base in the AFLP extension

will amplify a different subset of fragments. Therefore when AFLP amplifications are performed using a combination of primers, methylation statuses of CCGG sites in the whole genome can be screened. The sequence of the adapter and the adjacent restriction site serves as the binding site for primers (Figure 2.1).

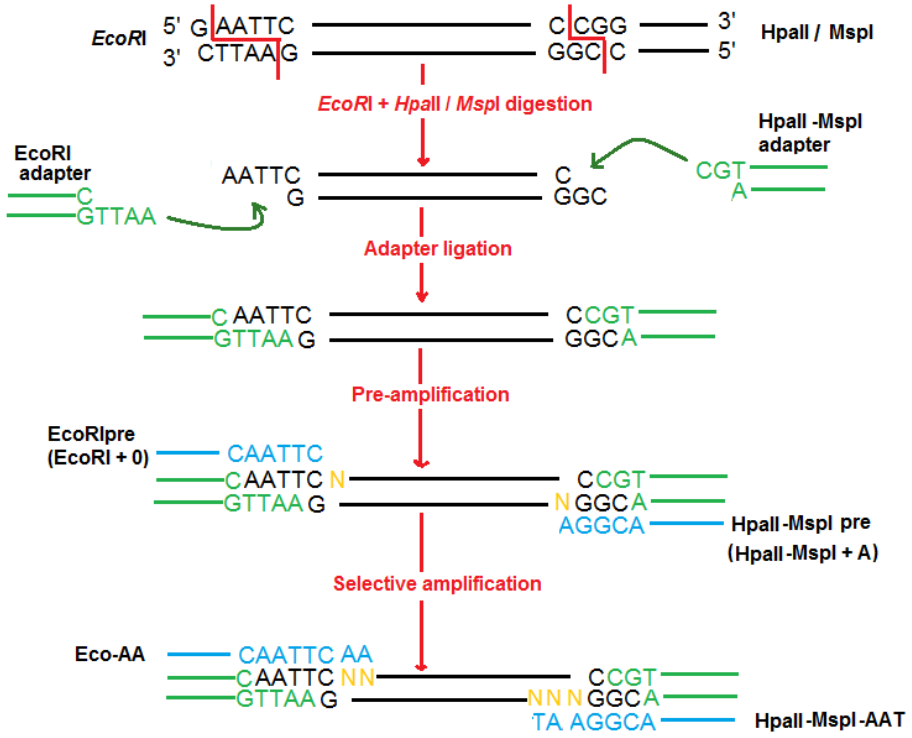

Figure 2.1: An illustration of MS-AFLP principle

Both HpaII and MspI isoschizomers recognize and cleave the same restriction site (5'-CCGG-3'), but have differential sensitivity to cytosine methylation. See Appendix figure A.1a.

The logic behind the use of two restriction enzymes is that the frequent cutter will generate smaller fragments that will amplify well and are in the optimal size range for separation on a denaturing gel. The use of a rare cutter will reduce the number of fragments produced and only the rare cutter-frequent cutter fragments will be amplified in subsequent steps (Vos et al., 1995). The pattern and size distribution of the restriction fragments implies the overall methylation level in each bee. Main steps of the MS-AFLP protocol are given below.

2.6.1 Restriction digestion

The MS-AFLP protocol was modified as according to Kronforst et al. 2008. 500ng of genomic DNA was digested with EcoRI and MspI (3µl of the target DNA, 0.05µl of EcoRI (20,000 units/ml), 0.25µl of MspI (20,000 units/ml), 1µl 10x NEBuffer 4 and 5.7µl of dH_20) and another 500ng with EcoRI and HpaII (3µl of the target DNA, 0.05µl of EcoRI (20,000 units/ml), 0.5µl of HpaII (10,000 units/ml), 1µl 10x NEBuffer 1 and 5.45µl of dH_20) at 37°C for 3 hours.

2.6.2 Adapter ligation

Restriction digested products of the two reactions, EcoRI-MspI and EcoRI-HpaII were then individually ligated with EcoRI adaptors (5µl of EcoRI-F and 5µl of EcoRI-R were mixed in a final concentration of 5pmol/µl and incubated at 65°C for 10 minutes) and HpaII-MspI adapters (25µl of HpaII-MspI-F and 25µl of HpaII-MspI-R were mixed in a final concentration of 50pmol/µl and incubated at 65°C for 10 minutes) respectively. See Appendix table A.1 for the sequences of all adapters and primers. 3µl of the digested product was combined with 7µl of the ligation reaction mixture (1µl of EcoRI adapter (5pmol), 1µl of HpaII-MspI adapter (50pmol), 0.25µl of T4 DNA ligase (400,000 units/ml), 1µl of 10x T4 ligase buffer (New England Biolabs) and 3.75µl of dH_20) at 37°C for 3 hours and then left overnight at room temperature. The ligation products were diluted with 100µl of dH_20 and used as the template for pre-amplification.

2.6.3 Pre-amplification

The first PCR (pre-amplification) used 1µl of ligation product with 1µl of each EcoRIpre and HpaII-MspIpre primers (10pmol/ml), and 7µl of the reaction mix (0.8µl of 2.5mM deoxynucleotide triphosphates (dNTPs), 1µl of 10x Paq5000 Hot Start Reaction Buffer, 0.3µl of Paq 5000 Hot Start DNA Polymerase (500 units), 0.8µl of 25mM $MgCl_2$, 4.1µl of sterile distilled H_20). The PCR conditions were 94°C for 2 minutes, followed by 20 cycles of 94°C for 30 seconds, 60°C for 1 minute and 72°C for 1 minute, followed by a final extension of 5 minutes at 72°C. 3µl of each

PCR product was run on 3% agarose gel (see section 2.4) and appearance of a smear of DNA on the gel indicated that the pre-amplification PCR was successful.

2.6.4 Selective amplification

Seven microlitres of PCR products were then diluted with 93µl of dH_20 and used as the template for selective amplification. During this step, one of 12 possible selective primer combinations (Figure 2.2) was used. This reduces the number of fragments visualized by gel electrophoresis to a usable number. The selective PCR reaction mixture was composed of 1µl of pre-amplified product, 1µl of HpaII-MspI primer (10pmol/ml), 1µl of EcoRI primer(10pmol/ml) and 7µl of reaction mix. PCR conditions used were 94°C for 2 minutes followed by 36 cycles (13 cycles of 30s at 94°C, 30s at 65°C (0.7°C reduction per cycle) and 1 minute at 72°C followed by 23 cycles of 30s at 94°C, 30s at 56°C and 1 minute at 72°C) followed by a final extension at 72°C for 5 minutes before a holding step at 4°C.

A summary of the MS-AFLP protocol described above is given in Figure 2.2

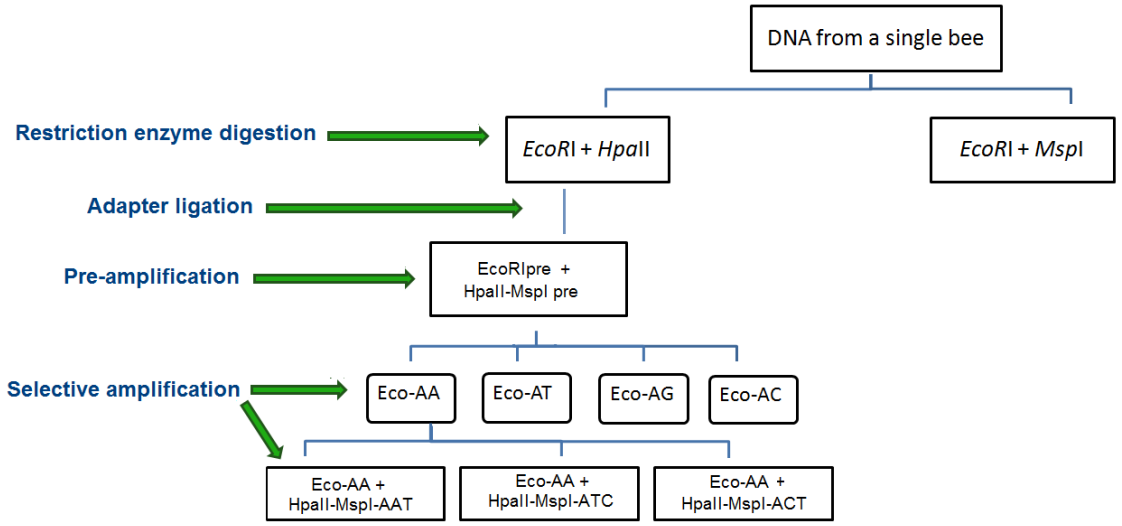

Figure 2.2: A flow diagram to show different primer, adapter and enzyme combinations used during the MS-AFLP protocol.

2.6.5 Gel staining and visualization of PCR products

The temperature of the running buffer (30mM TAE, Appendix table A.2) in the Origins electrophoresis system was brought up to 48°C prior to introduce 9% Poly(Nat) gels (Elchrom Scientific). PCR products from selective amplification were diluted with 100µl of sterilized water. 10µl of the diluted product was combined with 2µl of Elchrom loading dye and loaded in to a well on a 25 well 9% Poly(NAT) gel. The gels were run at 120V and 55°C for 81 minutes. The cooling pump was turned on after a delay of 2 minutes subsequent to loading.

The gel was then stained in the dark with SybrGold (Invitrogen) (1:10000 diluted in TAE) followed by a similar destaining step with 100 ml TAE for 30 minutes.

2.6.6 Band scoring and data analysis

Bands were scored as either present or absent using Gelanalyzer, 2010 (`http://www.gelanalyzer.com/download.html`; See Appendix figure A.1b). The resulting matrix was analysed using the R package MSAP version 2.15 which uses a combination of analysis of molecular variance (AMOVA) and Principle coordinate analysis (PCoA)(Pérez-Figueroa, 2013).

Chapter 3

Methylation differences during development and between worker reproductive castes

3.1 Introduction

The size of an eusocial insect colony varies greatly from a simple organization of four individuals living inside a twig (e.g allodapine bees, Schwarz et al. 2011) to a highly complex system comprising of millions of individuals (e.g. leafcutter ants, Hart and Ratnieks 2001). Despite the number of individuals present, a remarkable cooperation and division of labour is maintained within these colonies (Queller and Strassmann, 2009). Also in order to respond to external environmental challenges as an integral unit, its members show a great degree of phenotypic plasticity. That is although every individual in the colony shares the same genome they are capable of converting to morphologically, physiologically and behaviourally distinct phenotypes depending on the colony requirement (Weiner and Toth, 2012).

In eusocial societies, reproductive division of labour mainly exists between reproductive queens and sterile workers while non-reproductive division of labour exists between sterile workers that are specialized to perform nest maintenance, brood caring, foraging and defensive tasks important for the survival of the colony. However,

under some circumstances, workers begin to compete with the queen and each other over male production. This divides workers into two distinct castes; reproductive and non-reproductive (Wilson and Holldobler, 2005). Thus reproductive division of labour in a social insect colony is not just between the queen and workers, but it can also be between workers. The switch between sterility and reproduction in workers is a much more plastic process than queen-worker differentiation (Yagound et al., 2012).

Reproductive conflicts within a social insect colony vary depending on its species biology (Bourke and Ratnieks, 1999). Worker reproduction is rare in honeybees (Ratnieks, 1993). However, male production by workers is common in bumblebees. This makes *Bombus terrestris* a valuable model to test the importance of methylation in worker reproduction. The annual colony life cycle in the bumblebee is divided into a cooperative pre-competition phase when the queen has absolute reproductive dominance and a highly aggressive competition phase later in the season when workers and the queen compete over male production (Duchateau and Velthuis, 1988). If the queen dies or is removed, workers can be clearly differentiated into reproductive and non-reproductive subcastes by both their ovary development and aggressive behaviour (Amsalem and Hefetz, 2011). It seems likely that this extreme genomic flexibility is achieved via epigenetic modifications (Weiner and Toth, 2012).

Haig's kinship theory for the evolution of genomic imprinting predicts that there should be a conflict between maternally derived alleles and paternally derived alleles of loci involved with worker reproduction. That is, worker reproduction loci should be imprinted (Queller, 2003). The first step in testing this theory is to search for the molecular mechanism of genomic imprinting, methylation, at worker reproduction loci.

DNA methylation is one of the most widely conserved forms of epigenetic and genomic imprinting mechanism. It is associated with modulation of gene expression in eukaryotic organisms (Glastad et al., 2011). Methylation tags can be transmissible across generations through cell division (Bonasio et al., 2010a) and also can be reversed by appropriate stimuli such as heat and chemical agents (Christman, 2002, Feliciello et al., 2013).

Localization of methylation patterns vary substantially among taxa. In contrast to the genome-wide methylation found in vertebrates, methylation in insects is sparse and found mainly within the gene body (Zemach et al., 2010). In eukaryotic genomes, DNA methylation primarily occurs at CpG dinucleotides. Since methylated cytosines are frequently subjected to spontaneous deamination, CpG dinucleotides are gradually depleted in many methylated regions (Duncan and Miller, 1980). Therefore genes with low CpG content tend to be hypermethylated and vice versa (Bird, 1980). Recent honeybee (*Apis mellifera*) research has shown that the low CpG genes are important in biological processes such as gene expression, translation and metabolism whereas high CpG, hypomethylated genes are primarily involved in development (Steyaert, 2014, Zeng and Yi, 2010).

DNA methylation patterns can be tissue specific or consistent among different cell types. For example, compared to the other tissues in the body a mature sperm in humans is significantly hypomethylated (Rakyan et al., 2008). In mammals, genes with hypermethylated promoters are transcriptionally silenced whereas those with hypomethylated promoters show a broad range of gene expression (Smith and Meissner, 2013). Conversely in insects, most of the hypermethylated genes are broadly expressed and function as housekeeping genes whereas hypomethylated genes are only expressed in specific tissues (Foret et al., 2009). Housekeeping genes in both the honeybee and all seven sequenced ant genomes show enrichment for methylation, suggesting that methylation plays a common role in Hymenoptera (Simola et al., 2013). Another study found that DNA methylation affects alternative splicing and thereby contribute to a longer gene length and a slower rate of evolution in housekeeping genes (Flores et al., 2012). In relation to reproduction, different hymenopteran species may have evolved to utilize methylation, to imprint certain genes differentially, depending on whether they are inherited from a queen or a drone (Queller, 2003). Also it has been suggested that DNA methylation leads to production of caste-specific protein isoforms via alternative splicing (Glastad et al., 2011). Thus without altering the expression of different genes, it can promote the development of specific castes through the expression of different isoforms of the same gene.

Methylation systems are not ubiquitous among insects. CpG methylation is absent in flies, *Drosophila* (Lyko and Maleszka, 2011) and *Anopheles* (Elango et al., 2009),

although see Boffelli et al. 2014. Diminished levels of methylation are observed in the beetle, *Tribolium* (Feliciello et al., 2013). However, methylation systems in social insects appear to be more common (Kronforst et al., 2008). The honeybee (*Apis mellifera*) was the first insect found to have a fully functioning methylation system with all catalytically active orthologs of vertebrate *Dnmts* (Wang et al., 2006). Since then whole-genome sequencing of seven ant species has shown the DNA methylation toolkit is conserved across eusocial Hymenoptera: the carpenter ant, *Camponotus floridanus* (Bonasio et al., 2010b); Jerdon's jumping ant, *Harpegnathos saltator* (Bonasio et al., 2010b); the red harvester ant, *Pogonomyrmex barbatus* (Smith et al., 2011a); the Argentine ant, *Linepithema humile* (Smith et al., 2011b); the fire ant, *Solenopsis invicta* (Wurm et al., 2011) and two leaf-cutter species, *Acromyrmex echinatior* (Nygaard et al., 2011) and *Atta cephalotes* (Suen et al., 2011).

The evolution of eusociality and DNA methylation is most studied in *A. mellifera*. It plays a significant role in honeybee development (Elango et al., 2009, Lyko et al., 2010) and behaviour (Guan et al., 2013, Lockett et al., 2010). A female honeybee larva develops from a totipotent egg which has the potential to develop into either a queen or worker depending on differential nourishment they receive. This differential nourishment is thought to affect their *de novo* methylation. Altering methylation leads to a worker-destined larvae becoming a queen (Kucharski et al., 2008). Also reversion of DNA methylation patterns in the honeybee brain is involved in the switch between different types of behaviour and non-reproductive worker roles (Herb et al., 2012).

The main aim of this chapter was to examine the role of methylation in *B. terrestris* worker reproduction and development. First, the effect of methylation in formation of different reproductive castes was assessed. During this, methylation statuses of queenless reproductive and queenless and queen-right non-reproductive workers were compared using methylation-sensitive amplified fragment length polymorphism (MS-AFLP). The basic principle of MS-AFLP is given in section 2.6.

Towards the end of the above experiment, the *B. terrestris* colony, went through a social phase transition from pre-competition phase to competition phase. According to previous research, during this transition the reproductive status of colony workers changes dramatically (Alaux et al., 2007) and a significant increase of worker

oviposition is observable even after 2 days of the queen removal (Bloch and Hefetz, 1999). Methylation differences in queenless reproductive workers between the pre-competition and competition phase were assessed by MS-AFLP. During this analysis brain and ovary tissues were compared to check if any tissue-specific methylation difference occurred within them during this transition.

During the next set of experiments, a demethylating agent (Christman, 2002) was used to check if methylation is fundamentally involved in alteration of worker reproductive ability. Low doses of 5-aza-2'-deoxycytidine (Decitabine) was fed to adult and callow worker bees and their ovary development and aggression was assessed as indicators of these workers taking on a reproductive role (Amsalem and Hefetz, 2011).

Finally, to assess the degree of methylation difference cause by Decitabine, bee samples from the treatment and control experiments were compared using three different methylation analysis techniques: Methylation sensitive restriction enzymes (MSRE), Bisulfite treatment and Amplification of inter-methylated sites (AIMS).

3.2 Methods

3.2.1 DNA methylation differences between worker reproductive castes

3.2.1.1 Rearing of different bumblebee reproductive castes

Bumblebee callow workers (less than 1 day old) from a single colony were reared in 4 separate boxes, 5 workers per box. Another 5 callow workers, captured at the same time with bees in the above 4 boxes, were tagged with a coloured Opalith tag (Christian Graze KG, Germany) and released into the original queen-right colony. The maximum number of callow workers collected within a single day was limited to 10. Thus the total number of callow workers collected everyday was equally divided as box workers and colony workers. Similar aged callow workers which were released back to the colony were marked with a similar colour tag (Figure 3.1a) and various

coloured tags were used to differentiate these bee groups by age (Appendix table A.3).

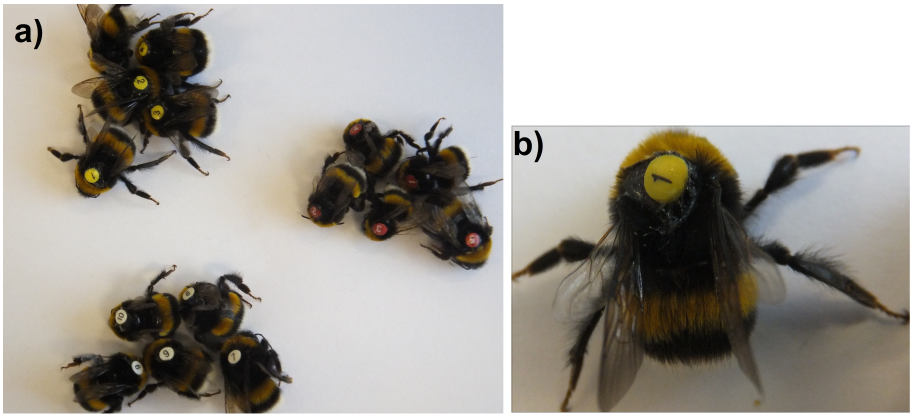

Figure 3.1: Colony workers marked with different coloured tags

Bees in the first 3 boxes were collected before the colony queen died whereas those reared in box 4 were collected after the queen died. After 6 days, same age bees were collected from the colony and dissected with their similar aged bees reared in the box.

Bees were sacrificed by freezing at -20°C for 30 minutes. The reproductive status of each worker was confirmed by examining the ovaries, see section 3.2.4.3. Any worker with discernible ovaries was classed as reproductive (oocyte length: mean+/- standard deviation = 0.6375 +/- 0.2459 mm), whereas non-reproducing bees had no discernible ovaries. The bees and their ovary samples were immediately stored at -80°C till DNA extractions were performed.

3.2.1.2 Methylation analysis of different reproductive worker castes

Bees which were reared in the first 3 boxes (collected before the colony queen died) were used for this experiment (Appendix table A.3). Genomic DNA was extracted using the QlAamp DNA Mini kit from the heads of bees belonging to 3 different groups: a reproductive worker (**RW**; 3 bees), a non-reproductive box worker (**BW**; 3 bees) and a non-reproductive queen-right colony worker (**CW**; 4 bees). DNA was eluted in distilled water instead of Buffer AE provided in the kit since it may affect the restriction digestion in subsequent steps.

The quantity and quality of DNA was assessed according to section 2.4. Methylation statuses of these 3 reproductive castes were analysed by MS-AFLP as described in section 2.6.

3.2.1.3 Sequencing of loci showing methylation differences

Bands of interest from the MS-AFLP analysis (e.g. Bands relevant for hemi or full methylation etc, see Appendix figure A.1b) were excised from 9% PolyNat gels under UV light, using BandPick^{TM} (Elchrom) and stored at -20°C. These gels are thermally stable for up to a temperature of 95°C and are also inert for a wide range of chemicals. I was unable to extract DNA directly from gels using a commercial gel cleanup kit. Therefore DNA extractions were performed using the below protocol as recommended by Elchrom Scientific, 2012 (http://www.elchrom.com/fileadmin/pdf/Catalogue-01:14.pdf).

DNA in an excised gel band was incubated with 50µl of distilled water at 65°C for 30 minutes. Quantification of DNA in the above extraction was carried out in two ways; with the use of NanoDrop 1000 Spectrophotometer and 3% agarose gel. Although I was able to measure the concentrations of DNA extractions using the NanoDrop, a blank gel picture indicated that the amount of DNA in these extractions was very low.

Therefore 5µl from each DNA extraction was PCR amplified with its correspondent Eco- and HpaII-MspI- primers, used during the selective amplification of the original MS-AFLP protocol (Appendix table A.3). Amplified products were then checked on 3% agarose gel and the presence of bands with the correct product size indicated that the PCR was successful. 50bp and 1kb DNA ladders (NEB) were used as the standard markers to estimate the product size and the concentration. DNA concentration in each sample was estimated by comparing their intensities with that of the reference and thereby the respective concentration given in the standard marker.

Gel bands with the correct product size were excised from the agarose gel and purified with the Promega Gel and PCR Clean-Up System prior to send for sequencing. However the DNA quantity after purification was too low to detect by agarose gel

electrophoresis indicating that there was a loss of the PCR product during purification. Therefore the PCR products were directly sent for a commercial clean up and sequencing (Source BioScience LifeSciences, UK) with their respective HpaII-MspI-primers.

The sequence results from PolyNat gels were analysed by NCBI BLASTx (`http://blast.ncbi.nlm.nih.gov/Blast.cgi`) to find the matching gene sequences in the *Bombus terrestris* genome (Table 3.3).

3.2.2 Tissue specific methylation differences between queenless reproductive workers

To compare the methylation status of different reproductive castes (reproductive and non-reproductive), it is essential to select a tissue which is common to all reproductive castes. Thus bee heads were selected in section 3.2.1.2. In this section I wished to examine methylation differences between different tissues of reproductive workers. Ovary is the main tissue that is involved in worker reproduction. In mammals, they are important as sites where maternal imprints are established during oogenesis (El-Maarri et al., 2001, Olek and Walter, 1997) and many differentially methylated regions are associated with ovary functions and development (Bartolomei and Tilghman, 1997).

The aims of this experiment were, (i) to compare methylation statuses among different tissues; ovaries and heads and also (ii) to check if any methylation difference occur in queenless reproductive workers during the colony transition from pre-competition to competition phase.

The annual life cycle of a *B. terrestris* colony goes through two distinct social phases: First, the 'pre-competition phase' (PCP) where reproduction is exclusive to the queen. Second, the 'competition phase' (CP) in which the queen starts laying male and gyne destined eggs and the mature workers in the colony begin to compete with the queen and also among each other over male parentage. As a result, aggressive interactions and oophagy among workers and between workers and the queen arise. In many cases the queen dies within the CP (van Honk and Hogeweg, 1981).

Callow workers for boxes 1, 2 and 3 were collected while the colony was queen-right i.e. before the queen died (during PCP) where as workers for box 4 were collected after the queen died (during CP). See Appendix tables A.3.

During the 7 days of bee rearing (see section 3.2.1.1), the 5 callow workers reared in box 4 (CP) were all reproductive whereas in the first 3 boxes (PCP), only a single reproductive worker was present per box (Appendix table A.3). The rest of the 4 workers in these boxes lacked any visible ovaries (i.e. non-reproductive workers).

Methylation analysis of reproductive worker tissues (ovaries and heads) belonging to pre-competition and competition phases was conducted using the MS-AFLP protocol as described in section 2.6.

DNA was extracted using the DNA Mini kit from ovary and head tissues of the 3 reproductive workers in boxes 1, 2, 3 and 4 reproductive workers in box 4: 3 head samples and 3 ovary samples (PCP), 4 head samples and 4 ovary samples (CP).

Due to the limitation of well number (25 wells) in a PolyNat gel, the 3 DNA samples of head tissues from reproductive workers in boxes 1, 2, 3 were run on a separate gel while the rest of all samples were run on a single gel. During DNA extraction, the initial digestion of ovary samples with proteinase K was limited to 1.5 hours (it's usually 3 hours for hard tissues like bee heads) to avoid any DNA degradation.

3.2.3 Methylation differences between different developmental stages of *B. terrestris* life cycle

Methylation analysis of different *B. terrestris* developmental stages (L2, L3, L4 larval stages, pupae and and callow workers) were carried out by Dr. Crisenthiya Clayton using AIMS analysis (see section 3.2.4.6). The resultant gel images from this experiment were reanalysed by me using the R package MSAP (Pérez-Figueroa, 2013). Bands were scored as either present or absent using Gelanalyzer, 2010 (`http://www.gelanalyzer.com/download.html`). Each gel image was comprised of 12 samples: 2 replicates of each L2, L3, L4 larval stage, 3 replicates of pupae and 3 replicates of callow workers.

3.2.4 Effect of DNA demethylation on *Bombus terrestris* worker reproduction

3.2.4.1 Decitabine as a demethylating agent

5-aza-2'-deoxycytidine (Decitabine) was used to alter methylation and to find its effect on ovary development and reproductive behaviour in queenless workers.

A stock solution of Decitabine was made by dissolving 5mg of Decitabine in 2ml of 1:1 v/v acetic acid : distilled water. Then the samples were aliquots in to 200µl volumes and stored in -80°C to prevent decomposition.

3.2.4.2 Rearing queenless mini colonies

Workers were reared in separate Perspex boxes, with its own supply of pollen and apiary syrup (Figure 3.2c). A 10mM non-lethal dosage of Decitabine (18.5ml) was added to the apiary syrup (20ml) of the test group while the control group was fed with unadulterated apiary syrup. As a preliminary test, both apiary solutions were coloured using a natural food colorant (green) and randomly a few bees were dissected after 24 hours to confirm that they were drinking the solution. Fresh apiary solution, Decitabine and pollen were provided everyday through the experiment.

One hundred workers from 3 different colonies were used for this experiment: preliminary experiment 1 (20 adult workers), preliminary experiment 2 (40 adult workers), callow worker experiment (40 callow workers).

During the preliminary experiment 1 and 2, the number of antennal segments (12) and abdominal segments (6) of each adult bee was counted to confirm that they are female bees. For a male bee these counts are 13 and 5 respectively (Prys-Jones and Corbet, 2003).

Preliminary experiment 1 20 adult workers (1 week old), were reared in two Perspex boxes, each box containing 10 bees. Bees in one box were fed Decitabine while the other box was used as the control.

Preliminary experiment 2 Preliminary experiment 1 was expanded with a new colony using 40 adult workers (1 week old), 20 bees as the test and 20 bees as the control. The 20 bees in each test and control group were subdivided into 4 groups such that each Perspex box contained 5 adult bees (Figure 3.2c). Unlike in the Preliminary experiment 1 where all workers attempted to reproduce, here only a single reproductive worker was present per box. Thereby the optimum number of workers required to observe the reproductive caste differentiation in a queenless colony was determined as five.

Callow worker experiment 40 callow workers (less than 24 hours in age) (Figure 3.2a) from a different colony were used, 20 bees as the test and 20 bees as the control. Newly emerged workers were collected every day and reared for 7 days in separate Perspex boxes as described in preliminary experiment 2 (Figure 3.2c). In order to distinguish callow workers from adult workers, all existing workers in the queen-right colony were marked with a white spot of nail varnish on their thorax.

Unlike callow workers, adult bees had to be reared for 3 weeks, before I observed nest construction or egg laying as an indicator of these workers initiating their ovary functions. Therefore, behavioural and methylation analyses were only carried out on callow workers, whereas ovary measurements were made on both adults and callows.

Figure 3.2: (a) A callow worker emerging from a cell. (b) *B. terrestris* callow worker (c) Rearing queenless mini colonies.

3.2.4.3 Dissections and measurements of ovarian development

Once egg-laying or nest construction behaviour was observed (7 and 21 days for callow and adults respectively), bees were sacrificed by freezing at -20^{0}C for 30 minutes. They were dissected by making two lateral incisions in the abdomen to observe ovary development. Dissections were carried out on a ice cooled surface in double distilled water.

Then the ovaries were removed and the average length of the largest oocyte in each of the two ovaries was measured as an index for their ovary development (Appendix tables A.5 and A.6). All measurements were obtained to the nearest 0.05mm with an eyepiece micrometer under a dissecting microscope. The length of the largest oocyte in bumblebees is tightly correlated with a worker's reproductive status (Foster et al., 2004, Geva et al., 2005). Always the terminal (basal) oocyte was observed as the largest oocyte in an ovary (Figure 3.3b).

Another 10 bees from each of the queen-right colony were also dissected to check their ovary development in the presence of queen.

3.2.4.4 Additional measurements

Apart from the mean length of the largest oocyte, wet weight of both ovaries in each bee and the weight of brood chambers (nests) produced by bees in each box (Figure 3.4) were also recorded to the nearest 0.0001g on a Sartorius Digital Scale BP 61S (Appendix tables A.5 and A.6). In preliminary experiment 1, only the wet weight of both ovaries was measured as an alternative indicator of the ovary size and thereby to assess the reproductive potential of workers (Appendix table A.4).

Bees and their ovary samples were immediately stored at -80°C until DNA extractions were performed to compare methylation status of the test and control groups. All dissections and measurements were taken approximately within 2 minutes from each bee.

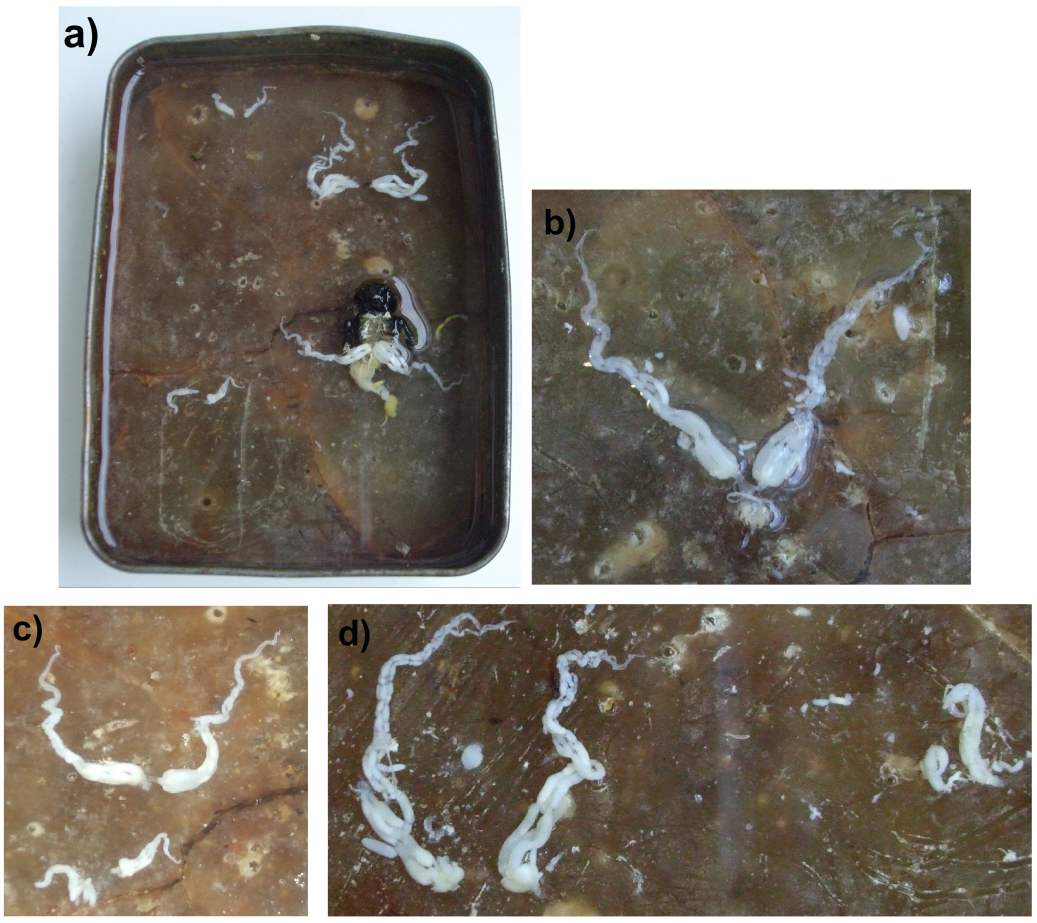

Figure 3.3: Developed and undeveloped ovaries of *B. terrestris* workers (a-d).

3.2.4.5 Behavioural observations

Aggression is a clear parameter of reproductive conflict among workers (Amsalem and Hefetz, 2010, 2011). The dominance hierarchy of bees is usually established through their overt agonistic behaviours (Cnaani et al., 2007, Duchateau, 1989). Hence, among numerous behaviours present in bees, only aggressive and threatening behaviours which come under this category were recorded. Three distinct behaviours were recorded as follows: a) 'Attack' included occurrence of one of the following behaviours; biting, pushing, head butting, dragging, wing pulling, struggling or an attempt to sting. b) 'Darting'; a sudden movement of a bee towards the direction of another bee but, without any body contact between the two bees. c) 'Humming/ Buzzing'; a series of rapid, short wing vibrations that produce a distinctive buzzing sound.

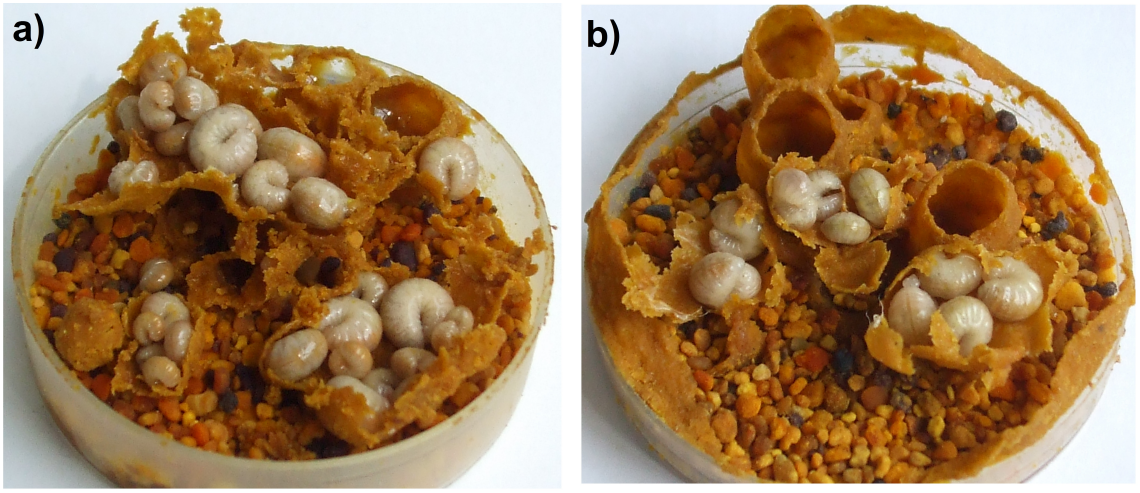

Figure 3.4: Brood produced by Decitabine treated (a) and control (b) group bees

Callow workers collected from the main colony were placed in separate boxes about 16 hours before the behavioural experiment started. This time allowed the bees to acclimatize to their new environmental conditions. Bees in each box were scanned for 10 minutes 3 times a day at fixed intervals (9.00 am, 1.00 pm and 5.00 pm) and the frequency of each behaviour recorded. Sampling of behaviour was carried out during 6 days of the rearing period. During observation times, mini-colonies were kept spaced (10cm) equally. An index of aggression was constructed as the unweighted sum of 'Attack', 'Darting' and 'Humming' that were observed during all observation bouts throughout the experiment.

3.2.4.6 Comparison of DNA methylation differences among Decitabine treated and non-treated groups

I used three techniques to attempt to measure DNA methylation differences here:

(a) Methyl sensitive restriction enzymes (MSRE)

(b) Bisulfite treatment

(c) Amplification of intermethylated sites (AIMS)

(a) Methylation sensitive restriction enzymes (MSRE) MSRE PCR (Figure 3.5) uses restriction enzymes that are sensitive to methylation of a cytosine at

a certain CpG site. If a cytosine is methylated, then the enzyme cannot cleave that DNA thus the sequence can be amplified with a suitable pair of primers. Conversely, if the cytosine of a CpG site is not methylated the enzyme can cleave the DNA. Therefore the sequence cannot be amplified by PCR. Hence by designing primers which cover the region of interest, the presence or absence of methylated CpG sites in that region can be determined by the presence or absence of a DNA band followed by enzymatic digestion (Roach and Hashimoto, 2007).

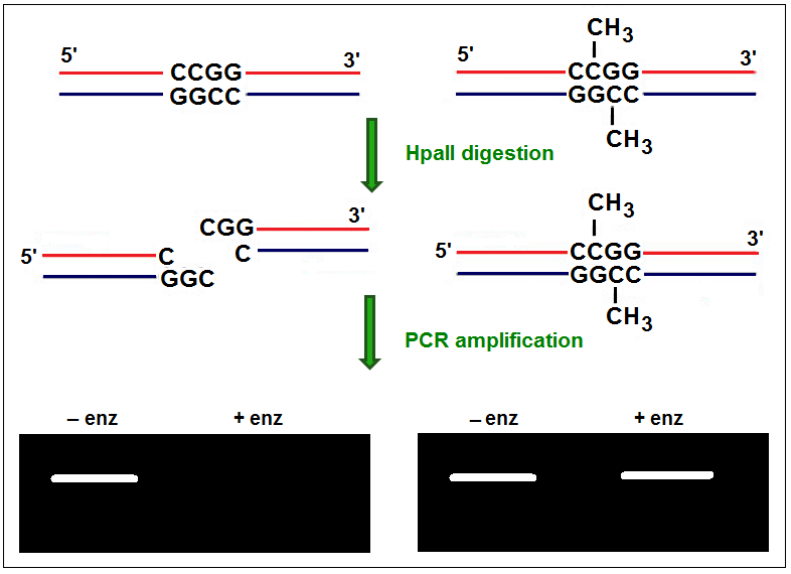

Figure 3.5: Principle of MSRE assay for methylation status at a specific CpG site. In the absence of methylation (left), DNA is cleaved by the enzyme thus cannot be amplified by PCR. In the presence of methylation (right) DNA remain intact for HpaII and can be amplified by a suitable primer pair.

Dynactin p62 was used as the candidate gene to compare among the Decitabine treated and control groups due to the presence of a high concentration of CpG sites and it has already been used for the same purpose in *Apis mellifera* (Kucharski et al., 2008).

The genomic DNA sequence of *Apis mellifera dynactin p62* (Gene ID: 725207) was blasted against *Bombus terrestris* nucleotide library and the *dynactin* equivalent gene in *Bombus terrestris* was found as *dynactin subunit-4 like* (Gene ID: 100649495).

According to Kucharski et al. 2008, methylation in *Apis dynactin p62* is exclusively present within the coding exonic regions. By an alignment between genomic DNA and mRNA sequences of *dynactin subunit-4 like*, sequences of exons were obtained.

Exon 6 included the highest number (9) of CpG sites as well as the HpaII restriction enzyme recognition site, CCGG.

Primer 3 (http://primer3.ut.ee/) was used to design primers for exon 6. Forward and revere primer sequences used were 5' CTACGTTGGAACAGCGATTG 3' and 5'TGCAGAATGTGAACAAAAATCC3' respectively.

Twenty adult workers from preliminary experiment- 2 and 20 callows from callow worker experiment (section 3.2.4.2) were used for MSRE analysis. Each 20 bee group was comprised of 10 Decitabine treated and 10 non-treated control bees. DNA extractions were performed with the DNA Micro kit (section 2.2.2), using a 10mg of tissue sample from each bee. 9µl (5ng/µl) of genomic DNA from each bee was digested with 1µl of NEB 10x PCR Buffer 1, 1µl of HpaII (10000 units/ml) and 4µl of dH_20. For each genomic DNA, a parallel sample omitting the enzyme was also prepared as the non-enzyme control. Both samples were incubated in a thermal cycler at 37°C for 3 hours. Then the enzyme was denatured by heating at 65°C for 10 - 20 minutes.

DNA was PCR amplified, immediately after the restriction digestion. 5µl of digested DNA was used for this with 12.5µl of YB-Taq 2x Buffer, 1.5µl of each *dynactin* forward and reverse primers (10pmol/ml), 0.5µl of 50mM $MgCl_2$ and 4µl of dH_20. PCR profile used was, 2 minutes at 94°C, 35 cycles of 30s at 94°C, 30s at 58°C, 1 minute at 72°C and a final extension of 5 minutes at 72°C. Amplified products were checked on a 3% agarose gel.

(b) Bisulfite sequencing Bisulfite sequencing is considered the gold-standard for analysing DNA methylation (Suzuki and Bird, 2008) as it produces results with a single-nucleotide resolution (Figure 3.6).

The aim of this experiment was to perform bisulfite treatment for a known gene region with many CpG sites, then to clone the region into *E.coli* and obtain sequences to compare methylation statuses among Decitabine treated and control groups.

Dynactin subunit-4 like was used as the candidate gene for this experiment (section 3.2.4.6a) and exonic sequences were obtained by an alignment between genomic DNA and mRNA sequences. Among the 10 exons found in this gene,

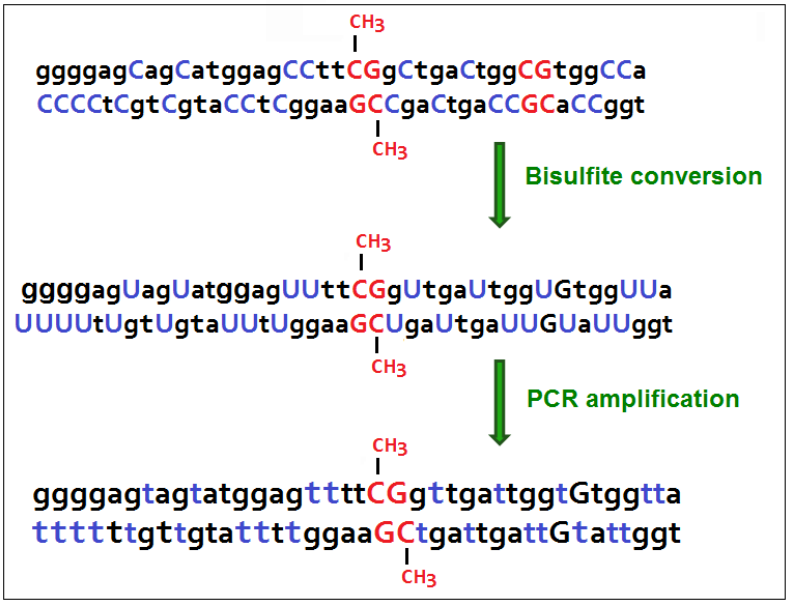

Figure 3.6: Schematic of sodium bisulfite conversion. Cytosine in a CpG dinucleotide (red) can be subjected to methylation changes. However, individual cytosines (blue) remain intact for methyltransferases. During the bisulfite treatment all un-methylated cytosines are converted into uracils and amplified as thymines whereas methylated cytosines are resistant to conversion thus remain protected. (Nucleotide sequence was adapted from German cancer research center, 2011; http://www.dkfz.de/gpcf/394.html).

exon 5, 6, 7 and 9 were selected for bisulfite treatment since they contained the highest number of CpG dinucleotides (Table 3.5) compared to the other exons. Primers were designed to these exonic sequences using MethPrimer program (http://www.urogene.org/methprimer/). All primers designed were 24-32 bases in length to achieve a maximum amplification of bisulfite treated DNA.

The optimum DNA fragment length for a successful PCR amplification after bisulfite treatment, using the EZ DNA Methylation-Gold^{TM} kit, is recommended as 150-300bp (http://www.zymoresearch.com/downloads/dl/file/id/69/d5020i.pdf). Amplicon length of each exon is given in table 3.5.

To identify the exon with the maximum number of methylated CpG sites, among exon 5, 6, 7 and 9, a preliminary experiment was conducted (Table 3.5). During this process, a DNA sample of *Bombus terrestris* worker was treated with bisulfite and amplified with their exon specific primers as given in Table 3.5. Optimum annealing temperature for each primer pair was obtained from a gradient PCR. Then the amplified products were run on a 3% agarose gel, correct bands were

excised, cleaned up with the Promega Wizard SV Gel and PCR Clean-Up System (section 2.5) and sequenced. Sequence results were analysed using Finch TV and Geneious version 7.3.0.

According to the sequencing results (Table 3.5), exon 6 contained the highest number of methylated cytosines. Thus it was selected to compare among the Decitabine treated and control bee groups.

Only the callow worker samples in section 3.2.4.2 were tested with bisulfite treatment.

Five hundred nanograms of DNA from Decitabine treated and control group bees were subjected to bisulfite conversion using the EZ DNA Methylation-Gold^{TM} Kit. These samples were then amplified with the methylation specific primers designed for exon 6 (Table 3.5). The PCR mixture was composed of 3µl of bisulfite treated DNA, 12.5µl YB-Taq 2x Buffer, 2µl of each forward (5'GATTGTAATAATTAGATGTATAAATAGAAA3') and reverse (5'CTTTACAAAATATAAACAAAAATCC3') primer (10pmol/ml), 3.5µl of dH_20 and 1.5µl of 10 mM $MgCl_2$. PCR profile used was as follows; at 94°C for 2 minutes followed by 40 cycles of 30s at 94°C, 1 minute at 53.7°C, 30s at 72°C and a 10 minutes final extension at 72°C. Due to the AT rich nature of bisulfite treated DNA, the number of PCR cycles used were increased up to 40 for a successful amplification of the product.

Amplified products were run on a 3% agarose gel. DNA that run as a single band with the correct product size were excised, excluding other secondary products and purified with Promega Wizard SV Gel and PCR Clean-Up System. Concentrations of purified DNA were measured with the NanoDrop and used for cloning in the subsequent steps. Direct sequencing of the purified PCR product may fail to read the entire target region. Thus DNA samples were cloned into a vector to obtain sample sequences with the full targeted gene region. This allows comparison of all 9 CpG sites among sample sequences by an alignment of sequences. In addition with cloning a proportionate number of clones with methylated cytosines can be obtained to compare among Decitabine treated and control samples.

Preparation of culture plates: LB medium was prepared with 10g BactoR-tryptone, 5g BactoR-yeast extract, 5g NaCl per liter of sterile distilled water. LB plates were made with 15g of Bioagar (BioGene) to each liter of LB medium. PH

was adjusted to 7 before and after autoclaving. Once the LB-agar medium was cooled to 50°C, ampicillin was added to a final concentration of 100g/ml and the medium was poured into petridishes under sterile conditions. The agar was left to harden for 30 minutes and the plates were stored at 4°C. 20µl of 50mg/ml X-Gal (dissolved in DMSO) and 100µl of 100mM IPTG were then spread over the surface of each plate and incubated for 30 minutes at 37°C to enhance antibiotic absorption.

Cloning and pGEM^{R} -T Easy Vector System: Cloning of samples were carried out using the pGEM^{R}-T Easy Vector Systems (Promega). Amount of DNA required for cloning was calculated using the following equation.

$$\frac{50\text{ng vector} \times 0.143\text{kb insert}}{3.0\text{kb vector}} \times \frac{3}{1} = \text{ng of insert}$$

Ligation reaction mixture was prepared according to Table 3.1. It was then incubated at room temperature for 1 hour and at 4°C overnight to obtain a maximum number of transfomants.

Then 2µl of the ligation reaction mixture was transferred into a sterile microcentrifuge tube placed on ice. 50µl of JM 109 High efficiency competent cells thawed on ice for 5 minutes were added to the ligation reaction mixture and incubated for 20 minutes. Cells were subjected to heat shock for 50 seconds in a thermal cycler at 42°C and placed on ice for 2 minutes. Then 950µl of LB was added to the ligation reaction transformation and incubated at 37°C in a shaking incubator (150 rpm/ a *g*-force of 20) for 1.5 hours. Finally, 100µl of this reaction was added to LB/ampicillin/x-Gal/IPTG plate and incubated overnight at 37°C. A blank plasmid (an empty pGEM^{R} -T Easy vector without any insert) was used as the background control. The vector ligated to the insert that comes with the kit was used as the positive control.

Checking the insert: According to pGEM^{R}-T Easy Vector System (Promega), a successful ligation of an insert DNA into a vector produces white colonies, while those with an empty (uncut) vector produce blue colonies.

A well isolated white colony from each sample plate was selected. Half of a colony was taken into a sterile toothpick and streak on a culture plate with LB/ampicillin/x-Gal/ IPTG (reference plate). The other half of the colony was extracted into a 50µl

Table 3.1: Components of ligation reaction mixture

Reagent	Standard reaction (µl)	Positive control (µl)	Background control (µl)
2X Rapid Ligation Buffer	5	5	5
pGEM^{R} -T Easy Vector (50ng)	1	1	1
Sample DNA	Xµl	-	-
Control insert DNA	-	2 (4ng/µl)	-
T4 DNA Ligase(3 units/µl)	1	1	1
Deionised water to a final volume of	10µl	10µl	10µl

of distilled water in a PCR tube and used to check the presence of the insert DNA in the vector by a colony PCR as follows. First the cells in the extraction were lysed in thermocycler at 100°C for 5 minutes and centrifuged at 16,000g for 5 minutes. Five microlitres of the supernatant was combined with 5µl YB-Taq 2x Buffer, 1µl (10pmol/ml) of each forward and reverse primer of M-13 (5' CGCCAGGGTTTTCCCAGTCACGAC 3' and 5' TCACACAGGAAACAGCTATGAC 3'), 0.5µl of 10Mm $MgCl_2$ and 1µl of dH_20. PCR profile used was included an initial denaturation at 94°C for 2 minutes, 35 cycles of 30s at 94°C, 1 minute at 55°C, 1 minute at 72°C and a final extension of 5 minutes at 72°C. Amplification products were analysed on a 1% agarose gel containing ethidium bromide.

Colonies with successful DNA inserts (3161 bp product size) were removed from the reference plate and cultured in falcon tubes with LB broth (5ml) and ampicillin (1.5ml) at 37°C overnight.

Isolating plasmid DNA and sequencing: Plasmid extractions from these bacterial cultures were performed with the GenElute^{TM} Plasmid Miniprep Kit (Sigma-Aldrich).

After confirming the presence of a successful insert in these plasmid samples, on a 1% agarose gel (Figure 3.14), they were sent for sequencing with the M-13 Forward primer. Three independent clones from each purified DNA sample were sequenced. Sequence alignments between clones and the insert DNA were performed with Geneious 7.0.3 version and the total number of methylated CpG sites present in Decitabine treated and control clones were counted (Appendix table A.7).

(c) Amplification of intermethylated sites (AIMS) AIMS is a genome-wide DNA methylation detection method. It relies on differential digestion of genomic DNA with a pair of methylation sensitive (SmaI) and insensitive (XmaI) isoschizomers, which produce DNA fragments with a methylated cytosine at each end (Frigola et al., 2002). The basic principle of the AIMS technique is given in Figure 3.7.

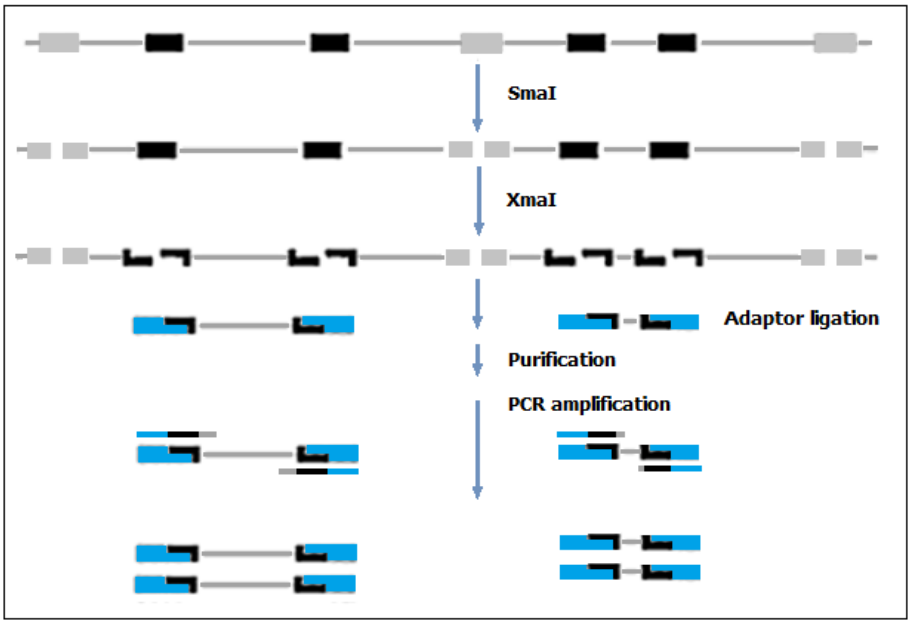

Figure 3.7: Principle of AIMS. Genomic DNA is represented by the first solid line with 7 CCCGGG sites, 3 unmethylated (grey boxes) and 4 methylated (black boxes). Both SmaI and XmaI recognize the CCCGGG site. SmaI cut the non-methylated CCCGGG sites, leaving blunt ends (CCC/GGG). The methylated sites remain unchanged (line 2). Then XmaI cuts only at methylated sites, leaving sticky ends, C/CCGGG overhang (line 3). Adaptors are ligated only to these sticky ends, which were derived from methylated cytosines (line 4). DNA fragments bordered by two ligated adaptors are amplified using adaptor specific primers, that bind to the adaptor and restriction site sequences followed by 1-4 arbitrarily chosen nucleotides. Thereby a methylation profile of loci in the whole genome can be visualized on a gel (Frigola et al., 2002).

Genomic DNA was extracted from 40 adult (preliminary experiment 2) and 40 callow bees (callow worker experiment) in section 3.2.4.2), using the phenol:chloroform method (see section 2.2.3). Half of an each bee was used to extract DNA. Concentrations of samples were measured using the NanoDrop 1000 Spectrophotometer and stored at -80°C.

Adaptors were prepared by incubating 25µl of the oligonucleotide Blue (100µM)(5'-ATTCGCAAAGCTCTGA-3') and 25µl of the oligonucleotide MCF (100µM) (5'-CCGGTCAGAGCTTTGCGAAT-3') at 65°C for 2 minutes followed by cooling to the room temperature for 1 hour in a thermocycler (Frigola et al., 2002).

DNA was digested first with the methylation sensitive restriction enzyme SmaI (1g of DNA, 0.1µl SmaI (20 000 units/ml), 1.5µl of NEB4 10x Buffer, and 3.4µl ddH_20) followed by incubating for 1 hour at 25°C in a thermocycler. This product was then digested with XmaI, methylation insensitive restriction enzyme, (0.5µl, 10 000 units/ml) for another hour at 37°C with 1.0µl of NEB4 10x Buffer, 0.5µl of bovine serum albumin (BSA) and 8µl of ddH_20 (Jorda et al., 2009).

Twenty five microlitres of this digested product was ligated to 20µl of adaptor (2nmol) with 8µl of 10x T4 ligase buffer and 2µl of T4 ligase (400 000 units/ml) and incubated at room temperature for 10 minutes. The enzymes were inactivated by incubating the samples for 10 minutes at 65°C (Jorda et al., 2009). The amplification of sequences with adapter-ligated products was conducted using the primer sets A (A1, A2), B (B1, A2) and C (C1, C2). (**A1**, Blue-CCGGGCTA; **A2**, Blue-CCGGGTGG; **B1**, Blue-CCGGGCTG; **C1**, Blue-CCGGGCGCG; **C2**, Blue-CCGGGCAAC).

Reaction volumes of 25µl were composed of 12.5µl of YB-Taq 2x reaction buffer, 1µl of each primer (10 pmol/ml), 3µl of DNA, 0.5µl of 10mM $MgCl_2$ and 7µl of dH_20. PCR with Primer set A and B were composed of of 30 two-step cycles; 15s at 94°C and 1.5 minutes at 74°C. PCR program for Primer set C was consisted of 30 three-step cycles ;15s at 94°C, 45s at 68°C and 1 minute at 74°C. All PCR cycles were preceded with a denaturing step of 95°C for 1 minute and ended with an extension of 72°C for 5 minutes. PCR products were checked on a 3% agarose gel according to section 2.4. Once the success of PCR amplification was verified, 7µl of PCR product was mixed with Elchrom loading buffer (2µl) and run on 9% poly(NAT) gels (Elchrom) using the Origins electrophoresis system at 120V for 81 minutes at 55°C. Staining and visualization of gels under UV light was carried out according to section 2.6.5.

According to Frigola et al. 2002, bands present in samples of one treatment group but absent in that of another treatment group were considered as hypermethylation whilst those with less intensity were concluded as hypomethylation. Bands were

scored as either present or absent using Gelanalyzer 2010. The resulting matrix was analysed using the R package MSAP which uses a combination of analysis of molecular variance (AMOVA) and Principle coordinate analysis (PCoA)(Pérez-Figueroa, 2013).

3.3 Results

3.3.1 DNA methylation differences between different worker reproductive castes

A total of 245 unique bands (loci) were present. One hundred and thirty six of them were methylation sensitive, that is they showed differences between the digests of HpaII and MspI (i.e. HPA+/MSP- and HPA-/MSP+). Thirty eight of the methylation sensitive loci were polymorphic, that is, they showed different banding patterns between individuals (Appendix figure A.1b). There was a significant difference between the methylation status of reproducing workers (RW) versus non-reproducing box workers (BW) (ϕ_{ST} = 0.3641, p = 0.0259). There were no significant differences between queenless reproducing workers (RW) and queen-right non-reproducing workers (CW) (ϕ_{ST} = 0.3572, p = 0.0963) nor between BW and CW workers (ϕ_{ST} = 0.06859, p = 0.2012).

Table 3.2 details the methylation levels of the three groups. The PCoA based on pairwise difference showed three groupings corresponding to their reproductive state (Figure 3.8). The first two axes explain a total of 51.7% of the variation.

3.3.1.1 Sequencing of selected PolyNAT gel bands

The sequence results from PolyNat gel bands were analysed by NCBI BLASTx (`http://blast.ncbi.nlm.nih.gov/Blast.cgi`) and the following matches of the *Bombus terrestris* genome were found (Table 3.3).

Table 3.2: Proportion of each banding type found in different reproductive castes.

Banding pattern	Methylation status	BW	CW	RW
HPA+/MSP+	Unmethylated	0.35478	0.40931	0.24755
HPA+/MSP-	Hemimethylation	0.05515	0.08088	0.07108
HPA-/MSP+	Methylation at internal cytosine	0.09191	0.10784	0.15686
HPA-/MSP-	Full methylation or absence of target	0.49816	0.40196	0.52451

Note: A reproductive worker (RW) a non-reproductive box worker (BW) and a non-reproductive queen-right colony worker (CW). + and - = Presence of a band and absence of a band in digestion with the restriction enzyme. For example, HPA+/MSP- denotes the presence of a bands in digestion with HpaII but absent when digested with MspI (See Appendix figure A.1b). HPA-/MSP- was counted as uninformative in MSAP. This is the more conservative approach.

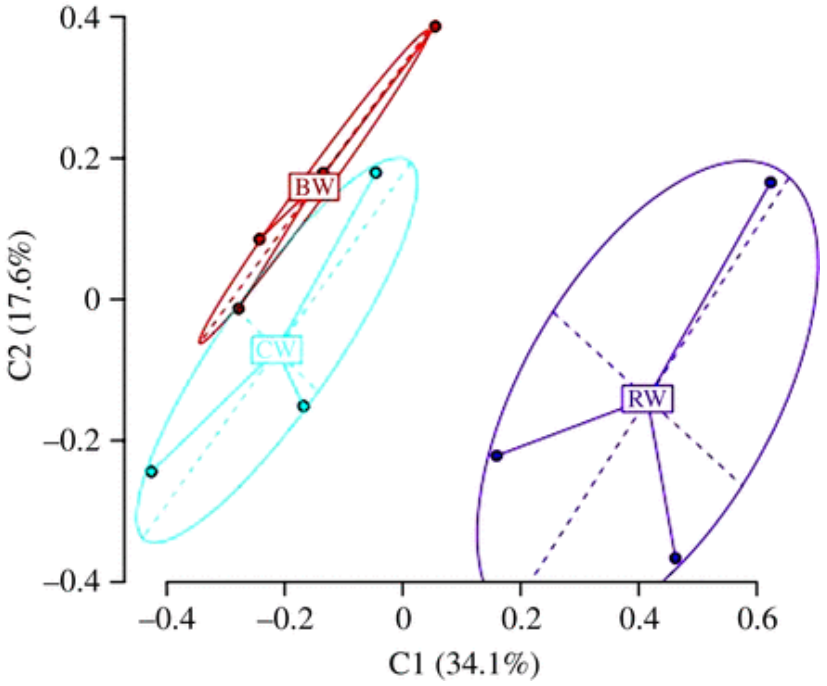

Figure 3.8: Principal coordinate analysis based on methylation status of loci for different reproductive castes. RW = reproductive workers. BW = non-reproductive queenless workers. CW = non-reproductive queen-right workers.

Table 3.3: Blast results of methylated DNA fragments

Primer combination and name of the AFLP fragment used	**Percentage match to *B.terrestris* genome**	**GenBank ID.**	**Sequence homology**	**Sequence (5'-3')**
AA-AAT (Q3)	95%	LOC100644761	hypothetical protein	CTTGTTTTCAATTTTNGCTTCC
AA-AAT (Q4)	91%	LOC100650157	"tumour susceptibility gene 101 protein-like"	ACAATAGAAGGAAGATTAAACAT
AA-ATC (P11)	91%	LOC100642674	LOW QUALITY PROTEIN: "dystrophin isoforms A/C/F/G/H-like"	GTGATGNNCAGTTCATGAACGGGATACTGG CTATGAATGGTAATCCAACCAGTCAATTAAA GTCGTTGATAATACAAATCCGTCCAGTTGA ATT
AG-ACT (S11)	92%	LOC100648711	"paraplegin-like"	GATAATCCNGCTGGCCTATTGATATGCTGAT GGTGAGACGTGTAGCGTGNNANTGCCCCCC ANTACGTATTGTCCGCCTGAATT
AT-ATC (A11)	92%	LOC100646148	"glutamate receptor 1-like"	TTGTCCACGCGCCCTGNATCAATTGGATTN NACAAGTTCTTGGCTCGGCTGCTGATCGCC GTTATTCCCATCCCGTTTTGAATATAACACG GCCTACTTCGCGATTCATAAGGCAAACTA CGCGTATGAATT
AC-AAT (H6)	96%	LOC100648549	"cytochrome c-like isoform 1"	GAATATGGGATGAAACAGATATTT
AC-AAT (H8)	90%	LOC100651733	"tubby-related protein 4-like"	TCAAGGATTTTTCCCGAAAGAAGAACTCGA

Chromatograms of these sequences were associated with two problems; presence of multiple peaks (3-4) at the same base position and the presence of unidentified bases, denoted by 'N'. Therefore sequencing of the rest of the gel bands didn't proceed.

3.3.2 Tissue specific methylation differences between queenless reproductive workers

One hundred and forty six unique bands (loci) were scored. Out of these 102 loci were methylation-susceptible which showed different digestion patterns between the two restriction enzymes, HpaII and MspI. Seventy eight loci of these were recorded as polymorphic, showing different banding patterns between individuals.

There was no significant difference of methylation between head and ovary tissues in reproductive workers at the pre competition phase (PCP) (boxes 1, 2, 3 Appendix table A.3) (ϕ_{ST} = 0.4507 , p = 0.0988) nor head-ovary tissue specific methylation difference in reproductive workers at the competition phase (CP) (box 4 Appendix table A.3) (ϕ_{ST} = 0.164, p = 0.1158).

There was a significant difference of methylation levels in heads of the queenless reproductive workers between the PCP and CP of the colony life cycle (ϕ_{ST} = 0.4009, p = 0.0294).

There was no significant difference of methylation in ovaries of reproductive workers between the PCP and CP of the colony (ϕ_{ST} = 0.04148, p = 0.3415).

Methylation levels of reproductive worker ovaries and heads during PCP and CP of the colony are summarized in Table 3.4.

The PCoA based on pairwise difference showed four groupings corresponding to different tissues and phases in their life cycle (Figure 3.9). The two axes explain a total of 43.7% of the variation.

Table 3.4: Proportion of each banding type found in head and ovary tissues of bees belong to different phases.

Banding pattern	Methylation status	Head PCP	Ovary PCP	Head CP	Ovary CP
HPA+/MSP+	Unmethylated	0.50000	0.30065	0.2132	0.3317
HPA+/MSP-	Hemimethylation	0.06209	0.15686	0.1593	0.1597
HPA-/MSP+	Methylation at internal cytosine	0.12092	0.08824	0.1814	0.1007
HPA-/MSP-	Full methylation or absence of target	0.31699	0.45425	0.4461	0.4079

Note: PCP= Pre-competition phase, CP= Competition phase, Head PCP= Head tissue of workers at PCP, Ovary PCP= Ovary tissue of workers at PCP. Head CP= Head tissue of workers at CP, Ovary CP= Ovary tissue of workers at CP. See Appendix figure A.1b for interpretation of banding patterns and their respective methylation statuses. HPA-/MSP- was counted as uninformative in MSAP. This is the more conservative approach.

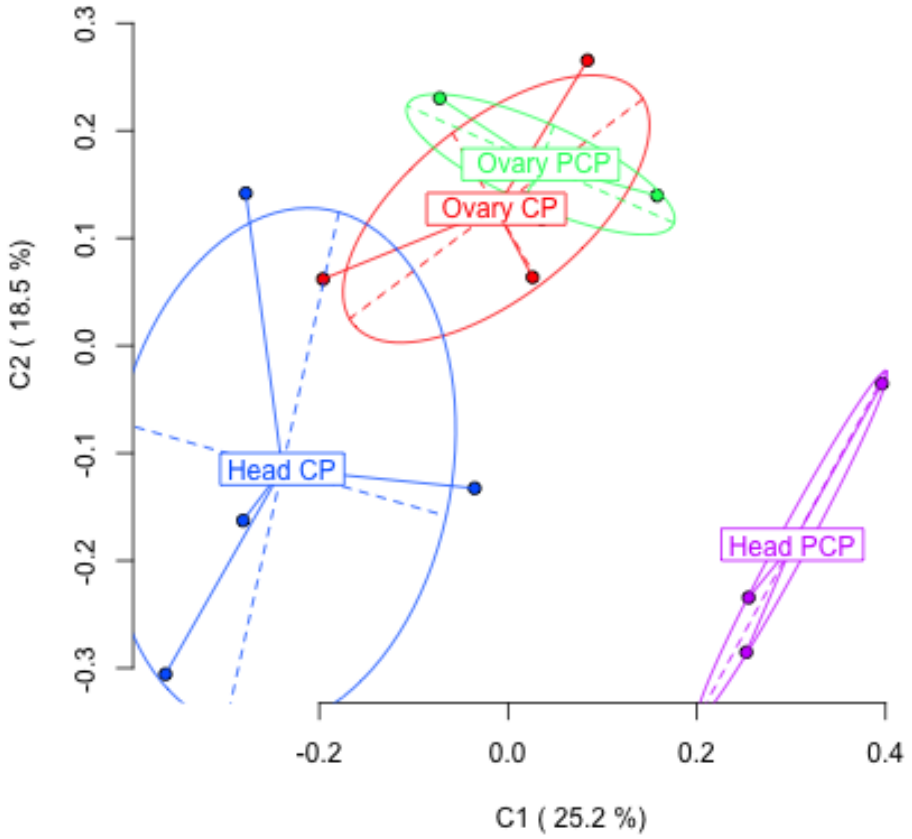

Figure 3.9: Principal coordinate analysis based on methylation status of loci for ovary-head tissues during PCP (Pre-competition phase before the colony queen died) and CP (During the competition phase after the colony queen died).

Workers which were collected during the CP of the colony were all reproductives (box 4 bees in Appendix table A.3). Conversely their similar aged colony workers, which were reared in the natal colony, all lacked visible ovaries.

3.3.3 Methylation differences between different developmental stages of *B. terrestris* life cycle

A total of 74 methylation susceptible loci belong to different developmental stages (L2, L3, L4 larval stages, pupae and and callow workers) were found during this analysis. 40% (30 out of 74) of the total loci examined showed a polymorphic banding pattern. That is those loci are unique to different individuals and developmental stages. However, pair-wise AMOVA results between different developmental stages was not significant (ϕ_{ST} = -0.1174 , p= 0.8103).

3.3.4 Effect of DNA demethylation on *Bombus terrestris* worker reproduction

3.3.4.1 Preliminary experiment 1

Eighty percent of the bees in both the test and control groups showed a development of ovaries (Appendix table A.4) suggesting that most of the workers within these mini-colonies were potentially reproductive. Also, a high level of aggression among bees was observed frequently indicating competition for dominance by bees. Instead of a single dominant worker, several workers in the mini-colony were sharing the reproduction.

3.3.4.2 Preliminary experiment 2

Decitabine had different effects on ovary development depending on whether the bees were callows or adults when placed in the box (lifestage : treatment: $F_{1,153}$ = 7.485, p = 0.006957). Decitabine had no effect on level of ovary development in adult bees ($F_{1,75}$ = 1.547, p = 0.217) (Appendix table A.5).

Unlike callow workers, adult bees were reared for 3 weeks, until nest construction or egg laying was observed indicating that these workers had functional ovaries.

3.3.4.3 Callow worker experiment

Decitabine had a significant effect on level of ovary development in callow bees ($F_{1,75}$ = 7.211, p = 0.00891; Figure 3.10) (Appendix table A.6). For callow workers, each control box had on average one bee out of five with developed ovaries (mean oocyte length more than or equal to 1 mm) (Duchateau and Velthuis, 1988) compared with three bees out of five with developed ovaries in Decitabine groups.

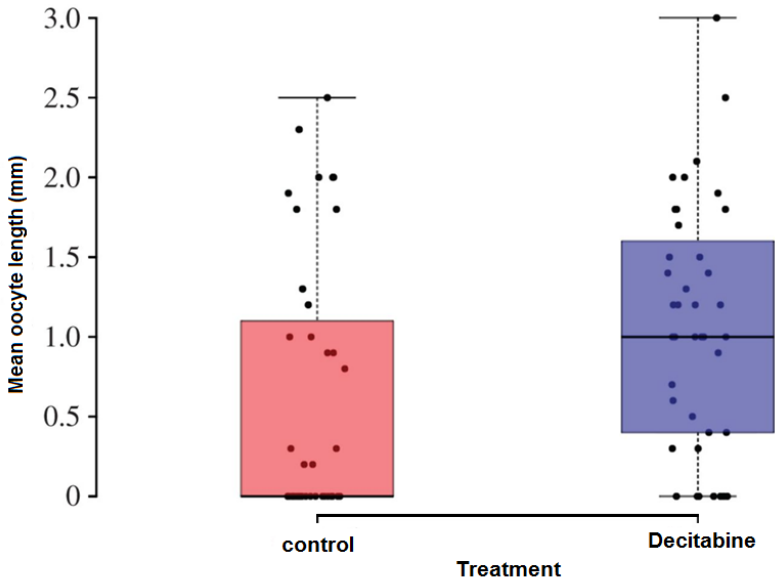

Figure 3.10: Ovary development between treatments for callow bees. Mean oocyte length is the average length of the two largest oocytes in each ovary of a bee. Dots are the individual data points.

3.3.4.4 Additional measurements

Dissections of bees reared in their original colonies showed a complete absence of ovary development. In the callow worker experiment, Decitabine treated group produced 40% more brood by weight compared to the control (Figure 3.4, Appendix table A.6).

3.3.4.5 Behavioural data

Aggression is a measure of the reproductive conflict among workers (Amsalem and Hefetz, 2011). The dominant hierarchy of bees is usually established through their overt agonistic behaviours. During the entire period of the experiment aggression of drug treated bees remain at a higher level compared to the non-treated bees (Figure 3.11). Aggression was analysed using a 2 way repeated measures ANOVA, where time was the repeated measure. Decitabine had a significant effect on level of aggression ($F_{1,143} = 32.17$, $p < 0.00001$). Time also had a significant effect ($F_{17,143} = 8.05$, $p < 0.00001$). There was no interaction effect between time and Decitabine ($F_{17,143} = 0.59$, $p = 0.8957$). i.e the time and the treatment are two independent variables from each other.

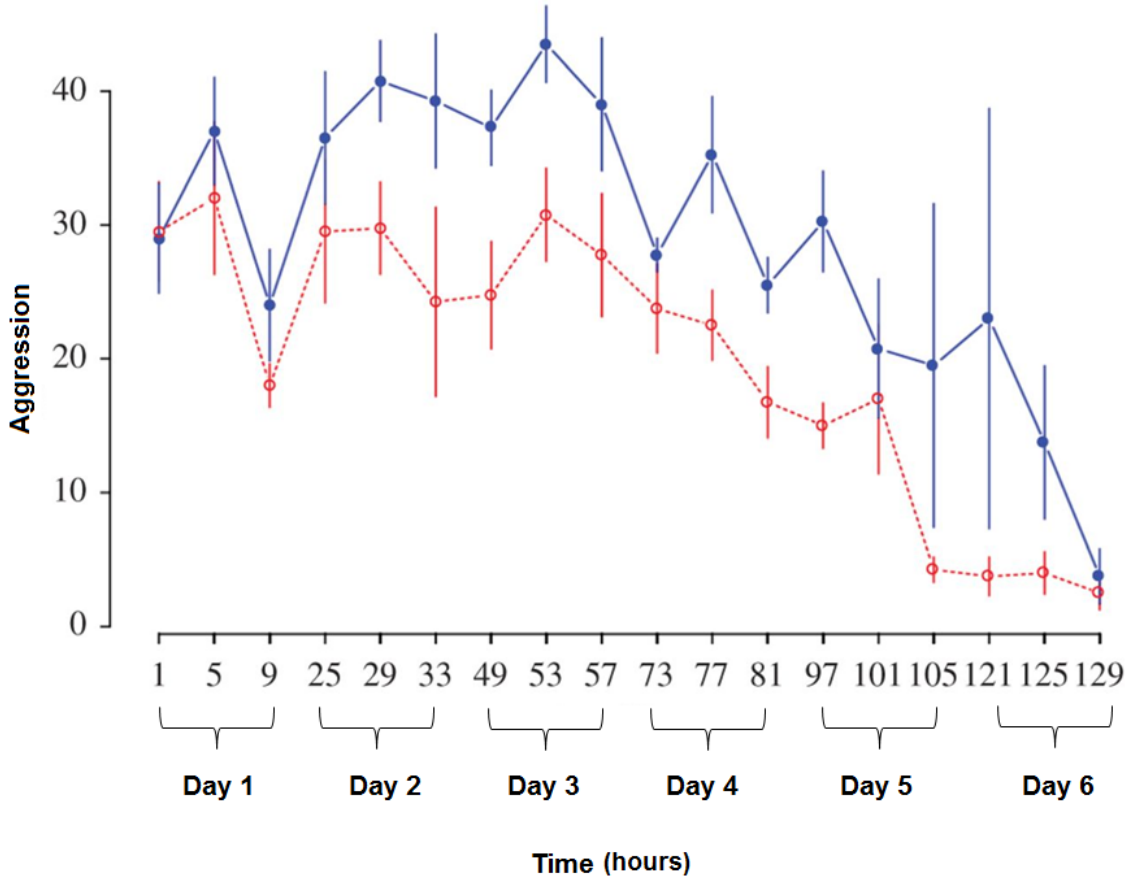

Figure 3.11: Changes in aggression over time between Decitabine and control callow bees. The filled circles (solid line) represent means of boxes treated with Decitabine. Open circles (dotted line) are the control boxes. Bars represent standard errors.

During day 1, relatively moderate amounts of aggressive behaviours among bees were observed. For the next 2 days, the overall aggression was held at a peak. Then from the 4th day onwards aggression level started to decline. Towards the 5th and 6th day, it decreased further and the workers were found working on separate locations of their nests and honey pots leading to a very little amount of interactions among them (Figure 3.11).

3.3.5 Comparison of DNA methylation differences among Decitabine treated and non-treated groups

3.3.5.1 Methylation sensitive restriction enzymes (MSRE)

In set 1, an absence of a band in the enzyme digested product in Decitabine treated sample (34D) was observed while the corresponding treatment in the control sample (31C) showed a complete band after PCR (Callow workers - Figure 3.12). A similar type of result is also shown in set 2 and set 3 samples where a very faint band was present for the enzyme treated Decitabine samples (31D and 21D). A converse result was observed for set 4 and 5 where both the treated and the control samples showed a presence of a band irrespective of the enzyme treatment.

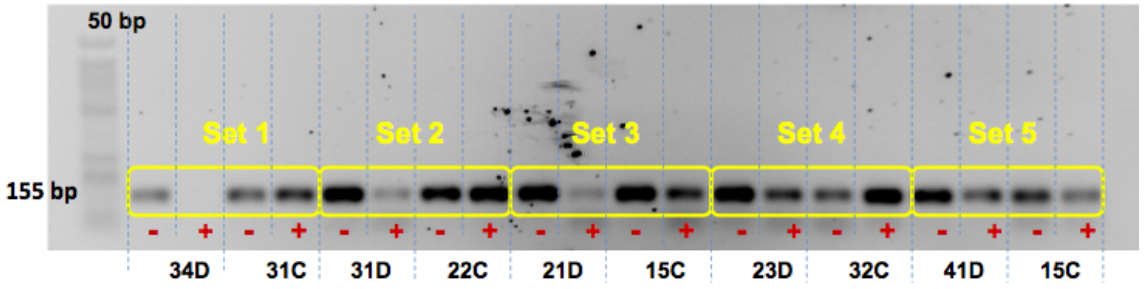

Figure 3.12: MSRE results for Decitabine treated and control callow workers. D = Decitabine treated, C = Non-treated control, + = Presence of HpaII enzyme in the reaction mixture, - = Absence of HpaII enzyme in the reaction mixture.

In adult workers, there is no difference between the enzyme digested samples of both the Decitabine treated and control samples. This was consistent among all 5 testing groups/sets (Figure 3.13).

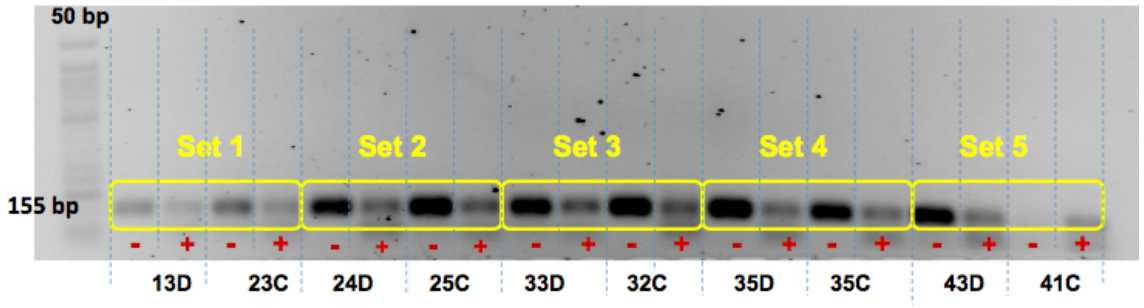

Figure 3.13: MSRE results for Decitabine treated and control adult workers. D = Decitabine treated, C = Non-treated control, + = Presence of HpaII enzyme in the reaction mixture, - = Absence of HpaII enzyme in the reaction mixture.

A clear conclusion on methylation status of the test and control groups was unable to made using MSRE. Thus bisulfite treatment was used as a quantification method of methylation levels in these groups.

3.3.5.2 Bisulfite treatment

According to the genomic DNA and mRNA alignment results, exon 8 showed a complete absence of CpG dinucleotides. Also, all cytosine residues present in the intronic regions in between exon 5 and 9 had no CpGs.

According to the sequencing results, 5 out of the 9 CpG sites present in exon 6 of *dynactin subunit-4 like* were methylated (Table 3.5). Thus exon 6 was selected for comparison among Decitabine treated and non-treated groups.

Both the gel quantification (Figure 3.14) and sequencing results (Appendix table A.7) of plasmid extractions suggested that the cloning of sample DNA in to *E. coli* vectors was successful.

Exon 6 was composed of 9 CpG sites. The corresponding primers of exon 6 (Table 3.5) were able to cover all 9 CpG sites present in exon 6. Thus the DNA insert in each clone sequence was composed of 9 CpG sites (Appendix table A.7). A summary of sequencing results for all clones for the two treatments is given in Table 3.6.

According to Table 3.6, the total number of methylated cytosines present in Decitabine treated samples (68) is slightly less than that of the control samples

Table 3.5: Total methylation levels in *dynactin subunit-4 like* exons.

Exon	Number of CpG sites present in the exon	Primers sequences	Amplicon length (in bp)	T_A	T_E	Number of methylated cytosines present
5	3	F: GTATAGGATATTTTATAAAGTTTGTTATTG	249	54	72	0
		R:TTTCTATTTATACATCTAATTATTACAATC				
6	9	F:CGTTGGAACAGCGATTGCAACAACCAGATGTA	153	54	72	5
		R:CTTTACAAAATATAAACAAAAATCC				
7	5	F:TTGTAAAGTTTAAGATTTAATTTGTTGTAT	160	53	72	2
		R:GTTTTGGTAAATTAAGTGAATTGTTT				
9	7	F:GTTTGTAGTAGAGGAAATAAAATTTATTAA	161	55	72	1
		R:AAAATTATACATACTTAAAATCATCTTAAA				

Note: T_A (^oC)= Annealing temperature, T_E (^oC)= Extension temperature.

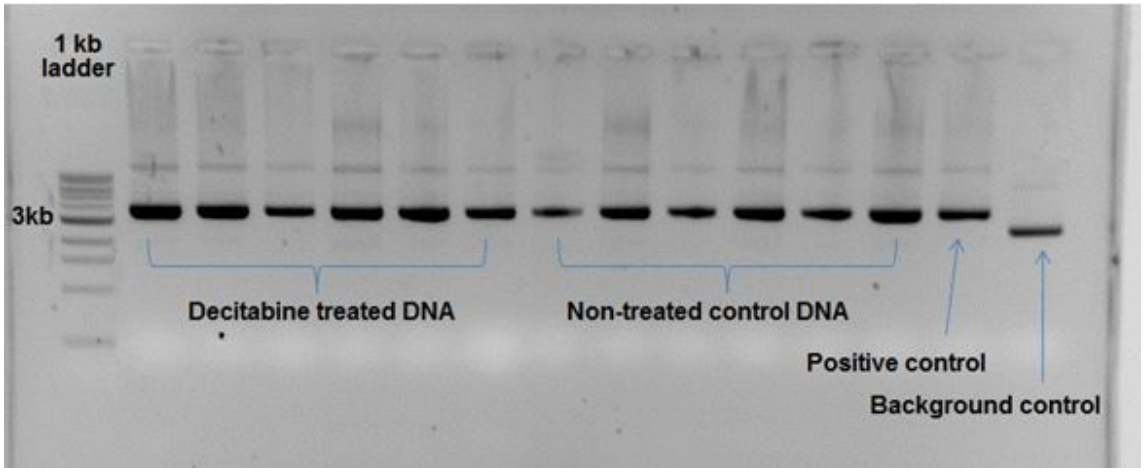

Figure 3.14: Plasmids/clones with Decitabine treated and control sample inserts.

(74). Statistically this is not a significant difference between the Decitabine treated and control samples ($\chi^2 = 6.583$, d.f. $= 8$, p >0.05).

As a clear reduction of methylation in Decitabine treated group was unobservable using the bisulfite treatment, I used amplification of intermethylated sites (AIMS) as an alternative method of comparing the methylation statuses among Decitabine treated and control groups.

Table 3.6: Total number of methylated cytosines at each CpG site in exon 6 of *dynactin subunit 4-like*

Group	CpG site 1	CpG2	CpG3	CpG4	CpG5	CpG6	CpG7	CpG8	CpG9	Total
Decitabine	17	17	15	0	8	0	2	0	9	68
Control	15	18	15	0	11	0	0	4	11	74

3.3.5.3 Amplification of intermethylated sites (AIMS) analysis

Based on AIMS data, Decitabine has a significant effect on methylation patterns in callow bees ($\phi_{ST} = 0.2227$, $p < 0.0001$). There were 62 methylation sensitive loci in total of which 54 were polymorphic. Forty-three loci had the same modal level of methylation in both the control and Decitabine-treated callow workers. Nine loci showed hypermethylation and 10 hypomethylation in Decitabine-treated callows compared with controls. The PCoA based on pairwise difference showed two groupings corresponding to whether the bees were treated with Decitabine or not (Figure 3.15).

AIMS analysis of adult workers didn't show any significant effect of Decitabine compared to its control group ($\phi_{ST} = 0.0149$, $p > 0.0001$).

3.4 Discussion

I found clear methylation differences between the genomes of queenless reproductive workers and queenless non-reproductive workers (Figure 3.8). This suggests the importance of methylation in reproductive caste formation in *B.terrestris* workers.

During this analysis, a consistent banding pattern among different worker types was observed at most loci. However individual specific banding patterns were also observed at certain loci (See Appendix figure A.1b). The amount and pattern of methylation vary with the type of tissue, age and environmental conditions (Cedar

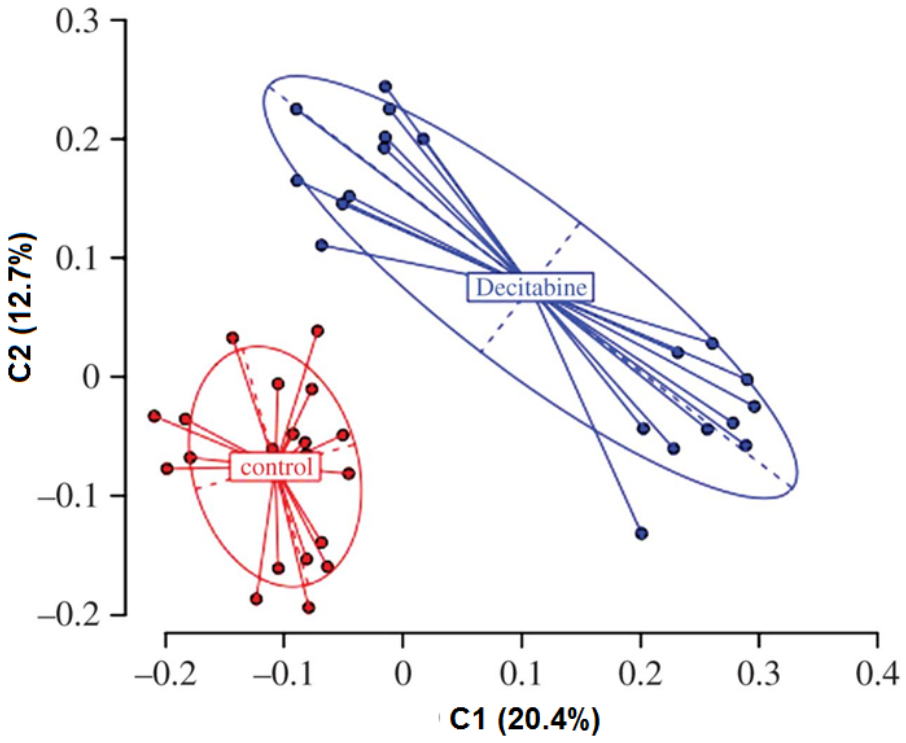

Figure 3.15: Principal coordinate analysis of AIMS data for callow bees.

and Bergman, 2012, Kucharski et al., 2008). During the experiment all these factors remained constant among all three worker types. Thus these unique banding patterns are probably due to their individual genetic polymorphisms.

MS-AFLP has been used in many epigenetic studies as a genome-wide methylation screening technique (Schrey et al., 2013). However, compared to more advanced techniques such as next generation sequencing it has limitations. MS-AFLP only detects methylation differences present at the restriction site (CCGG) of the frequent cutter endonucleases, HpaII and MspI. Cytosine methylation that is present in non-CpG sites such as CpA and CpT (Bonasio et al., 2012) is not detected by this method. In addition, during this study 109 loci out of the total 245 were recorded as uninformative. That is there were no bands for either HpaII or MspI digestions (HPA-/MSP-). See table 3.2. This could be due either to the target site being absent or it being hypermethylated (full methylation in both cytosines or in external cytosines). An absence of the target site could occur due to genetic variations such as point mutation to the restriction site or changes in the adjacent restriction sites. However, these types of variations are not very frequent in a genome and according to

my results 44% of the total number of loci were considered as uninformative (Table 3.2). Therefore most of them must have arisen due to hypermethylation. In most studies HPA-/MSP-banding pattern is considered as a hypermethylation (Gupta et al., 2012, Wang et al., 2011). In both these ways, the overall methylation level of the genome could have been underestimated. However, the significant methylation difference that I found between queenless reproductive workers and queenless non-reproductive workers excludes all these uninformative loci.

MS-AFLP provides a quantitative analysis of methylation statuses in a genome. However qualitative data such as the precise location of methylated loci in the genome cannot be obtained by basic MS-AFLP alone. For instance, in a gene whether methylation is associated with introns, exons or in UTRs etc. Unfortunately my attempt to sequence DNA fragments with unique methylation patterns was unsuccessful (Table 3.3). The sequencing peaks of the resulted chromatograms were not as clean as expected in a perfect chromatogram (Sanger Sequencing Troubleshooting Guide, 2011 http://www.agrf.org.au/assets/files/PDF%20documents/Troubleshooting%20-%20Sanger%20Sequencing.pdf) and several overlapping peaks were associated with many parts of the chromatogram, but with the correct base spacing. This could be due to contamination with adjacent DNA, when removing a DNA band from the PolyNat gel. Due to the thermo stability of PolyNat gels, I could not extract a sufficient amount of DNA from a gel band. Thus the extracted DNA had to be amplified prior to sequence. During PCR these contaminated sequences may also have amplified and appeared as multiple peaks in the chromatogram.

I then compared methylation statuses of tissues in reproductive workers belonging to different social phases. No tissue-specific methylation difference between ovary and head tissues of bees was discovered. Instead a significant methylation change in worker head tissues was observed during the colony transition from PCP to CP (Figure 3.9). This could be due to the inconsistency of the two gel runs (see section 3.2.2) or may be due to an actual effect of methylation. If it's the latter then, lack of ovary development in callow workers during PCP and activation of ovary development during CP, may be associated with differences in the methylation levels of bee brains. Previous research has also shown that the switch between different worker roles in honeybees is associated with methylation changes in the brain (Herb

et al., 2012). In addition, only about 70000 out of the 60 million cytosines present in the *Apis* genome are methylated and 80% of these methylation susceptible honeybee genes are expressed in the brain (Strachecka et al., 2012). Queen-induced worker sterility in *Bombus terrestris* is achieved by pheromones (Amsalem et al., 2015, Van Oystaeyen et al., 2014) and aggressive behaviour of the queen (Alaux et al., 2007). It is possible that this inhibition of worker reproduction by queen pheromones may be achieved via changes of methylation levels in worker brains and thereby differential expression of brain genes leading them to be sterile or reproductive. In honeybees, certain genes that are responsive to queens' pheromones are differentially methylated (Foret et al., 2012).

Older workers can impose a 'queen-like' effect on callow worker reproduction. All callow workers collected during the CP and reared in a box (box 4) had developed ovaries. But their similar aged callow workers reared in the natal colony (Appendix table A.3) lacked visible ovaries. This indicates that after the colony loses its queen, young workers will have the potential to become reproductive. But, as a result of competition among workers over male production, the older workers who started laying eggs in the colony may inhibit callow workers from initiating their reproduction. A similar result has previously been observed; Even after 5 days of the queen's death, young workers in the colony didn't become reproductive (van Honk et al., 1981). This is similar to the introduction of egg-laying workers in to a queen-right colony during the PCP, in which all of them will revert to sterility (Alaux et al., 2007).

Queenless workers whose genomes had experimentally altered methylation, were more aggressive (Figure 3.11) and more likely to develop ovaries compared to control queenless workers (Figure 3.10). Thus instead of a single dominant worker, several workers in Decitabine treated groups shared the reproduction. Otherwise, most of the time the α-worker shows the most aggressive behaviour and monopolizes the reproduction in the colony while other subordinate workers become her helpers to maintain the nest and brood (Amsalem and Hefetz, 2011).

Decitabine had no effect on bees that were adults at the beginning of the experiment. Only callows (bees less than 1 day old) were affected. Although it is exciting to think that this result is due to adult bees being developmentally fixed and unable to switch roles (Patalano et al., 2012), the adult insect cell division may be just

a technical artifact of the mechanism of Decitabine demethylation. During the s phase of DNA replication, 5-aza-2'-deoxycytidine (Decitabine) is converted into the triphosphate and is incorporated in place of CpG sites in DNA. DNA (cytosine-5)-methyltransferase 1 (Dnmt1) binds covalently on DNA at these modified CpG sites and becomes inactive. Lack of Dnmt1 blocks remethylation of hemi-methylated sites in daughter DNA, during the first round of DNA replication. Thereby Decitabine acts as a direct and irreversible inhibitor of Dnmt1 and cell division in the absence of Dnmt1 results in a progressive demethylation of DNA (Christman, 2002, Egger et al., 2004). Adult insects are considered post-mitotic (Finch, 1990) although see (Ward et al., 2008). If no cell division occurs in adult bees then Decitabine cannot be incorporated and there will be no effect on methylation on adult bees.

The effects of Decitabine are not due to a general toxicity effect. A more general toxic effect would be expected to act on adult bees as well. The provided dosage of Decitabine in the adult experiment showed no effect on bees (Appendix table A.5). This dosage is below the minimum used to test for genotoxic effects in *Drosophila* (Cunha et al., 2002). If it were merely a toxic effect I would expect to see reduced activity and reproduction. This was exactly the result when *B. terrestris* workers were exposed to chlorantraniliprole, a pesticide which is similar to Decitabine in toxicity but has no demethylating effect such as Decitabine (Smagghe et al., 2013). Instead, during my experiment I have observed increased aggression and more bees becoming reproductive when Decitabine is administered.

In adult workers, density had an effect on aggressiveness and thereby their ovary development. Compared to 5 worker groups where a single worker dominates reproduction, in 10 worker groups, on average 8 bees were contributed to reproduction (Appendix table A.4). In small groups such as in 5 bee groups the α-worker is capable of controlling the reproduction of her sub-ordinate workers. However, when the group size is large (e.g. 10 bees) the effectiveness of this control by means of aggressive behaviours and pheromones is less. Thus maintaining a dominant hierarchy among workers is difficult and most workers in the mini-colony tend to activate their reproduction.

The next step of this experiment was to analyse the amount of methylation difference required to bring these effects in reproductive workers. Both qualitative (MSRE)

and quantitative (bisulfite treatment and AIMS) methylation analysis techniques were used for this purpose.

A clear, consistent effect on methylation change in callow workers was not observed using MSRE. Decitabine is considered as a demethylating agent (Christman, 2002). Therefore less methylation is expected to be observed in Decitabine treated bees compared to their control group. Since HpaII is only sensitive to full methylation, the expected MSRE result can only be achieved if Decitabine treated bees show a hemi-methylation or no methylation whereas their control group shows a full methylation. This is represented in Figure 3.12, as an absence of a band (set 1) or as a very faint band (set 2 and 3) in the enzyme digested product of Decitabine treated sample while their corresponding treatment in the control sample shows a complete band after PCR. Compared to set 1, a contrast result was observed in sets 4 and 5. i.e. A band was present for the enzyme digested product in both Decitabine treated and control samples, indicating that there is no methylation difference between the treatment and the control. These two contrasting observations could be due to the heterogeneity of methylation statuses among two bees at the same CpG site (Kucharski et al., 2008). That is the same CpG site could be methylated in one bee while it is unmethylated in the other. If Decitabine completely removes the methylation tags in treated bees while their control group was in a hemi-methylation state, still it proves that Decitabine had a demethylating effect on those bees. However, these two methylation statuses cannot be differentiated by using HpaII, since under both these conditions HpaII cleaves the DNA and it is observed as an absence of a band in the enzyme digested product in both treated and control samples.

MSRE analysis for adult workers showed no difference between the Decitabine treated and control samples (Figure 3.13). Unlike AIMS and bisulfite treatment that could be applied in genome-wide scale, MSRE only detects the methylation status in a selected gene region at a specific restriction site. Therefore to obtain a clear conclusion on Decitabine demethylation using MSRE, it is important to check several genes with different methylation sensitive restriction enzymes.

According to bisulfite sequencing results the methylation reduction caused by Decitabine was not significant. However during the AIMS analysis a clear effect of Decitabine on methylation patterns in callow workers was observed (Figure 3.15).

But it is not an obvious reduction in overall methylation. This is due to the limitation of AIMS as a quantifier of exact methylation levels. As there is a PCR step, even if methylation was reduced, rather than completely removed, in a given locus, this locus would be classed as still methylated.

Previous research states that the genes involved in insect development and reproduction are hypomethylated (Sarda et al., 2012, Strachecka et al., 2012), which is the expected effect of Decitabine. I have observed clear effects of Decitabine on methylation which affects worker reproduction (increase ovary development and aggressive behaviour). However, a clear reduction of global methylation level by Decitabine wasn't observed. Thus the amount of methylation difference required to bring such effects may be less. A recent study has shown that Decitabine is not a global demethylator but rather demethylates specific and reproducible sites in human cancer cell lines (Hagemann et al., 2011).

Methylation changes are involved in queen-worker differentiation in honeybees (Kucharski et al., 2008). The development of a genetically identical embryo into either a queen or a worker has been compared to the transition from a totipotent stem cell to a fully differentiated cell type (Bonasio et al., 2012, Patalano et al., 2012). Recently it has been shown that bumblebee workers can reverse their reproductive status depending on the social context (Yagound et al., 2012). If a reproductive worker is returned to her natal colony she will regress back to sterility. However if she is placed in a foreign nest, she will remain fertile despite the presence of the resident queen. This chapter shows that the reproductive caste formation, an important aspect of eusociality in *B.terrestris* to be under epigenetic control. Therefore my work suggests, worker reproduction is influenced more by plastic epigenetic processes than those of queen-worker differentiation. Processes, it has been suggested, that could be analogous to somatic cell reprogramming and transdifferentiation (Bonasio et al., 2012).

DNA methylation patterns, in other words parental imprints that are established during gametogenesis can be transmissible to next generation. Thus they serve as important mediators of intragenomic conflict. Genomic imprinting in social insects is predicted to be arise as a results of within-genome conflicts (Haig, 2000). In the next two chapters I will discuss two experimental approaches I took to check the existence of monoallelic gene expression in the *Bombus terrestris* genome.

Chapter 4

Allele specific expression in the bumblebee, *Bombus terrestris*: candidate gene approach

4.1 Introduction

Normally an organism behaves as a cohesive entity which works towards a common goal of survival. However in reproduction, the genome of an individual organism can behave as two distinct units, the maternally derived part (matrigenes) and the paternally derived part (patrigenes) (Haig, 2000). Haig's kinship theory predicts that the different selectional pressures, to which the matrigene and patrigene are subjected, causes this intragenomic conflict. That is, a patrigene favours its offspring to take more resources from the mother, so that they are more likely to survive and reproduce, thus maximizing the patrigene's fitness. In contrast the matrigene's fitness depends on the number of offspring that carries it. Thus a matrigene tries to minimize resource provision, so the mother could invest in more offspring. As a result, the maternal and paternal genes show different expressions over reproductive traits of their offspring and each parental part of the genome uses imprinting in different ways to enhance its own inclusive fitness (Burt and Trives, 2006). Therefore reproduction loci are a potential place where imprinting could occur. Imprinted

genes that have been found in mammals are often involved in the regulation of embryonic growth and development (Reik and Walter, 2001).

Eusocial Hymenoptera are potential candidates where genomic imprinting could have evolved (see section 1.3). Also there is evidence to support monoallelic expression and parent-of-origin effects in eusocial Hymenoptera (see section 1.5). Here I used a candidate gene approach to search for parent-of-origin allele specific expression in bumblebee reproduction genes.

Cardoen et al. 2011 reported, 1292 genes that are differentially expressed between reproductive and non-reproductive workers in honeybees. This includes genes that are associated with the storage of maternal mRNA in oocytes, genes related to epigenetic control (e.g. *DNA methyltransferase 3*), genes known to respond to queen and brood pheromones, genes that are linked with steroid biosynthesis and insulin signaling pathways. All above gene categories are linked with reproduction. Thus they can serve as potential target sites where genomic imprinting could occur. The parental allele in which the imprint should be located depends on the role that the gene plays in worker reproduction. For example *Cabut*, one of the candidate genes tested in this chapter is involved in inhibition of worker reproduction via cell death and reduction of ovariole number (Cardoen et al., 2011). Thus under queen-right monandrous conditions, imprinting in this gene is expected to be observed in the paternally derived allele while the maternally derived allele is expressed.

The overall aim of this chapter was to search for parent-of-origin allele specific expression, in *Bombus terrestris* worker reproduction loci, under queen-right conditions. Reproduction loci were selected from 12 candidate genes that are previously implicated in enhancing and inhibiting worker reproduction. My first aim was to find maternal and paternal allelic variations present in workers at each selected gene locus. For this purpose genomic DNA of the queen and 5 workers from each bumblebee colony were compared using PCR coupled SSCP (single stranded conformation polymorphism) analysis. If any variation was observed during SSCP analysis, then the second aim was to sequence those loci to identify the single nucleotide polymorphisms (SNPs) that cause those variations. Compared with the queen's sequence, any additional SNP found in all of her worker sequences came from their father. Thus any heterozygous worker clearly shows that it carries both maternal and paternal alleles (Figure 4.2). Thirdly maternal and paternal allele specific primers were

designed to those heterozygous SNP regions and allele specific PCR was conducted to confirm the specificity of these primers on each parental allele. Then RT-qPCR was conducted using those allele specific primers to measure the amount of expression in each parental allele. This final aim was to compare the observed expression patterns of maternal and paternal alleles with that of the expected patterns for each gene, to see if they behave according to the predictions made by the kinship theory for a monandrous queen-right colony.

The theory predicts that in a monandrous queen-right colony such as in *B. terrestris*, at any locus that promotes/upregulates worker reproduction the maternal copy should be imprinted and the paternal copy should be expressed (see section 1.2, Figure 1.1). Similarly at any locus that enhance worker sterility should be expressed maternally and imprint paternally (Queller, 2003). The hypothesis of this study expects that at functionally different reproductive loci in *B. terrestris* workers, the expression pattern between parental alleles should vary.

The results of this chapter will not identify imprinted genes. If I discovered a gene which expresses either only the matrigene or the patrigene (i.e. monoallelically expressed gene) or expresses biallelically but differentially between parental alleles (allele specific expression), then by conducting a separate experiment on reciprocal crosses it could be confirmed whether this expression pattern is due to genomic imprinting or due to any other genetic phenomenon such as dominant-recessive effects, X chromosome inactivation etc.

4.2 Methods

4.2.1 Single strand confirmation polymorphism (SSCP)

SSCP has been used as a genotyping and mutation scanning method in many biomedical studies (Kozlowski and Krzyzosiak, 2001). It relies on the principle that the electrophoretic mobility of a single-stranded DNA molecule is dependent on its structure (nucleotide sequence) and size. In the absence of the complementary strand, DNA becomes unstable and reanneals to itself to form conformations; hairpins, pseudoknots and triple helices (Nielsen et al., 1995). These conformations

vary according to the primary sequence of the molecule, such that a single nucleotide difference in DNA could dramatically affect the strand's mobility through a gel due to its unique 3D structure (Figure 4.1). Therefore depending on the presence or absence of SNPs in a locus, the maternal and paternal alleles would produce a deviated mobility pattern on a gel (Figure 4.2).

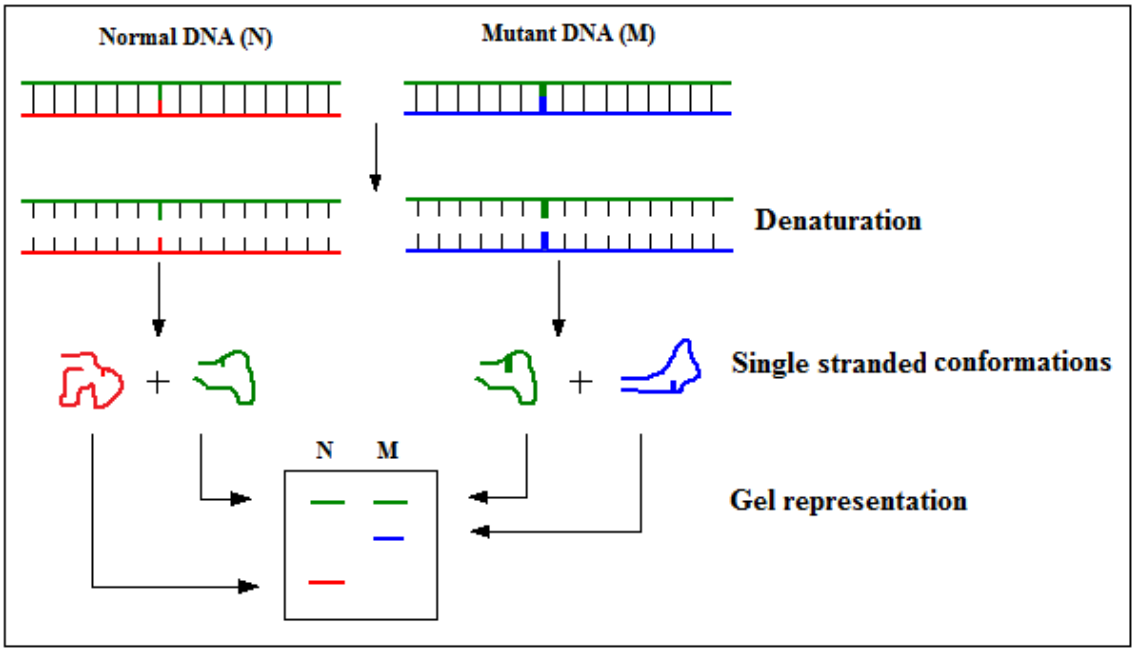

Figure 4.1: A schematic of SSCP mechanism for a locus with a SNP (blue strand in mutant DNA) and without a SNP (red strand in normal DNA).

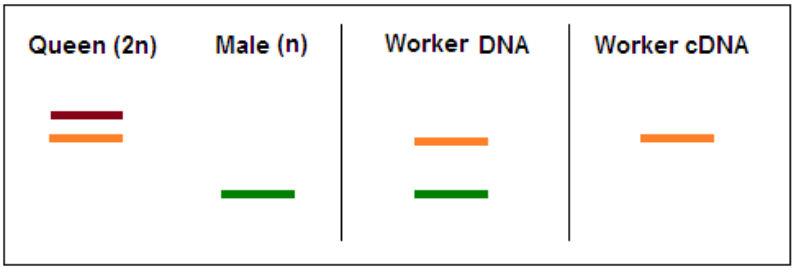

Figure 4.2: Separation of maternal and paternal alleles in a gel. The diploid queen has 2 alleles while the haploid drone has a single allele. Daughter workers receive one allele from each parent. Thus genomic DNA of the represented worker shows a heterozygous condition. If the maternal and paternal alleles are genotypically identical (homozygous workers) they will have identical electrophoretic mobilities, and thus cannot be separated in a gel representation. cDNA synthesised from this worker shows expression of the maternal allele while the paternal allele is not expressed, thus illustrating a monoallelic expression of the maternally derived allele.

4.2.2 Identification of candidate genes and designing primers

Candidate genes which are differentially expressed in the queen, reproducing workers and non-reproducing workers and also which are associated with the worker ovary activation and suppression were selected via an extensive literature search (Table 4.1).

Sequences for all selected candidate genes were obtained from *Apis mellifera* genome data, available in NCBI and Beebase databases.

Apis mellifera data was BLASTed against the *Bombus terrestris* Nucleotide library (NCBI) in order to find the same gene analogue in *Bombus terrestris*. Then, the mRNA and genomic DNA sequences were aligned to find out the exonic regions. Intron regions were excluded because in eukaryotic cells the mature mRNA is already spliced, thus the cDNA produced does not include introns. This means any polymorphism found in introns would be useless for my ultimate aim, identifying allele specific expression. Primers were designed to these exonic regions using Geneious Pro (version 5.5.6) and primer 3 version 0.4.0 (`http://frodo.wi.mit.edu`).

4.2.3 DNA Extraction

The queen and 5 randomly selected workers from each colony were used for SSCP analysis. Four colonies were used to test each candidate gene. Haploid drones are useless for the aim of this research, finding maternal and paternal allelic variations in a gene. Hence sex determination of bees was carried out according to section 3.2.4.2, to confirm that they are all female bees. Bees were flash frozen in liquid nitrogen and then stored at -80°C.

Genomic DNA for SSCP analysis was extracted from a half of an each queen and worker bee using the DNA Micro kit (section 2.2.2). The rest of the bee samples were stored at -80^{0}C for RNA extraction and qPCR analysis. Concentration of total genomic DNA was measured using the NanoDrop 1000 Spectrophotometer.

Table 4.1: Candidate genes selected from the literature search. NCBI gene ID is given within parenthesis.

Gene in *Apis millifera*	**Analogous gene in *Bombus terrestris***	**Related biological function**	**Reference**
Chymotrypsin	Chymotrypsin-1-like (LOC100648122)	- Upregulated in *B.terrestris* non-reproductive workers compared to reproductive workers	Pereboom et al. 2005
Gemini	Upstream binding protein 1-like (LOC100650338)	- Upregulated in reproductive workers	Jarosch et al. 2011
Cabut	Zinc finger protein 691-like (LOC100642767)	- Upregulated in non-reproducing workers - Ovarian regression	Cardoen et al. 2011
Ecdysone 20 monooxygenase	Ecdysone 20- monooxygenase-like (LOC100649449)	- Four fold upregulated in egg laying workers - Ovarian follicle cell development - Activation of ovary development - 20E biosynthesis	Cardoen et al. 2011
Yolkless	Vitellogenin receptor-like (LOC100649042)	- Vitellogenin receptor - Vitellogenin transport - 6.5 fold upregulated in ovaries of honeybee reproductive workers	Cardoen et al. 2011
Epidermal growth factor receptor	Epidermal growth factor receptor like (LOC100645521)	- Up-regulation of EGFR initiates ovary activation and oocyte maturation in queenless non- reproductive workers. - Down regulation of EGFR causes a complete worker sterility	Formesyn et al. 2014
Ribosomal Protein L26	60S Ribosomal Protein L26 like (LOC100648461)	- Differentially expressed in reproductive and non-reproductive workers - Upregulated in ovary suppressed workers	Thompson et al. 2007

Table 4.1: continued

Gene in *Apis millifera*	Analogous gene in *Bombus terrestris*	Related biological function	Reference
Odorant receptor 2	Or2 odorant receptor 2 / Queen mandibular pheromone (QMP) co-receptor (LOC100631089)	- Upregulated by exposure to QMP - Upregulated in sterile workers	Cardoen et al. 2011 Grozinger et al. 2007
Dop3 D2-like dopamine receptor	D2 like dopamine receptor (LOC100644210)	- Upregulated in non-reproductive workers - Expression of Dop3 is affected by the presence or absence of the queen. - Enhance apoptosis (cell death) in ovaries of queen-right workers	Vergoz et al. 2012
Megator	Megator TPR like nucleoprotein (LOC100645723)	- Upregulated in reproductive workers - Involved in honeybee caste determination - Downregulated when expose to QMP	Cardoen et al. 2011
Ecdysteroid-regulated gene E93/Mblk-1 transcription factor	Mushroom body large-type Kenyon cell specific protein 1-like (LOC100645656)	- 27- fold upregulated in reproductive workers - Important in the detection of QMP - Correlated with worker ovary size	Cardoen et al. 2011 Park et al. 2003 Hoover et al. 2003
Ecdysone inducible gene L2/ ImpL2	Neural/ectodermal development factor IMP-L2-like (LOC100645498)	- Homolog of vertebrate IGF-binding protein 7 - Upregulated in non-reproductive workers - Negative regulation of insulin signalling pathway and repress ovary activation - Germ band shortening	Cardoen et al. 2011 Grozinger et al. 2007

4.2.4 PCR amplifications

All primers (Table 4.2) were tested prior to running the SSCP analysis. For each PCR reaction a negative (without DNA) and a positive control (a known pair of primers which work with DNA obtained from any tissue) was also included. Actin, a housekeeping gene (NCBI gene ID: LOC100648624) which worked successfully for a wide range of annealing temperatures (55-60°C) was used as the positive control. Forward and reverse primers used for Actin were 5' ACCATGTACCCCGGTATTGC 3' and 5' GCGATGATCTTGATCTTGATGC 3' respectively. Distilled water was substituted for DNA in the negative control.

Primers which amplified the region of interest correctly were used for SSCP analysis. From a total reaction volume of 25µl (60ng of DNA from each queen and worker, 12.5µl YB-Taq 2x Buffer, 1.5µl of each forward and reverse primer (10µlM/µl) (Table 4.2), 1µl of 10mM $MgCl_2$ and 6.5µl of dH_20). 10µl was used to test PCR success and the remaining 15µl to run SSCP.

Various parameters such as volumes and concentrations of primers, $MgCl_2$ and DNA were changed in order to optimize PCR reactions (https://www.neb.com/tools-and-resources/usage-guidelines/guidelines-for-pcr-optimization-with-taq-dna-polymerase). The optimum annealing and extension temperatures were selected from a temperature gradient PCR.

PCR amplifications were accomplished with the following PCR protocol: an initial denaturation for 5 min at 94°C, 30 cycles of 30s at 94°C, 30s each at the relevant annealing and extension temperature followed by a final extension of 10 minutes and a holding step of 4°C. In all PCR amplifications the lids of PCR machines were maintained at 100°C before the samples were introduced to the machine.

Lab work for genes 6-12 (Table 4.2) was carried out by Bradley Toghill, a MSc student, under my supervision.

Each candidate gene was tested with 4 different bumblebee colonies. Genes 1-5 was tested with 4 bumblebee colonies and another 4 colonies were used to test genes 6-12. Only gene 4 was tested using all 8 colonies (Table 4.2).

Table 4.2: Primer sequences used for different candidate genes.

Gene number	Gene	Exon	Primer Sequence (5'-3')	Amplicon length (bp)
1	Chymotrypsin -1-like	Exon 1	F:CAAACTCGGAGGAAGAGCAC	176
			R:CATCATCGTGGATGAGTGGA	
		Exon 2	F:GGAGGAAGAGCACCAAACAA	169
			R:CATCATCGTGGATGAGTGGA	
		Exon 3	F:ACCGTCCACGCTGGAACGAA	224
			R:GCGTTTTGTCCGGCTGGGGA	
		Exon 4	F:AATCGGTGTCGTTTCCTTTG	223
			R:TCATCATTTTGCAATAAAGCAT	
		Exon 5	F:TTTATAATCAAAACTACTGGGATTTCA	166
			R:TGTGCTCTGCCTTGCAGTC	
		Exon 6	F:GCTTCCCTGAAGGCCAAAT	145
			R:ACACAGTGAGCTGCGGTGA	
		Exon 7	F:CGGTCCACTCATCCACGA	158
			R:CCGGTTGTTCTTCAGTTTCG	
		Exon 8	F:GGAGGAAGAGCACCAAACAA	132
			R:GCATGCACCTTCTCCAACTT	
		Exon 9	F:GCTGGAACGAATCAACTGAA	170
			R:CTACCGGTTGCCAATTTGAT	
		Exon 10	F:CAAATCGGTGTCGTTTCCTT	327
			R:GCAATAAAGCATGTACGTGTTAA	
		Exon 11	F:TCAAAACTACTGGGATTTCACG	218
			R:TGTGCTCTGCCTTGCAGTC	
		Exon 12	F:TGAACAACTTAAAAAGCATTACC	246
			R:TAGTTCCCCAGCCGGACA	
2	Upstream-binding protein 1 -like	Exon 1	F:TTAGGAACGCGATGAAGG	53
			R:ATAACACGAGCGAACGCTACA	
		Exon 2	F:TTGATAGTTTCAACGAAGTACATGC	447
			R:ACCGTCGTGTTTGGCTCAAT	
			F:TCAGCTTCGATACAGGACTGG	

Table 4.2: continued

		Exon 3	R:ATCACCTCAAACCCGCTCTA	192
		Exon 4	F:CCTCTGTTGAAGCATAGGGACT R:CCTAGACTCGTTATCGCCATC	68
		Exon 5	F:GACAGAACCGTGCAGTCGTA R:CCTTAGGTGAACTGCATTAGCA	251
		Exon 6	F:ATAAACGCCGACTTCCTTGG R:CGACAATACGCATTTGCTTC	206
		Exon 7	F:CTTCAAAATCTTCCCGCGATA R:ACGAGCATTGCGACGAAAG	120
		Exon 8	F:CAGTCCTCGTTGAGCTCGTT R:ACCACGTGTCACAGAGAACAA	205
		Exon 9	F:TTGGAATTTTGCTGATTTTCC R:CCTGCTCGCGTTACCATC	72
		Exon 10	F:CCGACGAAGAAAATTCTATTTACC R:GATCTGTCGGAAGCAATGGT	168
		Exon 11	F:CATCTCTCCTAGCATGTTTCTTCTC R:GATAAATACAACTCACCGAAGGA	190
		Exon 12	F:GATGGGGTCGCACCTGTT R:GGAGCTGGAATTCATCATGG	124
3	Zinc finger protein 691-like	Exon 1	F:GAGCCAACCGACCTGAGA R:GTGTTTGAATTGGGGTGAGG	218
		Exon 2	F:ATTGAACCAACGATTCAGTCTTC R:TTGCCAATTCACATTGTTCC	208
		Exon 3	F:CAGAACCAGCCCGTGTTT R:GAGCTCGTGGAAGAACCAAA	351
		Exon 4	F:CTGCGCCTACGACTACAACA R:AGACACTCGCGGATGAAGAC	345
		Exon 5	F:TTGAAGTGTGAAAATTGCAAGAA R:TCCTTAACACTTTTGCATAGCC	301
		Exon 6	F:CGTTATCGATTTACAAAAGTGTCTTG R:GGCATGTAGTGCAAACACAA	333
4	Ecdysone 20 -monooxygenase -like	Exon 1	F:GTACCATCCGGCGAGGAG R:GCATCGTGAACCTTGCTCAG	140

Table 4.2: continued

		Exon 2	F:AGATTTGAACAAACGATATGGAC R:CGTTGACCAAACCGAGATT	200
		Exon 3	F:GCACCTGTACTTTGATCTTGG R:TGTATATCTCGTCCTCGCTTTC	198
		Exon 4	F:CACCTTGGTGTTTCTATTCTACTTGA R:GGAATGATTCGGTGATACATGC	158
		Exon 5	F:ACGGTCCTGTTACACACTTGG R:GCCAAAATAAGCTGCAACG	191
		Exon 6	F:TTCAAACGAGTGCCATTGTC R:CCATTCCGAATCTACCCAGA	319
5	Vitellogenin receptor-like	Exon 1	F:TTACATGATACCAGCCGATCA R:CACCGCAGTTATCCTTCTCA	110
		Exon 2	F:GGCGAAGGTTTGATGATGTT R:TGCTGTAACAATGACCTCTCG	213
		Exon 3	F:AAAAGATATAGCAAAGAGGCCATT R:TTTGTAGACACTAAAACTTCCCAATTT	135
		Exon 4	F:AGCGATCAGCTAAGGGACAA R:ATCGTAAGGCTCGTCTTGGA	216
		Exon 5	F:GAGCGTAACATGCCAAAGAAA R:CTTGTTTCGGAAGACACAGTTC	215
		Exon 6	F:CCAAGTACAATCCCTGCACA R:TCGCTTCGTATGACGTATTCC	205
		Exon 7	F:TTTTGGAGCCAAACTTCTTG R:AACCATCTAAATTCGAGGACTCA	202
		Exon 8	F:TGGATACCCGAAAGATTCCA R:TTTTTCTTTGAAAAGCGTGGT	210
6	Epidermal growth factor receptor like	Exon 1	F:AATTTCACGTCGCCAGAACG R:AGAAGTTCTTGCAGGCGATG	229
		Exon 2	F: GAGCGGATGAGAAGACTTGC R: TGTTGTTTTCAGGACCGTCG	210
		Exon 3	F: AAGCCTTGGTAGACGCTGAT R: CGTTTCTGTGATTAGCGCCA	216
			F:CTTAACGAGCTGACTGTCGC	

Table 4.2: continued

		Exon 4	F: R:TTTACTGCACGTCGTTTCCC	209
7	60S Ribosomal Protein L26 like	Exon 1	F:ACTTGTTTCTTCTTCACGTCGA R:CGAATTGGCATAGAACGAACG	135
		Exon 2	F:ATTGACAGGAGGAGTAAAGGTAG R:ATTTATGAAGTTTCCATGGCTGC	95
8	Or2 odorant receptor 2	Exon 1	F:GCTGTGCATATTTGGAAACCG R:AGAGATCCAAAGAGACGGTGA	182
		Exon 2	F: GGTACTGGGAGCTATGGTCA R: TCGTTTAGTCACAGTTTCCAGC	216
		Exon 3	F:GCACTCAGCGATGTACCAAA R:GCTGTGCATATTTGGAAACCG	207
9	D2 like dopamine receptor	Exon 1	F:ATACCGGACCAATGCCTCTT R:CCATGATGATGCAAGGTATGTAG	85
		Exon 2	F: CAAGCTCAACTGGGTGGAAC R: CGTTCTTCTTGGGTTGCGAA	208
		Exon 3	F: GGGTGTCACTGCTTTCATCG R: TTGAGCTGTTCGATTGTCCG	199
		Exon 4	F:TTCTCAGCACTCTCACACCC R:TTTCTAACACCCCGACGCTC	183
10	Megator TPR like nucleoprotein	Exon 1	F:GTTTGGCACAACGTGATGAA R:TCAGCTTTTGCATTTGCTTC	229
		Exon 2	F: AGATGAAGCTTTGGCAGCTC R: GGCGCATTCTCTTGTAGCAT	256
		Exon 3	F: CAGAAGCTGCCCTTAATTCG R: TCGTGATGTAGCTCGGAGTG	249
		Exon 4	F: CGAAGGTGAAGGAGGTGGTA R: TTCCGACTGTTGTTGCTGTT	229
		Exon 5	F:TGCAAACGAACGGTGTAAGA R:TTCCGAAGAATTAGAAATCGACA	240

Table 4.2: continued

11	Mushroom body large-type Kenyon cell-specific protein 1-like (Mblk1-like)	Exon 1	F:TCGCATTCCAGAGGTGATGA R:CGCCTCTTCCTCGATGTTTG	217
		Exon 2	F: TGAATTTCCGCGACGTGATC R: TCGGTTTGTTCGCTTCTTGG	231
		Exon 3	F: AGATTCCGTCCTACAAGCCC R: GATCACGTCGCGGAAATTCA	212
		Exon 4	F:ACCGGCATTTTACACACCAC R:AGTCGTGCATGAATTTCCGG	211
12	Neural/ ectodermal development factor (IMP-L2-like)	Exon 1	F:GCCCCATATTCCGACAATGG R:CGCCTCTTCCTCGATGTTTG	195
		Exon 2	F: CCAGTACGCATTCTTTGTACGA R: TGTTAACCAGCAGCACCG	171
		Exon 3	F: GCAGAGCAAGTGGAAAACCA R: GGGTAGAGGAACGTGCTGAT	205
		Exon 4	F:CGAAAAGGAAGTGGCTCGAA R:GCGTTGTTGATTCTTAAACGCA	230

Note: F = forward primer, R = reverse primer. Exons where polymorphism was observed are marked in red.

The annealing temperatures used for *Ecdysone 20-monooxygenase-like*, *Mblk1-like* and *IMPL2-like* are 56.5°C, 59.5°C and 58.1°C respectively while the extension temperature used was 72°C for all three genes.

Prior to SSCP analysis, each PCR product (10µl) was checked on a 3% agarose gel and visualized according to section 2.4. If the correct size for the amplicon was obtained, then the rest of the sample (15µl) was used to proceed SSCP to find out the maternal and paternal allelic variations in that gene.

4.2.5 SSCP analysis

SSCP analysis was carried out according to Gasser et al. 2007 using GMA wide mini S-2x25 gels (Elchrom scientific), which have a resolution of 1 bp. Sample denaturing

solution was prepared by mixing 990µl of 95% formamide with 10µl of 1 M NaOH just prior to use. 4µl of the PCR product was denatured with 7µl of denaturing mixture, incubating in a thermocycler at 94°C for 10 minutes and immediately chilling on ice for 5 minutes.

The temperature of the running buffer (1x TAE) in the Origins gel tank was cooled to 9°C before the GMA gel was placed in the apparatus. 7µl of the denatured PCR product was mixed with 2µl of Elchrom loading dye and loaded in to a well on a 25 well GMA gel. Then the gels were subjected to electrophoresis at 72V and at a constant temperature of 9°C through a power pack. The cooling pump was switched off during sample loading and it was activated with a delay of 12 minutes allowing the samples to run a few distance front leaving the wells. Inside the tank, the gels were held in place by a catamaran frame. The electrophoretic running times were varied depending on the fragment size; 10 hours for 150 - 200bp fragment length, 12 hours for 200 - 250bp fragment length, 15 hours for 250 - 350bp fragment length and 17 hours for 350 - 450bp fragment length.

Following electrophoresis, the gels were stained for 30 minutes with SybrGold (invitrogen) (1:10000 diluted in TAE) on a rocking platform in dark and destained with 100ml of 1x TAE buffer for another 30 minutes. Then the plastic backing of the gels was removed and visualized under the UV light.

If any variation or polymorphic banding pattern among the queen and her 5 workers was observed during SSCP, another SSCP run was conducted to confirm the reproducibility of those results. Then genomic DNA of those queen and worker bees were amplified with their respective primers (Table 4.2) and PCR products were sent for commercial clean up and sequencing. Additionally, queen and worker DNA samples of a locus with no variation were also sequenced as a control to the above sequencing results.

All sequencing results were blasted against NCBI, *Bombus terrestris* nucleotide library to verify if the correct sequence was amplified. Then they were analysed using Geneious version 7.3.0 and a heterozygote analysis was performed to identify SNPs that distinguish the queen from her 5 workers in each colony.

After confirming the presence of clear SNPs in the sequences, allele specific PCR was conducted to identify the correct primers for gene expression analysis.

4.2.6 Allele specific PCR (ASP)

Allele specific primers were designed using Batch primer 3 program (http://probes.pw.usda.gov/batchprimer3/) to cover the SNPs identified in section 4.2.5. Two forward primers specific to maternal (F1) and paternal (F2) allele SNPs and a common reverse primer were designed. Genomic DNA of the queen and 5 heterozygous workers in each colony, were PCR amplified with these allele specific primers (Table 4.3). The PCR program used was as per section 4.2.4 and PCR products were checked on a 3% agarose gel. With allele specific primers, only the allele which includes the relevant SNP is amplified. Therefore amplification of genomic DNA with allele specific primers further confirms the maternal and paternal alleles identified during heterozygote analysis in section 4.2.5. Primers which amplified the SNP region successfully were used for qPCR analysis.

Amplification with allele specific primers was unsuccessful for *Mblk1-like*. Thus a GMA gel expression analysis was conducted with the cDNA synthesised from workers which showed allelic polymorphisms during SSCP for *Mblk1-like* (Figure 4.6c). If either the maternal or the paternal allele is not-expressed in *Mblk1-like*, it should be shown in the gel results (Figure 4.7).

4.2.7 Differential expression of maternal and paternal alleles

After confirming maternal and paternal alleles by ASP, RNA was extracted from those bees (queen and 5 heterozygous workers in colony 5, see Table 4.3) and cDNA was synthesized, to test these alleles further for their monoallelic gene expression. The remaining bee samples stored at -80°C (section 4.2.3) were used for RNA extractions.

4.2.7.1 RNA extraction

RNA extractions were performed with QIAGEN RNeasy Mini Kit.

A 30mg sample of frozen tissue from -80°C was ground with a motor and pestle on dry ice. The ground tissue was transferred into a RNase free eppendorf and

Table 4.3: Allele specific primers used for gene expression analysis.

Gene	Colony number (variation seen)	SNP position in chromatogram	Primer sequence (5'-3')	T_A (°C)	Product size (bp)
Ecdysone 20-monooxygenase-like	3 (Figure 4.5a)	Figure 4.5c at 48bp	F1: GCGAGGCCGTAAAGTGTATC	59.3	96
			F2: GCGAGGCCGTAAAGTGTATT		
			R: ACCCAAATGTCGACCCAAGA		
	5 (Figure 4.5b)	Figure 4.5c at 48bp	F1: GCGGAAGCCGTCAGG	60.7	190
			F2: TTAGCGGAAGCCGTCAGA		
			R1: CGAATACTGGCGCGAGATG		
		Figure 4.5c at 48bp	**F1: GCGGAAGCCGTCAGG**	58	34
			F2: TTAGCGGAAGCCGTCAGA		
			R2: GCGAGGCCGTAAAGTGTAT		
Ecdysone internal reference primers			**F: GATTTAGCGGAAGCCGTCAG**	59	36
			R: GCGAGGCCGTAAAGTGTAT		
IMPL2-like	5 (Figure 4.6a)	Figure 4.6b at 253bp	**F1: ACTTGCCAAGCCAAGTCTG**	59.5	205
			F2: CACTTGCCAAGCCAAGTCTA		
			R: TTCGAGCCACTTCCTTTTCG		
IMPL2-like internal reference primers			**F: CTACACTTGCCAAGCCAAGTCT**	59.5	207
			R: TTCGAGCCACTTCCTTTTCG		
Mblk1-like	7 (Figure 4.6c)	Figure 4.6d at 49bp	F1: TCTCTCTCTCTCTTTCTCCCTCTG	58	45
			F2: TCTCTCTCTCTCTTTCTCCCTCTC		
			R1: GTGTACATGCTTGAACCGAAA		
		Figure 4.6d at 51bp	F1: TCTCTCTCTCTCTTTCTCCCTCTGTT	58.2	47
			F2: TCTCTCTCTCTCTTTCTCCCTCTGTA		
			R1: GTGTACATGCTTGAACCGAAA		
		Figure 4.6d at 49bp	F1: TCTCTCTCTCTCTTTCTCCCTCTG	59.6	171
			F2: TCTCTCTCTCTCTTTCTCCCTCTC		
			R2: CGCGTTCGCTTTCATAGTCG		
		Figure 4.6d at 51bp	F1: TCTCTCTCTCTCTTTCTCCCTCTGTT	58.4	169
			F2: CTCTCTCTTTCTCCCTCTCTA		
			R: CGCGTTCGCTTTCATAGTCG		

Note: F1= Forward primer 1, F2= Forward primer 2, R= Common reverse primer. The SNP present is marked in red and is located at the 3' end of each forward primer. Primers which amplified the SNP region successfully (marked in blue) were used for qPCR analysis. T_A = Annealing temperature.

600µl of pre-made β-Mercaptoethanol (10µl) and Buffer RLT (1ml) mixture was added to that. Samples were homogenized by passing the lysate about 5 times through a blunt 20-gauge needle fitted to an RNase-free syringe. Then the lysate was centrifuged for 3 minutes at 20,000g and the supernatant was transferred in to a new microcentrifuge tube.

An equal volume of 70% ethanol was added to that and 700µl of this mixture was

transferred in to an RNeasy spin column placed in a 2ml collection tube. The spin column assembly was centrifuged for 15s at 8000g. Sample volumes that exceeded 700µl were centrifuged in the same RNeasy spin column and the flow-through after each centrifugation was discarded. Followed by another two centrifugation steps (15s at 8000g) with Buffer RW1 (700µl) and Buffer RPE (500µl), the column was spun for 2 minutes at 8000g. Next, it was placed in a new collection tube and centrifuged at 20,000g for 1 minute to evaporate all residual ethanol. The old collection tube with the flow-through was discarded and the column was transferred into a new eppendorf tube. Fifty microlitres of RNase-free water was added to the column membrane and centrifuged for 1 minute at 8000g to elute RNA. This last step was repeated to maximize the yield of total RNA in the final volume.

4.2.7.2 DNase I treatment and cDNA synthesis

Any DNA contamination present in the above RNA extractions were eliminated according to Amplification Grade DNase I Kit protocol (Sigma-Aldrich), prior to synthesize cDNA.

Eight microliters of RNA sample was mixed with 1µl of 10X reaction buffer and 1µl of DNase I (1 unit/µl) in a RNase free PCR tube and incubated at room temperature for 15 minutes. Then 1µl of stop solution was added to that to inactivate the DNase I. Finally, the samples were incubated at 70°C for 10 minutes and chilled on ice. The remaining RNA samples were stored at -80°C to avoid degradation. Duplicate samples were made to use for both +RT and -RT reactions (with and without Reverse Transcriptase respectively).

After measuring the concentrations of DNase treated RNA by NanoDrop Spectrophotometer, they were used to synthesize cDNA, using the Tetro cDNA synthesis Kit (Bioline). The priming premix was prepared on ice in a RNase-free reaction tube according to Table 4.4.

After mixing the components gently by pipetting, they were incubated at 45°C for 30 minutes. The reaction was terminated by incubating at 85°C for 5 minutes and immediately chilling on ice. Synthesized cDNA was stored at -80°C.

Table 4.4: Components and volumes of PCR priming mix.

Component	Volume (µl)
DNase treated RNA (1µg)	n
Primer: Oligo (dT)18	1
10mM dNTP mix	1
5x RT Buffer	4
Ribosafe RNase Inhibitor	1
Tetro Reverse Transcriptase (200units/µl)	1
DEPC-treated water	Up to a final volume of 20µl

Note: n = Volume of RNA extraction required to synthesis 1µg of total cDNA. DEPC treated water was substituted for reverse transcriptase in -RT reactions.

4.2.7.3 Allele specific quantitative PCR

Reference primers were designed according to Gineikiene et al. 2009. A common forward primer was designed to the same target SNP sequence, a few bases upstream excluding the SNP position, leaving the same common reverse primer previously used with SNP primers (see reference sequences in Table 4.3). The reference primers measure the total expression of the gene, whereas the target SNP primers measure the amount of allele expression due to the SNP. Thus the expression difference between the reference and SNP primers would be the relative expression due to the SNP.

Each heterozygous locus was composed of 3 different reactions; maternal (F1), paternal (F2) and reference (Table 4.3). Three replicate samples were run for each reaction. -RT reactions and water negative controls were also set up for each reaction to test for genomic contaminations and primer dimer formation respectively. Altogether 60 reactions were prepared for each gene from the 5 workers. i.e. for each worker 12 +RT reactions (3 x F1-reverse, 3 x F2-reverse, 3 x Reference-reverse) and 3 -RT reactions (1 x F1-reverse, 1 x F2-reverse, 1 x Reference-reverse). All reactions were prepared by a Corbett robotics machine, in 96 well qPCR plates (Thermo Scientific, UK). Q-PCR reaction mix (20µl) was composed of 1µl of diluted cDNA (50ng/µl), 1µl of forward and reverse primer (5µM/µl each, Table 4.3), 10µl 2X SYBR Green JumpStart Taq ReadyMix (Sigma Aldrich, UK) and 7µl ddH_20.

Then, the qPCR plates were removed from the Corbett robotics machine, covered with PCR caps and placed in a PTC-200 MJ thermocycler, set up on Opticon software (v4.7.97.A) for Windows 2000 Professional. The qPCR profile used was comprised of; 4 minutes at 95°C denaturation followed by 40 cycles of; 30 s at 95°C, 30 s at the relevant annealing temperature (Table 4.3) and 30 s at 72°C and a final extension of 5 minutes at 72°C.

4.2.7.4 Data analysis

SYBR Green is a fluorescent dye that binds to double stranded DNA. Fluorescence emitted during qPCR is proportional to the amount of double stranded DNA present in the reaction. The point at which the reaction shows a detectable increase in fluorescence due to the formation of PCR products is defined as the cycle threshold (Ct), i.e. the greater the PCR product available in the reaction, the smaller the number of cycles (Ct value) required for the fluorescent signal to cross the threshold.

Relative quantification of the amount of a target template with that of a reference template in the sample helps to calculate the amount of expression due to the SNP. Mean value for each triplicate set (maternal, paternal and reference reactions) were calculated by the Opticon software. Mean expression (-delta Ct) values for each parental allele in each worker bee was calculated as follows:

-delta Ct(maternal) = Mean reference Ct value - Mean maternal allele Ct value

-delta Ct(paternal) = Mean reference Ct value - Mean paternal allele Ct value

-Delta Ct ratios for the maternal allele of the 5 workers were grouped together and compared to the corresponding -delta Ct ratios for the paternal allele in each gene to check whether the overall expression difference between parental alleles is significant.

The Shapiro-Wilk normality test was used to check if the data was normally distributed. P values that were equal to or greater than p=0.05 indicated as a normal distribution and allows parametric assumptions for data analysis. A matched paired t-test was performed to check if the allele specific expression values were significantly different among the two parental alleles.

4.3 Results

4.3.1 SSCP analysis

Exon coverage for each gene is given in Table 4.5. The number of genes analysed was given priority over the percentage of a gene covered.

Table 4.5: SSCP exon coverage.

Gene name	Exon coverage (%)	Presence /absence of variation
Chymotrypsin-1-like	92	Absent
Upstream-binding protein 1-like	93	Absent
Zinc finger protein 691-like	70	Absent
Vitellogenin receptor-like	30	Absent
Epidermal growth factor receptor like	21	Absent
60S Ribosomal Protein L26 like	32	Absent
Or2 odorant receptor 2	25	Absent
D2 like dopamine receptor	54	Absent
Megator TPR like nucleoprotein	17	Absent
Ecdysone 20-monooxygenase-like (Edy20-like)	37	Present
Mushroom body large-type Kenyon cell-specific protein 1-like	35	Present
Neural/ectodermal development factor IMP-L2-like	47	Present

Note: Percentage exon coverage for each gene was calculated as, the sum of all tested amplicon lengths as a fraction of the total length of mRNA.

Variation was not present in 9 genes out of the 12 candidate genes (Table 4.5) tested, i.e. the queen and her workers shared the same banding pattern (homozygous), since there was no difference in sequences between the maternal and paternal alleles (Figures 4.3 and 4.4). Except for 3 exons which showed a heterozygous banding pattern among queen and workers, the rest of all exons tested using SSCP (Table

4.2) showed a homozygous banding pattern. Thus only a single gel picture from each gene is represented in Figures 4.3 and 4.4.

Exon 3 in *Odorant receptor 2* was used as a control to compare with heterozygous sequences. An identical banding pattern between the queen and her workers was observed in the control reaction (Figure 4.4i) indicating that these workers are homozygous at that locus. Their sequencing results further confirms this fact showing no sequence variation between the queen and worker chromatograms (Figure 4.4j).

Compared with the queen, workers in colony 3, 5 and 7 showed a heterozygous banding pattern in 3 genes. i.e they show both polymorphic paternal bands (present only in workers) and monomorphic maternal bands (shared between the queen and her workers); *Ecdysone 20-monooxygenase-like* (*Edy20-like*), *IMPL2-like* (Figure 4.6a) and *Mblk1-like* (Figure 4.6c). For *Edy20-like* the same heterozygous banding pattern was observed in two tested colonies; colony 3 (Figure 4.5a) and colony 5 (Figure 4.5b).

4.3.2 Identifying the SNPs in polymorphic loci

Heterozygous banding patterns were observed in SSCP gels for 3 genes. Clear SNPs that explain these heterozygosities were identified in their sequencing results; (a) In *Edy20-like* the entire queen sequence is homozygous, displaying a single fluorescent peak at each base position (Figure 4.5c). At the SNP (48^{th} base pair) while the queen is homozygous showing a guanine (G), all 5 of her workers show double peaks corresponding to both guanine (G) and adenine (A) bases. This proves that all workers are heterozygous at that locus and they all have received allele G from their mother (maternal allele) and allele A from their father (paternal allele). (b) A similar SSCP banding pattern and a chromatogram was resulted for *IMPL2-like* (Figure 4.6a and b). (c) In *Mblk1-like*, several closely associated SNPs (hyper-variable region) were observed (Figure 4.6d) and at each of these SNP position the heterozygosity of all 5 workers remained consistent. Due to the presence of clear, consistent SNPs in worker sequences allele specific primers were able to be designed for each parental allele of the above 3 genes for expression analysis.

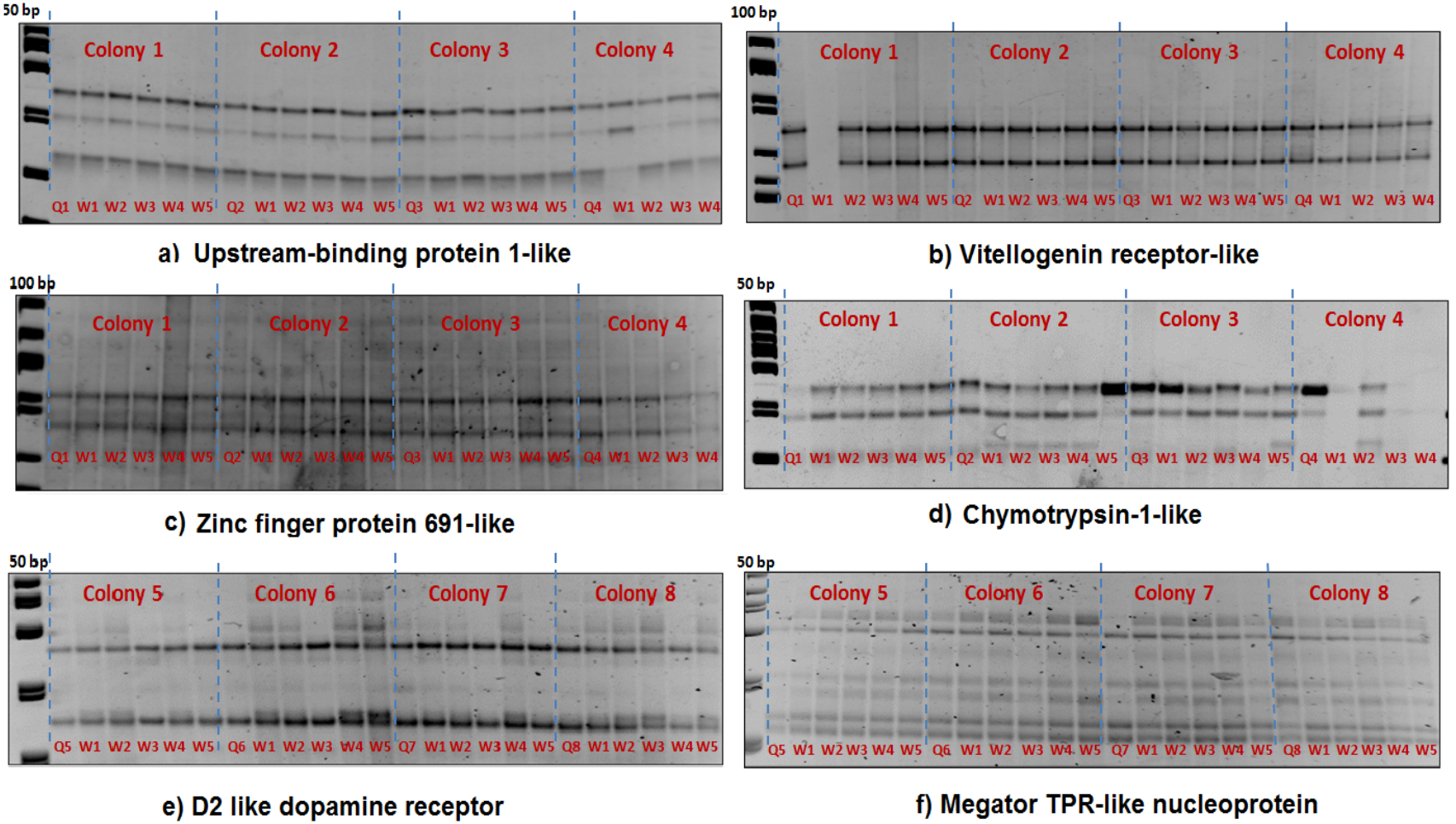

Figure 4.3: SSCP gel results of six genes (a - f) with no queen-worker variations (homozygous banding patterns). The queen (Q) and 5 workers (W1-W5) are represented in each colony.

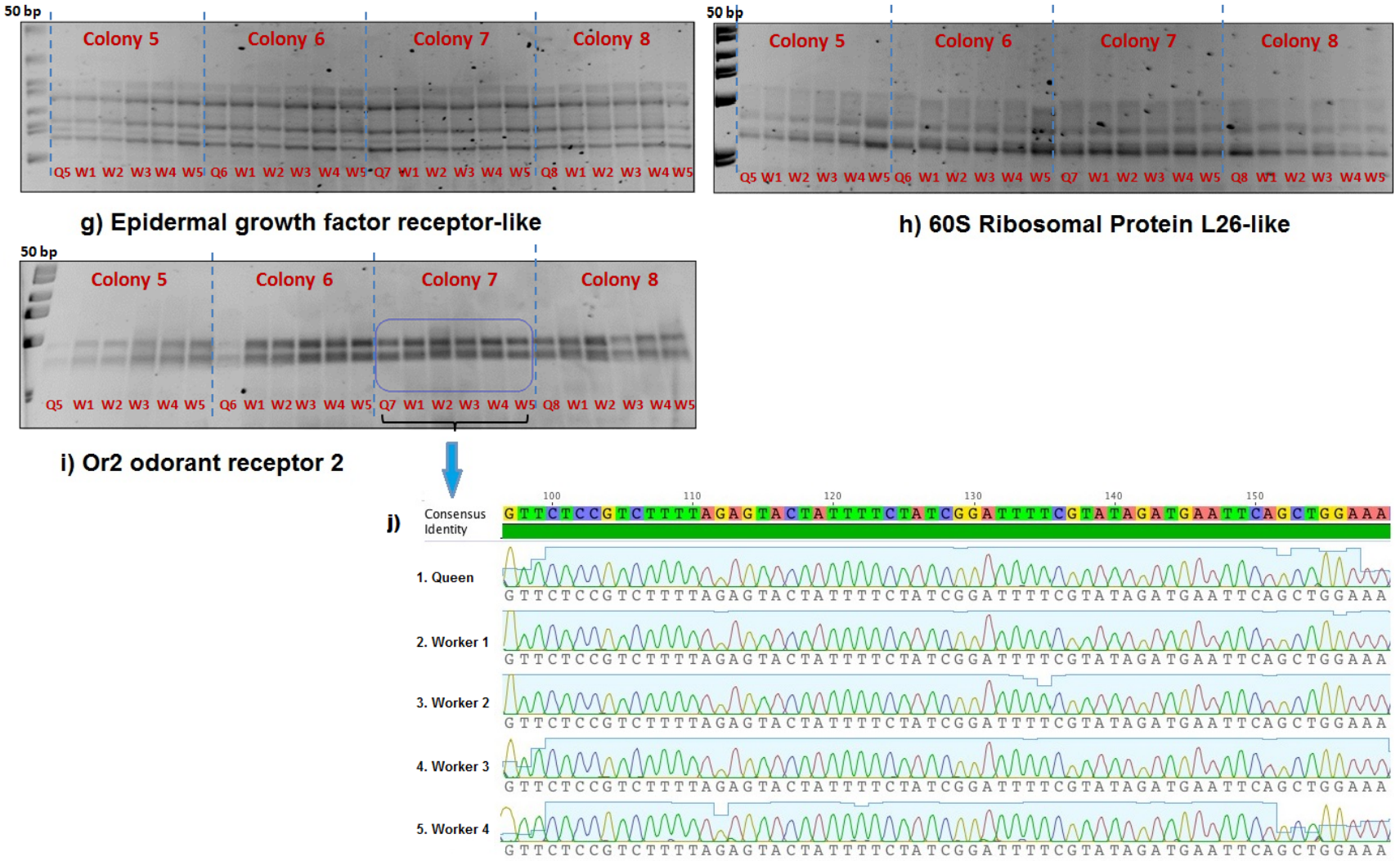

Figure 4.4: Homozygous banding patterns among the queen and her workers for 3 genes (g,h,i). Sequencing results at the represented locus in colony 7, *Odorant receptor 2* shows no variation between queen and worker sequences - control (j).

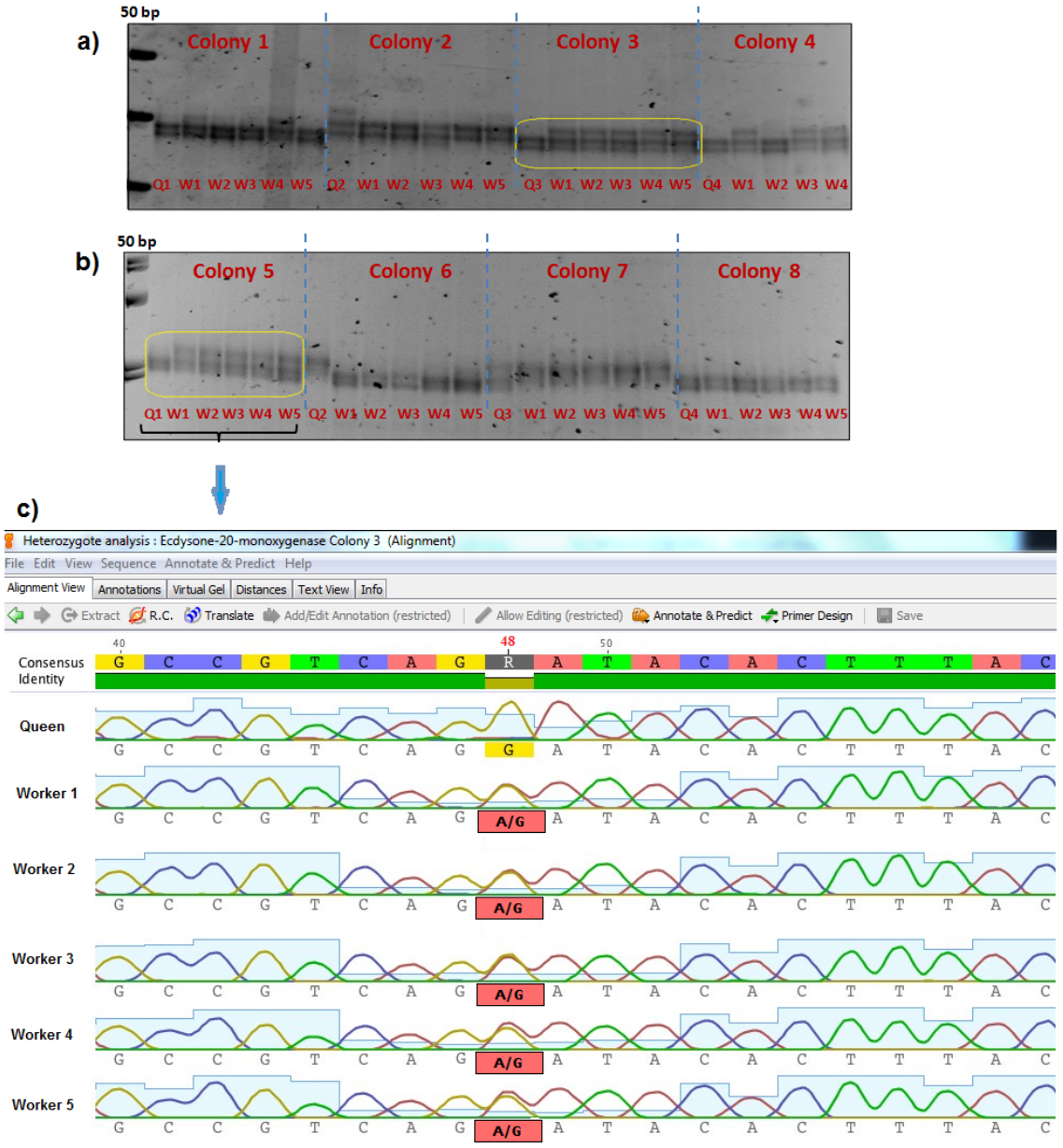

Figure 4.5: Allelic polymorphisms in *Edy20-like*; SSCP gel results for colony 3 (a) and colony 5 (b) show allelic polymorphism between the queen and her workers. i.e. Each worker in colony 3 and 5 shows a unique band (paternal) that is not present in the queen. Sequencing results indicate that the queen is homozygous G at the SNP position while the workers are heterozygous with double peaks for both A and G bases (c).

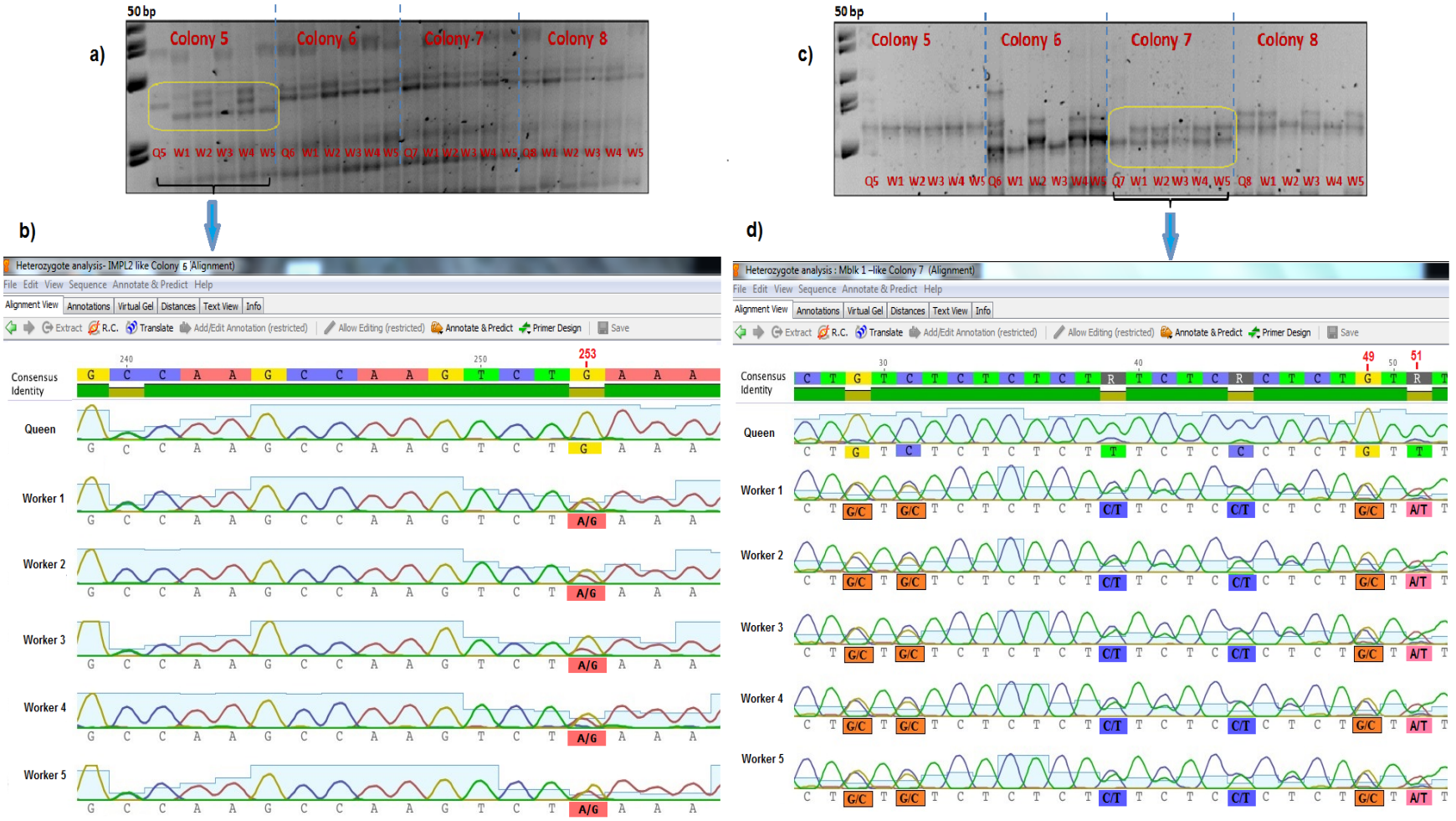

Figure 4.6: SSCP gels showing allelic polymorphisms and sequencing results showing heterozygous SNPs for *IMPL2-like* (a and b), *Mblk1-like* (c and d) genes. In *IMPL2-like* (b), the queen is homozygous G while all 5 of her workers are heterozygous with both A and G. *Mblk1-like* (d) shows a hyper variable region with many closely associated repeated SNPs.

4.3.3 Allele specific PCR

a) *Edy20-like* and *IMPL2-like*

Allele specific primers designed for *Edy20-like* and *IMPL2-like* worked successfully with genomic DNA to produce the expected fragment lengths, 34bp and 205bp respectively (Table 4.3, Figure 4.8a and d). Hence they were used for gene expression analysis. In addition, allele specific PCR of genomic DNA for *Edy20-like* (Figure 4.8a) and *IMPL2-like* (Figure 4.8d) also showed the presence of both maternal (G) and paternal (A) alleles in all workers. This further confirms the heterozygosity of workers. Their queen samples show only allele G (maternal band) and lack of the paternal band (allele A) in queens (Figure 4.8a and d) which also confirms the specificity of the allele specific primers designed.

As with the genomic DNA, allele specific PCR of cDNA also showed the expression of both maternal and paternal alleles in these workers (Figures 4.8b and e). Thus it confirms that monoallelic expression is not present at the tested loci of *Edy20-like* and *IMPL2-like*. Therefore using allele specific qPCR, expression differences of the two parental alleles were analysed.

-RT reactions of *Edy20-like* and *IMPL2-like* used for qPCR analysis were also checked in an agarose gel. The resulted blank gel (Figures 4.8c and f) confirmed that there was no genomic contamination present in the cDNA samples prepared.

b) *Mblk1-like*

Due to hyper variability of the selected *Mblk1-like* locus (Figure 4.6d), amplification using any of the designed allele specific primer sets (see table 4.3) was unsuccessful. Thus qPCR analysis for *Mblk1-like* was unable to be conducted. cDNA-SSCP expression analysis, the alternative method used for qPCR showed the expression of both maternal and paternal alleles. Thus monoallelic expression is not present at this selected locus of *Mblk1-like* (Figure 4.7).

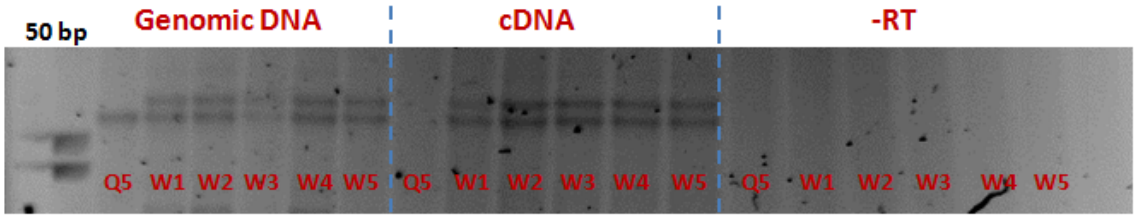

Figure 4.7: SSCP gel showing the expression of parental alleles in *Mblk1-like*; Genomic DNA shows the queen is homozygous while all 5 workers are heterozygous. cDNA shows the expression of both parental alleles. Blank -RT reactions confirm that there was no genomic DNA contamination in the cDNA samples

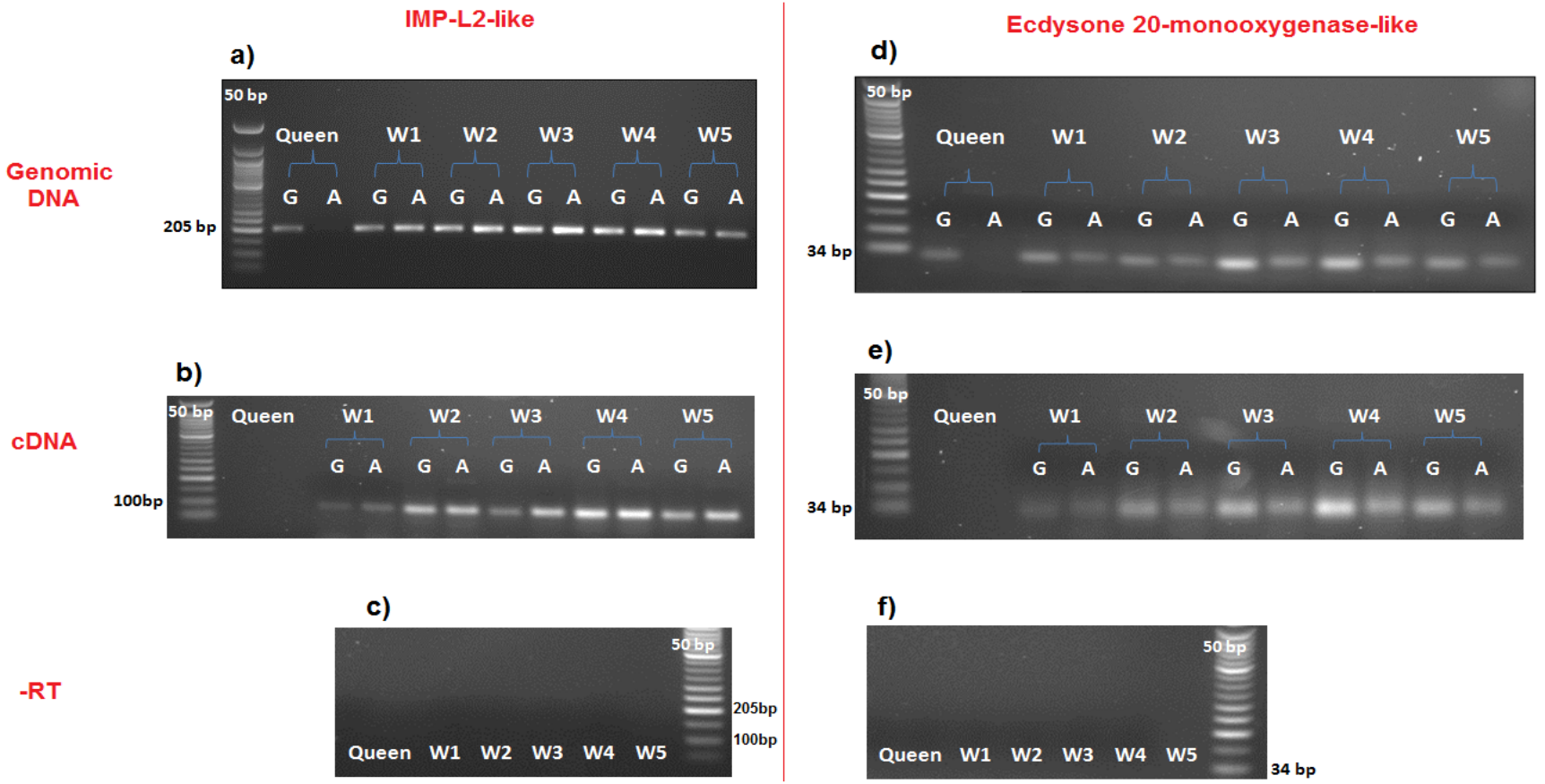

Figure 4.8: Gel results of allele specific PCRs with genomic DNA (a and d) and cDNA (b and e) for *IMPL2-like* and *Edy20-like*. Presence of both maternal and paternal bands with genomic DNA confirms worker heterozygosity (a and d). Presence of both alleles with cDNA confirms biallelic expression of parental genes (b and e). W = Worker, G = Maternal allele, A = Paternal allele. -RT results (c and f) confirm that there are no genomic contamination present in the samples.

4.3.4 Allele specific qPCR

Table 4.6: qPCR data of *Edy20-like* showing the relative expression of parental genes in each worker bee.

Sample	delta Ct		Normaility test passed?	Matched paired t-test p value
	Maternal allele	Paternal allele		
W1	-6.76	-5.6	Yes at p=0.127	0.0378
W2	-5.76	-5.57		
W3	-5.17	-4.77		
W4	-6.28	-5.09		
W5	-5.87	-5.59		

Compared to the matrigene, patrigene showed an increased expression in *Edy20-like*. This pattern was consistent among all 5 workers (Figure 4.9) and the expression difference between the two parental alleles was statistically significant (t = 2.94, df = 4, p = 0.0378, Table 4.6).

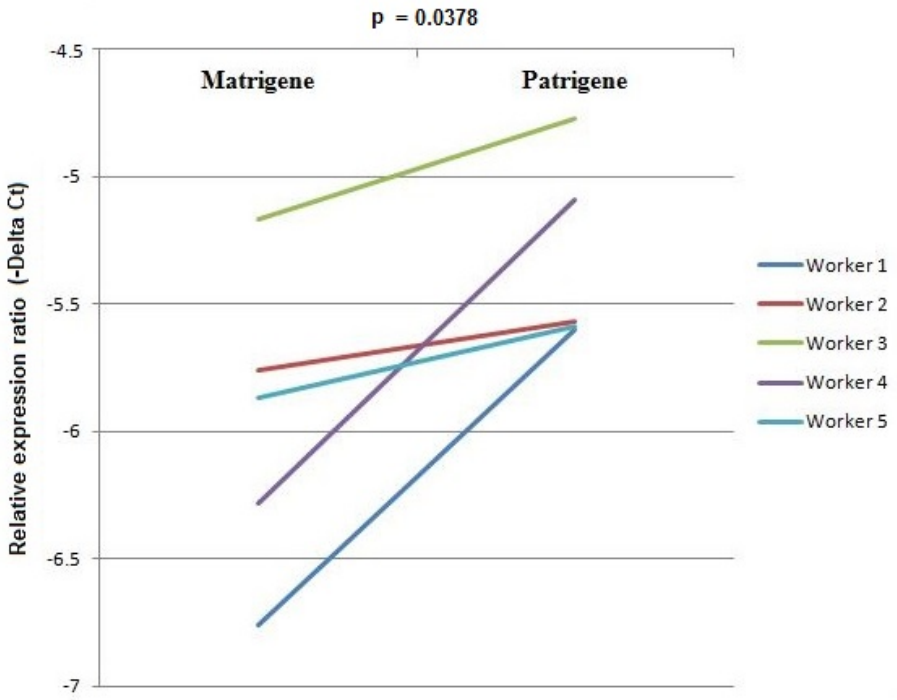

Figure 4.9: Interaction plot displaying the relative expression (-delta Ct) of matrigene and patrigene of *Edy20-like*. Each coloured line represent an individual bee.

Table 4.7: qPCR data of *IMPL2-like* showing the relative expression of parental genes in each worker bee.

Sample	delta Ct		Normaility test	Matched paired t-test
	Maternal allele	Paternal allele	passed?	p value
W1	-3.03	-1.42	Yes at p=0.3403	0.2758
W2	-1.19	-2.01		
W3	-2.45	-1.27		
W4	-1.83	-1.58		
W5	-2.15	-1.43		

For *IMPL2-like*, compared to the patrigene, a higher expression of the matrigene was observed only in worker 2. This pattern was reversed for the rest of the workers (Figure 4.9). Overall expression difference between the matrigene and patrigene in *IMPL2-like*, was not significant (t = 1.40, df = 4, p = 0.2758, Table 4.6).

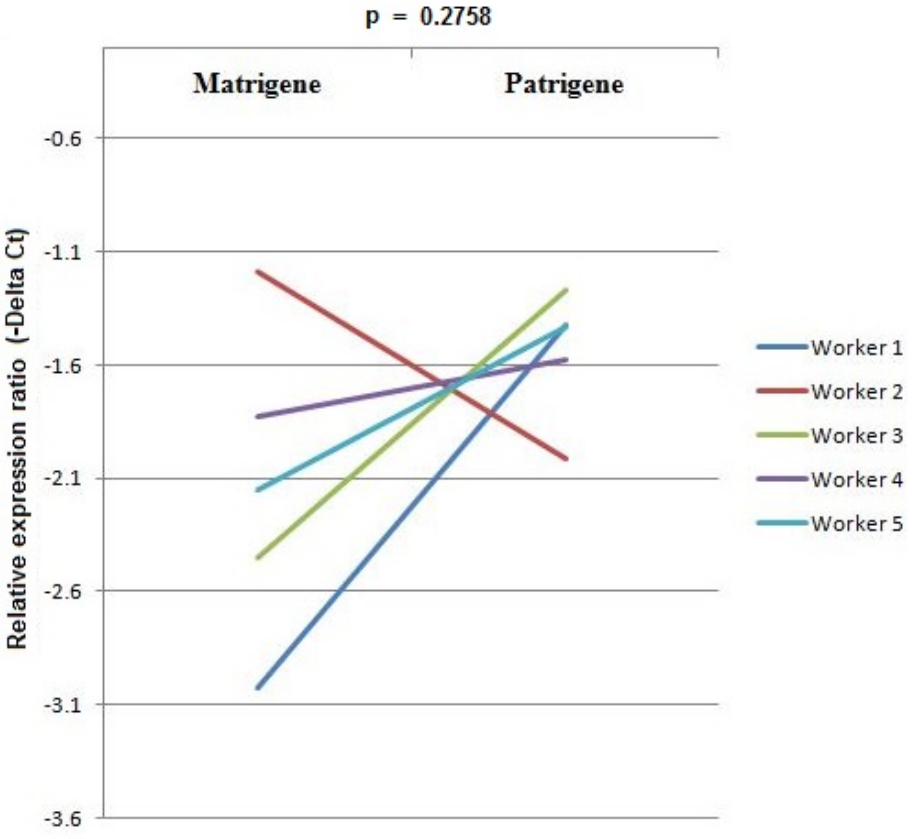

Figure 4.10: Interaction plot displaying the relative expression (-delta Ct) of matrigene and patrigene of *IMPL2-like*. Each coloured line represent an individual bee.

4.4 Discussion

The overall aim of this chapter, searching for apparent parent-of-origin allele specific expression in *Bombus terrestris* worker reproduction loci was successful. The observed expression patterns of parental alleles are exactly the way they should be as predicted by the theory. For a queen-right monandrous colony, the theory predicts that at a locus which enhances worker reproduction the matrigene should be silenced and patrigene should be expressed (Queller, 2003). Instead of a complete silencing of an allele, I observed for a locus (*Edy20-like*) that enhances worker reproduction the matrigene is less expressed compared to the patrigene.

Out of 12 genes examined during this study, allelic polymorphisms were observed only in 3 genes; *Mblk1-like*, *IMPL2-like* and *Edy20-like*. All three of these genes belong to the same ecdysteroid regulated gene family. However they are linked with ovary functions in different ways; *Mblk1-like* and *Edy20-like* are involved in increasing worker reproduction whereas *IMPL2-like* is involved in regulating worker sterility. Generally, ecdysteroids have been identified as key regulators of *B.terrestris* worker reproduction (Geva et al., 2005).

Mblk1-like is a brain expressed ecdysteroid induced gene which has been implicated in the detection of queen mandibular pheromone (QMP) in honeybees (Cardoen et al., 2011). QMP is important for workers to detect the queen-right condition. Thereby it suppress worker reproduction and enables the maintenance of cooperation among colony members (Although see Yagound et al. 2012, since the reproductive statuses of workers could be varying in the presence of a foreign colony queen). Hence in the absence of QMP, workers may upregulate *Mblk1-like* expression and enhance their reproduction. Compared to the sterile workers, 27 fold upregulated expression of *Mblk1-like* has been observed in reproductive workers (Kocher et al., 2010). Therefore *Mblk1-like* expression is positively correlated with worker reproduction. In a queen-right colony, if imprinting is present in this gene, it should be in the maternally derived allele (see section 1.2). However, I was unable to carry out qPCR analysis to measure the precise expression of each parental allele in *Mblk1-like*, as amplification by allele specific primers wasn't successful. This is because the locus in which the allelic variations were observed during SSCP (Figure 4.6c) is a hyper variable region. i.e. a highly polymorphic region with closely associated,

repeated SNPs (Figure 4.6d). Therefore the designed allele specific primers may not have been specific enough to bind to the DNA for a successful PCR amplification. However cDNA analysis on SSCP gel (Figure 4.7), showed the expression of both parental alleles. The expression difference between maternal and paternal alleles may be significant. Unfortunately I was unable to measure it during this study.

Edy20-like has a similar function to *Mblk1-like*. In addition they are also functionally linked with each other; Targeted inhibition of *Edy20-like* decreases the induction of *Mblk1-like* expressed proteins (Cardoen et al., 2011). Hence *Edy20-like* provides an alternative target to test the same assumptions of the research hypothesis, although obtaining quantitative measurements of allele specific gene expression wasn't successful for *Mblk1-like*.

Sequencing results of *Edy20-like* (Figure 4.5c) for the queen showed a single peak for allele G thus the queen is homozygous G. In contrast, all 5 workers were heterozygous at the same SNP position showing double peaks for alleles A and G. Thus they have received the allele A from the drone (paternal allele) and allele G from the queen (maternal allele). Four-fold upregulation of *Edy20-like* was observed in egg laying honeybee workers compared to non-reproductive workers (Cardoen et al., 2011). This indicates a positive correlation between its expression and worker reproduction. Similar to the predictions made for *Mblk1-like*, if genomic imprinting is present in *Edy20-like*, it is expected to be observed in the maternally derived allele. Thus the paternally derived allele should be more expressed than the maternally derived allele. This expression pattern was exactly observed in all 5 workers (Figure 4.9) tested for exon 3 (Table 4.2) in *Edy20-like*. Thus monoallelic expression wasn't discovered in *Edy20-like*. Instead allele specific expression, which supports the predictions made by the theory, was observed.

Ecdysone-inducible gene L2, (*B. terrestris*, *IMPL2-like* analogous gene in the honeybee) is linked with reproductive inhibition of workers. It functions similarly to an insulin like peptide and negatively regulates insulin signaling pathways to repress ovary activation. Increased expression of this gene has been found in sterile honeybee workers compared to reproductive workers (Cardoen et al., 2011, Grozinger et al., 2007). All workers tested for *IMPL2-like* (Figure 4.6b) were heterozygous with double peaks (A and G) at the SNP position. The queen was homozygous G. Therefore the maternal and paternal alleles were confirmed as G and A respectively.

According to Haig's theory any locus that inhibits worker reproduction should be imprinted at the paternally derived allele and expressed in the maternally derived allele. This was observed in worker 2 during qPCR analysis, as an increased expression of the matrigene compared to the patrigene (Figure 4.10, Table 4.7). However, the expression difference between these two parental alleles was not significant. The rest of the workers tested for *IMPL2-like* showed a contrast expression pattern, where patrigene was more expressed than the matrigene.

Therefore it seems that these results go against the predictions of the Haig's theory. *IMPL2-like* is an ecdysteroid inducible gene which regulates worker reproduction potential in a synergistic way at the start of a feedback system. Hence its expression and role in worker reproduction may be more complicated than I suggested in the previous paragraph. Therefore any prediction for its expression also may be complicated. Unlike *IMPL2-like*, ecdysteroids such as *Edy20-like* play a central role in *B.terrestris* reproduction. For example there is a strong association between reproduction, level of juvenile hormone and the level of ecdysteroids in bumblebees. In addition, ecdysteroids are key compounds involved in ovary activation, regulating agonistic behaviour and establishing the dominance hierarchy in workers and queens (Geva et al., 2005). I suggest this makes it a better candidate than *IMPL2-like* for this study purpose.

Use of SSCP to find genomic imprinting is challenging. SSCP detects variation up to 500bp fragment size with a high resolution of 1bp. However, the sensitivity of SSCP decreases when the fragment length exceed 200bp. Thus medium length fragments around 200bp were used for this analysis (Weber et al., 2005). Therefore covering the full exome using SSCP is a time consuming and labor intensive process. On the other hand finding genomic imprinting is extremely rare. Also SNPs in exons are expected to be rare. Out of 66 loci tested, SNPs were found only in 3 loci (Table 4.2).

One way around the lack of SNPs might be to examine the untranslated regions (UTRs). Selecting target sites for SSCP analysis in UTRs may increase the potential of finding SNPs in worker reproduction loci. As the name implies UTRs are not translated into proteins. This means they are not as visible to selection as the gene coding exons and indeed many variations have been observed in both 3' and 5' UTR mRNA transcripts (Lytle et al., 2007). Hence variation between the maternally and

paternally derived alleles will have a greater chance to occur in UTRs. If target sites for the SSCP analysis were selected in UTRs, then many polymorphic regions could be found among the queen and workers. Thus the potential number of genes that could proceed to gene expression analysis would be increased.

Finding monoallelic expression or genomic imprinting is very rare and imprinting in the human genome only has 1% chance to occur (Luedi et al., 2007). During this study the number of genes analysed was given more preference over the extent of the gene covered. However, in certain candidate genes (e.g. *Chymotrypsin-1-like*, *Upstream-binding protein 1-like*, *Zinc finger protein 691-like*, see Table 4.5), 70-93% of the exome length was checked during the SSCP analysis. However, allelic polymorphism wasn't observed in any of these genes. In addition genomic imprinting could vary among tissues and also even within the same gene it could vary between different isoforms (Flores et al., 2012). During my analysis whole bodies of the bee were used for DNA extractions. Hence gene expression patterns observed during this analysis should represent the overall expression of all body tissues. In *B. terrestris* the ovary has been identified as the primary site of ecdysteroid synthesis (Geva et al., 2005). Therefore ovary might be a better target site where genomic imprinting could occur in genes such as *Edy20-like*.

As another future implication, it would be interesting to check the extent of methylation present in each parental allele of *Edy20-like* (e.g. bisulfite sequencing). Genomic imprinting in mammals has been linked with DNA methylation (Varriale, 2014). Although genomic imprinting is yet to be discovered, DNA methylation has been found in social Hymenoptera (Wang et al., 2006). Hence the lower expression of matrigene observed in *Edy20-like* may be correlated with DNA methylation. If a pattern such as monoallelic methylation was observed, it would further support the role methylation plays in parent-of-origin allele specific expression patterns.

In addition, the expression patterns of parental alleles, in *Edy20-like* and *IMPL2-like* can be checked for honeybees, thereby to test the predictions made by the theory for a polyandrous mating system.

During this study allele specific expression that is associated with *Bombus terrestris* worker reproduction and sterility was found in two genes (*Edy20-like* and *IMPL2-like*). At a reproduction enhancing locus (e.g. *Edy20-like*), the theory predicts that

the patrigene should be expressed and matrigene should be silenced. According to this study in *Edy20-like*, patrigene has been found expressed more, than the matrigene. Therefore these results match with the predictions made by the theory for a queen-right monandrous colony. However, to confirm the allele specific expression patterns observed in this study are exactly due to parent-of-origin effects, a reciprocal cross experiment (see section 8.2.1) would need to be carried out.

Clearly the candidate gene approach is limited in its application. Unlike the candidate gene approach which uses targeted gene regions, next generation sequencing technology provides much more sophisticated methods to conduct methylation and gene expression analysis in genome-wide scale. A next generation approach used to check allele specific DNA methylation and monoallelic expression of genes in the *Bombus terrestris* genome is discussed in the next chapter.

Chapter 5

Monoallelic methylation and allele specific expression in the bumblebee, *Bombus terrestris*: Next generation sequencing approach

5.1 Introduction

Most genes in diploid organisms are biallelically expressed. However in some genes, only a single allele is transcriptionally active. This is known as monoallelic expression (Eckersley-Maslin and Spector, 2014). The inactivate allele is selected either randomly (e.g. X-chromosome inactivation and some autosomal genes) or predetermined in a parent-of-origin dependent manner (genomic imprinting, Fedoriw et al. 2012). DNA methylation is a key mechanism which governs monoallelic gene expression via the regulation of transcription and alteration of chromatin structure (see section 1.6, Varriale 2014). In humans the two alleles in most imprinted genes are differentially methylated. Thus in genomic imprinting, DNA methylation serves as an epigenetic tag defining the parental identity. It has also been found that most monoallelically expressed genes in mammals are monoallelically methylated

(Milani et al., 2009). In mammals, at imprinted loci, the transcriptionally silenced allele is significantly more methylated than the active, expressed allele. For example in humans at the *H19/Igf2* imprinted locus, the paternal allele is silenced and is hypermethylated compared to the expressed maternal allele (Steyaert, 2014).

Eusocial Hymenoptera are a potential candidate where genomic imprinting could have evolved (see section 1.5). Yet imprinting in eusocial insects has not been discovered. In recent years, whole-genome sequence data of the honeybee and seven species of ants have become available (see section 3.1). In addition, genome-wide DNA methylation maps have been obtained from different castes of 3 ant species; *Camponotus floridanus*, *Harpegnathos saltator* (Bonasio et al., 2012, Simola et al., 2013) and *Solenopsis invicta* (Hunt et al., 2013). As a result the presence of monoallelic methylation and monoallelic expression have been discovered in *C.floridanus* and *H.saltator* (Bonasio et al., 2012). However very little is known about the functional role of monoallelically expressed genes and genome-wide locations of methylated loci in eusocial Hymenoptera.

In my previous work, I have observed the phenotypic effects of DNA methylation on *B. terrestris* worker reproduction (chapter 3). Then using a candidate gene approach, I found apparent parent-of-origin allele specific expression in targeted reproductive genes of *B. terrestris* (chapter 4). In this chapter my aim was to conduct a genome-wide survey to find monoallelically methylated and monoallelically expressed genes that could potentially be imprinted in *B. terrestris*.

An integrated approach of methylated DNA immunoprecipitation sequencing (MeDIP-seq), methylation-sensitive restriction enzyme sequencing (MRE-seq) and ribonucleic acid sequencing (RNA-seq) was used for this purpose. Compared to the other expensive, genome-wide DNA-methylation profiling methods (e.g. bisulfite and methylC-seq), this has been recommended as a cost effective method which also has a high capacity to generate accurate and qualitative data (Harris et al., 2010, Zhang et al., 2013). The theoretical background of main steps included in this experiment are given below (Figure 5.1).

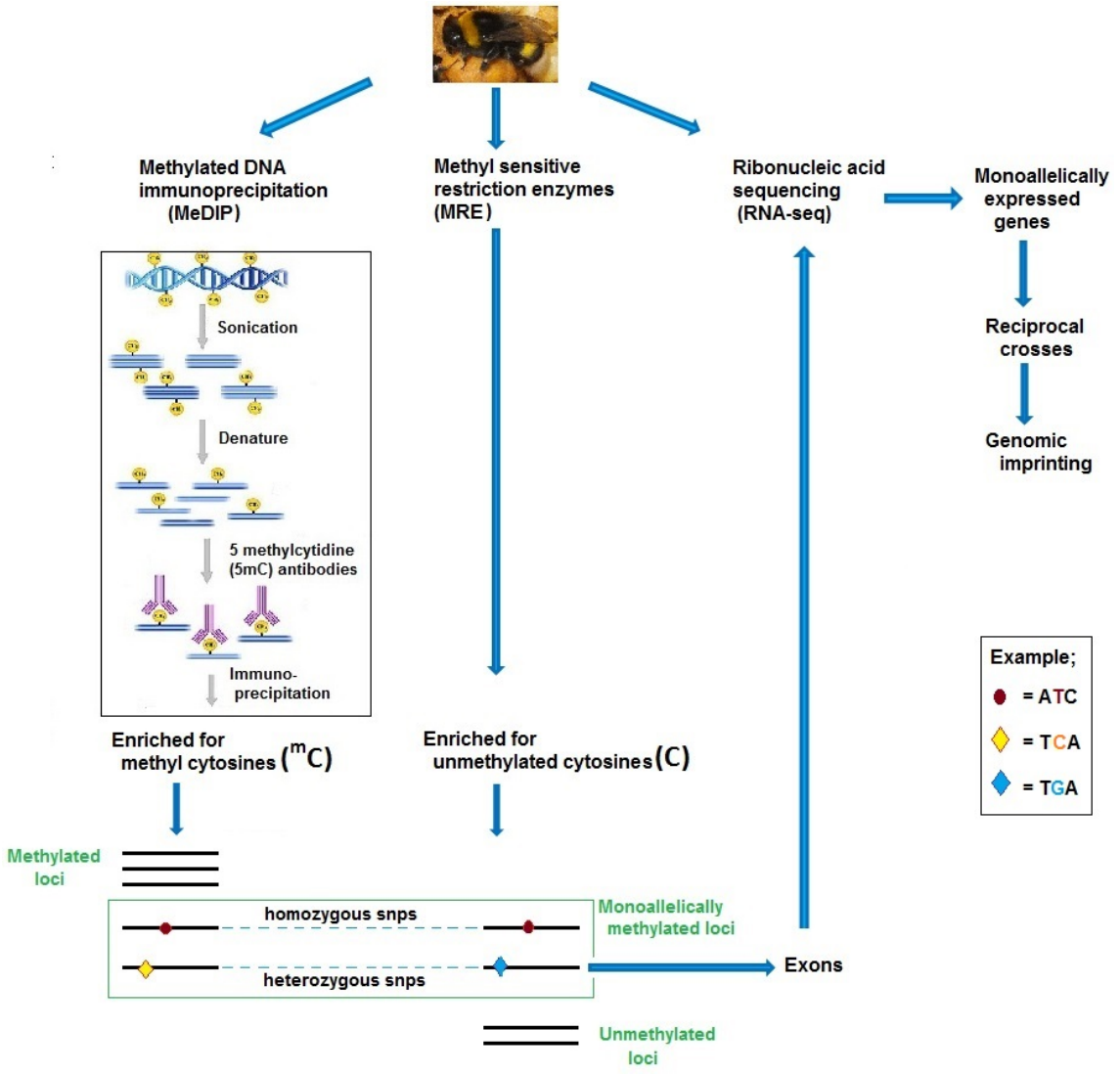

Figure 5.1: A flow diagram illustrating main steps of the next generation sequencing protocol. MeDIP-seq, MRE-seq and RNA-seq libraries were constructed using DNA and RNA extracted from a single worker bee. When MeDIP-Seq and MRE-seq reads were aligned, methylated loci (reads that present only in the MeDIP library), unmethylated loci (reads that present only in the MRE library) and monoallelically methylated loci (present in both MeDIP-seq and MRE-seq libraries) can be identified. Monoallelically methylated loci, could be homozygous (identical alleles of a gene) or heterozygous (different alleles of a gene). Monoallelic expression could occur at both homozygous and heterozygous loci. During this study we have focused only on heterozygous loci. Exonic sequences (introns are not included in the mRNA transcripts) which contain the SNPs that are responsible for the heterozygosity of alleles were compared with the RNA-seq library. Among the two heterozygous SNPs/alleles, if only a single allele is present in the RNA-seq library, that indicates a monoallelic expression. Reciprocal crosses (see section 8.2.1) need to be carried out on those monoallelically expressed genes to confirm that the observed expression is exactly due to genomic imprinting.

MeDIP is a genome enrichment technique that is used to enrich for methylated DNA sequences. It consists of isolating methylated DNA fragments via an antibody that specifically recognizes 5-methylcytidine (5mC). First genomic DNA is randomly sheared in to fragments by sonication. Then they are denatured to produce single-stranded DNA to enhance the binding affinity of antibodies in subsequent steps. Next methylated DNA fragments are immunoprecipitated with a monoclonal antibody (Figure 5.1). Following digestion of antibodies with proteinase K, methyl enriched DNA can be isolated for further analysis. The resulting enrichment of methylated DNA can be used as an input to high-throughput DNA detection methods. This produces a genome-wide readout of the methylation status of cytosines (Mohn et al., 2009).

MRE is a complementary technique to MeDIP. It identifies unmethylated CpG sites via a series of DNA digestions with methyl-sensitive restriction enzymes such as HpaII (C/CGG), Hin6I (G/CGC) and AciI (C/CGC). As in MeDIP-seq, the isolated unmethylated DNA can be combined with next generation sequencing methods to obtain a genome-wide profile of unmethylated cytosines (Maunakea et al., 2010).

When MeDIP-seq (methylated DNA) and MRE-seq (unmethylated DNA) reads are aligned with each other, any locus which is common to both these reads would represent monoallelic methylation (Harris et al., 2010). In other words, at that gene locus one allele is methylated while the other allele is unmethylated. Nucleotide sequence of alleles at such a locus could be identical (homozygous) or different among the two alleles (heterozygous) at specific SNPs. See Figure 5.1. Therefore both homozygous and heterozygous alleles could be monoallelically methylated. During this study I have focused only on heterozygous alleles.

Exonic sequences with heterozygous SNPs that are corresponding to monoallelic methylation, were compared with the RNA-seq library. Intronic regions were excluded because the mature mRNA in eukaryotes is already spliced, thus the cDNA produced does not include introns. This means any SNP present in introns would not be included in the RNA-seq library. Among the two alleles (each allele from MeDIP-seq and MRE-seq library) with heterozygous SNPs, if only one allele is present in the RNA-seq library that indicates a monoallelic expression of the gene (Harris et al., 2010). If both the alleles are present in the RNA-seq library that gene is biallelically expressed.

Out of the two heterozygous alleles used for RNA-seq analysis, we know which allele was methylated (present in the MeDIP-seq library) and which was unmethylated (present in the MRE-seq library). Therefore monoallelic methylation can be related to monoallelic expression observed in the RNA-seq analysis. If I were able to find monoallelic methylation and/or monoallelically expressed genes in the *B. terrestris* genome, that will not confirm the presence of genomic imprinting in the bumblebee. To confirm that those monoallelic expressions were exactly due to parent-of-origin effects (genomic imprinting), a reciprocal cross experiment (see section 8.2.1) will need to be carried out for those monoallelically expressed loci.

5.2 Methods

5.2.1 Sample preparation

A bumblebee colony was purchased from Koppert Biological Systems, UK and reared in the laboratory according to section 2.1.

RNA and DNA was extracted from a single, five day old whole bee (female worker).

DNA extraction DNA was extracted using an ethanol precipitation method as follows. First, DNA extraction buffer was prepared by mixing 20ml 1M TRIS (pH 7.5), 5ml 0.5M EDTA, 12.5ml 2M NaCl and 57.5ml dH20. Then this solution mixture was autoclaved and 5ml of 10% SDS was added to that.

Four hundred microlitres of DNA extraction buffer and add 7µl of 20mg/ml proteinase K were added to the tissue in an eppendorf tube, vortexed for 15s and incubated in dry heat block at 55°C overnight (16 hours approximately). Then the samples were heated at 92°C for 10 minutes and and vortexed for 10 - 15s. It was then centrifuged for 2 minutes at full speed. The supernatant (300µl) was pipetted into a fresh eppi tube and 300µl of isopropanol was added into that. It was mixed for 3-5 times and placed in the -80°C freezer. After 30 minutes the samples were taken out of he freezer. Once the samples were defrosted they were centrifuged for 10 minutes at 13000*g*. Then all the supernatant was discarded until a dry pellet remained in the tube. 198µl of 70% EtOH was added to the pellet, gently flicked

the tube and centrifuged for 2 minutes at 13000g. The pellet was dried again by discarding all the supernatant. The tube was covered with a micropore tape and left to dry the pellet fully for 2 hour. 100µl of ddH_20 was added into the pellet and left overnight at 4°C to resuspend DNA into the solution.

RNA extraction RNA extraction was performed according to Sigma-Aldrich, TRI Reagent protocol.

Every 100mg of tissue was homogenized with 1ml of TRI reagent(TRIzol) in a sterile, RNase free microcentrifuge tube. Then the samples were incubated for 5 minutes at room temperature followed by centrifugation at 12000g for 10 minutes. The supernatant was transferred in to a fresh eppendorf tube and 200µl of chloroform which is not contained any isoamyl alcohol was added to that. The sample tube was shaken vigorously for 15s and left at room temperature for 15 minutes for phase separation. It was centrifuged at 12000g for 15 minutes and the sample was separated into 3 clear phases; a red organic phase (Protein) in the bottom,an inter-phase pellet (DNA) in the middle and a colourless aqueous phase (RNA) at the top.

RNA isolation The top aqueous phase was transferred into a new eppendorf tube without any contamination of the inter-phase. This sample was mixed with 500µl of isopropanol and left at room temperature for 10 minutes. It was centrifuged at 12000g for 10 minutes. RNA was precipitated as a pellet at the bottom of the tube. Next, all the supernatant was discarded and the RNA pellet was washed with 1ml of 75% ethanol and vortexing for 15 seconds. It was then centrifuged at 7500g for 5 minutes, all ethanol was discarded and the tube was left inverted for 10 minutes to get the RNA pellet dry. 20µl of RNase free water was added to the pellet and incubated at 55°C for 10 minutes. RNA samples were immediately stored at -80°C and all centrifugation steps were performed at 4°C. During the procedure, all equipment used were treated with Diethylpyrocarbonate (DEPC) treated water.

Determination of concentration and purity of samples DNA concentration was measured by NanoDrop 1000 Spectrophotometer (Table 5.1). The quality and the concentration of RNA was assessed by Agilent 2100 Bioanalyzer (Figure 5.2).

Table 5.1: Quantification of DNA and RNA

Sample	Concentration (ng/µl)	Volume (µl)	TOTAL amount of nucleic acid extracted(µg)
DNA	1075	100	96
RNA	3065	20	55

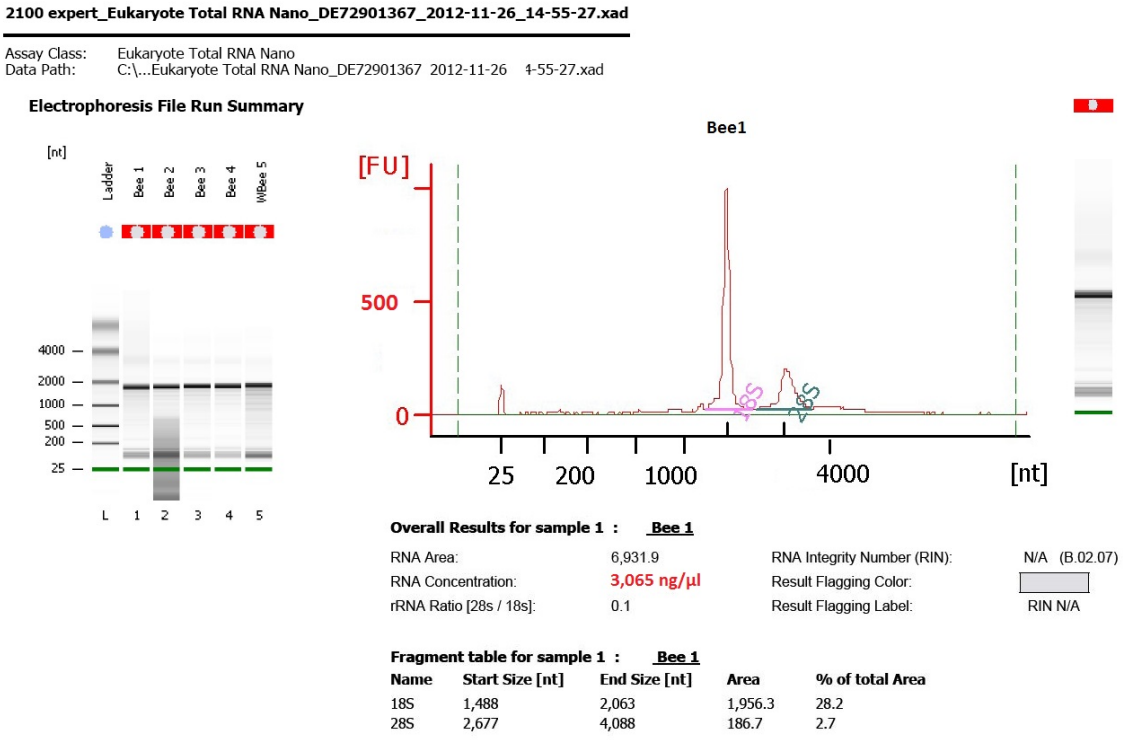

Figure 5.2: Quality and quantity of RNA used for RNA-seq analysis (Bioanalyzer results).

5.2.2 Next generation sequencing

Three libraries were prepared from this bee. These were MeDIP-seq and MRE-seq libraries on the DNA sample and one amplified short insert cDNA library with size of 150-400 bp using RNA. Both the MeDIP-seq and MRE-seq library preparations are based on previously published protocols (Harris et al., 2010). MeDIP-seq uses monoclonal antibodies against 5-methylcytosine to enrich for methylated DNA independent of DNA sequence. MRE-seq enriches for unmethylated cytosines by

using methylation-sensitive enzymes that cut only restriction sites with unmethylated CpGs. Each library was individually indexed. Sequencing was performed on an Illumina HiSeq 2000 instrument (Illumina, Inc.) by the manufacturer's protocol (Eurofins Medigenomix GmbH, Ebersberg, Germany). Multiplexed 100 base paired-read runs were carried out yielding 9390 Mbp for the MeDIP-seq library, 11597 Mbp for the MRE-seq library and 8638 Mbp for the RNA-seq library.

5.2.3 Bioinformatic analysis

Bioinformatic analysis of these data was carried out by Kate Lee, Bioinformatics and Biostatistics Analysis Support Hub (BBASH), University of Leicester. The main steps included in the analysis are described below.

Monoallelic methylation and expression We searched for genes that were monoallelically methylated (present in both methylation libraries), heterozygous and monoallelically expressed (only one allele present in the RNA-seq library). As we will also know which allele is methylated (present in MeDIP-seq) and which is unmethylated (present in MRE-seq), we can also relate methylation status to monoallelic expression.

Alignment and bam refinement mRNA reads were aligned to the *Bombus terrestris* genome assembly (AELG00000000) using Tophat (Kim et al., 2013) and converted to bam files with Samtools (Li et al., 2009). Reads were labelled with the AddOrReplaceReadGroups.jar utility in Picard (`http://picard.sourceforge.net/`). The MRE-seq and MeDIP-seq reads were aligned to the genome using BWA mapper (`http://bio-bwa.sourceforge.net/`) (Li and Durbin, 2009). The resultant sam alignments were soft-clipped with the CleanSam.jar utility in Picard and converted to bam format with Samtools. The Picard utility AddOrReplaceReadGroups.jar was used to label the MRE and MeDIP reads which were then locally re-aligned with GATK (DePristo et al., 2011, McKenna et al., 2010). PCR duplicates for all bams (mRNA, MeDIP and MRE) were marked with the Picard utility Markduplicates.jar.

Identifying regions of interest and integrating data Coverage of each data type was calculated using GATK DepthofCoverage (McKenna et al., 2010). Only regions with a read depth of at least six in each of the libraries (RNA-seq, MeDIP-seq and MRE-seq) was used. Heterozygotes were identified using Samtools mpileup and bcftools on each data set separately (Li and Durbin, 2009) and results were merged with vcf tools (Danecek et al., 2011). CpG islands were identified using CpG island searcher (Takai and Jones, 2002). Regions of mRNA with overlaps of MeDIP, MRE, CpG islands and monoallelic SNPs were identified with custom perl scripts.

5.3 Results

Connection between monoallelic methylation and allele specific expression

The MRE library yielded 11597 Mbp (57,987,374 reads), the MeDIP library was 9390 Mbp (46,952,052 reads) and the RNA-seq library was 8638 Mbp (43,190,661 reads).

We found 593 loci that were present in both MeDIP-seq and MRE-seq libraries indicating that in these genes only one allele is methylated (monoallelic methylation). Of these, in 19 genes only one allele was present in the RNA-seq library indicating that these 19 genes are monoallelically expressed (Table 5.2). Of the nineteen, fourteen had the MeDIP allele expressed, while five had the MRE-seq allele expressed. These were blasted against the nr/nt database (blastn). Four returned no hits. Another four returned non-informative hits ("hypothetical proteins"). The remaining eleven are discussed in section 5.4.

This includes two genes that are important in worker reproduction: (i) *bicaudal D*, a gene which has previously been shown to be differentially methylated in eggs and sperm of the honeybee (Figure 5.3a). (ii) *ecdysone receptor* which belong to the same ecdysteroid family as *Ecdysone 20 monooxygenase* that has been identified in chapter 4 as a gene with apparent parent-of-origin allele specific expression (Figure 5.3b). During the current study we found that these two genes are monoallelically methylated and monoallelically expressed.

Table 5.2: Monoallelically methylated and monoallelically expressed genes in the *Bombus terrestris* genome.

Gene number	Accession number mRNA(start:end)	Sequence description from blast2go	Length (bp)	Allele expressed
1	AELG01000543.1|2414 |4660|	*Apis mellifera* ecdysone receptor transcript variant misc_rna	2246	MRE-seq
2	AELG01001021.1|2595|3944|	*Apis dorsata* protein yippee-like 1 like transcript variant mrna	1349	MeDIP-seq
3	AELG01001021.1|2651|4107|	*Apis mellifera* potassium voltage-gated channel protein shaker 1 like transcript variant mrna	1456	MeDIP-seq
4	AELG01004618.1|50434|51141|	*Apis mellifera* ras GTPase-activating protein nGAP-like transcript variant mrna	707	MeDIP-seq
5	AELG01000977.1|210957|213433|	*Apis mellifera* centrosomal and chromosomal factor-like transcript variant mrna	2477	MeDIP-seq
6	AELG01000623.1|2|3526|	*Bombus terrestris* slit homolog 2 protein-like mrna	3524	MeDIP-seq
7	AELG01000544.1|153356|154477|	*Bombus terrestris* methionine aminopeptidase 1-like mrna	1121	MeDIP-seq
8	AELG01000969.1|141301|142147|	*Bombus terrestris* excitatory amino acid transporter 4-like partial mrna	846	MRE-seq
9	AELG01003672.1|34174|36205|	*Bombus terrestris* calmodulin-lysine N-methyltransferase-like mrna	2031	MRE-seq
10	AELG01004467.1|2976|7852|	*Bombus terrestris* elongation of very long chain fatty acids protein 6-like transcript variant 2 mrna	4876	MeDIP-seq
11	AELG01005399.1|62869|63510|	*Bombus terrestris* bicaudal D-related protein homolog mrna	641	MeDIP-seq
12	AELG01000620.1|45487|46965|	PREDICTED:*Bombus terrestris* hypothetical protein LOC100651168 (LOC100651168) mrna	1478	MeDIP-seq
13	AELG01002224.1|26170|28244|	*Bombus terrestris* hypothetical LOC100650069 miscrna	2074	MeDIP-seq
14	AELG01002224.1|30371|30980|	*Bombus terrestris* hypothetical LOC100650069 miscrna	609	MeDIP-seq
15	AELG01003249.1|27|707|	*Populus tricocarpa* clone pop018-complete sequence	743	MeDIP-seq
16	AELG01001796.1|10244|11803|	N/A	1559	MeDIP-seq
17	AELG01002621.1|84719|85551|	N/A	832	MRE-seq
18	AELG01004342.1|151574|152335|	N/A	761	MeDIP-seq
19	AELG01006475.1|1|1769|	N/A	1768	MeDIP-seq

Note: Sequence accession number, sequence name, length of the RNA fragment and the allele expressed in the RNA-seq library are illustrated in the table. Out of the 19 monoallelically methylated and monoallelically expressed genes found, four were returned as non-informative hits or "hypothetical proteins" (marked in red). Another four genes were returned as no hits (N/A; marked in green). The remaining eleven linked with *Bombus terrestris* physiology and discussed in this chapter are marked in blue.

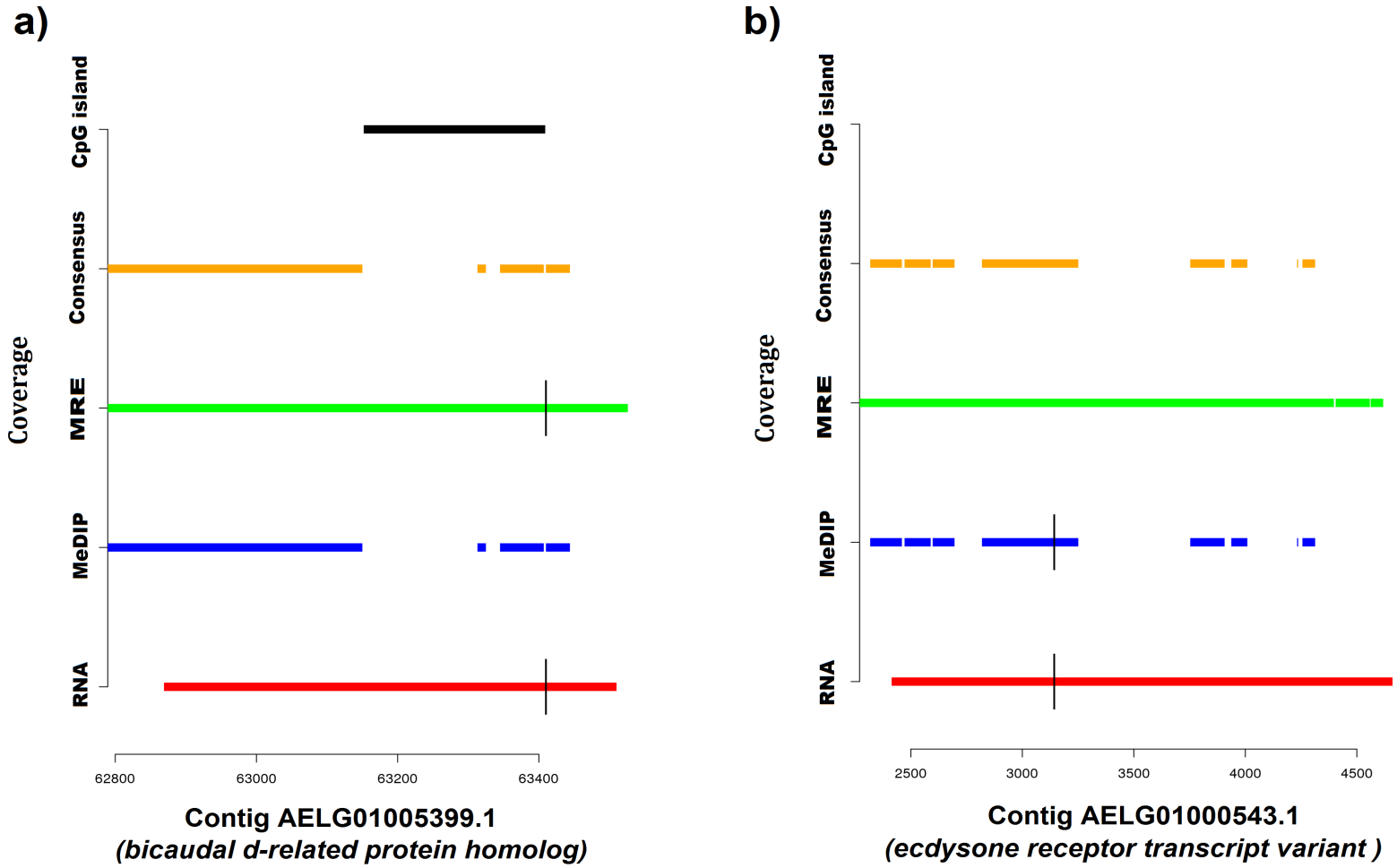

Figure 5.3: Alignment of MeDIP-, MRE- and RNA- sequencing reads for (a) *bicaudal D-related protein homolog* and (b) *ecdysone receptor transcript variant*. Overlaps among MeDIP-seq and MRE-seq indicate monoallelically methylated regions. RNA fragments which intersect with these monoallelically methylated regions are represented by consensus. Vertical black lines represent the locations of monoallelic SNPs which also indicate the methylation status of the expressed allele.

5.4 Discussion

Monoallelically methylated and monoallelically expressed genes discovered in the *B. terrestris* genome, excluding those of which returned no or non-informative hits, are discussed below. See table 5.2.

Ecdysone receptor

Ecdysteroids are involved in metamorphosis and oogenesis in non-social insects (Takeuchi et al., 2007). In various social insect species they have been shown to be involved in ovary activation and dominance hierarchy in workers and queens (Geva et al., 2005). In bumblebees, there is a strong association between behavioural dominance, reproduction and higher levels of juvenile hormone and ecdysteroids (Geva et al., 2005). We found the MRE-seq allele of *Ecdysone receptor* (AELG01000543.1) to be expressed (Figure 5.3b). In *Drosophila melanogaster*, ecdysone receptor interacts with ecdysone to activate a series of ecdysteroid genes (Takeuchi et al., 2007). In honeybees, *Ecdysone receptor* is expressed in the brain mushroom bodies of both workers and queens and ovaries of queens (Takeuchi et al., 2007).

Bicaudal D-related protein homolog

We found the MeDIP allele of *bicaudal D-related protein homolog* (AELG01005399.1) to be expressed in our sample (Figure 5.3a). Bicaudal is involved in embryonic pattern formation in *Drosophila* (Markesich et al., 2000). It is thought to be involved in the differentiation between soldiers and workers in the termite *Reticulitermes flavipes* (Scharf et al., 2003). Intriguingly, *bicaudal protein D* has been shown to be methylated more in eggs than sperm in a recent paper comparing the methylation differences between these in honeybees (Drewell et al., 2014).

Yippee-like 1

We found the MeDIP allele of *yippee-like 1* (AELG01001021.1) was expressed. Yippie was first discovered in *Drosophilia*. It is an intracellular protein with a zinc-finger like domain. It interacts with hemolin in insects (Roxstrom-Lindquist and Faye, 2001). Hemolin is well known in its antimicrobial role but seems also to be developmentally regulated (Roxstrom-Lindquist et al., 2005). The steroid hormone 20-hydroxyecdysone (20E) activates the expression of *Hyalophora cecropia* Hemolin in

the fat body of diapausing pupae (Roxstrom-Lindquist et al., 2005). DNA methylation of a CpG island near the *yippie-like 3* promoter in humans represents a possible epigenetic mechanism leading to decreased gene expression in tumours (Kelley et al., 2010).

Slit homolog 2 protein-like

The MeDIP allele of *slit homolog 2 protein-like* (AELG01000623.1) was expressed. Slit (AELG01000623.1) is produced by midline glia in insects and is involved in cell projection during development (Rothberg et al., 1990). All three human Slits were found to be hypermethylated in hepatocellular carcinoma cell lines (Zheng et al., 2009).

Methionine aminopeptidase 1-like

Methionine aminopeptidases catalyse N-terminal methionine removal, a cellular process required for proper biological activity, subcellular localization, and eventual degradation of many proteins (Leszczyniecka et al., 2006). We found the MeDIP allele of *methionine aminopeptidase 1-like* (AELG01000544.1) was expressed. MAP1D in humans was found to be potentially oncogenic (Leszczyniecka et al., 2006).

Shaker

Shaker (MeDIP allele expressed, AELG01001021.1) is involved in the operation of potassium ion channel. *Shaker* expression was correlated with foraging experience in honeybees (Lutz et al., 2012). *Shaker* expression was upregulated in sterile versus reproductive honeybee workers (Cardoen et al., 2011). Again, this was thought to be due to foraging behaviour differences.

Centrosomal and chromosomal factor-like

We found that the MeDIP allele of *centrosomal and chromosomal factor-like* (AELG01000977.1) was expressed. The Drosophila homolog CORTO is required for proper condensation of mitotic chromosomes and progression through mitosis (Salvaing et al., 2003).

Excitatory amino acid transporter 4-like

The MRE-seq allele of *excitatory amino acid transporter 4-like* (AELG01000969.1) was expressed. Excitatory amino acid transporters are neurotransmitter transporters. In mammals, excitatory amino acid transporters terminate the excitatory signal by removal of glutamate from the neuronal synaptic cleft. *Excitatory amino acid transporter 3* expression was upregulated in sterile versus reproductive honeybee workers (Cardoen et al., 2011). It is thought to be involved in the regulation of flight muscle contraction. *Excitatory amino acid transporter 1* expression differences was associated with worker - queen differentiation in the paper wasp *Polistes metricus* (Toth et al., 2014).

Ras GTPase-activating protein nGAP-like

The MeDIP allele of *ras GTPase-activating protein nGAP-like* (AELG01004618.1) was expressed in our sample. *Ras GTPase-activating protein 1* was found to be upregulated in reproductive compared to sterile honeybee workers (Cardoen et al., 2011). It is involved in oocyte meiosis.

Calmodulin-lysine N-methyltransferase-like

Calmodulin-lysine N-methyltransferase catalyses the trimethylation of a lysine residue of calmodulin. Calmodulin is a ubiquitous, calcium-dependent, eukaryotic signalling protein with a large number of interactors. The methylation state of calmodulin causes phenotypic changes in growth and developmental processes (Magnani et al., 2010). The MRE-seq allele of *calmodulin-lysine N-methyltransferase-like* (AELG01003672.1) was expressed in our study.

Elongation of very long chain fatty acids protein 6-like

Only the MeDIP allele of *elongation of very long chain fatty acids protein 6-like* (AELG01004467.1) was expressed in our sample. Elongation of very long chain fatty acid proteins are involved in the elongation of fatty acids from the diet into very long chain fatty acids (Sassa and Kihara, 2014). The timing of the upregulation of fatty acid metabolism was found to be different in queen and worker honeybees (Li et al., 2010).

When the different functional roles of above eleven genes are considered, they become ideal examples to support many aspects of genomic imprinting.

Firstly, Haig's kinship theory predicts that parental alleles could be in conflict during reproduction. Thus reproductive loci are potential places where imprinting could occur. We found that *bicaudal D-related protein homolog*, a gene which is differentially methylated among the two parental gametes in honeybees (Drewell et al., 2014), is monoallelically expressed. Thus this is evidence for different parental expression interests during worker reproduction. Drewell et al. 2014 further supports our findings since it states that during gametogenesis in the honeybee, a substantial proportion of CpGs are methylated in a paternal specific manner. Including *bicaudal protein D* they reported 381 genes that are significantly different in CpG methylation among egg and sperm in the honeybee.

In addition, several other genes found in our study serve as important candidates of reproductive division of labour in eusocial insects. For example, *Shaker* in honeybee workers is associated with non-reproductive roles such as foraging. *Excitatory amino acid transporter 3* and *Ras GTPase-activating protein 1* show an upregulation in sterile workers compared to the reproductive workers. *Ecdysone receptor* which belongs to the ecdysteroid family is a key compound for maintaining reproductive dominance among the queen and workers.

Secondly, imprinted genes found in mammals are often involved in embryonic growth and development (Reik and Walter, 2001). Some monoallelically expressed genes found in this study (e.g. *slit homolog 2 protein-like* and *calmodulin-lysine N-methyltransferase-like*) are linked with embryogenesis and also with differential development.

Thirdly, almost all imprinted genes in mammals are associated with differentially methylated regions or differentially methylated domains (Fedoriw et al., 2012), suggesting a strong correlation between DNA methylation and genomic imprinting. All monoallelically expressed genes found in this study are monoallelically methylated. Thus this study also provides evidence for a link between monoallelic methylation and monoallelic expression.

Fourthly, loss of imprinting and high promoter methylation are associated with carcinogenesis (Wang, 2012). Genes such as MAPID and *yippee-like 1*, found in this study are related with oncogenesis in humans.

In conclusion, monoallelically methylated and monoallelically expressed loci found in this study are exactly where they would be predicted to be for imprinted genes. Thus these results provide evidence for the feasibility of evolution of genomic imprinting in eusocial hymenopteran societies. By carrying out an analysis on reciprocal colonies (section 8.2.1) it can be confirmed whether these genes are actually imprinted in *B. terrestris* or only that they are expressed monoallelically.

As there are conflicts between maternal and paternal components of an individual's genome there are numerous conflicts between individuals who share the same genome. Colony members of eusocial insects are a remarkable example for the above phenomenon. Most of these conflicts arise between queen and workers and among workers during reproduction as a result of favouring their own reproductive fitness. The final outcome of these conflicts could vary with the biology of each species.

A comparative analysis of two selected sociobiological factors (relatedness and colony size) on worker male production in 90 eusocial hymenopteran species is discussed in the next chapter.

Chapter 6

Effects of sociobiological factors on worker male production in eusocial Hymenoptera

6.1 Introduction

The origin of eusociality is one of the major transitions in evolution (Maynard-Smith and Szathmary, 1995). One of the key characteristics of eusociality is reproductive division of labour where some colony members are involved in reproduction and others are specialized for non-reproductive tasks such as nest maintenance, brood caring, foraging and defense (Wilson and Holldobler, 2005). The queen in social insect colonies usually holds the reproductive dominancy. However, workers are capable of laying unfertilized eggs which can develop into males if reared. As a result there are conflicts among colony members for their reproductive fate. Conflicts such as aggressive behavior, brood destruction or sometimes even matricide could affect the overall social organization in the colony (Bourke, 1988). Thus avoiding self reproductive interests of colony members and minimizing costly conflicts among them is important for a colony to function efficiently and to survive.

Worker policing or mutual prevention of reproduction among colony members is one mechanism of maintaining this balance. Worker policing is achieved via selective

removal or oophagy of worker-laid eggs and aggressive behaviour towards reproductively/ovary activated nest mates which in turn force them to regress their ovary functions (Ratnieks and Helantera, 2009).

According to the relatedness hypothesis, occurrence of worker policing behavior is determined by the relative relatedness of workers to worker produced sons versus workers to queen produced sons. In other words, parentage of males is mainly determined by the colony kin structure (Hamilton, 1964). A queen is more related to her sons ($r = 0.5$) than to grandsons ($r = 0.25$) thus the queen is selected to carry out such policing to inhibit worker fertility (Trivers and Hare, 1976). From the workers perspective, they prefer to produce males themselves because relatedness to their own sons ($r = 0.5$) is higher than relatedness to any other male. When the number of males mated with the queen increases, workers' relatedness to another workers'/sisters' son (full- and half-nephews, $r < 0.25$) decreases, but their relatedness to their brothers (queen produced sons, $r = 0.25$) remains unchanged. Therefore workers are selected to rear brothers over nephews and to police each other's reproduction (Ratnieks, 1988). Thereby the reproductive division of labour in the colony is forced towards the queen with the majority of males in the colony being queen produced sons. Policing is best studied in the honeybee (*Apis mellifera*), where 99% of all worker-laid eggs are policed by other workers and only 0.01 - 0.1% of the workers try to activate their reproduction in the presence of the queen (Wenseleers and Ratnieks, 2006a). In contrast in a monandrous mating system, where the queen mates only to a single drone, workers are more related to their sisters' sons (nephews, $r = 0.375$) than to the queen's sons (brothers, $r = 0.25$). Thus rearing a sister's son increases their inclusive fitness and theoretically worker policing should be less prevalent. Supporting this, worker reproduction has been observed to be more common in monandrous species (e.g. bumblebees and stingless bees) than in polyandrous species (Crozier and Pamilo, 1996, Ratnieks et al., 2006). However, behavioural observations suggest that worker policing in monandrous species is not as rare as expected (van Honk and Hogeweg, 1981). In addition a higher degree of cooperation among the queen and workers as seen in a polyandrous species (*Apis mellifera*) also exists in monandrous species (Ratnieks, 1988) with the parentage of the majority of males in the colony still being held by the queen.

Hammond and Keller 2004 explained these controversies stating that the male

parentage and occurrence of worker-policing in social Hymenoptera cannot be supported by kin relatedness alone. Rather the cost of worker reproduction which affects the overall colony efficiency also plays an important role in determining the parentage of males. For example, aggressive fights for reproduction and laying a higher number of eggs than that can be reared are a waste of energy and productive time for other nest tasks. Wenseleers and Ratnieks 2006a, using 90 hymenopteran species to extend the analysis of male parentage presented by Hammond and Keller 2004, showed that there is a strong effect of kin relationships on male parentage conflicts and worker policing. The relatedness difference between workers to workers' sons versus queen's sons (r_{diff}) can be used as a variable when making predictions about worker male production based on colony kin interactions (Wenseleers and Ratnieks, 2006a). According to Wenseleers and Ratnieks 2006a study the proportion of worker produced males was significantly lower (only 0.2% males in 15 species) when workers were related more to the queen's sons. This implies strong policing among workers in reproduction. In 75 species in which workers were related more to workers' sons, an average of 14% of the males were workers' sons.

However, they further state that there are significant unexplained variations in these results. For example in several species with intraspecific variation (*Myrmica tahoensis*, *Vespula rufa* and *Leptothorax acervorum*, see Appendix table A.8) when workers were more closely related to the queens sons than to other workers sons ($r_{\text{diff}}<0$), a higher percentage of males observed were worker derived offspring. This is the opposite of what would be predicted by the relatedness hypothesis.

Interestingly, the above contrasting results were observed in these species when their colonies are polygynous but not when they are monogynous. Thus these variations could be due to differences in sociobiology of eusocial Hymenoptera. In other words, the proportion of male production by workers could vary with the colony social structure comprising of factors such as the number of queens in the colony and relative relatedness among those queens. For instance, some colonies are founded by a single queen (monogyny) while there are several nestmate queens in other colonies (polygyny). Thus the queen number could vary among species and sometimes among colonies within the same species. Queens in polygynous colony could be closely related full sisters (e.g *Polybioides tabidus*, Henshaw et al. 2000 and *Polistes fuscatus*

variatus, Metcalf 1980) or may be completely unrelated to each other (colony usurpation/invasion by foreign queens; e.g. *Scaptotrigona postica*, Paxton et al. 2003 and *Leptothorax nylanderi*, Foitzik and Herbers 2001). Occasionally the existing queen can be replaced by older workers (queen supersedure). In such cases the new queens are the daughters of the previous founder queen (e.g. *Augochlorella striata*, Mueller et al. 1994 and *Lasioglossum zephyrum*, Crozier et al. 1987). Likewise differences in relatedness among nestmate queens could also affect worker reproduction capacity in the colony. In addition, queen/s can be mated to a single male (monandry) or several males (polyandry). Therefore, worker reproduction could also depend on species specific physiological aspects such as the queen mating frequency. Overall, the different degrees of kin relationships that exist among colony members could be a collective effect of all above factors.

Overall relatedness difference (r_{diff}) among members in an eusocial hymenopteran colony is a cumulative effect of both maternal and paternal components. Factors such as queen number in the colony and relatedness among those queens represent the maternal effects on offspring whereas the queen mating frequency which determines the number of fathers contributing to the progeny represents the paternal component. Wenseleers and Ratnieks 2006a provides a detailed comparative analysis of the effect of overall relatedness differences among colony members in determining the proportion of male production by workers. In this chapter my aim was to analyse how variations in maternally contributed factors and paternally contributed factors individually affect that relatedness difference and thereby to determine the proportion of males produced by workers.

A secondary aim was to integrate colony size into this analysis. Reproductive conflicts could vary with the colony size. Reproductive dominancy of workers via aggressive behaviours and pheromones may be more effective when the colony size is small. In contrast, laying eggs and probability of developing those eggs into adult males is low in a high worker policing environment such as in large colonies. Thus depending on the colony size the percentage of worker produced males in the colony could vary (Bourke, 1988). Hence the colony size i.e the number of adult workers in a matured colony was also used as another variable factor in assessing the male production of workers. This was initially done by Hammond and Keller 2004 using

50 species of eusocial Hymenoptera. In this chapter I was aim to repeat their analysis with 90 species of eusocial Hymenoptera used for the comparative analysis in Wenseleers and Ratnieks 2006a.

During the analysis, my first expectation was to analyse the correlation between percentage worker produced males (WPM) and each parental component (maternal and paternal) of the difference between a worker's relatedness to a queen's son and to another worker's son (r_{diff}). Then to assess their partial correlations to find out the independent effect of each parental component on WPM. For example, the effect of paternal factors of r_{diff} on WPM when the maternal effect is excluded etc. The next aim was to repeat the above analysis for colony size.

6.2 Methods

Ninety eusocial hymenopteran species (Appendix table A.8) given in Wenseleers and Ratnieks 2006a, Table A2, representing stingless bees, bees, wasps and ants were used for this analysis. Data analyses were performed with R package version 2.15.

Data on the following factors which affect worker male production were collected for each species of the 90 Hymenopterans (Appendix table A.8) via an extensive literature survey; queen number and relatedness among colony queens (maternal effects), queen mating frequency (paternal effect) and number of adult workers in a mature colony (colony size). Data on the percentage of worker produced males for each species were obtained from (Wenseleers and Ratnieks, 2006a), Table A2.

r_{diff} is the difference in relatedness between workers to worker produced males (r_{w}-r_{wm}) and workers to queen produced males (r_{w}-r_{qm}). If workers are more related to the queens sons than to the worker's sons then r_{diff} is negative ($r_{\text{diff}}<0$) and the majority of males in the colony are produced by the queen. In contrast, when r_{diff} is positive ($r_{\text{diff}} >0$), more males in the colony should be worker produced sons (Hammond and Keller, 2004). Therefore r_{diff} provides a good predictor variable about worker reproduction.

As a check to the accuracy of my data and calculations, I first plotted r_{diff} Vs percentage worker produced males (WPM) using Wenseleers and Ratnieks 2006a, Table A2 data (Figure 6.2).

According to Pamilo 1991, r_{diff} is given by, $r_{diff} = (r_w\text{-}r_{wm})$ - $(r_w\text{-}r_{qm})$. $r_w\text{-}r_{wm} = (2+k)/(8kNe)$, $r_w\text{-}r_{qm} = [1+G(Ne\text{-}1)]/4Ne$, where k is the effective queen mating frequency (paternal effect on WPM), Ne is the effective number of queens and G is the relatedness among those queens (maternal effect on WPM).

Therefore, $r_{diff} = (2+k)/(8kNe)$ - $[1+G(Ne\text{-}1)]/4Ne$

Factors responsible for maternal effect and paternal effect on WPM were isolated via an elaboration of the above equation (Figure 6.1). The maternal and paternal effect for each species was calculated as 1/4k and (1/8Ne)-[(GNe +G)/4Ne] respectively (Figure 6.1; Appendix table A.8).

I first carried out three simple pairwise correlations between each independent variable and WPM; Maternal effect Vs WPM, paternal effect Vs WPM and colony size Vs WPM.

Then partial correlations were calculated for the following variable combinations to identify the effect of each individual component on WPM:

a) Paternal effect Vs WPM while the maternal effect remained constant. b) Paternal effect Vs WPM while the colony size remained constant. c) Paternal effect Vs WPM while both the maternal effect and colony size remained constant. d) Maternal effect Vs WPM while the paternal effect remained constant. e) Maternal effect Vs WPM while the colony size remained constant. f) Maternal effect Vs WPM while both the paternal effect and colony size g) Colony size Vs WPM while the paternal effect remained constant. h) Colony size Vs WPM while the maternal effect remained constant. i) Colony size Vs WPM while both the paternal and maternal effects remained constant.

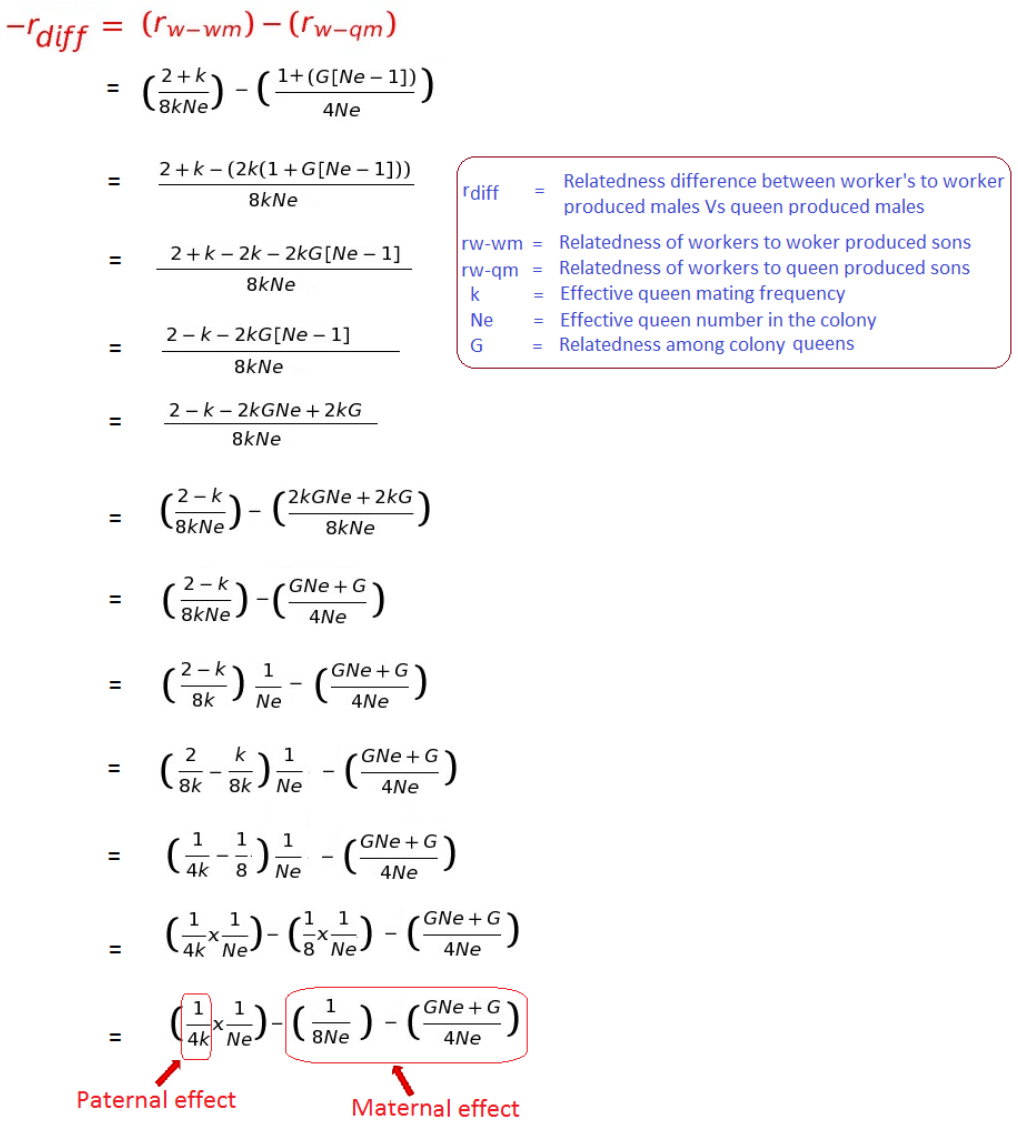

Figure 6.1: Contribution of maternal and paternal components on worker reproduction.

6.3 Results

The graph of r_{diff} Vs WPM plotted using Wenseleers and Ratnieks 2006a, Table A2 data shows an increased worker male production when $r_{\text{diff}}>0$. i.e when workers are related more to worker produced sons. A lower percentage of worker reproduction is shown when $r_{\text{diff}}<0$. i.e when workers are related more to the queen produced sons (Figure 6.2).

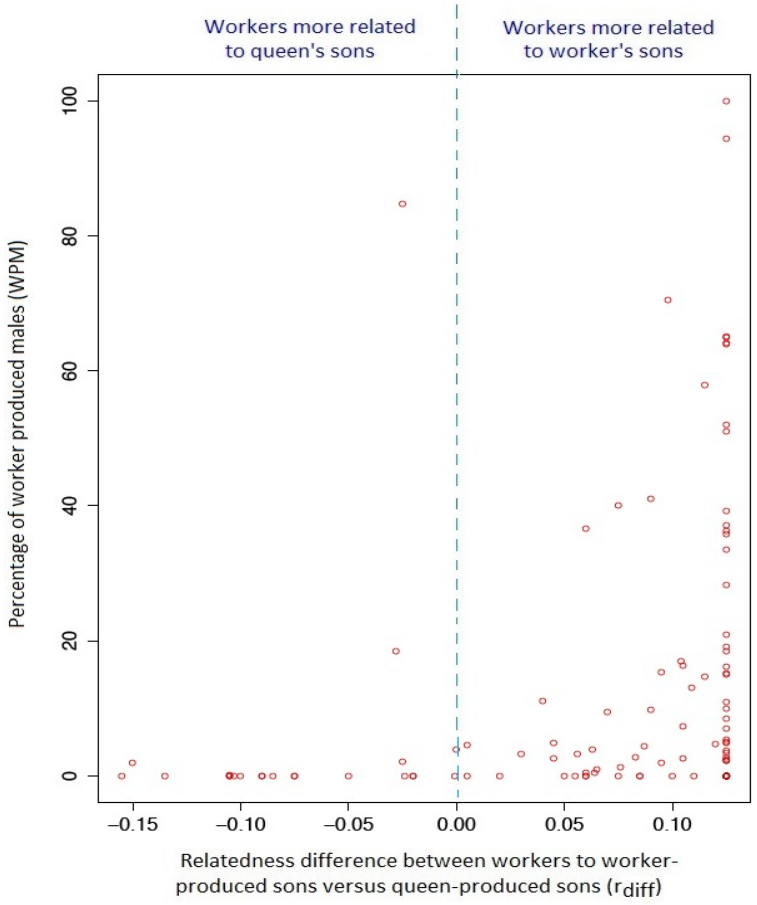

Figure 6.2: r_{diff} Vs WPM in 90 species of eusocial Hymenoptera. Proportion of worker produced males increases when workers are related more to the worker-produced males than to the queen-produced males.

6.3.1 Simple correlations

There was a strong correlation between the paternal effect and percentage male production of workers; ρ = -0.2659, n = 121, p-value = 0.0031, Spearman's rho (Paternal effect Vs WPM). Worker male production decreases with the increase of k, the queens mating frequency (Figure 6.3a). In other words, Figure 6.3a shows that the worker reproduction is rare in polyandrous colonies compared to the monandrous colonies.

There was no clear correlation between the maternal effect and worker male production; ρ = 0.0015, n = 121, p-value = 0.9866, Spearman's rho (Maternal effect Vs WPM) (Figure 6.3b).

According to Figure 6.3c, an increased worker reproduction was present in small sized colonies compared to larger colonies. It further shows that worker reproduction is nearly absent in large colonies and percentage worker reproduction reaches zero when the number of adult workers in the colony exceeded 2000 approximately. However, statistically the correlation between the colony size and worker male production didn't appear as significant; ρ = -0.1456, n = 121, p-value = 0.1112, Spearman's rho (Colony size Vs WPM).

6.3.2 Partial correlations

When the effect of maternal contribution was held constant, there was a strong negative correlation between the paternal effect and worker male production; ρ = -0.2712, n = 121, p-value = 0.0022, Spearman's rho (Paternal effect Vs WPM)$_{\text{maternal effect}}$. Factors in subscript were held constant.

When the effect of colony size was held constant, there was a strong correlation between the paternal effect and worker male production; ρ = 0.2347, n = 121, p-value = 0.0087, Spearman's rho (Paternal effect Vs WPM)$_{\text{colony size}}$.

Paternal effect and WPM show a strong correlation between them when both the maternal effect and colony size were held constant; ρ = 0.2231, df = 115, p-value = 0.0156, Spearman's rho (Paternal effect Vs WPM)$_{\text{maternal effect and colony size}}$.

There was no significant correlation between the maternal effect and worker male production either when the paternal effect held constant; ρ = -0.0553, n = 121, p-value = 0.5467, Spearman's rho (Maternal effect Vs WPM)$_{\text{paternal effect}}$ or colony size held constant ρ = -0.0254, n = 121, p-value = 0.7819, Spearman's rho (Maternal effect Vs WPM)$_{\text{colony size}}$.

Also there was no statistically significant correlation between the maternal effect and WPM when both the paternal effect and colony size were constant; ρ = -0.0175, df = 115, p-value = 0.8510, Spearman's rho (Maternal effect Vs WPM)$_{\text{paternal effect and colony size}}$.

There was no significant correlation between the colony size and worker male production when the paternal effect was held constant; ρ = -0.0689, n = 121, p-value = 0.4528, Spearman's rho (Colony size Vs WPM)$_{\text{paternal effect}}$, or maternal effect was constant; ρ = -0.1477, n = 121, p-value = 0.1045, Spearman's rho (Colony size Vs WPM)$_{\text{maternal effect}}$, or when both the maternal and paternal effects were constant; ρ = -0.0417, df = 115, p-value = 0.6547, Spearman's rho (Colony size Vs WPM)$_{\text{maternal effect and paternal effect}}$.

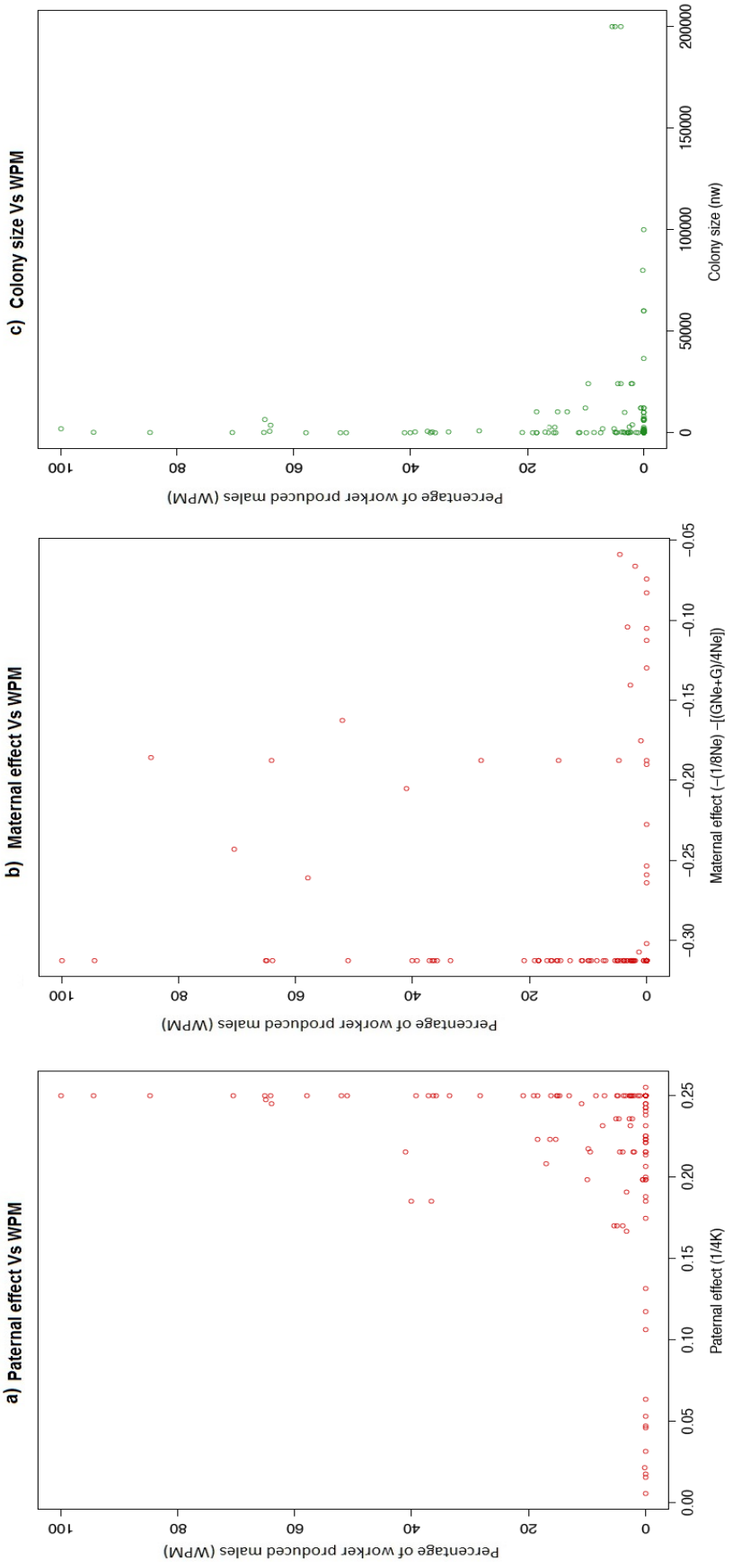

Figure 6.3: Change of percentage worker produced males with paternal effect, maternal effect and colony size.

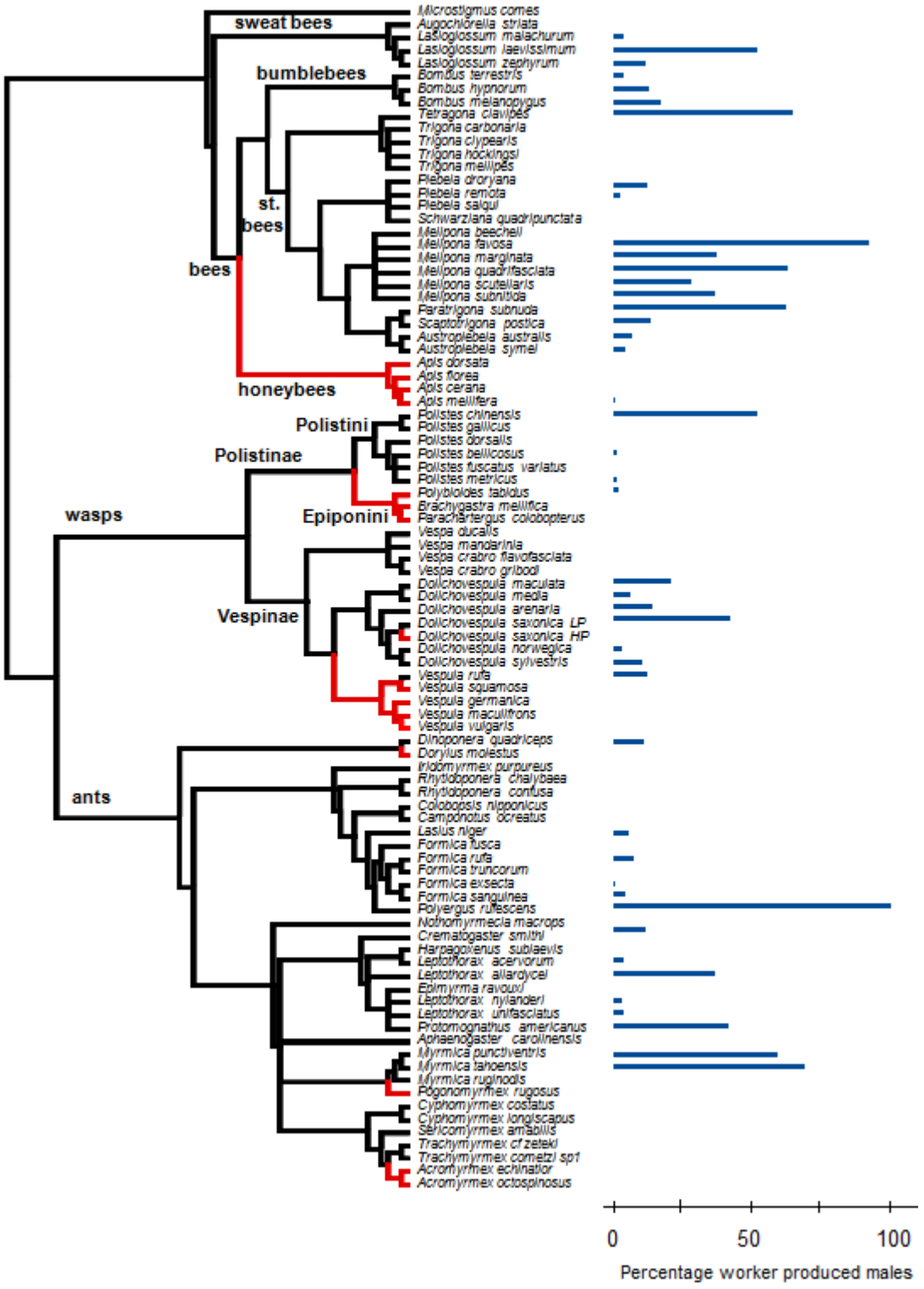

Figure 6.4: Percentage male production of workers in 90 hymenopteran species used for the comparative analysis. Branches marked in red represent $r_{\text{diff}}<0$ and a strong worker policing is predicted in these species. Branches marked in black represent $r_{\text{diff}}>0$ and a wide range of percentage worker male production is shown in these species depending on the biology of each species. Phylogenetic tree was adapted from Wenseleers and Ratnieks 2006a.

6.4 Discussion

The effect of two selected sociobiological factors on worker male production in eusocial Hymenoptera was analysed in this chapter. The relatedness difference between workers to queen's sons and another worker's son (r_{diff}) and colony size were taken into account. The overall effect of r_{diff} was assessed as two separate components; effect on r_{diff} by maternally contributed factors $r_{\text{diff}\,(maternal)}$ and paternally contributed factors $r_{\text{diff}\,(paternal)}$.

According to my results, maternal factors have no significant effect on worker male production; Figure 6.3b. This may be because most (75/90) hymenopteran species included in this analysis are monogynous (Appendix table A.8). Thus overall variation in r_{diff} due to the changes in queen number and queen-queen relatedness is low. In some polygynous species even though there are several queens occupying a colony, the majority of offspring usually belong to a single founder queen (e.g. in *Polistes spp.* 78% of the brood is produced by the alpha foundress while beta foundresses contribute only for 22%; Arevalo et al. 1998, Metcalf 1980). Therefore the contribution of maternal factors to the overall relatedness difference in such colonies is more similar to that of a monogynous colony. In addition, queens in a polygynous colony may be derived from the same mother or the founding queen thus may share the same set of maternal genes. For example, in species such as *Parachartergus spp.*, first the number of queens in the polygynous colony is reduced until a single queen and then new queens are started to be produced by the alpha queen to restore polygyny (Henshaw et al., 2000, Mueller et al., 1994). Thus their relatedness to each other's offspring cannot be changed significantly. However if these queens mated to several different males that would introduce new genes in to the colony leading to differences in relatedness among different queen derived workers and those workers to their male offspring. Thus the influence of maternal factors on worker reproduction may less obvious compared to that of paternal factors.

Variation in paternity frequency showed a significant effect on worker male production. Even when the effects of maternal contribution and colony size were removed by partial correlation, it remained as a strong correlation. As predicted, in a polyandrous colony, workers are more closely related to the mother than to each other. Thus increased worker policing and less worker reproduction is expected to be seen.

My results support the above fact showing a strong negative correlation between the queen's mating frequency (k) and worker male production (Figure 6.3a). This means that, paternal effect (i.e. 1/4k) is maximum (0.25) when the queen's mating frequency is 1 (monandrous). The highest percentage of worker male production was also observed when the paternal effect is 0.25 (Figure 6.3a).

Monandry is considered to be the ancestral state in the Hymenoptera and has evolved independently in 8 eusocial insect lineages (Hughes et al., 2008a). Sterile workers in monandrous species are capable of reversing into reproductives. Thus they can compete with the queen for reproduction and even replace her position in the colony. Occurrence of polyandry in eusocial Hymenoptera is not as common as monandry. Polyandry evolved secondarily after workers lost the ability to mate and reproduce (reproductive totipotency; Hughes et al. 2008a. As a result of this more cooperation and less conflict is ensured among colony members during reproduction. Polyandry could be an added advantage particularly to larger colonies and perennial colonies since it enhances the genetic heterogeneity among colony members. This can help the colony to resist diseases and survive (Oldroyd and Fewell, 2007). However, it could also bring cost effects on the queen such as high risk of predation, parasitism, energy expense etc.

I found no significant correlation between worker reproduction and colony size. This may be because the overall number of species which have a higher colony size in the data set is relatively low (larger colony sizes for example 10,000 $\leq$ is represented only by 11 species out of a total of 90 species used for the analysis). However, just by looking at the figure 6.3c there does appear to be a trend. That is a higher worker reproduction was observed for species with small colony size. This may be because due to the small number of workers present in small sized colonies, aggressive fights and competition for reproduction among workers is relatively low compared to larger colonies with many individuals. Thus a higher proportion of workers is capable of achieving their reproductive dominance. In contrast, larger colonies showed little or nearly absent worker reproduction (Figure 6.3c). Theoretically this can be explained in several ways: If many of the workers in such a colony started to reproduce it would be at a cost to other colony tasks. Also, increasing the number of males in the colony via worker reproduction will disrupt the balance of colony sex ratio (Ratnieks et al., 2006). A female biased sex ratio or a balance in sex ratio of both females (workers

produced by the queen) and males (produced by both the queen and workers) is essential for colony maintenance and survival (Wenseleers et al., 2013). Thus less worker reproduction is more beneficial for large colonies to function efficiently than for small colonies.

Due to this preponderance of monogyny in the data, my model with split maternal and paternal contribution does not improve on Wenseleer's original model with a single r_{diff}. It may be interesting to increase the proportion of polygynous species in the data set so that the paternal and maternal contribution may be more clearly separate.

In conclusion, variations in paternity frequency show a significant influence on worker male production. Effects on relatedness due to maternal factors or colony size are not statistically significant. However it is important to note that these conclusions were made based only on the results of partial correlations - the independent effect of each factor (i.e. either the maternal effect or paternal effect or colony size) excluding the effects other factors and the effects of phylogenetic relationships of species wasn't taken in to account (see Figure 6.4). In other words, variation in percentage worker male production could not only be due to parental (maternal and paternal) effects and colony size but also due to the evolutionary relationships that exist between species (phylogeny). Therefore the phylogenetically independent contrast analysis is required for a more accurate conclusion. Due to time constraints I did not do this. When the effect of phylogeny is removed the significant effect of paternity frequency on worker male production and the non-significant effect of maternal factors and colony size on worker male production I observed in the current analysis could likely be changed.

For example queens of the common wasp *Vespula vulgaris* are multiply mated and a high worker policing has been observed as expected (Foster and Ratnieks, 2001). However in *Vespa crabro*, although the queen shows monandry/facultative polyandry workers police each other's reproduction leading the majority of males in the colony still being queen produced (Foster et al., 2002). Hence relatedness and colony size alone cannot support these contraries and they may be best explained by a phylogenetic point of view. That is closely related taxa (such as the above two wasp species) which may utilize similar habitats and similar life histories may share similar patterns of worker male production and policing behaviour.

Chapter 7

Role of DNA methylation in behavioural phase polyphenism of the desert locust, *Schistocerca gregaria*

7.1 Introduction

Epigenetic modification of DNA by cytosine methylation has emerged as an important mechanism for tailoring behavioural phenotypes to the prevailing environmental conditions (Herb et al., 2012, Lyko et al., 2010). Genome-wide analyses of both vertebrate and invertebrate taxa indicate that the environmentally induced differences in behaviour correlate with differences in the DNA methylation landscape (Ernst et al., 2015, Kelly et al., 2012, Kucharski et al., 2008, Srinivasan and Brisson, 2012, Szyf et al., 2008). However, the place of DNA methylation in the mechanistic chain of events which leads from environmental signals to changes in the behavioural phenotype is not well understood. For example, do changes in the brain methylome affect behavioural change, or do they instead serve to consolidate changes that first arose through other mechanisms etc.

Phenotypic plasticity is particularly common in insects, a fact that has been implicated in their evolutionary success (Simpson et al. 2011; also see section 1.8). A striking example is provided by phase polyphenism in locusts. Locusts are grasshoppers (Acrididae) that respond to changes in population density by transforming between two extreme phenotypes known as the solitarious and gregarious phase (Pener and Simpson, 2009, Uvarov, 1966). The two phases differ profoundly in morphology and colouration, in endocrine, metabolic and reproductive physiology, and most importantly, in their behaviour. Solitarious locusts are cryptic and shy, and avoid other locusts; gregarious locusts are active and mobile and seek out the presence of conspecifics, causing them to aggregate in swarms that can wreck agriculture on a continental scale. Several distantly related Acridid species show phase polyphenism, with migratory locusts (*Locusta migratoria*) and desert locusts (*Schistocerca gregaria*) being amongst the most extreme and economically relevant. The sole direct driver of phase change is the presence or absence of conspecifics. Solitarious desert locusts acquire gregarious behaviour within the space of a few hours of forced crowding (nymphs: Roessingh and Simpson 1994; adults: Bouaichi et al. 1995). The converse process of solitarisation of long-term gregarious locusts is markedly slower. At least in desert locusts, phase state reflects both individual history and transgenerational epigenetic inheritance (Islam et al., 1994). Although the transgenerational effect on behaviour and colouration is largely reversible within a life time, it takes several successive generations of crowding or isolation to establish the phase extremes.

In *S. gregaria,* 1.6-1.9% of all genomic cytosines and over 3% of cytosines in genomic sequences that map on ESTs are methylated (Boerjan et al., 2011, Falckenhayn et al., 2013). These values are very high by insect standards - more than ten times higher than in the honeybee, a species where methylation is implicated in pronounced caste polyphenism (Herb et al., 2012, Lyko et al., 2010). Unlike some insects such as flies (Glastad et al., 2014a), *L. migratoria* and *S. gregaria* have retained orthologs of DNA (cytosine-5)-methyltransferase 1 (*Dnmt1*) although *Dnmt3* appears absent (Falckenhayn et al., 2013). The claim by Wang et al. 2014 that a *Dnmt3* ortholog is present in the *L. migratoria* genome appears unfounded, since blast-searching the genome with honeybee *Dnmt3* as a query does not find any convincing matches. A reduced representation bisulfite sequencing (RRBS) analysis by the same authors identified about 90 genes as differentially methylated between solitarious and gregarious 4th instar *L. migratoria* nymphs (Wang et al., 2014). Different numbers are quoted in

different parts of the paper - over '90' in the main text, '89' in the supplemental text and '90' in Table S19. These findings clearly suggest some role for DNA methylation in behavioural phase change.

Locust phase change provides a model for answering fundamental questions about the functional role of DNA methylation in behavioural plasticity. There are three distinct but not mutually exclusive scenarios for the role of DNA methylation in behavioural phase polyphenism: (a) it may be part of the effector cascade that initiates behavioural change; (b) it may underpin the consolidation of behavioural state with prolonged crowding within a locust's lifetime; (c) it may mediate transgenerational epigenetic inheritance.

To begin to address these questions, I compared DNA methylation patterns in the CNS of adult desert locusts (*S. gregaria,*) with identical parental histories, but different individual social histories, using methylation-sensitive amplified fragment length polymorphisms (MS-AFLP) analysis.

7.2 Methods

7.2.1 Locust rearing

Desert locusts (*Schistocerca gregaria* Forskål, 1775) were obtained from an established colony at the Department of Biology, University of Leicester that had been maintained under crowded conditions for many generations. Solitarious-phase locusts were produced from this gregarious stock by isolating them immediately upon hatching and rearing them in individual cages under visual, tactile and olfactory isolation from conspecifics (Roessingh et al., 1993). All locusts were maintained on an identical diet of fresh seedling wheat and dry wheat germ under a 12:12 hour photoperiod.

7.2.2 Experimental cohorts and treatments

All locusts in this study were sexually mature virgin adults sacrificed 17-21 days after the final moult. Long-term gregarious (LTG) locusts were removed from the gregarious colony as final larval instars, sexed, and set up as one all-male and one all-female cohort of 40 individuals each in separate tanks (length × width × height = $40 \times 30 \times 25$ cm^3) in the controlled environment room that also housed the solitarious locusts. All solitarious locusts were sibling offspring from a single gregarious mother (first-generation solitarious, 1GS). A total of $N = 24$ locusts were split into three treatment groups of four males and four females each ($n = 8$): (i) eight 1GS locusts that never experienced crowding (solitarious); (ii) eight long-term gregarious locusts; and (iii) eight behaviourally gregarised 1GS locusts. These behaviourally gregarious locusts were produced by placing 1GS locusts in the tank that housed the 40 LTG virgins of the same sex (i.e. four 1GS males with the 40 LTG males; and four 1GS females with the 40 LTG females). Crowding lasted for the final 24 hours before sacrifice, with unrestrained physical interaction between all locusts in the crowding tanks. Locusts were sacrificed by decapitation and immediate dissection under ice-cold saline. The brain (including the optic lobes, but not the retinae) and the chain of thoracic ganglia (prothoracic, mesothoracic and metathoracic ganglion) were dissected out and cleaned from contaminant tissue in ice-cold saline and snap-frozen on dry ice.

All locust tissue samples used for the following methylation analysis were provided by Dr. Swidbert Ott, Department of Biology, University of Leicester.

7.2.3 Preliminary experiment

To find the optimum amount of material (locust brains) required for MS-AFLP analysis, a preliminary experiment was conducted. DNA was extracted from brains of 4 male and 4 female locusts using the DNA Micro kit (section 2.2.2). The testing libraries were produced as follows; 1 male locust, 1 female locust, 2 pooled males, 2 pooled females and a combined sample of a single male and a single female brains. The original MS-AFLP protocol (section 2.6) was carried out with the above samples

for two selected Eco-HpaII/MspI primer combinations; AG-ACT and AC-ATC (see supplementary table A.1).

According to the preliminary experiment results, pooling of samples was required to obtain a sufficient amount of DNA for the analysis and pooling of locusts from both sexes was selected to represent the overall methylation in locusts belong to a particular phases.

7.2.4 Methylation analysis by MS-AFLP

Differences in DNA methylation patterns were detected by analysis of methylation-sensitive amplified fragment length polymorphism (MS-AFLP; see section 2.6) in $n = 4$ independent samples per treatment group, for a total of $N = 12$ MS-AFLP primer combinations (Figure 2.2; Supplementary table A.1). Each sample comprised the pooled brains and thoracic ganglia from one arbitrarily chosen male and one arbitrary female within the same treatment group. DNA was extracted with the QlAamp DNA Micro Kit following the manufacturers instructions (QIAGEN).

7.3 Results

Using MS-AFLP analysis, I compared the methylation differences between solitarious, behaviourally gregarious and gregarious locusts. The four possible banding patterns and their biological explanations are given in Table 7.1. One hundred and two unique bands (loci) were scored. Out of these, 99 loci were recorded as methylation-susceptible. That is they showed different digestion patterns between the two restriction enzymes, HpaII and MspI. Fifty three of these loci were recorded as polymorphic, meaning that they showed different banding patterns between individual samples (See Appendix figure A.1b).

Overall I found significant methylation differences between the treatment groups (ϕ_{ST}= 0.2086, p = 0.01). I then examined the pair-wise differences between groups. There was a significant difference in methylation between the solitarious and gregarious phases (ϕ_{ST}= 0.3952, p = 0.0303). Compared to the solitarious samples (14%

of loci unmethylated), gregarious samples were much less methylated (33% of loci unmethylated), see (HPA+/MSP+) in table 7.1.

The gregarious samples were again less methylated than the behaviourally gregarious form (ϕ_{ST}= 0.2084, p = 0.0272) (gregarious 33% of loci unmethylated compared to behaviourally gregarious 17%), see (HPA+/MSP+) in table 7.1.

The methylation status of the solitarious and behaviourally gregarious phase samples were very similar, with 14% and 17% of their loci unmethylated respectively (ϕ_{ST} = -0.1135, p = 0.9446) see Table 7.1

Table 7.1: Proportion of each banding type found in locusts of different behavioural phases.

Banding pattern	Methylation status	BG	Gr	So
HPA+/MSP+	Unmethylated	0.1667	0.3283	0.1389
HPA+/MSP-	Hemimethylation	0.1364	0.1566	0.1010
HPA-/MSP+	Methylation at internal cytosine	0.1263	0.1818	0.1288
HPA-/MSP-	Full methylation or absence of target	0.5707	0.3333	0.6313

Note: BG = Behaviourally gregarious, Gr = Gregarious, So = Solitarious. + and - = Presence of a band and absence of a band in digestion with the restriction enzyme. For example; HPA+/MSP- denotes the presence of a band in digestion with HpaII but absent when digested with MspI (See Appendix figure A.1b). HPA-/MSP- was counted as hypermethylated in MSAP.

The Principal coordinate analysis based on pairwise methylation difference showed two groupings corresponding to their physical phase polymorphism with behaviourally gregarious locusts overlapping the solitarious (Figure 7.1). The first two axes explain a total of 49.9% of the total variation.

7.4 Discussion

The main aim of this study was to determine if there was a role for DNA methylation in the behavioural phase change in desert locusts, *Schistocerca gregaria*. According to my results, 99 out of the 102 loci found were methylation-susceptible. This indicates

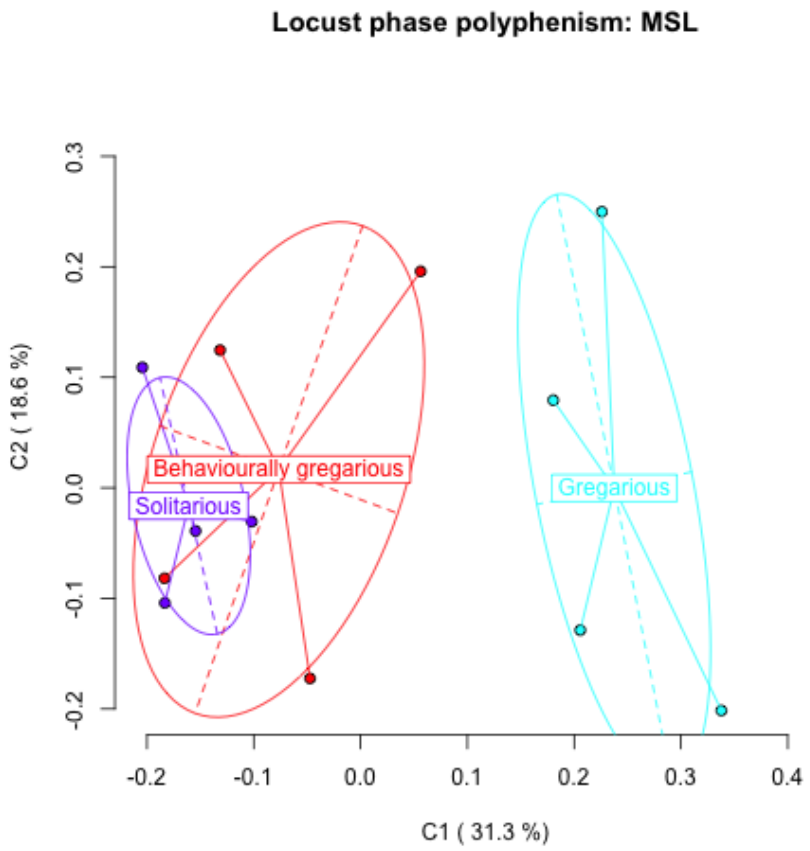

Figure 7.1: Principal coordinate analysis based on methylation status of loci for locusts in different behavioural phases

that, in the CpG sequence context, the *S. gregaria* methylome is characterized by high levels of DNA methylation.

In addition, there was a pronounced difference in the global methylation pattern between the solitarious and the gregarious phase (Figure 7.1). Our solitarious-phase locusts were the direct offspring of long-term gregarious parents. Therefore, rearing the offspring of chronically crowded locusts in social isolation leads to pronounced alterations in the global methylation pattern within an individual's life-time. We did not compare the methylation patterns between first and later generations of isolation, and we therefore cannot rule out that DNA methylation has a role in trans-generational epigenetic programming of phase state.

The most important finding of this study is that crowding adult locusts for 24 hours, a period that is sufficient for establishing gregarious behaviour (Anstey et al.,

2009), has no significant effect on their methylome. Figure 7.1 illustrates this non-significant difference as an overlap between solitarious and behaviourally gregarious locusts.

Increasing permanence of the gregarious state in locusts depends on the time spent in crowded condition (Roessingh et al., 1993, Simpson et al., 1999). DNA methylation is a possible mechanism for this consolidation. The solitarious to fully gregarious transition is a gradual process which occurs through an intermediate state of behaviourally gregarious phase. i.e. after 24 hours of crowding solitarious locusts turn into fully gregarious, in behaviour. My results indicate that this transition to fully gregarious behaviour occurs without significant changes in the DNA methylation landscape of the CNS. The pronounced difference found in the fully established gregarious phase must therefore occur at a later stage.

In nymphs of *S. gregaria*, touch by other locusts and the combination of seeing and smelling other locusts are each sufficient to initiate gregarious behaviour. For touch stimuli, the hind legs have been identified as the principal sites of gregarising sensory input (Rogers et al., 2003). Once such a stimulus is received, that information is passed to the CNS by sensory neurons. In desert locusts, serotonin has been identified as an important chemical of neuronal plasticity which induce behavioral gregarization (Anstey et al., 2009). Our results show that initiation of gregarious behaviour does not require epigenetic modification. However, the neural methylome of locusts in behaviourally gregarious state is subjected to major changes during the transition into the fully gregarious state. Therefore DNA methylation may provide a consolidation mechanism by which neurochemically mediated rapid changes in behaviour become more stable with time. Thus the overall process of transition among two contrast phases may occur due a combination of effects of neurochemical and epigenetic pathways. However, it is interesting to note that the same duration (24 hours) of isolation is not sufficient to cause appreciable solitarisation in chronically crowded locusts.

The CNS is the principal site which regulates behaviour in organisms. Another study found that in desert locusts overall brain size is 30% larger when they are in gregarious phase than to the solitarious phase. This may be because when foraging as a swarm gregarious locusts may face a higher intraspecific competition for food, thus a higher integration between the brain and other sensory organs is essential

(Ott and Rogers, 2010). Anatomical differences such as the increase of brain size could be a result of long-term, phase-dependent differential gene expression. My results suggest that fully established behavioural changes begins without any direct involvement of methylation in the CNS.

Evidence for the effect of DNA methylation in caste determination has been shown in the honeybee by knockdown studies, where depletion of Dnmt3 induces the development of queen-like phenotypes (Kucharski et al., 2008). In support the majority of differentially methylated genes in the honeybee brain have been found to be hypomethylated in queens (Foret et al., 2012). Likewise, pharmaceutical compounds and RNA interference (RNAi) have been successfully used in eusocial insects to manipulate gene expression (genes to over express, knockdown or express in a tissue specific manner) and thereby to alter their caste fate and behaviour. This includes several species of ants (*Cerapachys biroi*, *Monomorium pharaonis*, *Harpegnathos saltator*); bees (*Apis mellifera*, *Lasioglossum albipes*); wasps (*Polistes metricus*) and termites (*Zootermopsis nevadensis*) (Yan et al., 2014).

Phase transition in locusts is a reversible process (Simpson et al., 1999, Wang and Kang, 2014). This suggests that pharmacology and RNAi knockdown techniques may be able in locusts to alter the methylation status in one phase that would lead to an alternative phase. According to my results, locusts in the solitarious phase show a higher overall methylation level than in the gregarious phase (see Table 7.1). This means that DNA demethylation by feeding a chemical such as Decitabine or an RNAi knockdown experiment against Dnmt3 may be able to induce the transition from solitarious to gregarious phase. On the other hand, feeding of a methyl enriched diet to gregarious locusts may be able to trigger their transition back to solitarious phase. Therefore artificial manipulation of methylation levels may be useful to gain an insight to the exact role of methylation play in behavioural phase polyphenism of locusts. Also it would generate information useful for locust swarm prevention and sustainable management.

A set of conserved differentially methylated genes that are responsible for caste specific gene expression have been found in Formicinae (e.g. *C. floridanus*) and Ponerinae (e.g. *H. saltator*) subfamilies of ants (Bonasio et al., 2012). It would be interesting to see whether this also applies in grasshoppers: *L. migratoria* and *S. gregaria* that belong to two different subfamilies. In other words, methylation may

be targeted at a certain set of genes that are differentially expressed or alternatively spliced to produce distinct phenotypic effects between phases. By an analysis of the methylome and transcriptome in the migratory locust genome, Wang et al. 2014 revealed that long term phase differences correlate with extensive differences in DNA methylation, gene expression and alternative splicing in the CNS. They have found over 90 genes that are differentially methylated between long-term solitarious and long-term gregarious locusts. In addition they also have found 45 genes in brain samples that are differentially spliced between the two phases. So in future it would be interesting to investigate whether the genes identified by Wang et al. 2014 are also differentially methylated and/or spliced between phases in *S. gregaria*.

Methodological considerations Consistent with other insect taxa, over 90% of 5-methylcytosine residues in *S. gregaria* genome has been found in the CpG dinucleotide context (Bonasio et al., 2012, Falckenhayn et al., 2013, Lyko et al., 2010). Thus MS-AFLP, the methylation profiling technique used in my study should cover methylation status of the majority of the genome. However, cytosine methylation that is present in non-CpG sites such as CpA and CpT (Bonasio et al., 2012) is not detected by MS-AFLP. Thus bisulfite sequencing and pyrosequencing methods that detect methylated cytosines present in both CpG and non-CpG contexts (Pinney, 2014), may ensure a complete measurement of all methylcytosines present in the genome. However, the locust genome is large for an insect. The genome size of *Locusta migratoria* is 6500Mb approximately; i.e twice as long as the human genome and 36 times as the *Drosophila melanogaster* genome and 27 times as the *Apis mellifera* genome. The *S. gregaria* genome is 8550Mb in size, i.e. it is even 50% larger than that of *L. migratoria* (Camacho et al., 2014). Hence, whole-genome bisulfite sequencing would be a challenging and expensive task. Thus reduced representation bisulfite sequencing (RRBS) provides a good alternative, since it reduces the amount of nucleotides needed to be sequenced to 1% of the genome (Meissner et al., 2005) and allows genome-wide methylation profiling, minimizing the high cost and depth of sequencing.

In conclusion, this study discovered that the solitarious-gregarious fully established behavioural phase polyphenism in desert locust is associated with major epigenetic shifts in their neural methylome.

Chapter 8

General discussion

In humans, genomic imprinting plays a significant role in fetal development and diseases including several types of common cancers (Bbosa et al., 2013, Haig and Graham, 1991). However it is extremely rare. So far, out of approximately 25 000 protein encoding genes discovered in the human genome, only about 100 have been identified as imprinted (Pliushch et al., 2010). This is because the occurrence of genomic imprinting could be a gene, tissue, individual, caste or an age specific process. Sometimes even within the same gene it could vary among different isoforms (Flores et al., 2012, Li-Byarlay et al., 2013). Along with the small size of certain insect genomes (e.g. compared to the 3000Mb of human genome, *Bombus terrestris* genome size is only 249Mb) with easy manipulation of epigenetic status, an insect model could become a valuable asset to understand this phenomenon in mammals. In addition, this phenomenon also serves as the subject of a major predictive evolutionary theory, the Haig's kinship theory of genomic imprinting. Thus better understanding of Haig's theory would also provide an insight to many other important aspects of sociobiology such as social evolution and inclusive fitness.

Social insects have been recognized for more than a decade as potential candidates for the evolution of genomic imprinting. However so far no imprinted genes have been discovered in any social insect (Drewell et al., 2012) and very little work has been done on this topic. This may be due to technical limitations, lack of a proper experimental pipeline and a model organism to test this phenomenon.

The main aim of my PhD was to begin to establish a social insect model to study genomic imprinting. Although it is still to be confirmed that my experimental observations were due to genomic imprinting, my research based on the bumblebee, *Bombus terrestris* demonstrates the first evidence for the feasibility of genomic imprinting in the bumblebee.

A summary of the main findings and a brief background of research for each chapter is given below.

8.1 A summary of the results

Chapter 3: Methylation differences during development and between worker reproductive castes

Haig's kinship theory predicts that genomic imprinting in mammals has evolved during reproduction, and is associated with fetal development, via the regulation of maternal resource provision to offspring. Thus reproduction loci could be under imprinted control and also linked with development (Haig, 2000). Therefore the first aim of this thesis, addressed in chapter 3, was to check for the presence of the molecular mechanism required for genomic imprinting in *Bombus terrestris* and also to confirm that this mechanism is involved in its reproduction and development.

Methylation is common epigenetic mechanism for genomic imprinting in mammals (Gehring, 2013). It is also a widespread phenomenon in social insects (Kronforst et al., 2008, Wang et al., 2006). Using a genome-wide methylation profiling method (MS-AFLP), I found that a functional methylation system which is involved in reproductive caste formation, development and social behaviour is present in *Bombus terrestris* (Amarasinghe et al., 2014). The main findings of chapter 3 are as follows: (i) Under the queenless condition, brain methylomes of reproducing and non-reproducing worker castes are clearly different. (ii) Alteration of methylation (by Decitabine) can cause a sterile worker to turn into a reproductive worker with increased aggressive behaviour and ovary development. (iii) However, comparison of methylation profiles of bees with experimentally altered methylation (Decitabine treated) and normal condition (control) were not significantly different from each

other. This indicates that such a dramatic shift of reproductive status in Decitabine treated bees may be associated with reduction of methylation only at targeted sites in specific genes, cells or gene isoforms, but not due to a global reduction of methylation in the bee body. (iv) In a separate sub experiment of this chapter, I found that during the colony transition from the pre-competition phase to the competition phase, the brain methylome of bees changed significantly.

After confirming the presence of DNA methylation in *Bombus terrestris*, I then used two different technical approaches to identify genes that potentially be imprinted in the bumblebee genome.

Chapter 4: Allele specific expression in the bumblebee, *Bombus terrestris*: candidate gene approach

Imprinting is the differential expression of alleles based on their parent-of-origin. Using a combined approach of SSCP-qPCR on targeted candidate genes, I found differential allele specific expression in *B. terrestris* worker reproduction loci, that are apparently due to parent-of-origin of those alleles. Here I say "apparently" since the reason for the observed allele specific expression, is still left to be confirmed as genomic imprinting (see section 8.2.1). This is because the same expression pattern could be due to other genetic phenomena such as dominant-recessive effects. During this experiment, I further discovered that maternally and paternally derived alleles of *Ecdysone 20- monooxygenase-like*, a gene which plays a central role in *Bombus terrestris* reproduction (Geva et al., 2005) shows the exact expression pattern predicted by Haig's kinship theory for a monandrous queen-right colony (Queller, 2003).

Chapter 5: Monoallelic methylation and allele specific expression in the bumblebee, *Bombus terrestris*: Next generation sequencing approach

Methylation has been found to be associated with monoallelic expression in social insects (Bonasio et al., 2012). Using a genome-wide next generation sequencing approach, I found that monoallelic methylation is present in the bumblebee and it is also associated with monoallelic expression of genes. However, I cannot confirm whether it was the matrigene or patrigene that was expressed monoallelically, since I did not know the parental genotypes of the worker bee used for this experiment.

Overall, the genes that expressed in an allele specific manner in chapter 4 (*Edy20-like*, *IMPL2-like* and *Mblk1-like*) and the 19 genes that show monoallelic methylation and monoallelic expression in chapter 5 (see table 5.2) showed the exact predicted characteristics (functional role, expression pattern, link with methylation etc.; see section 5.4) of mammalian imprinted genes. Thus they would serve as important candidates for future research on genomic imprinting in social insects. Then, the next important step which I couldn't cover in this thesis is to conduct reciprocal crosses to confirm the presence of genomic imprinting. A combined approach of reciprocal crosses with genetically distant bumblebee colonies and RNA-seq analysis, to measure the expression of the matrigenic and patrigenic alleles in the reciprocally crossed bees would provide an independent validation method for this purpose (see section 8.2.1).

Finding genomic imprinting in social insects would be a major advance in evolutionary biology for both current and future research. I conclude that *Bombus terrestris* is a valuable test model for further research of genomic imprinting due to the following findings of my PhD: (i) Presence of the main molecular mechanism (DNA methylation) of genomic imprinting; (ii) Presence of monoallelic methylation, the most common methylation pattern found in mammalian imprinted genes (Steyaert, 2014); (iii) Presence of expected gene expression patterns (monoallelic expression and apparent parent-of-origin allele specific expression) that are in accordance with the predictions made by Haig's kinship theory, for different social contexts (e.g. queen-right monandrous etc.), in predicted body tissues (e.g. egg and sperm gametes) and physiological processes (e.g. development, reproduction, oncogenesis). In addition, DNA methylation can be considered as a key component of a complex, integrated epigenetic network in social insect genomes.

Chapter 6: Effects of sociobiological factors on worker male production in eusocial Hymenoptera

As described above, the main aim of my PhD was to find the parent-of-origin specific gene expression in *Bombus terrestris*, using worker reproduction as a test. In reproduction, workers are capable of laying unfertilized eggs which can develop into males if reared (Duchateau and Velthuis, 1988).

In previous chapters I studied how reproduction in a worker is controlled by the maternally and paternally derived components of it's genome. In this chapter I addressed what sociobiological factors contribute to those maternally and paternally derived components and how worker male production is affected by each parental component individually. Using 90 eusocial hymenopteran species Wenseleers and Ratnieks 2006a, showed that there is a strong effect of relatedness difference on male parentage conflicts and worker policing. I analysed the overall relatedness difference in their analysis as two separate factors; maternal r_{diff} and paternal r_{diff}. As an additional factor colony size was also incorporated in to this study.

My results showed that the relatedness difference derived due to paternity/queen mating frequency has a significant influence on worker male production. However, the r_{diff} derived due to maternal factors (queen number in a colony and relatedness among them) and the colony size didn't show any statistically significant effects on worker male production.

Chapter 7: Role of DNA methylation in behavioural phase polyphenism of the desert locust, *Schistocerca gregaria*

Reproductive caste formation and division of labour are among the key characteristics of eusociality. DNA methylation is a common epigenetic modification which allows them this phenotypic plasticity. As the extraordinary ecological success of eusocial Hymenoptera depends on their caste specific reproductive morphs, so too do locusts use extreme phase dependent phenotypes in order to respond to changing environmental conditions.

Wang et al. 2014 reports about differentially methylated genes among solitarious and gregarious phases in the migratory locust, *Locusta migratoria*.

However the current study is the first evidence that the phase-related behavioural polyphenism between solitarious and gregarious phases in the desert locust, *Schistocerca gregaria* are associated with major epigenetic shifts in their neural methylome. The solitarious to the permanent gregarious transition is a gradual process which depends on the time spent in crowded condition. I found that the behaviourally gregarious phase is an intermediate state of solitarious and gregarious phases. i.e. morphology wise and neural methylome wise they are similar to the solitarious phase

but behaviourally similar to the gregarious phase. This indicates that the transition of solitarious to fully gregarious behaviour occurs without significant changes in the DNA methylation landscape of the CNS. The pronounced differences between the fully established phases must therefore occur at a later stage.

8.2 Future implications

Overall content discussed in this thesis is comprised of three main subjects areas: genomic imprinting, DNA methylation, worker reproduction in eusocial Hymenoptera. Future directions and further work for each area is discussed below.

8.2.1 Further work on genomic imprinting

Reciprocal crosses are recommended as future work to confirm that the gene expression patterns observed in chapter 4 and 5 were due to genomic imprinting.

Use of reciprocal crosses to differentiate patrigenes and matrigenes

The first step of this experiment is to produce several pairs of reciprocally crossed colonies. A pair of reciprocal colonies is where a male and a queen from one colony are mated reciprocally to a queen and a male from another colony to produce a pair of complementary descendant colonies in which bees have reciprocal matrigenes and patrigenes. For example a reciprocal cross between strain A queen x strain B drone versus strain B queen x strain A drone would produce a genetically identical progeny, AB (Table 8.1). However when the allele-specific gene expression is measured by RNA-seq, if only a single allele (either allele A or allele B) is expressed in the resultant F1 progeny that would confirm the parent-of-origin allele specific expression. In other words, that is genomic imprinting. Likewise Table 8.1 illustrates, 6 possible reciprocal crosses based on different genotype combinations that the parental bees could occupy. However, if the descendants of a reciprocally crossed colonies are homozygous (e.g. genotype AA), distinguishing maternal and paternal alleles would be unfeasible. Therefore production of heterozygous reciprocal colonies is important.

Table 8.1: Outline of all the possible crosses that may occur at any given locus.

No.	Cross type	Mother	Father	Offspring	Matrigene	Patrigene
1	Different homozygotes (AAxBB)	AA	B	AB	A	B
		BB	A	BA	B	A
2	Different heterozygotes (ABxCD)	AB	C	AC or BC	A or B	C
		AB	D	AD or BD	A or B	D
		CD	A	CA or DA	C or D	A
		CD	B	CB or DB	C or D	B
3	Homozygote x different heterozygote (AAxCD)	AA	C	AC	A	C
		AA	D	AD	A	D
		CD	A	CA or DA	C or D	A
4	Homozygote x same heterozygote (AA x AB)	AA	A	AA	A	A
		AA	B	AB	A	B
		AB	A	AA or BA	A or B	A
5	Same heterozygotes (ABxAB)	AB	A	AA or BA	A or B	A
		AB	B	AB or BB	A or B	B
6	Same homozygotes (AAxAA)	AA	AA	AA	A	A

Note: Where there are multiple rows per cross type, this represents the reciprocal crosses with the colonies swapping who provides the males or the females. The dark grey boxes show crosses where there are identical patrigenes and matrigenes so these loci would not be useful for genomic imprinting analysis. In the light grey crosses the heterozygote offspring would still be useful as we will know what the patrigene is. All other crosses provide unambiguous information.

Production of reciprocally crossed colonies As discussed in chapter 5, using MeDIP-seq and MRE-seq, we found 593 monoallelically methylated sites that were heterozygous in one randomly selected bee. This suggests that any two colonies reciprocally mated will produce offspring that are heterozygous at many loci. Selection of geographically distant colonies to produce virgin queens and drones required for the reciprocal crosses will further maximize the probability of the parental generation possessing different alleles at any given locus. Thus it will increase the chance of any given locus being heterozygous in the descendant colonies as well. Choosing colonies from disparate locations will also ensure less inbreeding among populations which lead the colonies to be homozygous.

Removal of the queen from a colony after the start of the third brood (this is roughly on the eighth day) will stimulate the emergence of new queens from the existing brood (Pereboom et al., 2005). Similarly housing callow workers together in groups of 5, in the absence of the queen initiate them laying drone/male destined eggs

(Amsalem and Hefetz, 2011). These newly emerged queens (8 day old) and drones (6-9 day old) can be crossed by placing them in a net enclosure (Amin et al., 2012). Once mated, queens will be placed in hibernation at 4°C for 45 days (Gosterit and Gurel, 2009). Males will be snap frozen and stored at -80°C to be used for sequencing later, to confirm the paternal alleles of the F1 progeny. After hibernation, a queen will be induced to begin a colony by anaesthetizing with carbon dioxide for thirty minutes and placing in a starting box with a tagged callow worker from a donor colony. Standard bee rearing conditions as described in section 2.1 will be provided. Once a colony (i.e. the F1 progeny) is established (>20 workers), five day old workers will be used for RNA-seq analysis to find out which parental allele is expressed. Sequencing of the queen's RNA, will confirm the maternally expressed alleles. Since we know the genotype of both the queen and drone, at a particular locus, we can identify which parental allele is expressed in the RNA-seq library.

Several recent papers also describe the application of RNA-seq to search for novel imprinted genes in mammals (DeVeale et al., 2012, Gregg et al., 2010, Okae et al., 2012, Wang and Clark, 2014) and flowering plants (Gehring, 2013).

Due to the high cost of sequencing and also as the first attempt of testing our next generation sequencing experimental setup and bioinformatic pipeline, my thesis only includes data of whole-genome sequencing of a single worker bee (chapter 5). However with the above RNA-seq analysis and our bioinformatic pipeline this could be expanded into several reciprocally crossed bees. Thus apart from understanding the expression patterns discovered in my thesis, it would also help to find novel imprinted genes in the *B. terrestris* genome that may express in an individual specific manner.

8.2.2 DNA methylome analysis

Although DNA methylation is mitotically heritable and stable over time, patterns of methylation can change in response to cell differentiation, environmental influences and disease (Bonasio et al., 2010a). Also depending on the extent of methylation present, expression of a gene could be in a range between full expression to complete silence. Therefore accurate mapping of DNA methylation is vital in understanding

the precise relationship between gene expression and physiological conditions and thereby for treatments (cancer therapies; ageing Weidner et al. 2014) and further research (evolutionary biology, epigenetics, ecotoxicological studies Vandegehuchte and Janssen 2011 etc.) purposes.

During this thesis I have used three basic types of methylation analysis techniques: bisulfite treatment; methylation sensitive digestion of DNA with restriction enzymes (MS-AFLP, MSRE, AIMS); and DNA immunoprecipitation with antibodies (MeDIP) combined with next generation sequencing. Reduced representation bisulfite sequencing (RRBS) provides a combined approach of all above techniques; bisulfite treatment, restriction enzymes and next generation sequencing. Thus it is a more sophisticated genome-wide methylation profiling method compared to MS-AFLP and also can be used as an alternative to MeDIP/MRE/RNA-seq analysis discussed in chapter five.

During RRBS, DNA is digested with the restriction enzyme, MspI which cleaves DNA at CCGG sites, the CpG enriched regions in the genome. The digested DNA fragments are ligated to specific adapters and DNA libraries with appropriate length fragments are produced. Then, they are bisulfite treated, amplify with suitable primers which are complementary to the adapter sequences and sequenced on an Illumina platform. Thus it is capable of producing a reduced representative sample of the whole genome that is enriched for CpGs, reducing the complexity of sequencing results (Gu et al., 2011). Thereby genes that are differentially methylated in a genome can be identified. These data can also be linked with gene expression data (RNA-seq) to find out which gene is upregulated or downregulated at specific worker reproduction loci.

There are many modern techniques which are used to interrogate genomic and epigenomic data (Bock et al., 2010, Harris et al., 2010). Among them bisulfite-based methods are considered the gold standard for a quantitative methylation analysis. However a qualitative analysis about different methylation statuses of each cytosine (mCs/methyl cytosines; hmCs/hydroxymethyl cytosines; and mCG/methyl cytosine guanine dinucleotides) that a full methylome is comprised of, cannot be distinguished even by bisulfite based methods (Beck, 2010). This is because after bisulfite conversion, 5hmC is not deaminated to thymine, but converted to cytosine

5-methylenesulfonate, which is read as cytosine during sequencing. Thus methylated cytosine and 5-hydroxymethylcytosine cannot be discriminated (Krueger et al., 2012).

Nanopore (Clarke et al., 2009) and single molecule real-time sequencing (SMRT), where epigenetic and other base modification data are gathered in the sequencing process, as an integral part of the sequencing workflow are more promising techniques for both quantitative and qualitative analyses of full methylomes. These newer NGS platforms have numerous advantageous including longer read length, less bias during template preparation, better accuracy and higher speed (Flusberg et al., 2010).

8.2.3 Worker reproduction in eusocial Hymenoptera

Although a higher level of cooperation and less competition are essential components of eusociality each individual favours its own reproductive fitness, over the others. In advanced eusocial Hymenoptera maintaining the cooperation between colony tasks, is believed to be driven by strict policing among nest mates (Wenseleers and Ratnieks, 2006b). Another idea is that high relatedness is essential to begin reproductive behaviours in workers (Hamilton, 1964).

My study was only focused on two potential factors that could affect worker male production: i.e. the relatedness asymmetries derived due to different maternal and paternal effects and colony size. However, the sociobiology of Hymenoptera is complex and there are many other factors which would be equally important and interconnected with each other in determining the exact proportion of worker produced males in a colony. For example, queen policing, worker matricide, colony sex ratio are some of those factors (Ratnieks et al., 2006).

Queen policing is more effective in small colonies such as in bumblebees while it is less practical to police all worker reproduction in larger colonies (Duchateau, 1989); Worker matricide, i.e. the killing of the queen by workers towards the end of the colony life cycle, greatly enhances egg laying of workers. In the bumblebee *Bombus terrestris*, the queen removes nearly all worker-laid eggs (Duchateau and Velthuis, 1988). Worker matricide enables all these eggs to be raised to adulthood, increasing the percentage male number in the colony: Maintaining a balanced colony sex ratio

is another important factor for a colony to survive. In Vespidae wasps both male and female offspring are reared in cells Wenseleers and Ratnieks 2006a. Thus eliminating a worker-laid egg from a cell would increase the chance of a queen's offspring to rear in that cell. Likewise worker male production is closely associated with several other aspects of eusociality. Therefore further work in combining above factors to the current study is recommended.

8.2.4 Relationship between ploidy level, body size and mating systems

Haplodiploidy can influence several aspects of male development in eusocial Hymenoptera. It leads to numerous kin relationship asymmetries among colony members and thereby determines the percentage male production in the colony (Hamilton, 1964). Also there is evidence for the effects of different ploidy levels in determining drone size in eusocial Hymenoptera. For example in polyandrous species such as *Apis mellifera* (Woyke, 1986) and *Solenopsis invicta* (Ross and Fletcher, 1985), male body size has been observed to increase with increased ploidy level. In other words, body size of a triploid (3n) male is bigger than that of a diploid (2n) and a haploid (n) male. According to the Haig's kinship theory, in a polyandrous mating system, a patrigene acts more selfishly than the matrigene (because several patrigenes compete with each other over maternal resources). In triploids with two sets of patrigenes, this effect is higher than diploids with one set of paternal chromosomes, than haploids with zero paternal contribution. As a result triploid size is bigger than diploids and haploids.

Conversely in a monandrous mating system the patrigene acts altruistically. Thus the opposite effect should be observed in a polyandrous mating system. i.e. increased ploidy level should be associated with a decrease in male body size. *Bombus terrestris* provides an ideal example for this showing the body size of a triploid (3n) male is smaller than that of a diploid (2n) and a haploid (n) (Figure 1.2b; Ayabe et al. 2004). In *Bombus atratus* diploid males show smaller testes size than haploids (Kerr, 1974).

Glastad et al. 2014b strengthened this idea in epigenetic point of view and provide the first evidence for the relationship between ploidy level and DNA methylation. They have found similar DNA methylation levels in individuals with similar ploidy levels and also have found a positive correlation between DNA methylation and gene expression level. Thus they suggest that DNA methylation may act as a compensation mechanism for the gene expression differences derived due to differing ploidy levels in individuals. In other words, differences in DNA methylation may help to maintain the appropriate gene expression patterns in individuals with different genome copy numbers.

Other research suggests that imprinting could be evolved in males to reduce their body size and thereby to minimize colony resource intake. This is because they replace the working caste in the colony but like workers, they don't contribute to colony tasks. Therefore they will be a considerable cost to colony productivity. However much less information is available on this topic and much of it are only theoretical postulations. Therefore the relationship between ploidy level, mating system and male body size would also be an interesting topic for future research.

The overall conclusions and contributions that my thesis has made to the field of insect epigenetics and genomic imprinting are summarized in the next section.

Chapter 9

Conclusions and contributions

My PhD discovered the following as a basis for an insect model of epigenetics and genomic imprinting.

(a) The presence of monoallelic methylation in the bumblebee, *Bombus terrestris*.

(b) The link between monoallelic methylation with monoallelic gene expression in *Bombus terrestris*.

(c) The presence of apparent parent-of-origin allele specific expression in loci associated with *Bombus terrestris* worker reproduction.

(d) The presence of a functional methylation system which is involved in reproductive caste formation and social behaviour in *Bombus terrestris*.

(e) The occurrence of a clear methylation difference in the brain *Bombus terrestris* workers during the colony transition from pre-competition to competition phase.

(f) The significant impact of variation in paternity i.e. the variation derived due to queen mating frequency, on worker male production in eusocial Hymenoptera.

(g) The first evidence that the phase-related behavioural polyphenism in the desert locust, *Schistocerca gregaria* is mediated by pronounced differences in the DNA methylation landscape of their CNS.

Appendix

Table A.1: Sequences of ligation adapters, pre and selective amplification primers.

Adapter / Primer	Sequence (5' - 3')
Ligation	
EcoRI-F	CTCGTAGACTGCGTACC
EcoRI-R	AATTGGTACGCAGTCTAC
HpaII-MspI-F	GACGATGAGTCTAGAA
HpaII-MspI-R	CGTTCTAGACTCATC
Pre-amplification	
EcoRIpre (EcoRI + 0)	GACTGCGTACCAATTC
HpaII-MspI pre (HpaII-MspI + A)	GATGAGTCTAGAACGGA
Selective amplification	
Eco-AA	GACTGCGTACCAATTCAA
Eco-AT	GACTGCGTACCAATTCAT
Eco-AG	GACTGCGTACCAATTCAG
Eco-AC	GACTGCGTACCAATTCAC
HpaII-MspI-AAT	GATGAGTCTAGAACGGAAT
HpaII-MspI-ACT	GATGAGTCTAGAACGGACT
HpaII-MspI-ATC	GATGAGTCTAGAACGGATC

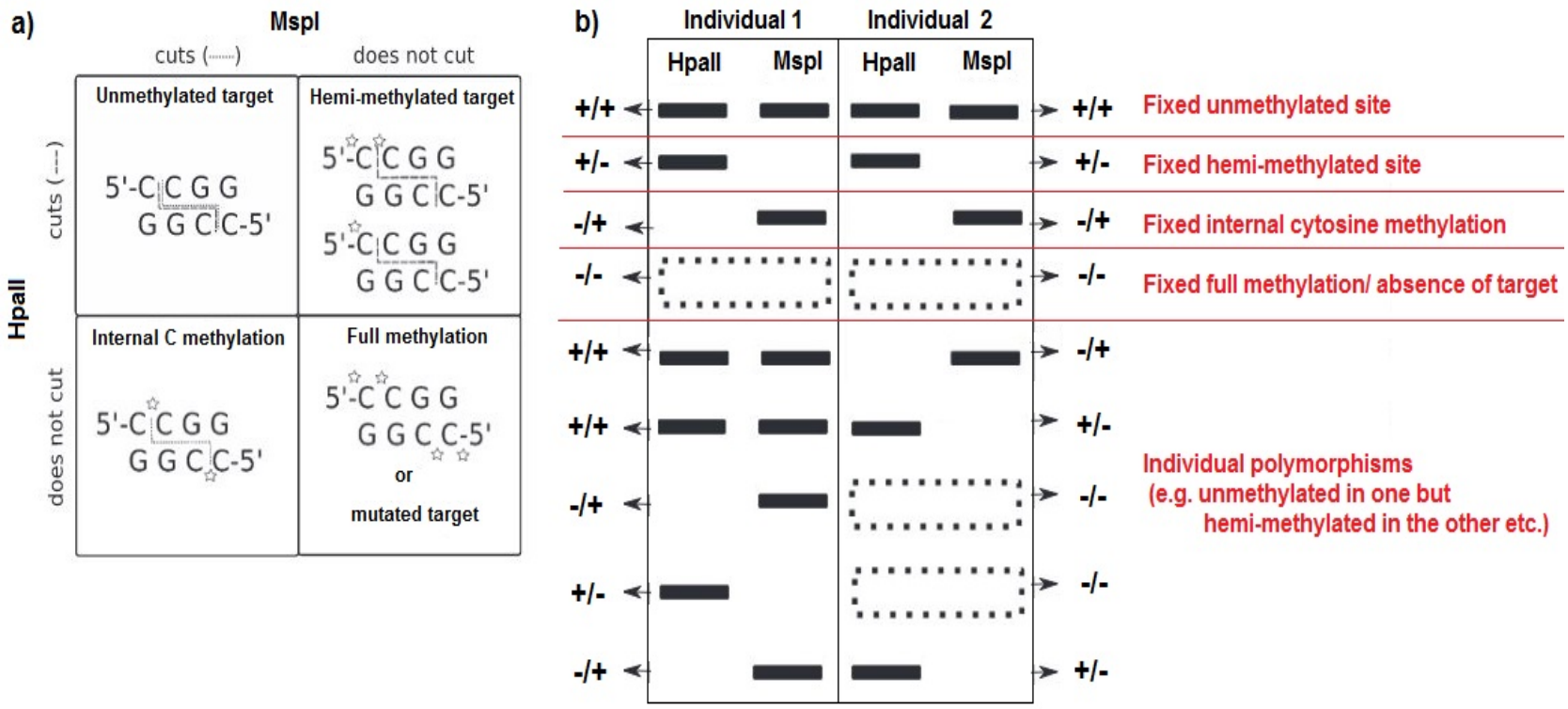

Figure A.1: a) Differential methylation sensitivities of HpaII and MspI at the CCGG target site. Stars indicate methylation at cytosines. Dash lines represent the cleavage pattern. HpaII cleave at unmethylated sites and hemi-methylated (i.e. only one strand is methylated) sites. MspI cleaves at unmethylated sites and at internal cytosine methylated sites. Neither of the enzymes cleave if the restriction site is fully methylated (i.e. both strands are methylated) at both external and internal cytosines. Adapted from Pérez-Figueroa 2013. b) Comparison of MS-AFLP profiles of two individual samples. Depending on the presence (+) or absence (-) of a fragment during the gel electrophoresis, AFLP banding patterns can be interpreted as follows;

A fragment that is present with both restriction enzymes indicates an unmethylated site (HpaII+/MspI+). A fragment that is present with only HpaII indicates a hemi-methylated site (HpaII+/MspI-). A fragment that is present with only MspI indicates a site which is methylated at its internal cytosine (HpaII+/MspI-). Absence of a fragment with both HpaII and MspI indicates (showed by dashed lines) a fully methylated site or absence of the target site due to a mutation (HpaII-/MspI-). Certain banding patterns could be arise due to individual polymorphisms. i.e. a fragment could be present with one enzyme in one individual but with the other enzyme in the other individual. For example, a banding pattern such as (HpaII-/MspI+) in individual 1 and (HpaII+/MspI-) in individual 2 indicates that, the restriction site is methylated in individual 1, but hemi-methylated in individual 2 etc (Pérez-Figueroa, 2013).

Table A.2: Components used to prepare 40x TAE stock solution.

Components	**Amount used to prepare 1 litre**
Trishydroxymethylaminomethane	145.37 g
$Na_2EDTA.2H_20$ (Diaminoethanetetra-acetic acid disodium salt)	11.16 g
Glacial acetic acid	34.4 ml
Distilled water	965 ml

Note: 50 ml of 40x TAE was then diluted with 2 liters of dH_20 prepare 30mM TAE.

Table A.3: Results of bee dissections.

Queenless box workers						**Queen-right colony workers**	
Box number and bee number	Measurements					Bee number	Presence of ovaries
	largest oocyte length (mm)		Mean oocyte length (mm)	Wet weight of both ovaries (g)	Weight of brood (g)		
	Left ovary	Right ovary					
Box1-bee1	-	-			2.3037	Bee1	-
Box1-2	-	-				Bee2	-
Box1-3	-	-				Bee3	-
Box1-4	0.9	1.0	0.95	0.0075		Bee4	-
Box1-5	-	-				Bee5	-
Box2-1	1.0	1.0	1.0	0.0225	1.1460	Bee1	-
Box2-2	-	-				Bee2	-
Box2-3	-	-				Bee3	-
Box2-4	0.5	0.4	0.45	0.0016		Bee4	-
Box2-5	-	-				Bee5	-
Box3-1	-	-			1.2732	Bee1	-
Box3-2	-	-				Bee2	-
Box3-3	0.7	0.9	0.8	0.0450		Bee3	-
Box3-4	-	-				Bee4	-
Box3-5	-	-				Bee5	-
Box4-1	1.0	0.9	0.95	0.0056	1.1185	Bee1	-
Box4-2	1.0	1.2	1.1	0.0440		Bee2	-
Box4-3	1.0	0.8	0.9	0.0100		Bee3	-
Box4-4	1.2	1.1	1.15	0.0326		Bee4	-
Box4-5	1.3	1.2	1.25	0.0173		Bee5	-

Note: '-' = no discernible ovaries. Yellow, red, white, and green tags were used to differentiate bees by age. For example; 5 bees in box 1 were similar in age with the yellow colour tagged bees. Box 1,2,3 = collected during the pre competition phase, Box 4 = collected during the competition phase.

Table A.4: Comparison of ovary sizes among Decitabine treated and control samples (Preliminary experiment 1)

Bee number	Total weight of both ovaries (g)	
	Decitabine treated group	Control group
1	0.0815	0.1441
2	0.0691	0.0826
3	0.0327	0.0201
4	0.0449	0.0616
5	0.0954	0.0932
6	0.0208	0.0886
7	0.0698	0.0405
8	0.0235	0.0219
9	No discernible ovaries	0.0388
10	No discernible ovaries	No discernible ovaries

Table A.5: Comparison of ovary development among Decitabine treated and control samples of adult workers (Preliminary experiment - 2).

Decitabine treated group							Control						
Box no.	Bee no.	Largest oocyte length (mm)		Mean oocyte length (mm)	Both ovary weight (g)	Weight of brood (g)	Box no.	Bee no.	Largest oocyte length (mm)		Mean oocyte length (mm)	Both ovary weight (g)	Weight of brood (g)
		left ovary	right ovary						left ovary	right ovary			
1	1	-	-			0.0076	1	1	-	-			-
	2	-	-					2	-	-			
	3	-	-					3	-	-			
	4	-	-					4	-	-			
	5	1.20	0.70	0.95	0.0144			5	-	-			
2	1	-	-			-	2	1	3.20	2.80	3.00	0.0369	5.2389
	2	-	-					2	-	-			
	3	-	-					3	-	-			
	4	-	-					4	1.80	2.10	1.95	0.0315	
	5	-	-					5	-	-			
3	1	-	-			3.5350	3	1	-	-			6.2992
	2	0.70	1.00	0.85	0.0467			2	2.30	2.10	2.20	0.0438	
	3	0.80	0.80	0.80	0.0581			3	-	-			
	4	0.30	0.10	0.20	0.0071			4	-	-			
	5	-	-					5	-	-			
4	1	1.10	1.00	1.05	0.0178	4.1867	4	1	2.10	2.00	2.05	0.0930	2.2057
	2	-	-					2	-	-			
	3	0.50	0.40	0.45	0.0107			3	-	-			
	4	-	-					4	-	-			
	5	0.70	0.90	0.80	0.0380			5	-	-			

Note: ‘ - ’denotes the absence of discernible ovaries.

Table A.6: Comparison of ovary development among Decitabine treated and control samples of callow workers (Callow worker experiment).

Decitabine treated group							Control					
Box no.	Bee no.	Largest oocyte length (mm)		Mean oocyte length (mm)	Both ovary weight (g)	Weight of brood (g)	Box no.	Bee no.	Largest oocyte length (mm)		Mean oocyte length (mm)	Both ovary weight (g)	Weight of brood (g)
		left ovary	right ovary						left ovary	right ovary			
1	1	2.5	2.0	2.25	0.0324	0.5145	1	1	-	-			3.4062
	2	3.0	-	1.50	0.0423			2	1.2	0.8	1.00	0.0095	
	3	-	-					3	0.2	0.3	0.25	0.0025	
	4	-	-					4	1.3	-	0.65	1.0170	
	5	0.5	0.3	0.40	0.0012			5	2.3	2.5	2.40	0.0823	
2	1	1.8	1.4	1.60	0.0287	1.9389	2	1	-	-			0.9127
	2	2.1	2.0	2.05	0.0264			2	1.8	2.0	1.90	0.0175	
	3	1.5	1.7	1.60	0.0103			3	-	-			
	4	1.8	1.0	1.40	0.0081			4	-	-			
	5	1.2	1.5	1.35	0.0176			5	-	-			
3	1	1.9	1.4	1.65	0.0225	3.2527	3	1	2.0	1.8	1.90	0.0136	0.8886
	2	0.4	0.6	0.50	0.0006			2	0.3	0.2	0.25	0.0097	
	3	1.0	0.9	0.95	0.0077			3	-	-			
	4	1.2	1.0	1.10	0.0262			4	1.0	0.9	0.95	0.0074	
	5	-	-					5	1.9	2.0	1.95	0.0165	
4	1	1.2	1.8	1.50	0.0212	1.872	4	1	-	-			0.1898
	2	1.3	1.0	1.15	0.0370			2	-	-			
	3	0.3	0.4	0.35	0.0007			3	-	-			
	4	1.0	1.2	1.10	0.0147			4	1	0.9	0.95	0.0116	
	5	1.0	0.7	0.85	0.0120			5	-	-			

Note: ' - 'denotes the absence of discernible ovaries.

Table A.7: Number of methylated CpG sites present in Decitabine treated (Aza) and non-treated control (C) sample clones.

Clone	Methylation status of cytosine at various CpG sites									Total Number of Methylated CpG sites in the insert region
	CpG1 (4386bp)	CpG2 (4710bp)	CpG3 (4721bp)	CpG4 (4780bp)	CpG5 (4783bp)	CpG6 (4821bp)	CpG7 (4826bp)	CpG8 (4832bp)	CpG9 (4841bp)	
Aza 1.1	M	M	N	U	U	N	U	N	N	3
Aza 1.2	M	M	N	U	U	U	U	U	U	2
Aza 1.3	Unsuccessful cloning									
Aza 2.1	Unsuccessful cloning									
Aza 2.2	M	M	M	U	U	U	U	U	M	4
Aza 2.3	M	M	M	U	U	U	U	U	M	4
Aza 3.1	M	M	M	N	U	U	U	U	U	3
Aza 3.2	N	N	M	N	N	U	U	U	M	1
Aza 3.3	M	M	M	U	M	N	U	N	M	5
Aza 4.1	M	M	M	U	M	N	U	N	M	5
Aza 4.2	M	M	M	U	M	N	U	N	M	5
Aza 4.3	M	M	M	U	M	U	U	U	U	4
Aza 5.1	M	M	M	N	M	U	U	N	M	5
Aza 5.2	M	M	M	U	U	U	U	U	U	3
Aza 5.3	M	M	M	N	N	U	M	U	N	4
Aza 6.1	M	M	M	U	M	N	U	N	N	4
Aza 6.2	M	M	M	U	U	U	M	U	U	4
Aza 6.3	M	M	M	U	U	U	U	U	M	4
Aza 7.1	M	M	M	U	M	U	U	U	M	5
Aza 7.2	M	M	N	U	M	N	U	N	N	3
Total no. of ^{M}C	17	17	15	0	8	0	2	0	9	68

Table A.7: continued

Clone	Methylation status of cytosine at various CpG sites									Total Number of Methylated CpG sites in the insert region
	CpG1 (4386bp)	CpG2 (4710bp)	CpG3 (4721bp)	CpG4 (4780bp)	CpG5 (4783bp)	CpG6 (4821bp)	CpG7 (4826bp)	CpG8 (4832bp)	CpG9 (4841bp)	
Control 1.1	Unsuccessful cloning									
C 1.2	M	M	N	N	U	N	N	M	U	3
C 1.3	N	M	M	N	N	U	U	U	U	2
C 2.1	M	M	M	U	M	N	U	N	M	5
C 2.2	Unsuccessful cloning									
C 2.3	M	M	M	N	M	U	U	N	M	5
C 3.1	M	M	M	U	U	U	U	U	U	3
C 3.2	M	M	M	U	U	U	U	M	U	4
C 3.3	N	M	N	U	M	U	N	M	M	4
C 4.1	U	M	N	U	M	U	U	M	M	4
C 4.2	M	M	M	U	U	U	U	U	M	4
C 4.3	M	M	M	U	M	U	U	U	U	4
C 5.1	M	M	M	U	M	N	U	N	M	5
C 5.2	M	M	M	U	M	U	U	U	U	4
C 5.3	M	M	M	U	M	N	U	N	M	5
C 6.1	M	M	M	U	M	N	U	N	M	5
C 6.2	M	M	M	U	M	N	U	N	M	5
C 6.3	M	M	M	N	M	U	U	U	M	5
C 7.1	M	M	M	U	U	U	U	U	M	4
C 7.2	M	M	M	U	U	U	U	U	U	3
Total no. of ^{M}C	15	18	15	0	11	0	0	4	11	74

Note: M = Methylated (remained as a cytosine/^{M}C), U=Unmethylated (converted into a Tymine), N=Unidentified / Identified as A or G.

Table A.8: Colony kin structure, maternal effect, paternal effect and percentage worker produced males of 90 eusocial Hymenoptera

Taxon and species	Kin structure	Study	Ne	G	k	r_{diff}	WPM	nw	Paternal effect	Maternal effect	References
SWEET BEES											
Augochlorella striata	Monogynous		1	0.5	1	0.125	0	7	0.25	-0.375	Mueller et al. 1994
Lasioglossum laevissimum	Monogynous		1.25	0.5	1	0.125	52	2.5	0.25	-0.325	Packer and Owen 1994
Lasioglossum malachurum	Monogynous		1	0.5	1	0.12	4.7	33	0.25	-0.375	Richards et al. 2005
Lasioglossum zephyrum	Monogynous		1	0.5	1	0.125	15	20	0.25	-0.375	Crozier et al. 1987; Kukuk and May 1991
BUMBLEBEES											
Bombus hypnorum	Monogynous	Study 1	1	1	1.12	0.075	0	28.5	0.2232	-0.625	Paxton et al. 2001b
B. hypnorum	Monogynous	Study 2	1	1	1.12	0.125	18.5	28.5	0.2232	-0.625	Brown et al. 2003
B. hypnorum	Monogynous	Study 3	1	1	1.12	0.105	16.4	28.5	0.2232	-0.625	
B. hypnorum	Monogynous	Study 4	1	1	1.12	⎵.010	0	28.5	0.2232	-0.625	Paxton et al. 2001b
B. hypnorum	Monogynous	All	1	1	1.12	0.095	15.4	28.5	0.2232	-0.625	
Bombus melanopygus	Monogynous		1	1	1	0.125	19.1	34	0.25	-0.625	Owen and Plowright, 1980, 1982
Bombus terrestris	Monogynous	Study 1	1	1	1	0.125	3.8	350	0.25	-0.625	Lopez Vaamonde et al. 2004
Bombus terrestris	Monogynous	Study 2	1	1	1	0.125	4.8	350	0.25	-0.625	Alaux et al. 2004
Bombus terrestris	Monogynous	Study 3	1	1	1	0.125	2.3	350	0.25	-0.625	Velthuis et al. 2002
Bombus terrestris	Monogynous	Study 4	1	1	1	0.125	3.4	350	0.25	-0.625	
STINGLESS BEES											
Austroplebeia australis	Monogynous		1	1	1	0.125	7	2000	0.25	-0.625	Drumond et al. 2000
Austroplebeia symei	Monogynous		1	1	1.06	0.125	5	2000	0.2358	-0.625	Palmer et al. 2002
Melipona beecheii	Monogynous		1	1	1.16	0.125	0	1192	0.2155	-0.625	Paxton et al. 2001a
Melipona favosa	Monogynous		1	1	1	0.125	94.5	204	0.25	-0.625	Sommeijer et al. 1999; Chinh et al. 2003
Melipona marginata	Monogynous		1	1	1	0.125	37.1	750	0.25	-0.625	Toth et al. 2002b
Melipona quadrifasciata	Monogynous		1	0.5	1	0.125	64.2	750	0.25	-0.375	Toth et al. 2002b
Melipona scutellaris	Monogynous		1	0.5	1	0.125	28.3	1000	0.25	-0.375	Toth et al. 2002b
Melipona subnitida	Monogynous	Study 1	1	1	1	0.125	39.3	450	0.25	-0.625	Contel and Kerr 1976
M. subnitida	Monogynous	Study 2	1	1	1	0.125	33.6	450	0.25	-0.625	Koedam et al. 1999, 2005
M. subnitida	Monogynous	Study 3	1	1	1	0.125	36.3	450	0.25	-0.625	Toth et al. 2002a
Paratrigona subnuda	Monogynous		1	1	1.02	0.125	64	3750	0.2451	-0.625	Toth et al. 2002a
Plebeia droryana	Monogynous	Study 1	1	1	1	0.125	0	2700	0.25	-0.625	Toth et al. 2002b
P. droryana	Monogynous	Study 2	1	1	1	0.125	16.2	2700	0.25	-0.625	Machado et al. 1984
P. droryana	Monogynous	Study 3	1	1	1	0.125	15.2	2700	0.25	-0.625	
Plebeia remota	Monogynous		1	1	1	0.125	2.4	2900	0.25	-0.625	Toth et al. 2002b
Plebeia saiqui	Monogynous		1	1	1	0.125	0	1500	0.25	-0.625	Toth et al. 2002b
Scaptotrigona postica	Monogynous	Study 1	1	1	1	0.109	13.1	10375	0.25	-0.625	Paxton et al. 2003
S. postica	Monogynous	Study 2	1	1	1	0.125	18.5	10375	0.25	-0.625	Toth et al. 2002b
S. postica	Monogynous	Study 3	1	1	1	0.115	14.7	10375	0.25	-0.625	
Schwarziana quadripunctata	Monogynous		1	1	1	0.125	0	1650	0.25	-0.625	Toth et al. 2003

Table A.8: continued

Tetragona clavipes	Monogynous		1	1	1.01	0.125	65	6500	0.2475	-0.625	Toth et al. 2002b
Trigona carbonaria	Monogynous		1	1	1	0.125	0	10000	0.25	-0.625	Green and Oldroyd 2002
Trigona clypearis	Monogynous		1	1	1.04	0.125	0	500	0.2404	-0.625	Palmer et al. 2002
Trigona hockingsi	Monogynous		1	1	1.16	0.125	0	6500	0.2155	-0.625	Palmer et al. 2002
Trigona mellipes	Monogynous		1	1	0.98	0.125	0	2000	0.2551	-0.625	Palmer et al. 2002
HONEYBEES											
Apis cerana	Monogynous		1	1	14.1	_.075	0	6884	0.0177	-0.625	Palmer and Oldroyd 2000; Oldroyd et al. 2001
Apis dorsata	Monogynous		1	1	44.2	_.105	0	36630	0.0057	-0.625	Palmer and Oldroyd 2000; Wattanachaiyingcharoen et al. 2002
Apis florea	Monogynous		1	1	7.9	_.105	0	6271	0.0316	-0.625	Palmer and Oldroyd 2000; Halling et al. 2001
Apis mellifera	Monogynous		1	1	11.6	_.105	0.1	80000	0.0216	-0.625	Visscher 1989; Estoup et al. 1994; Palmer and Oldroyd 2000
SPHECID WASPS											
Microstigmus comes	Monogynous		1	0.51	1	0.085	0	18	0.25	-0.38	Ross and Matthews 1989
Brachygastra mellifica	Polygynous		398	0.66	1	_.135	0	7951	0.25	-0.16573	Hastings et al. 1998
Parachartergus colobopterus	Polygynous		46	0.57	1	_.155	0	393	0.25	-0.14832	Henshaw et al. 2000a
Polybioides tabidus	Polygynous		34	0.5	1	_.150	2	4000	0.25	-0.13235	Henshaw et al. 2002
Polistes bellicosus	Monogynous		1.6	0.67	1	0.065	1	15.75	0.25	-0.35031	Arevalo et al. 1998
Polistes chinensis	Monogynous		1	1	1	0.125	51.1	46	0.25	-0.625	Tsuchida et al. 2003; Saigo and Tsuchida 2004;
Polistes dorsalis	Monogynous		1	1	1	0.125	0	14.5	0.25	-0.625	Arevalo et al. 1998
Polistes fuscatus variatus			1.86	0.5	1.02	0.02	0	76	0.2451	-0.25941	Metcalf 1980
Polistes gallicus	Monogynous		1	0.66	1	0.125	0	25	0.25	-0.455	Strassmann et al. 2003
Polistes metricus			1.03	1	1	0.076	1.3	86	0.25	-0.61408	Metcalf 1980
VESPINAE WASPS											
Dolichovespula arenaria	Monogynous		1	1	1.2	0.104	17	378	0.2083	-0.625	Foster and Ratnieks 2001b
Dolichovespula maculata	Monogynous		1	1	1	0.125	20.9	181	0.25	-0.625	Foster et al. 2001
Dolichovespula media	Monogynous		1	1	1.08	0.105	7.4	74	0.2315	-0.625	Foster et al. 2001
Dolichovespula norwegica	Monogynous		1	1	1.08	0.105	2.6	44	0.2315	-0.625	Foster et al. 2001
Dolichovespula saxonica	Monogynous	Study 1	1	1	1.35	0.075	40	69	0.1852	-0.625	Foster and Ratnieks 2000
D. saxonica	Monogynous	Study 2	1	1	1.35	_.020	0	69	0.1852	-0.625	Foster and Ratnieks 2000
D. saxonica	Monogynous	All	1	1	1.35	0.06	36.7	69	0.1852	-0.625	Foster and Ratnieks 2000
Dolichovespula sylvestris	Monogynous		1	1	1.15	0.09	9.8	76	0.2174	-0.625	Foster et al. 2001
Vespa crabro flavofasciata	Monogynous		1	1	1.13	0.11	0	125	0.2212	-0.625	Takahashi et al. 2004b
Vespa crabo gribodi	Monogynous		1	1	1.11	0.085	0	550	0.2252	-0.625	Foster et al. 2000
Vespa ducalis	Monogynous		1	1	1	0.125	0	44	0.25	-0.625	Takahashi et al. 2002

Table A.8: continued

Vespa mandarinia	Monogynous		1	1	1.03	0.1	0	221	0.2427	-0.625	Takahashi et al. 2004a
Vespula germanica	Monogynous		1	1	2.35	-.024	0	1000	0.1064	-0.625	Goodisman et al. 2002
Vespula maculifrons	Monogynous		1	1	5.43	-.090	0	1000	0.046	-0.625	Ross 1986
Vespula rufa	Monogynous	Study 1	1	1	1.5	0.056	3.3	57	0.1667	-0.625	Wenseleers et al. 2005a
V. rufa	Monogynous	Study 2	1	1	1.5	-.028	18.5	57	0.1667	-0.625	Wenseleers et al. 2005a
V. rufa	Monogynous	All	1	1	1.5	0.04	11.1	57	0.1667	-0.625	Wenseleers et al. 2005a
Vespula squamosa	Monogynous		1	1	2.13	-.050	0	1000	0.1174	-0.625	Ross 1986
Vespula vulgaris	Monogynous		1	1	1.9	0.005	0	**1000**	0.1316	-0.625	Foster and Ratnieks 2001a
ANTS											
Acromyrmex echinatior			1.35	1	5.3	-.075	0	7000	0.0472	-0.52778	Bekkevold et al. 1999
Acromyrmex octospinosus	Monogynous		1	1	3.93	-.085	0	60000	0.0636	-0.625	Boomsma et al. 1999; Villesen et al. 1999
Aphaenogaster carolinensis	Monogynous		1	1	1	0.125	0	250	0.25	-0.625	Crozier 1974
Camponotus ocreatus	Monogynous		1.06	1	1.25	0.125	0	147	0.2	-0.60377	Goodisman and Hahn 2004
Colobopsis nipponicus	Monogynous		1	1	1.03	0.125	0	1000	0.2427	-0.625	Hasegawa 1994
Crematogaster smithi	Monogynous		1	1	1.02	0.125	11	165	0.2451	-0.625	Heinze et al. 2000
Cyphomyrmex costatus	Monogynous		1	1	1	0.125	0	300	0.25	-0.625	Villesen et al. 2002; Villesen and Boomsma 2003
Cyphomyrmex longiscapus	Monogynous		1	1	1.05	0.125	0	300	0.2381	-0.625	Villesen et al. 2002; Villesen and Boomsma 2003
Dinoponera quadriceps	Monogynous		1	1	1	0.125	8.5	80	0.25	-0.625	Monnin and Peeters 1997
Dorylus molestus	Monogynous		1.4	1	15.9	-.103	0	100000	0.0157	-0.51786	Kronauer et al. 2006
Epimyrma ravouxi	Monogynous/ polygynous		1	1	1	0.125	0	40	0.25	-0.625	Hammond and Keller 2004
Formica exsecta	Monogynous/ polygynous	Study 1	1	1	1.26	0.125	10	12179	0.1984	-0.625	Pamilo and Rosengren 1983
F. exsecta	Monogynous/ polygynous	Study 2	1	1	1.26	0.06	0	12179	0.1984	-0.625	Sundstrom et al. 1996; Walin et al. 1998
F. exsecta	Monogynous/ polygynous	Study 3	1	1	1.26	0.064	0.5	12179	0.1984	-0.625	
F. exsecta	Monogynous/ polygynous	Study 4	1	1	1.26	-.020	0	12179	0.1984	-0.625	Sundstrom et al. 1996; Walin et al. 1998
F. exsecta	Monogynous	All	1	1	1.26	0.06	0.4	12179	0.1984	-0.625	
Formica fusca	Monogynous		3.33	0.53	1.11	-.0008	0	1500	0.2252	-0.20983	Helantera and Sundstrom 2005
Formica rufa	Monogynous		1	1	1.47	0.125	5.3	200000	0.1701	-0.625	Walin et al. 1998
F. rufa	Monogynous		1	1	1.47	0	3.9	200000	0.1701	-0.625	Walin et al. 1998
F. rufa	Polygynous		1	1	1.47	0.045	4.8	200000	0.1701	-0.625	Walin et al. 1998
Formica sanguinea	Monogynous/ polygynous		3.25	0.52	1.31	0.03	3.2	10000	0.1908	-0.20846	Pamilo 1982; Pamilo and Rosengren 1983
Formica truncorum	Monogynous		1	1	1.43	0.05	0	10000	0.1748	-0.625	Sundstrom 1994

Table A.8: continued

Harpagoxenus sublaevis	Monogynous/ polygynous		1	1	1.02	0.125	0	83	0.2451	-0.625	Bourke et al. 1988
Iridomyrmex purpureus	Monogynous		1.46	1	1	0.125	0	60000	0.25	-0.50685	Halliday 1983
Lasius niger	Monogynous	Study 1	1	1	1.16	0.07	9.4	24247	0.2155	-0.625	van der Have et al. 1988
L. niger	Monogynous	Study 2	1	1	1.16	0.095	1.9	24247	0.2155	-0.625	Fjerdingstad et al. 2002, 2003
L. niger	Monogynous	Study 3	1	1	1.16	0.087	4.4	24247	0.2155	-0.625	
L. niger	Monogynous	Study 4	1	1	1.16	..025	2.1	24247	0.2155	-0.625	Fjerdingstad et al. 2002, 2003
L. niger	Monogynous/ polygynous	All	1	1	1.16	0.063	3.9	24247	0.2155	-0.625	
Leptothorax acervorum	Monogynous	Study 1	1	1	1.06	0.125	2.3	84.2	0.2358	-0.625	Chan and Bourke 1994
L. acervorum	Polygynous	Study 2	3.6	0.26	1.06	0.005	4.6	84.2	0.2358	-0.11778	Chan and Bourke 1994
L. acervorum	Monogynous/ polygynous	All	2.1	0.6	1.06	0.083	2.8	84.2	0.2358	-0.28095	Chan and Bourke 1994
Leptothorax allardycei	Monogynous		1	1	1	0.125	35.8	24	0.25	-0.625	Cole 1981
Leptothorax nylanderi	Monogynous		1	1	1	0.045	2.5	37	0.25	-0.625	Foitzik 1998; Foitzik and Heinze 2001
Leptothorax unifasciatus	Monogynous		1	1	1	0.125	2.7	116	0.25	-0.625	Hammond and Keller 2004
Myrmica punctiventris	Polygynous		1.38	1	1	0.115	58	63.8	0.25	-0.52174	Herbers and Mouser 1998
Myrmica ruginodis	Polygynous		2.5	0.5	1.08	0.06	0	600	0.2315	-0.225	Walin et al. 1998
Myrmica tahoensis	Monogynous	Study 1	1	1	1	0.125	65.1	175	0.25	-0.625	Evans 1998
M. tahoensis	Polygynous	Study 2	1.7	0.75	1	..025	84.8	175	0.25	-0.37132	Evans 1998
M. tahoensis	Monogynous/ polygynous	All	1.1	0.78	1	0.098	70.6	175	0.25	-0.48591	Evans 1998
Nothomyrmecia macrops	Monogynous		1	1	1.33	0.055	0	60	0.188	-0.625	Sanetra and Crozier 2001
Pogonomyrmex rugosus	Monogynous		1	1	4.71	..090	0	6000	0.0531	-0.625	Hammond et al. 2004
Polyergus rufescens	Monogynous		1	1	1	0.125	100	2000	0.25	-0.625	Brunner et al. 2005
Protomognathus americanus	Monogynous		1	0.57	1.16	0.09	41	6.3	0.2155	-0.41	Foitzik and Herbers 2001
Rhytidoponera chalybaea	Monogynous		1	1	1	0.125	0	100	0.25	-0.625	Ward 1983
Rhytidoponera confusa	Monogynous		1	1	1	0.125	0	100	0.25	-0.625	Ward 1983
Sericomyrmex amabilis	Monogynous		1	1	1.13	0.125	0	1500	0.2212	-0.625	Villesen et al. 2002; Villesen and Boomsma 2003
Trachymyrmex cornetzi sp.1	Monogynous		1	1	1.21	0.125	0	300	0.2066	-0.625	Villesen et al. 2002; Villesen and Boomsma 2003
Trachymyrmex cf. zeteki	Monogynous		1	1	1.17	0.125	0	300	0.2137	-0.625	Villesen et al. 2002; Villesen and Boomsma 2003

Note: Ne = number of effective queens per colony, G = relatedness among colony queens, k = queen mating frequency, WPM = percentage of worker produced males, r_{diff} = relatedness difference between workers to worker produced sons and queen produced sons, nw = colony size (number of adult workers in a matured colony), Paternal effect = 1/4k, Maternal effect = -[(1/8Ne)-((GNe +G)/4Ne)]. Data obtained from different sources are colour coded as follows; Ne and k (marked in green) were obtained from Hughes et al. 2008a,b. Colony size (marked in red) were obtained from Hammond and Keller 2004. Colony size (marked in blue) were obtained from Jaffe et al. 2014. Data on WPM was obtained from Wenseleers and Ratnieks 2006a. Data on Ne and G were obtained from the references given in the table.

Bibliography

Alaux, C., Boutot, M., Jaisson, P., and Hefetz, A. Reproductive plasticity in bumblebee workers (*Bombus terrestris*) - reversion from fertility to sterility under queen influence. *Behavioral Ecology and Sociobiology*, 62:213–222, December 2007. ISSN 0340-5443. doi: 10.1007/s00265-007-0455-6. URL ://000250832800007. 2.

Amarasinghe, H. E., Clayton, C. I., and Mallon, E. B. Methylation and worker reproduction in the bumblebee (*Bombus terrestris*). *Proceedings of the Royal Society B: Biological Sciences*, 281(1780), April 2014. doi: 10.1098/rspb.2013.2502. URL http://rspb.royalsocietypublishing.org/content/281/1780/20132502.abstract.

Amin, M. R., Bussiere, L. F., and Goulson, D. Effects of male age and size on mating success in the bumblebee *Bombus terrestris*. *Journal of Insect Behavior*, 25(4):362–374, July 2012. ISSN 0892-7553, 1572-8889. doi: 10.1007/s10905-011-9306-4. URL http://link.springer.com/article/10.1007/s10905-011-9306-4.

Amsalem, E. and Hefetz, A. The appeasement effect of sterility signaling in dominance contests among *Bombus terrestris* workers. *Behavioral Ecology and Sociobiology*, 64(10):1685–1694, 2010. ISSN 0340-5443. doi: 10.1007/s00265-010-0982-4. URL http://dx.doi.org/10.1007/s00265-010-0982-4.

Amsalem, E. and Hefetz, A. The effect of group size on the interplay between dominance and reproduction in *Bombus terrestris*. *PLoS ONE*, 6(3):e18238, 2011. doi: 10.1371/journal.pone.0018238. URL http://dx.doi.org/10.1371%2Fjournal.pone.0018238.

Amsalem, E., Grozinger, C. M., Padilla, M., and Hefetz, A. Chapter Two - The Physiological and Genomic Bases of Bumble Bee Social Behaviour. In Kent, A. Z.

a. C. F., editor, *Advances in Insect Physiology*, volume 48 of *Genomics, Physiology and Behaviour of Social Insects*, pages 37–93. Academic Press, 2015. URL http://www.sciencedirect.com/science/article/pii/S0065280615000090.

Anstey, M. L., Rogers, S. M., Ott, S. R., Burrows, M., and Simpson, S. J. Serotonin mediates behavioral gregarization underlying swarm formation in desert locusts. *Science*, 323(5914):627–630, January 2009. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1165939. URL http://www.sciencemag.org/content/323/5914/627.

Arevalo, E., Strassmann, J. E., and Queller, D. C. Conflicts of interest in social insects: Male production in two species of *Polistes*. *Evolution*, 52(3):797, June 1998. ISSN 00143820. doi: 10.2307/2411273. URL http://www.jstor.org/discover/10.2307/2411273?uid=3738032&uid=2129&uid=2&uid=70&uid=4&sid=21104681257311.

Ayabe, T., Hoshiba, H., and Ono, M. Cytological evidence for triploid males and females in the bumblebee, *Bombus terrestris*. *Chromosome Research: An International Journal on the Molecular, Supramolecular and Evolutionary Aspects of Chromosome Biology*, 12(3):215–223, 2004. ISSN 0967-3849.

Barlow, D. P., Stoger, R., Herrmann, B. G., Saito, K., and Schweifer, N. The mouse insulin-like growth factor type-2 receptor is imprinted and closely linked to the Tme locus. *Nature*, 349(6304):84–87, January 1991. ISSN 0028-0836. doi: 10.1038/349084a0.

Barlow, D. P. and Bartolomei, M. S. Genomic imprinting in mammals. *Cold Spring Harbor Perspectives in Biology*, 6(2):a018382, February 2014. ISSN , 1943-0264. doi: 10.1101/cshperspect.a018382. URL http://cshperspectives.cshlp.org/content/6/2/a018382.

Bartolomei, M. S. and Tilghman, S. M. Genomic imprinting in mammals. *Annual Review of Genetics*, 31(1):493–525, 1997. doi: 10.1146/annurev.genet.31.1.493. URL http://dx.doi.org/10.1146/annurev.genet.31.1.493.

Bbosa, G., Kitya, D., Odda, J., and Ogwal-Okeng, J. Aflatoxins metabolism, effects on epigenetic mechanisms and their role in carcinogenesis. *Health*, 5, 2013. doi: doi:10.4236/health.2013.510A1003.

Beck, S. Taking the measure of the methylome. *Nature Biotechnology*, 28(10): 1026–1028, October 2010. ISSN 1087-0156. doi: 10.1038/nbt1010-1026. URL http://www.nature.com/nbt/journal/v28/n10/full/nbt1010-1026.html.

Beeler, S. M., Wong, G. T., Zheng, J. M., Bush, E. C., Remnant, E. J., Oldroyd, B. P., and Drewell, R. A. Whole-genome DNA methylation profile of the Jewel Wasp (*Nasonia vitripennis*). *G3: Genes|Genomes|Genetics*, 4(3): 383–388, December 2013. ISSN 2160-1836. doi: 10.1534/g3.113.008953. URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3962478/.

Bird, A. P. DNA methylation and the frequency of CpG in animal DNA. *Nucleic Acids Research*, 8(7):1499–1504, April 1980. ISSN 0305-1048. URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC324012/.

Bloch, G. and Hefetz, A. Regulation of reproduction by dominant workers in bumblebee (*Bombus terrestris*) queen-right colonies. *Behavioral Ecology and Sociobiology*, 45(2):125–135, February 1999. ISSN 0340-5443, 1432-0762. doi: 10.1007/s002650050546. URL http://link.springer.com.ezproxy3.lib.le.ac.uk/article/10.1007/s002650050546.

Bock, C., Tomazou, E. M., Brinkman, A. B., Muller, F., Simmer, F., Gu, H., Jager, N., Gnirke, A., Stunnenberg, H. G., and Meissner, A. Quantitative comparison of genome-wide DNA methylation mapping technologies. *Nature Biotechnology*, 28 (10):1106–1114, October 2010. ISSN 1546-1696. doi: 10.1038/nbt.1681.

Boerjan, B., Sas, F., Ernst, U. R., Tobback, J., Lemiere, F., Vandegehuchte, M. B., Janssen, C. R., Badisco, L., Marchal, E., Verlinden, H., Schoofs, L., and De Loof, A. Locust phase polyphenism: Does epigenetic precede endocrine regulation? *General and Comparative Endocrinology*, 173(1):120–128, August 2011. ISSN 1095-6840. doi: 10.1016/j.ygcen.2011.05.003.

Boffelli, D., Takayama, S., and Martin, D. I. K. Now you see it: Genome methylation makes a comeback in *Drosophila*. *BioEssays*, pages n/a–n/a, 2014. ISSN 1521-1878. doi: 10.1002/bies.201400097. URL http://dx.doi.org/10.1002/bies.201400097.

Bonasio, R. The role of chromatin and epigenetics in the polyphenisms of ant castes. *Briefings in Functional Genomics*, 13(3):235–245, May 2014. ISSN 2041-2649, 2041-2657. doi: 10.1093/bfgp/elt056. URL http://bfg.oxfordjournals.org.ezproxy4.lib.le.ac.uk/content/13/3/235.

Bonasio, R., Tu, S., and Reinberg, D. Molecular Signals of Epigenetic States. *Science*, 330(6004):612–616, October 2010a. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1191078. URL http://www.sciencemag.org/content/330/6004/612.

Bonasio, R., Zhang, G., Ye, C., Mutti, N. S., Fang, X., Qin, N., Donahue, G., Yang, P., Li, Q., Li, C., Zhang, P., Huang, Z., Berger, S. L., Reinberg, D., Wang, J., and Liebig, J. Genomic comparison of the ants *Camponotus floridanus* and *Harpegnathos saltator*. *Science*, 329(5995):1068–1071, August 2010b. doi: 10.1126/science.1192428. URL http://www.sciencemag.org/content/329/5995/1068.abstract.

Bonasio, R., Li, Q., Lian, J., Mutti, N. S., Jin, L., Zhao, H., Zhang, P., Wen, P., Xiang, H., Ding, Y., Jin, Z., Shen, S. S., Wang, Z., Wang, W., Wang, J., Berger, S. L., Liebig, J., Zhang, G., and Reinberg, D. Genome-wide and caste-specific DNA methylomes of the ants *Camponotus floridanus* and *Harpegnathos saltator*. *Current Biology*, 22(19):1755 – 1764, 2012. ISSN 0960-9822. doi: http://dx.doi.org/10.1016/j.cub.2012.07.042. URL http://www.sciencedirect.com/science/article/pii/S0960982212008676.

Bongiorni, S., Cintio, O., and Prantera, G. The relationship between DNA methylation and chromosome imprinting in the coccid *Planococcus citri*. *Genetics*, 151 (4):1471–1478, April 1999. ISSN 0016-6731.

Bouaichi, A., Roessingh, P., and Simpson, S. J. An analysis of the behavioural effects of crowding and re-isolation on solitary-reared adult desert locusts (*Schistocerca gregaria*) and their offspring. *Physiological Entomology*, 20 (3):199–208, 1995. ISSN 1365-3032. doi: 10.1111/j.1365-3032.1995.tb00002.x. URL http://onlinelibrary.wiley.com/doi/10.1111/j.1365-3032.1995.tb00002.x/abstract.

Bourke, A. F. G. Worker reproduction in the Higher Eusocial Hymenoptera. *Quarterly Review of Biology*, 63:291–311, 1988. ISSN 0033-5770. doi: 10.1086/415930.

Bourke, A. F. G. Hamilton's rule and the causes of social evolution. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 369 (1642):20130362, May 2014. ISSN 0962-8436, 1471-2970. doi: 10.1098/rstb.2013.0362. URL http://rstb.royalsocietypublishing.org/content/369/1642/20130362.

Bourke, A. F. G. and Ratnieks, F. L. W. Kin conflict over caste determination in social Hymenoptera. *Behavioral Ecology and Sociobiology*, 46(5):287–297, October 1999. ISSN 0340-5443, 1432-0762. doi: 10.1007/s002650050622. URL http://link.springer.com/article/10.1007/s002650050622.

Burt, A. and Trives, R. *Genes in Conflict*. Harvard University Press, Cambridge, MA, 2006.

Camacho, J. P. M., Ruiz-Ruano, F. J., Martin-Blazquez, R., Lopez-Leon, M. D., Cabrero, J., Lorite, P., Cabral-de Mello, D. C., and Bakkali, M. A step to the gigantic genome of the desert locust: chromosome sizes and repeated DNAs. *Chromosoma*, December 2014. ISSN 1432-0886. doi: 10.1007/s00412-014-0499-0.

Capinera, J. L. Polyphenism. In Capinera, J. L., editor, *Encyclopedia of Entomology*, pages 2995–2996. Springer Netherlands, January 2008. ISBN 978-1-4020-6242-1, 978-1-4020-6359-6. URL http://link.springer.com/referenceworkentry/10.1007/978-1-4020-6359-6_3057.

Cardinal, S. and Danforth, B. N. The antiquity and evolutionary history of social behavior in bees. *PLoS ONE*, 6(6), June 2011. ISSN 1932-6203. doi: 10.1371/journal.pone.0021086. URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3113908/.

Cardoen, D., Wenseleers, T., Ernst, U. R., Danneels, E. L., Laget, D., De Graaf, D. C., Schoofs, L., and Verleyen, P. Genome-wide analysis of alternative reproductive phenotypes in honeybee workers. *Molecular Ecology*, 20(19):4070–4084, October 2011. ISSN 1365-294X. doi: 10.1111/j.1365-294X.2011.05254.x. URL http://onlinelibrary.wiley.com/doi/10.1111/j.1365-294X.2011.05254.x/abstract.

Cedar, H. and Bergman, Y. Programming of DNA methylation patterns. *Annual Review of Biochemistry*, 81(1):97–117, 2012. doi: 10.1146/annurev-biochem-052610-091920. URL http://dx.doi.org/10.1146/annurev-biochem-052610-091920.

Christman, J. K. 5-azacytidine and 5-aza-2'-deoxycytidine as inhibitors of DNA methylation: Mechanistic studies and their implications for cancer therapy. *Oncogene*, 21(35):5483–5495, August 2002. ISSN 0950-9232. doi: 10.1038/sj.onc.1205699.

Clarke, J., Wu, H.-C., Jayasinghe, L., Patel, A., Reid, S., and Bayley, H. Continuous base identification for single-molecule nanopore DNA sequencing. *Nature Nanotechnology*, 4(4):265–270, April 2009. ISSN 1748-3387. doi: 10.1038/nnano.2009.12. URL http://www.nature.com/nnano/journal/v4/n4/full/nnano.2009.12.html.

Cnaani, J., Wong, A., and Thomson, J. D. Effect of group size on ovarian development in bumblebee workers (Hymenoptera: Apidae: *Bombus*). *Entomologia Generalis*, 29(2-4):305–314, 2007. URL http://dx.doi.org/10.1127/entom.gen/29/2007/305.

Crozier, R. H., Smith, B. H., and Crozier, Y. C. Relatedness and population structure of the Primitively Eusocial Bee *Lasioglossum zephyrum* (Hymenoptera: Halictidae) in Kansas. *Evolution*, 41(4):902, July 1987. ISSN 00143820. doi: 10.2307/2408898. URL http://www.jstor.org/discover/10.2307/2408898?uid=3738032&uid=2131&uid=2129&uid=2&uid=70&uid=4&sid=21105275441823.

Crozier, R. H. and Pamilo, P. *Evolution of social insect colonies. Sex allocation and kin selection.* Oxford series in Ecology and Evolution. Oxford University Press. Oxford, UK, 1996.

Cunha, K. S., Reguly, M. L., Graf, U., and de Andrade, H. H. R. Somatic recombination: a major genotoxic effect of two pyrimidine antimetabolitic chemotherapeutic drugs in *Drosophila melanogaster*. *Mutation research*, 514(1-2):95–103, February 2002. ISSN 0027-5107.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., and

Durbin, R. The variant call format and VCFtools. *Bioinformatics*, 27(15):2156–2158, August 2011. ISSN 1367-4803, 1460-2059. doi: 10.1093/bioinformatics/btr330. URL http://bioinformatics.oxfordjournals.org/content/27/15/2156.

Das, P. M. and Singal, R. DNA methylation and cancer. *Journal of Clinical Oncology*, 22(22):4632–4642, November 2004. ISSN 0732-183X, 1527-7755. doi: 10.1200/JCO.2004.07.151. URL http://jco.ascopubs.org/content/22/22/4632.

DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., Philippakis, A. A., del Angel, G., Rivas, M. A., Hanna, M., McKenna, A., Fennell, T. J., Kernytsky, A. M., Sivachenko, A. Y., Cibulskis, K., Gabriel, S. B., Altshuler, D., and Daly, M. J. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, 43(5):491–498, May 2011. ISSN 1061-4036. doi: 10.1038/ng.806. URL http://www.nature.com/ng/journal/v43/n5/full/ng.806.html.

DeVeale, B., van der Kooy, D., and Babak, T. Critical evaluation of imprinted gene expression by RNA-seq: A new perspective. *PLoS Genet*, 8(3):e1002600, March 2012. doi: 10.1371/journal.pgen.1002600. URL http://dx.doi.org/10.1371/journal.pgen.1002600.

Dixon, A. F. G. Aphid ecology: Life cycles, Polymorphism, and Population regulation. *Annual Review of Ecology and Systematics*, 8(1):329–353, 1977. doi: 10.1146/annurev.es.08.110177.001553. URL http://dx.doi.org/10.1146/annurev.es.08.110177.001553.

Drewell, R. A., Lo, N., Oxley, P. R., and Oldroyd, B. P. Kin conflict in insect societies: a new epigenetic perspective. *Trends in ecology & evolution*, 27(7):367–373, July 2012. ISSN 0169-5347. doi: 10.1016/j.tree.2012.02.005.

Drewell, R. A., Bush, E. C., Remnant, E. J., Wong, G. T., Beeler, S. M., Stringham, J. L., Lim, J., and Oldroyd, B. P. The dynamic DNA methylation cycle from egg to sperm in the honey bee *Apis mellifera*. *Development*, 141(13):2702–2711, July 2014. ISSN 0950-1991, 1477-9129. doi: 10.1242/dev.110163. URL http://dev.biologists.org/content/141/13/2702.

Duchateau, M. Agonistic behaviours in colonies of the bumblebee *Bombus terrestris*. *Journal of Ethology*, 7(2):141–151, 1989. ISSN 0289-0771. doi: 10.1007/BF02350036. URL http://dx.doi.org/10.1007/BF02350036.

Duchateau, M. and Velthuis, H. Development and reproductive strategies in *Bombus terrestris* colonies. *Behaviour*, 107(3-4):186–207, 1988. doi: 10.1163/156853988X00340.

Duncan, B. K. and Miller, J. H. Mutagenic deamination of cytosine residues in DNA. *Nature*, 287(5782):560–561, October 1980. ISSN 0028-0836.

Eckersley-Maslin, M. A. and Spector, D. L. Random monoallelic expression: regulating gene expression one allele at a time. *Trends in Genetics*, 30(6):237–244, June 2014. ISSN 0168-9525. doi: 10.1016/j.tig.2014.03.003. URL http://www.sciencedirect.com/science/article/pii/S0168952514000523.

Egger, G., Liang, G., Aparicio, A., and Jones, P. A. Epigenetics in human disease and prospects for epigenetic therapy. *Nature*, 429(6990):457–463, May 2004. ISSN 0028-0836. doi: 10.1038/nature02625. URL http://dx.doi.org/10.1038/nature02625.

El-Maarri, O. b., Buiting, K., Peery, E., Kroisel, P., Balaban, B., Wagner, K., Urman, B., Heyd, J., Lich, C., Brannan, C., Walter, J. g., and Horsthemke, B. Maternal methylation imprints on human chromosome 15 are established during or after fertilization. *Nature Genetics*, 27(3):341–344, 2001. URL http://www.scopus.com/inward/record.url?eid=2-s2.0-0035090961&partnerID=40&md5=39a56732f4f75c897419d9156238599f. cited By (since 1996)142.

Elango, N., Hunt, B. G., Goodisman, M. A. D., and Yi, S. V. DNA methylation is widespread and associated with differential gene expression in castes of the honeybee, *Apis mellifera*. *Proceedings of the National Academy of Sciences*, 106(27):11206–11211, July 2009. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.0900301106. URL http://www.pnas.org/content/106/27/11206.

Erick Greene, Michael Canfield, and Adam Ehmer. Developmental flexibility, Phenotypic plasticity, and Host plants. In *Phenotypic Plasticity of Insects*. Science

Publishers, 2009. ISBN 978-1-57808-423-4. URL http://www.crcnetbase.com/doi/abs/10.1201/b10201-5.

Ernst, U. R., Hiel, M. B. V., Depuydt, G., Boerjan, B., Loof, A. D., and Schoofs, L. Epigenetics and locust life phase transitions. *The Journal of Experimental Biology*, 218(1):88–99, January 2015. ISSN 0022-0949, 1477-9145. doi: 10.1242/jeb.107078. URL http://jeb.biologists.org/content/218/1/88.

Falckenhayn, C., Boerjan, B., Raddatz, G., Frohme, M., Schoofs, L., and Lyko, F. Characterization of genome methylation patterns in the desert locust *Schistocerca gregaria*. *The Journal of Experimental Biology*, 216(Pt 8):1423–1429, April 2013. ISSN 1477-9145. doi: 10.1242/jeb.080754.

Fedoriw, A., Mugford, J., and Magnuson, T. Genomic imprinting and epigenetic control of development. *Cold Spring Harbor Perspectives in Biology*, 4(7):a008136, July 2012. ISSN , 1943-0264. doi: 10.1101/cshperspect.a008136. URL http://cshperspectives.cshlp.org/content/4/7/a008136.

Feliciello, I., Parazajder, J., Akrap, I., and Ugarkovic, D. First evidence of DNA methylation in insect *Tribolium castaneum*: environmental regulation of DNA methylation within heterochromatin. *Epigenetics: official journal of the DNA Methylation Society*, 8(5):534–541, May 2013. ISSN 1559-2308. doi: 10.4161/epi.24507.

Ferguson-Smith, A. C. Genomic imprinting: the emergence of an epigenetic paradigm. *Nature Reviews Genetics*, 12(8):565–575, August 2011. ISSN 1471-0056. doi: 10.1038/nrg3032. URL http://www.nature.com.ezproxy4.lib.le.ac.uk/nrg/journal/v12/n8/full/nrg3032.html.

Field, L. M., Lyko, F., Mandrioli, M., and Prantera, G. DNA methylation in insects. *Insect Molecular Biology*, 13(2):109–115, 2004. ISSN 1365-2583. doi: 10.1111/j.0962-1075.2004.00470.x. URL http://onlinelibrary.wiley.com/doi/10.1111/j.0962-1075.2004.00470.x/abstract.

Finch, C. The comparative biology of senescence. In *Longevity, senescence, and the genome*. Chicago, IL: University of Chicago Press., 1990. URL http://scholar.google.com/scholar_lookup?title=Longevity%2C%20senescence%2C%20and%20the%20genome&author=CE%20Finch&publication_year=1990.

Flores, K., Wolschin, F., Corneveaux, J. J., Allen, A. N., Huentelman, M. J., and Amdam, G. V. Genome-wide association between DNA methylation and alternative splicing in an invertebrate. *BMC Genomics*, 13(1):480, September 2012. ISSN 1471-2164. doi: 10.1186/1471-2164-13-480. URL http://www.biomedcentral.com/1471-2164/13/480/abstract.

Flusberg, B. A., Webster, D. R., Lee, J. H., Travers, K. J., Olivares, E. C., Clark, T. A., Korlach, J., and Turner, S. W. Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nature Methods*, 7(6):461–465, June 2010. ISSN 1548-7091. doi: 10.1038/nmeth.1459. URL http://www.nature.com/nmeth/journal/v7/n6/full/nmeth.1459.html.

Foitzik, S. and Herbers, J. M. Colony structure of a Slavemaking ant. ii. frequency of slave raids and impact on the host population. *Evolution*, 55(2):316–323, February 2001. ISSN 1558-5646. doi: 10.1111/j.0014-3820.2001.tb01296.x. URL http://onlinelibrary.wiley.com/doi/10.1111/j.0014-3820.2001.tb01296.x/abstract.

Foret, S., Kucharski, R., Pittelkow, Y., Lockett, G. A., and Maleszka, R. Epigenetic regulation of the honey bee transcriptome: unravelling the nature of methylated genes. *BMC Genomics*, 10(1):472, October 2009. ISSN 1471-2164. doi: 10.1186/1471-2164-10-472. URL http://www.biomedcentral.com/1471-2164/10/472/abstract.

Foret, S., Kucharski, R., Pellegrini, M., Feng, S., Jacobsen, S. E., Robinson, G. E., and Maleszka, R. DNA methylation dynamics, metabolic fluxes, gene splicing, and alternative phenotypes in honey bees. *Proceedings of the National Academy of Sciences*, 109(13):4968–4973, March 2012. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1202392109. URL http://www.pnas.org/content/109/13/4968.

Formesyn, E. M., Cardoen, D., Ernst, U. R., Danneels, E. L., Van Vaerenbergh, M., De Koker, D., Verleyen, P., Wenseleers, T., Schoofs, L., and de Graaf, D. C. Reproduction of honeybee workers is regulated by epidermal growth factor receptor signaling. *General and Comparative Endocrinology*, 197:1–4, February 2014. ISSN 0016-6480. doi: 10.1016/j.ygcen.2013.12.001. URL http://www.sciencedirect.com/science/article/pii/S0016648013004681.

Foster, K. R., Gulliver, J., and Ratnieks, F. L. W. Worker policing in the European hornet *Vespa crabro*. *Insectes Sociaux*, 49(1):41–44, March 2002. ISSN 0020-1812, 1420-9098. doi: 10.1007/s00040-002-8277-z. URL http://link.springer.com/article/10.1007/s00040-002-8277-z.

Foster, K. R. and Ratnieks, F. L. W. Convergent evolution of worker policing by egg eating in the honeybee and common wasp. *Proceedings of the Royal Society of London B: Biological Sciences*, 268(1463):169–174, January 2001. ISSN 0962-8452, 1471-2954. doi: 10.1098/rspb.2000.1346. URL http://rspb.royalsocietypublishing.org/content/268/1463/169.

Foster, R. L., Brunskill, A., Verdirame, D., and O'Donnell, S. Reproductive physiology, dominance interactions, and division of labour among bumblebee workers. *Physiological Entomology*, 29(4):327–334, 2004. ISSN 1365-3032. doi: 10.1111/j.0307-6962.2004.00388.x. URL http://dx.doi.org/10.1111/j.0307-6962.2004.00388.x.

Fraga, M. F., Ballestar, E., Paz, M. F., Ropero, S., Setien, F., Ballestar, M. L., Heine-SuÃśer, D., Cigudosa, J. C., Urioste, M., Benitez, J., Boix-Chornet, M., Sanchez-Aguilera, A., Ling, C., Carlsson, E., Poulsen, P., Vaag, A., Stephan, Z., Spector, T. D., Wu, Y.-Z., Plass, C., and Esteller, M. Epigenetic differences arise during the lifetime of monozygotic twins. *Proceedings of the National Academy of Sciences of the United States of America*, 102(30):10604–10609, July 2005. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.0500398102. URL http://www.pnas.org/content/102/30/10604.

Frigola, J., Ribas, M., Risques, R.-A., and Peinado, M. A. Methylome profiling of cancer cells by Amplification of Inter-Methylated Sites (AIMS). *Nucleic Acids Research*, 30(7):e28–e28, April 2002. doi: 10.1093/nar/30.7.e28. URL http://nar.oxfordjournals.org/content/30/7/e28.abstract.

Fu, Y., Chen, J., Pang, B., Li, C., Zhao, J., and Shen, K. EZH2-Induced H3k27me3 is associated with epigenetic repression of the ARHI Tumor-Suppressor gene in Ovarian cancer. *Cell Biochemistry and Biophysics*, pages 1–8, July 2014. ISSN 1085-9195, 1559-0283. doi: 10.1007/s12013-014-0168-1. URL http://link.springer.com/article/10.1007/s12013-014-0168-1.

Gasser, R. B., Hu, M., Chilton, N. B., Campbell, B. E., Jex, A. J., Otranto, D., Cafarchia, C., Beveridge, I., and Zhu, X. Single-strand conformation polymorphism (SSCP) for the analysis of genetic variation. *Nature Protocols*, 1(6): 3121–3128, January 2007. ISSN 1754-2189. doi: 10.1038/nprot.2006.485. URL http://www.nature.com/nprot/journal/v1/n6/full/nprot.2006.485.html.

Gehring, M. Genomic imprinting: Insights from plants. *Annual Review of Genetics*, 47(1):187–208, 2013. doi: 10.1146/annurev-genet-110711-155527. URL http://dx.doi.org/10.1146/annurev-genet-110711-155527.

Geva, S., Hartfelder, K., and Bloch, G. Reproductive division of labor, dominance, and ecdysteroid levels in hemolymph and ovary of the bumblebee *Bombus terrestris*. *Journal of Insect Physiology*, 51(7):811 – 823, 2005. ISSN 0022-1910. doi: http://dx.doi.org/10.1016/j.jinsphys.2005.03.009. URL http://www.sciencedirect.com/science/article/pii/S0022191005000764.

Gineikiene, E., Stoskus, M., and Griskevicius, L. Single nucleotide polymorphism-based system improves the applicability of quantitative PCR for chimerism monitoring. *The Journal of Molecular Diagnostics : JMD*, 11(1):66–74, January 2009. ISSN 1525-1578. doi: 10.2353/jmoldx.2009.080039. URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2607568/.

Glastad, K. M., Hunt, B. G., Yi, S. V., and Goodisman, M. a. D. DNA methylation in insects: on the brink of the epigenomic era. *Insect Molecular Biology*, 20(5):553–565, 2011. ISSN 1365-2583. doi: 10.1111/j.1365-2583.2011.01092.x. URL http://onlinelibrary.wiley.com/doi/10.1111/j.1365-2583.2011.01092.x/abstract.

Glastad, K. M., Hunt, B. G., and Goodisman, M. A. Evolutionary insights into DNA methylation in insects. *Current Opinion in Insect Science*, 1:25–30, July 2014a. ISSN 2214-5745. doi: 10.1016/j.cois.2014.04.001. URL http://www.sciencedirect.com/science/article/pii/S2214574514000029.

Glastad, K. M., Hunt, B. G., Yi, S. V., and Goodisman, M. A. D. Epigenetic inheritance and genome regulation: Is DNA methylation linked to ploidy in haplodiploid insects? *Proceedings. Biological Sciences / The Royal Society*, 281(1785): 20140411, June 2014b. ISSN 1471-2954. doi: 10.1098/rspb.2014.0411.

Goll, M. G., Kirpekar, F., Maggert, K. A., Yoder, J. A., Hsieh, C.-L., Zhang, X., Golic, K. G., Jacobsen, S. E., and Bestor, T. H. Methylation of tRNAAsp by the DNA Methyltransferase Homolog Dnmt2. *Science*, 311(5759):395–398, January 2006. doi: 10.1126/science.1120976. URL http://www.sciencemag.org/content/311/5759/395.abstract.

Gosterit, A. and Gurel, F. Effect of different diapause regimes on survival and colony development in the bumblebee, *Bombus terrestris*. *Journal of Apicultural Research*, pages 279–283, October 2009. ISSN 00218839. doi: 10.3896/IBRA.1.48.4.08. URL http://www.ibra.org.uk/articles/20090911_8.

Gregg, C., Zhang, J., Weissbourd, B., Luo, S., Schroth, G. P., Haig, D., and Dulac, C. High-resolution analysis of parent-of-origin allelic expression in the mouse brain. *Science (New York, N.Y.)*, 329(5992):643–648, August 2010. ISSN 1095-9203. doi: 10.1126/science.1190830.

Grozinger, C. M., Fan, Y., Hoover, S. E. R., and Winston, M. L. Genome-wide analysis reveals differences in brain gene expression patterns associated with caste and reproductive status in honey bees (*Apis mellifera*). *Molecular Ecology*, 16(22):4837–4848, November 2007. ISSN 1365-294X. doi: 10.1111/j.1365-294X.2007.03545.x. URL http://onlinelibrary.wiley.com/doi/10.1111/j.1365-294X.2007.03545.x/abstract.

Gu, H., Smith, Z. D., Bock, C., Boyle, P., Gnirke, A., and Meissner, A. Preparation of reduced representation bisulfite sequencing libraries for genome-scale DNA methylation profiling. *Nat. Protocols*, 6(4):468–481, March 2011. ISSN 1754-2189. doi: 10.1038/nprot.2010.190. URL http://dx.doi.org/10.1038/nprot.2010.190.

Guan, C., Barron, A. B., He, X. J., Wang, Z. L., Yan, W. Y., and Zeng, Z. J. A comparison of digital gene expression profiling and methyl DNA Immunoprecipitation as methods for gene discovery in honeybee (*Apis mellifera*) behavioural genomic analyses. *PLoS ONE*, 8(9):e73628, September 2013. doi: 10.1371/journal.pone.0073628. URL http://dx.doi.org/10.1371/journal.pone.0073628.

Gupta, V., Bijo, A., Kumar, M., Reddy, C., and Jha, B. Detection of epigenetic variations in the protoplast-derived germlings of *Ulva reticulate* using Methylation

Sensitive Amplification Polymorphism (MSAP). *Marine Biotechnology*, 14(6): 692–700, 2012. ISSN 1436-2228. doi: 10.1007/s10126-012-9434-7. URL http://dx.doi.org/10.1007/s10126-012-9434-7.

Guzman-Novoa, E., Hunt, G. J., Page, R. E., Uribe-Rubio, J. L., Prieto-Merlos, D., and Becerra-Guzman, F. Paternal effects on the Defensive Behavior of Honeybees. *Journal of Heredity*, 96(4):376–380, July 2005. ISSN 0022-1503, 1465-7333. doi: 10.1093/jhered/esi038. URL http://jhered.oxfordjournals.org/content/96/4/376.

Hagemann, S., Heil, O., Lyko, F., and Brueckner, B. Azacytidine and decitabine induce gene-specific and non-random DNA Demethylation in human cancer cell lines. *PLoS ONE*, 6(3):e17388, 2011. doi: 10.1371/journal.pone.0017388. URL http://dx.doi.org/10.1371%2Fjournal.pone.0017388.

Haig, D. The Kinship Theory of Genomic Imprinting. *Annual Review of Ecology and Systematics*, 31(1):9–32, 2000. doi: 10.1146/annurev.ecolsys.31.1.9. URL http://dx.doi.org/10.1146/annurev.ecolsys.31.1.9.

Haig, D. and Graham, C. Genomic imprinting and the strange case of the insulin-like growth factor II receptor. *Cell*, 64(6):1045–1046, March 1991. ISSN 0092-8674. doi: 10.1016/0092-8674(91)90256-X. URL http://www.sciencedirect.com/science/article/pii/009286749190256X.

Hamilton, W. D. The genetical evolution of social behaviour. i. *Journal of Theoretical Biology*, 7(1):1–16, July 1964. ISSN 0022-5193. doi: 10.1016/0022-5193(64)90038-4. URL http://www.sciencedirect.com/science/article/pii/0022519364900384.

Hammond, R. L. and Keller, L. Conflict over Male Parentage in Social insects. *PLoS Biol*, 2(9):e248, August 2004. doi: 10.1371/journal.pbio.0020248. URL http://dx.doi.org/10.1371/journal.pbio.0020248.

Harris, R. A., Wang, T., Coarfa, C., Nagarajan, R. P., Hong, C., Downey, S. L., Johnson, B. E., Fouse, S. D., Delaney, A., Zhao, Y., Olshen, A., Ballinger, T., Zhou, X., Forsberg, K. J., Gu, J., Echipare, L., O'Geen, H., Lister, R., Pelizzola, M., Xi, Y., Epstein, C. B., Bernstein, B. E., Hawkins, R. D., Ren, B., Chung, W.-Y., Gu, H., Bock, C., Gnirke, A., Zhang, M. Q., Haussler, D., Ecker, J. R.,

Li, W., Farnham, P. J., Waterland, R. A., Meissner, A., Marra, M. A., Hirst, M., Milosavljevic, A., and Costello, J. F. Comparison of sequencing-based methods to profile DNA methylation and identification of monoallelic epigenetic modifications. *Nature Biotechnology*, 28(10):1097–1105, October 2010. ISSN 1087-0156. doi: 10.1038/nbt.1682. URL http://www.nature.com/nbt/journal/v28/n10/abs/nbt.1682.html.

Hart, A. G. and Ratnieks, F. L. Task partitioning, division of labour and nest compartmentalisation collectively isolate hazardous waste in the leafcutting an *Atta cephalotes*. *Behavioral Ecology and Sociobiology*, 49(5):387–392, 2001. ISSN 0340-5443. doi: 10.1007/s002650000312. URL http://dx.doi.org/10.1007/s002650000312.

Hattori, M., Kishida, O., and Itino, T. Soldiers with large weapons in predator-abundant midsummer: phenotypic plasticity in a eusocial aphid. *Evolutionary Ecology*, 27(5):847–862, January 2013. ISSN 0269-7653, 1573-8477. doi: 10.1007/s10682-012-9628-5. URL http://link.springer.com/article/10.1007/s10682-012-9628-5.

Henckel, A. and Arnaud, P. Genome-wide identification of new imprinted genes. *Briefings in Functional Genomics*, 9(4):304–314, July 2010. ISSN 2041-2649, 2041-2657. doi: 10.1093/bfgp/elq016. URL http://bfg.oxfordjournals.org/content/9/4/304.

Henshaw, M. T., Strassmann, J. E., and Queller, D. C. The independent origin of a queen number bottleneck that promotes cooperation in the African swarm-founding wasp, *Polybioides tabidus*. *Behavioral Ecology and Sociobiology*, 48(6):478–483, November 2000. ISSN 0340-5443, 1432-0762. doi: 10.1007/s002650000264. URL http://link.springer.com/article/10.1007/s002650000264.

Herb, B. R., Wolschin, F., Hansen, K. D., Aryee, M. J., Langmead, B., Irizarry, R., Amdam, G. V., and Feinberg, A. P. Reversible switching between epigenetic states in honeybee behavioral subcastes. *Nature Neuroscience*, 15(10):1371–1373, October 2012. ISSN 1097-6256. doi: 10.1038/nn.3218. URL http://www.nature.com/neuro/journal/v15/n10/full/nn.3218.html.

Hoover, S., Keeling, C., Winston, M., and Slessor, K. The effect of queen pheromones on worker honey bee ovary development. *Naturwissenschaften*, 90(10):477–480, 2003. ISSN 0028-1042. doi: 10.1007/s00114-003-0462-z. URL http://dx.doi.org/10.1007/s00114-003-0462-z.

Horsthemke, B. and Wagstaff, J. Mechanisms of imprinting of the Prader-willi/Angelman region. *American Journal of Medical Genetics Part A*, 146A(16):2041–2052, August 2008. ISSN 1552-4833. doi: 10.1002/ajmg.a.32364. URL http://onlinelibrary.wiley.com.ezproxy4.lib.le.ac.uk/doi/10.1002/ajmg.a.32364/abstract.

Hughes, W. O. H., Ratnieks, F. L. W., and Oldroyd, B. P. Multiple paternity or multiple queens: two routes to greater intracolonial genetic diversity in the eusocial Hymenoptera. *Journal of Evolutionary Biology*, 21(4):1090–1095, July 2008a. ISSN 1420-9101. doi: 10.1111/j.1420-9101.2008.01532.x. URL http://onlinelibrary.wiley.com/doi/10.1111/j.1420-9101.2008.01532.x/abstract.

Hughes, W. O. H., Oldroyd, B. P., Beekman, M., and Ratnieks, F. L. W. Ancestral Monogamy shows Kin Selection Is Key to the Evolution of Eusociality. *Science*, 320(5880):1213–1216, May 2008b. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1156108. URL http://www.sciencemag.org/content/320/5880/1213.

Hunt, B. G., Glastad, K. M., Yi, S. V., and Goodisman, M. A. D. Patterning and regulatory associations of DNA methylation are mirrored by Histone modifications in Insects. *Genome Biology and Evolution*, 5(3):591–598, January 2013. ISSN , 1759-6653. doi: 10.1093/gbe/evt030. URL http://gbe.oxfordjournals.org/content/5/3/591.

Islam, M. S., Roessingh, P., Simpson, S. J., and Mccaffery, A. R. Effects of population density experienced by parents during mating and oviposition on the phase of hatchling desert locusts, *Schistocerca gregaria*. *Proceedings of the Royal Society of London B: Biological Sciences*, 257(1348):93–98, July 1994. ISSN 0962-8452, 1471-2954. doi: 10.1098/rspb.1994.0099. URL http://rspb.royalsocietypublishing.org/content/257/1348/93.

Iwasa, Y. and Pomiankowski, A. Sex specific X Chromosome expression caused by genomic imprinting. *Journal of Theoretical Biology*, 197(4):487–495, April 1999. ISSN 0022-5193. doi: 10.1006/jtbi.1998.0888. URL http://www.sciencedirect.com/science/article/pii/S0022519398908888.

Jaffe, R., Pioker-Hara, F. C., dos Santos, C. F., Santiago, L. R., Alves, D. A., de M. P. Kleinert, A., Francoy, T. M., Arias, M. C., and Imperatriz-Fonseca, V. L. Monogamy in large bee societies: a stingless paradox. *Naturwissenschaften*, 101(3):261–264, March 2014. ISSN 0028-1042, 1432-1904. doi: 10.1007/s00114-014-1149-3. URL http://link.springer.com/10.1007/s00114-014-1149-3.

Jarosch, A., Stolle, E., Crewe, R. M., and Moritz, R. F. A. Alternative splicing of a single transcription factor drives selfish reproductive behavior in honeybee workers (*Apis mellifera*). *Proceedings of the National Academy of Sciences*, 108 (37):15282–15287, September 2011. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1109343108. URL http://www.pnas.org/content/108/37/15282.

Jorda, M., Rodriguez, J., Frigola, J., and Peinado, M. Analysis of DNA Methylation by Amplification of Inter-Methylated Sites (AIMS). In Tost, J., editor, *DNA Methylation*, volume 507 of *Methods in Molecular Biology*, pages 107–116. Humana Press, 2009. ISBN 978-1-934115-61-9. URL http://dx.doi.org/10.1007/978-1-59745-522-0_9.

Jurkowska, R. Z., Jurkowski, T. P., and Jeltsch, A. Structure and Function of Mammalian DNA Methyltransferases. *ChemBioChem*, 12(2):206–222, 2011. ISSN 1439-7633. doi: 10.1002/cbic.201000195. URL http://dx.doi.org/10.1002/cbic.201000195.

Keller, L. and Genoud, M. Extraordinary lifespans in ants: a test of evolutionary theories of ageing. *Nature*, 389(6654):958–960, October 1997. ISSN 0028-0836. doi: 10.1038/40130. URL http://www.nature.com.ezproxy4.lib.le.ac.uk/nature/journal/v389/n6654/abs/389958a0.html.

Kelley, K., Miller, K. R., Todd, A., Kelley, A., Tuttle, R., and Berberich, S. J. YPEL3, a p53-regulated gene that induces cellular senescence. *Cancer research*, 70

(9):3566–3575, May 2010. ISSN 0008-5472. doi: 10.1158/0008-5472.CAN-09-3219. URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2862112/.

Kelly, S. A., Panhuis, T. M., and Stoehr, A. M. Phenotypic plasticity: molecular mechanisms and adaptive significance. *Comprehensive Physiology*, 2(2):1417–1439, April 2012. ISSN 2040-4603. doi: 10.1002/cphy.c110008.

Kerr, W. E. Sex determination in bees. III. Caste determination and genetic control in *Melipona*. *Insectes Sociaux*, 21(4):357–367, December 1974. ISSN 0020-1812, 1420-9098. doi: 10.1007/BF02331565. URL http://link.springer.com/article/10.1007/BF02331565.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S. L. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biology*, 14(4):R36, April 2013. ISSN 1465-6906. doi: 10.1186/gb-2013-14-4-r36. URL http://genomebiology.com/2013/14/4/R36/abstract.

Kocher, S. D., Ayroles, J. F., Stone, E. A., and Grozinger, C. M. Individual variation in pheromone response correlates with reproductive traits and brain gene expression in worker honeybees. *PLoS ONE*, 5(2):e9116, February 2010. doi: 10.1371/journal.pone.0009116. URL http://dx.doi.org/10.1371/journal.pone.0009116.

Kozlowski, P. and Krzyzosiak, W. J. Combined SSCP/duplex analysis by capillary electrophoresis for more efficient mutation detection. *Nucleic Acids Research*, 29 (14):e71, July 2001. ISSN 0305-1048. URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC55818/.

Kronforst, M. R., Gilley, D. C., Strassmann, J. E., and Queller, D. C. DNA methylation is widespread across social Hymenoptera. *Current Biology*, 18 (7):R287 – R288, 2008. ISSN 0960-9822. doi: http://dx.doi.org/10.1016/j.cub.2008.02.015. URL http://www.sciencedirect.com/science/article/pii/S0960982208001607.

Krueger, F., Kreck, B., Franke, A., and Andrews, S. R. DNA methylome analysis using short bisulfite sequencing data. *Nature Methods*, 9(2):145–151, February

2012. ISSN 1548-7091. doi: 10.1038/nmeth.1828. URL http://www.nature.com.ezproxy4.lib.le.ac.uk/nmeth/journal/v9/n2/full/nmeth.1828.html.

Kucharski, R., Maleszka, J., Foret, S., and Maleszka, R. Nutritional control of reproductive status in honeybees via DNA methylation. *Science*, 319:1827–1830, March 2008. ISSN 0036-8075. URL ://000254394000047. 5871.

Leszczyniecka, M., Bhatia, U., Cueto, M., Nirmala, N. R., Towbin, H., Vattay, A., Wang, B., Zabludoff, S., and Phillips, P. E. MAP1d, a novel methionine aminopeptidase family member is overexpressed in colon cancer. *Oncogene*, 25 (24):3471–3478, June 2006. ISSN 0950-9232. doi: 10.1038/sj.onc.1209383.

Li, E. Chromatin modification and epigenetic reprogramming in mammalian development. *Nat Rev Genet*, 3(9):662–673, September 2002. ISSN 1471-0056. doi: 10.1038/nrg887. URL http://dx.doi.org/10.1038/nrg887.

Li, H. and Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)*, 25(14):1754–1760, July 2009. ISSN 1367-4811. doi: 10.1093/bioinformatics/btp324.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)*, 25(16):2078–2079, August 2009. ISSN 1367-4811. doi: 10.1093/bioinformatics/btp352.

Li, J., Wu, J., Begna Rundassa, D., Song, F., Zheng, A., and Fang, Y. Differential protein expression in honeybee (*Apis mellifera L.*) larvae: Underlying Caste Differentiation. *PLoS ONE*, 5(10):e13455, October 2010. doi: 10.1371/journal.pone.0013455. URL http://dx.doi.org/10.1371/journal.pone.0013455.

Li-Byarlay, H., Li, Y., Stroud, H., Feng, S., Newman, T. C., Kaneda, M., Hou, K. K., Worley, K. C., Elsik, C. G., Wickline, S. A., Jacobsen, S. E., Ma, J., and Robinson, G. E. RNA interference knockdown of DNA methyl-transferase 3 affects gene alternative splicing in the honey bee. *Proceedings of the National Academy of Sciences*, 110(31):12750–12755, July 2013. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1310735110. URL http://www.pnas.org/content/110/31/12750.

Libbrecht, R., Schwander, T., and Keller, L. Genetic components to Caste allocation in a Multiple-Queen Ant Species. *Evolution*, 65(10):2907–2915, October 2011. ISSN 1558-5646. doi: 10.1111/j.1558-5646.2011.01348.x. URL http://onlinelibrary.wiley.com.ezproxy3.lib.le.ac.uk/doi/10.1111/j.1558-5646.2011.01348.x/abstract.

Lillycrop, K. A., Hoile, S. P., Grenfell, L., and Burdge, G. C. DNA methylation, ageing and the influence of early life nutrition. *Proceedings of the Nutrition Society*, 73(03):413–421, August 2014. ISSN 1475-2719. doi: 10.1017/S0029665114000081. URL http://journals.cambridge.org/article_S0029665114000081.

Lockett, G. A., Helliwell, P., and Maleszka, R. Involvement of DNA methylation in memory processing in the honeybee. *Neuroreport*, 21(12):812–816, August 2010. ISSN 1473-558X. doi: 10.1097/WNR.0b013e32833ce5be.

Luedi, P. P., Dietrich, F. S., Weidman, J. R., Bosko, J. M., Jirtle, R. L., and Hartemink, A. J. Computational and experimental identification of novel human imprinted genes. *Genome Research*, 17(12):1723–1730, December 2007. ISSN 1088-9051. doi: 10.1101/gr.6584707.

Lutz, C. C., Rodriguez-Zas, S. L., Fahrbach, S., and Robinson, G. E. Transcriptional response to foraging experience in the honeybee mushroom bodies. *Developmental Neurobiology*, 72(2):153–166, February 2012. ISSN 1932-8451. doi: 10.1002/dneu.20929. URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3256269/.

Lyko, F. and Maleszka, R. Insects as innovative models for functional studies of DNA methylation. *Trends in Genetics*, 27(4):127–131, January 2011. ISSN 0168-9525. doi: 10.1016/j.tig.2011.01.003. URL http://www.cell.com/article/S0168952511000047/abstract.

Lyko, F., Foret, S., Kucharski, R., Wolf, S., Falckenhayn, C., and Maleszka, R. The honey bee epigenomes: Differential methylation of brain DNA in queens and workers. *PLoS Biol*, 8(11):e1000506, November 2010. doi: 10.1371/journal.pbio.1000506. URL http://dx.doi.org/10.1371/journal.pbio.1000506.

Lytle, J. R., Yario, T. A., and Steitz, J. A. Target mRNAs are repressed as efficiently by microRNA-binding sites in the 5' UTR as in the 3' UTR. *Proceedings of the*

National Academy of Sciences of the United States of America, 104(23):9667–9672, June 2007. ISSN 0027-8424. doi: 10.1073/pnas.0703820104.

Magnani, R., Dirk, L. M. A., Trievel, R. C., and Houtz, R. L. Calmodulin methyltransferase is an evolutionarily conserved enzyme that trimethylates Lys-115 in calmodulin. *Nature Communications*, 1:43, July 2010. doi: 10.1038/ncomms1044. URL http://www.nature.com/ncomms/journal/v1/n4/full/ncomms1044.html.

Mandrioli, M. and Volpi, N. The genome of the lepidopteran *Mamestra brassicae* has a vertebrate-like content of methyl-cytosine. *Genetica*, 119(2):187–191, October 2003. ISSN 0016-6707.

Markesich, D. C., Gajewski, K. M., Nazimiec, M. E., and Beckingham, K. bicaudal encodes the *Drosophila* beta NAC homolog, a component of the ribosomal translational machinery. *Development (Cambridge, England)*, 127(3):559–572, February 2000. ISSN 0950-1991.

Maunakea, A. K., Nagarajan, R. P., Bilenky, M., Ballinger, T. J., D'Souza, C., Fouse, S. D., Johnson, B. E., Hong, C., Nielsen, C., Zhao, Y., Turecki, G., Delaney, A., Varhol, R., Thiessen, N., Shchors, K., Heine, V. M., Rowitch, D. H., Xing, X., Fiore, C., Schillebeeckx, M., Jones, S. J. M., Haussler, D., Marra, M. A., Hirst, M., Wang, T., and Costello, J. F. Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature*, 466(7303):253–257, July 2010. ISSN 0028-0836. doi: 10.1038/nature09165. URL http://www.nature.com/nature/journal/v466/n7303/full/nature09165.html.

Maynard-Smith, J. and Szathmary, E. *The major transitions in evolution*. Oxford, UK: Oxford University Press., 1995.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M. A. The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, 20(9):1297–1303, September 2010. ISSN 1088-9051, 1549-5469. doi: 10.1101/gr.107524.110. URL http://genome.cshlp.org/content/20/9/1297.

Meissner, A., Gnirke, A., Bell, G. W., Ramsahoye, B., Lander, E. S., and Jaenisch, R. Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Research*, 33(18):5868–5877, 2005. ISSN 0305-1048. doi: 10.1093/nar/gki901. URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1258174/.

Messerschmidt, D. M., Knowles, B. B., and Solter, D. DNA methylation dynamics during epigenetic reprogramming in the germline and preimplantation embryos. *Genes & Development*, 28(8):812–828, April 2014. ISSN 0890-9369, 1549-5477. doi: 10.1101/gad.234294.113. URL http://genesdev.cshlp.org/content/28/8/812.

Metcalf, R. A. Sex ratios, Parent-offspring conflict, and Local competition for mates in the social wasps *Polistes metricus* and *Polistes variatus*. *American Naturalist - AMER NATURALIST*, 116(5), 1980. doi: 10.1086/283655.

Milani, L., Lundmark, A., Nordlund, J., Kiialainen, A., Flaegstad, T., Jonmundsson, G., Kanerva, J., Schmiegelow, K., Gunderson, K. L., Lonnerholm, G., and Syvanen, A.-C. Allele-specific gene expression patterns in primary leukemic cells reveal regulation of gene expression by CpG site methylation. *Genome Research*, 19(1): 1–11, January 2009. ISSN 1088-9051, 1549-5469. doi: 10.1101/gr.083931.108. URL http://genome.cshlp.org/content/19/1/1.

Moczek, A. P. Phenotypic plasticity and diversity in insects. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1540):593–603, February 2010. ISSN 0962-8436. doi: 10.1098/rstb.2009.0263. URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2817146/.

Mohn, F., Weber, M., Schubeler, D., and Roloff, T. C. Methylated DNA immunoprecipitation (MeDIP). *Methods in Molecular Biology (Clifton, N.J.)*, 507:55–64, 2009. ISSN 1064-3745. doi: 10.1007/978-1-59745-522-0_5.

Mueller, U. G., Eickwort, G. C., and Aquadro, C. F. DNA fingerprinting analysis of parent-offspring conflict in a bee. *Proceedings of the National Academy of Sciences of the United States of America*, 91(11):5143–5147, May 1994. ISSN 0027-8424. URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC43948/.

Nielsen, D. A., Novoradovsky, A., and Goldman, D. SSCP primer design based on single-strand DNA structure predicted by a DNA folding program. *Nucleic Acids Research*, 23(12):2287–2291, 1995. ISSN 0305-1048, 1362-4962. doi: 10.1093/nar/23.12.2287. URL zotero://attachment/123/.

Nygaard, S., Zhang, G., Schiott, M., Li, C., Wurm, Y., Hu, H., Zhou, J., Ji, L., Qiu, F., Rasmussen, M., Pan, H., Hauser, F., Krogh, A., Grimmelikhuijzen, C. J. P., Wang, J., and Boomsma, J. J. The genome of the leaf-cutting ant *Acromyrmex echinatior* suggests key adaptations to advanced social life and fungus farming. *Genome Research*, 21(8):1339–1348, August 2011. ISSN 1088-9051, 1549-5469. doi: 10.1101/gr.121392.111. URL http://genome.cshlp.org/content/21/8/1339.

Ogawa, K. and Miura, T. Aphid polyphenisms: trans-generational developmental regulation through viviparity. *Frontiers in Physiology*, 5, January 2014. ISSN 1664-042X. doi: 10.3389/fphys.2014.00001. URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3900772/.

Okae, H., Hiura, H., Nishida, Y., Funayama, R., Tanaka, S., Chiba, H., Yaegashi, N., Nakayama, K., Sasaki, H., and Arima, T. Re-investigation and RNA sequencing-based identification of genes with placenta-specific imprinted expression. *Human Molecular Genetics*, 21(3):548–558, February 2012. ISSN 1460-2083. doi: 10.1093/hmg/ddr488.

Oldroyd, B. P. and Fewell, J. H. Genetic diversity promotes homeostasis in insect colonies. *Trends in Ecology & Evolution*, 22(8):408–413, August 2007. ISSN 0169-5347. doi: 10.1016/j.tree.2007.06.001.

Oldroyd, B. P., Allsopp, M. H., Roth, K. M., Remnant, E. J., Drewell, R. A., and Beekman, M. A parent-of-origin effect on honeybee worker ovary size. *Proceedings. Biological Sciences / The Royal Society*, 281(1775):20132388, January 2014. ISSN 1471-2954. doi: 10.1098/rspb.2013.2388.

Olek, A. and Walter, J. The pre-implantation ontogeny of the H19 methylation imprint. *Nature genetics*, 17(3):275–276, 1997. URL http://www.scopus.com/inward/record.url?eid=2-s2.0-0031279892&

partnerID=40&md5=df11f39fd3d2af1aac5dcc66b0b82f71. cited By (since 1996)152.

Ott, S. R. and Rogers, S. M. Gregarious desert locusts have substantially larger brains with altered proportions compared with the solitarious phase. *Proceedings. Biological Sciences / The Royal Society*, 277(1697):3087–3096, October 2010. ISSN 1471-2954. doi: 10.1098/rspb.2010.0694.

Pamilo, P. Evolution of colony characteristics in social insects. II. number of reproductive individuals. *The American Naturalist*, 138(2):412–433, 1991.

Pardo-Manuel de Villena, F., de la Casa-Esperon, E., and Sapienza, C. Natural selection and the function of genome imprinting: beyond the silenced minority. *Trends in genetics: TIG*, 16(12):573–579, December 2000. ISSN 0168-9525.

Park, J. M., Kunieda, T., and Kubo, T. The activity of Mblk-1, a Mushroom Body-selective Transcription Factor from the honeybee, is modulated by the Ras/MAPK pathway. *Journal of Biological Chemistry*, 278(20):18689–18694, 2003. doi: 10.1074/jbc.m300486200. URL http://dx.doi.org/10.1074/jbc.m300486200.

Patalano, S., Hore, T. A., Reik, W., and Sumner, S. Shifting behaviour: epigenetic reprogramming in eusocial insects. *Current Opinion in Cell Biology*, 24(3):367 – 373, 2012. ISSN 0955-0674. doi: http://dx.doi.org/10.1016/j.ceb.2012.02.005. URL http://www.sciencedirect.com/science/article/pii/S0955067412000312. Nucleus and gene expression.

Paxton, R. J., Bego, L. R., Shah, M. M., and Mateus, S. Low mating frequency of queens in the stingless bee *Scaptotrigona postica 1* and worker maternity of males. *Behavioral Ecology and Sociobiology*, 53(3):174–181, February 2003. ISSN 0340-5443, 1432-0762. doi: 10.1007/s00265-002-0561-4. URL http://link.springer.com/article/10.1007/s00265-002-0561-4.

Pener, M. P. and Simpson, S. J. Locust phase polyphenism: An update. *Advances in Insect Physiology - ADVAN INSECT PHYSIOL*, 36:1–272, 2009. doi: 10.1016/S0065-2806(08)36001-9.

Pereboom, J. J. M., Jordan, W. C., Sumner, S., Hammond, R. L., and Bourke, A. F. G. Differential gene expression in queen-worker caste determination in

bumble-bees. *Proceedings of the Royal Society B: Biological Sciences*, 272(1568): 1145–1152, June 2005. ISSN 0962-8452, 1471-2954. doi: 10.1098/rspb.2005.3060. URL http://rspb.royalsocietypublishing.org/content/272/1568/1145.

Pérez-Figueroa, A. MSAP: a tool for the statistical analysis of methylation-sensitive amplified polymorphism data. *Molecular Ecology Resources*, 13(3):522–527, 2013. ISSN 1755-0998. doi: 10.1111/1755-0998.12064. URL http://dx.doi.org/10.1111/1755-0998.12064.

Pinney, S. E. Mammalian Non-CpG Methylation: Stem cells and Beyond. *Biology*, 3(4):739–751, November 2014. ISSN 2079-7737. doi: 10.3390/biology3040739. URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4280509/.

Pliushch, G., Schneider, E., Weise, D., El Hajj, N., Tresch, A., Seidmann, L., Coerdt, W., Muller, A. M., Zechner, U., and Haaf, T. Extreme methylation values of imprinted genes in human abortions and stillbirths. *The American Journal of Pathology*, 176(3):1084–1090, March 2010. ISSN 0002-9440. doi: 10.2353/ajpath.2010.090764. URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2832130/.

Prokhortchouk, E. and Hendrich, B. Methyl-CpG binding proteins and cancer: are MeCpGs more important than MBDs? *doi:10.1038/sj.onc.1205631*, 21 (35), August 2002. doi: 10.1038/sj.onc.1205631. URL http://www.nature.com.ezproxy4.lib.le.ac.uk/onc/journal/v21/n35/full/1205631a.html.

Prys-Jones, O. and Corbet, S. *Bumblebees Naturalists' Handbooks 6*. The Richmond Publishing Co. Ltd., England., 2003.

Queller, D. C. Theory of genomic imprinting conflict in social insects. *BMC Evolutionary Biology*, 3:art. no.–15, 2003. URL ://000188122100015.

Queller, D. C. and Strassmann, J. E. The many selves of social insects. *Science*, 296:311–313, April 2002. URL ://000175000300046. 5566.

Queller, D. C. and Strassmann, J. E. Beyond society: the evolution of organismality. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1533): 3143–3155, November 2009. ISSN 0962-8436, 1471-2970. doi: 10.1098/rstb.2009.0095. URL http://rstb.royalsocietypublishing.org/content/364/1533/3143.

Rakyan, V. K., Down, T. A., Thorne, N. P., Flicek, P., Kulesha, E., Graf, S., Tomazou, E. M., Backdahl, L., Johnson, N., Herberth, M., Howe, K. L., Jackson, D. K., Miretti, M. M., Fiegler, H., Marioni, J. C., Birney, E., Hubbard, T. J. P., Carter, N. P., Tavare, S., and Beck, S. An integrated resource for genome-wide identification and analysis of human tissue-specific differentially methylated regions (tDMRs). *Genome Research*, 18(9):1518–1529, September 2008. ISSN 1088-9051, 1549-5469. doi: 10.1101/gr.077479.108. URL http://genome.cshlp.org/content/18/9/1518.

Ratnieks, F. L. W. Egg-laying, egg-removal, and ovary development by workers in queenright honey bee colonies. *Behavioral Ecology and Sociobiology*, 32(3):191–198, March 1993. ISSN 0340-5443, 1432-0762. doi: 10.1007/BF00173777. URL http://link.springer.com/article/10.1007/BF00173777.

Ratnieks, F. L. W. Reproductive harmony via mutual policing by workers in Eusocial Hymenoptera. *American Naturalist*, 132(2), 1988. doi: 10.1086/284846.

Ratnieks, F. L. W. and Helantera, H. The evolution of extreme altruism and inequality in insect societies. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1533):3169–3179, November 2009. ISSN 0962-8436. doi: 10.1098/rstb.2009.0129. URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2781879/.

Ratnieks, F. L., Foster, K. R., and Wenseleers, T. Conflict Resolution in Insect Societies. *Annual Review of Entomology*, 51(1):581–608, 2006. doi: 10.1146/annurev.ento.51.110104.151003. URL http://www.annualreviews.org/doi/abs/10.1146/annurev.ento.51.110104.151003.

Reik, W. and Walter, J. Genomic imprinting: parental influence on the genome. *Nature Reviews Genetics*, 2(1):21–32, January 2001. ISSN 1471-0056. doi: 10.1038/35047554. URL http://www.nature.com.ezproxy4.lib.le.ac.uk/nrg/journal/v2/n1/full/nrg0101_021a.html.

Roach, H. I. and Hashimoto, K. PCR-based methods to determine DNA methylation status at specific CpG sites using methylation-sensitive restriction enzymes. 2007. URL http://eprints.soton.ac.uk/66394/3/indexcodes.txt.

Roessingh, P. and Simpson, S. J. The time-course of behavioural phase change in nymphs of the desert locust, *Schistocerca gregaria*. *Physiological Entomology*, 19(3):191–197, 1994. ISSN 1365-3032. doi: 10.1111/j.1365-3032.1994.tb01042.x. URL http://onlinelibrary.wiley.com/doi/10.1111/j.1365-3032.1994.tb01042.x/abstract.

Roessingh, P., Simpson, S. J., and James, S. Analysis of phase-related changes in behaviour of Desert locust nymphs. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 252(1333):43–49, April 1993. ISSN 0962-8452, 1471-2954. doi: 10.1098/rspb.1993.0044. URL http://rspb.royalsocietypublishing.org/content/252/1333/43.

Rogers, S. M., Matheson, T., Despland, E., Dodgson, T., Burrows, M., and Simpson, S. J. Mechanosensory-induced behavioural gregarization in the desert locust *Schistocerca gregaria*. *The Journal of Experimental Biology*, 206(Pt 22):3991–4002, November 2003. ISSN 0022-0949.

Ross, K. G. and Fletcher, D. J. C. Genetic origin of male diploidy in the fire ant, *Solenopsis invicta* (Hymenoptera: Formicidae), and its evolutionary significance. *Evolution*, 39(4):888, July 1985. ISSN 00143820. doi: 10.2307/2408688. URL http://www.jstor.org/discover/10.2307/2408688?uid=3738032&uid=2129&uid=2&uid=70&uid=4&sid=21104680324491.

Rothberg, J. M., Jacobs, J. R., Goodman, C. S., and Artavanis-Tsakonas, S. slit: an extracellular protein necessary for development of midline glia and commissural axon pathways contains both EGF and LRR domains. *Genes & Development*, 4 (12a):2169–2187, December 1990. ISSN 0890-9369, 1549-5477. doi: 10.1101/gad.4.12a.2169. URL http://genesdev.cshlp.org/content/4/12a/2169.

Roxstrom-Lindquist, K. and Faye, I. The *Drosophila* gene Yippee reveals a novel family of putative zinc binding proteins highly conserved among eukaryotes. *Insect Molecular Biology*, 10(1):77–86, February 2001. ISSN 0962-1075.

Roxstrom-Lindquist, K., Assefaw-Redda, Y., Rosinska, K., and Faye, I. 20-hydroxyecdysone indirectly regulates hemolin gene expression in *Hyalophora cecropia*. *Insect Molecular Biology*, 14(6):645–652, December 2005. ISSN 1365-2583. doi: 10.1111/j.1365-2583.2005.00593.x. URL http://onlinelibrary.wiley.com/doi/10.1111/j.1365-2583.2005.00593.x/abstract.

Salvaing, J., Lopez, A., Boivin, A., Deutsch, J. S., and Peronnet, F. The *Drosophila* corto protein interacts with polycomb-group proteins and the GAGA factor. *Nucleic Acids Research*, 31(11):2873–2882, June 2003. ISSN 1362-4962.

Sambrook, J. and Russell, D. *Molecular Cloning: A laboratory manual*, volume 3. Cold Spring Harbor Laboratory Press, New York, 3rd edition, 2001.

Sarda, S., Zeng, J., Hunt, B. G., and Yi, S. V. The evolution of invertebrate gene body methylation. *Molecular Biology and Evolution*, page mss062, February 2012. ISSN 0737-4038, 1537-1719. doi: 10.1093/molbev/mss062. URL http://mbe.oxfordjournals.org/content/early/2012/02/10/molbev.mss062.

Sassa, T. and Kihara, A. Metabolism of very long-chain fatty acids: Genes and pathophysiology. *Biomolecules and Therapeutics*, 22(2):83–92, 2014. ISSN 1976-9148. doi: 10.4062/biomolther.2014.017.

Scharf, M. E., Wu-Scharf, D., Pittendrigh, B. R., and Bennett, G. W. Caste and development-associated gene expression in a lower termite. *Genome Biology*, 4 (10):R62, September 2003. ISSN 1465-6906. doi: 10.1186/gb-2003-4-10-r62. URL http://genomebiology.com/2003/4/10/R62/abstract.

Schrey, A. W., Alvarez, M., Foust, C. M., Kilvitis, H. J., Lee, J. D., Liebl, A. L., Martin, L. B., Richards, C. L., and Robertson, M. Ecological Epigenetics: Beyond MS-AFLP. *Integrative and Comparative Biology*, April 2013. doi: 10.1093/icb/ict012. URL http://icb.oxfordjournals.org/content/early/2013/04/12/icb.ict012.abstract.

Schwarz, M. P., Tierney, S. M., Rehan, S. M., Chenoweth, L. B., and Cooper, S. J. B. The evolution of eusociality in allodapine bees: workers began by waiting. *Biology Letters*, 7(2):277–280, April 2011. doi: 10.1098/rsbl.2010.0757. URL http://rsbl.royalsocietypublishing.org/content/7/2/277.abstract.

Simola, D. F., Wissler, L., Donahue, G., Waterhouse, R. M., Helmkampf, M., Roux, J., Nygaard, S., Glastad, K. M., Hagen, D. E., Viljakainen, L., Reese, J. T., Hunt, B. G., Graur, D., Elhaik, E., Kriventseva, E. V., Wen, J., Parker, B. J., Cash, E., Privman, E., Childers, C. P., Munoz-Torres, M. C., Boomsma, J. J., Bornberg-Bauer, E., Currie, C. R., Elsik, C. G., Suen, G., Goodisman, M. A., Keller, L., Liebig, J., Rawls, A., Reinberg, D., Smith, C. D., Smith, C. R., Tsutsui, N., Wurm, Y., Zdobnov, E. M., Berger, S. L., and Gadau, J. Social insect genomes exhibit dramatic evolution in gene composition and regulation while preserving regulatory features linked to sociality. *Genome Research*, 23(8):1235–1247, August 2013. ISSN 1088-9051. doi: 10.1101/gr.155408.113. URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3730098/.

Simpson, S. J., McCaffery, A. R., and Hagele, B. F. A behavioural analysis of phase change in the desert locust. *Biological Reviews*, 74(4):461–480, 1999. ISSN 1469-185X. doi: 10.1111/j.1469-185X.1999.tb00038.x. URL http://onlinelibrary.wiley.com/doi/10.1111/j.1469-185X.1999.tb00038.x/abstract.

Simpson, S. J., Sword, G. A., and Lo, N. Polyphenism in insects. *Current biology: CB*, 21(18):R738–749, September 2011. ISSN 1879-0445. doi: 10.1016/j.cub.2011.06.006.

Smagghe, G., Deknopper, J., Meeus, I., and Mommaerts, V. Dietary chlorantraniliprole suppresses reproduction in worker bumblebees. *Pest Management Science*, 69(7):787–791, 2013. ISSN 1526-4998. doi: 10.1002/ps.3504. URL http://onlinelibrary.wiley.com/doi/10.1002/ps.3504/abstract.

Smith, C. R., Smith, C. D., Robertson, H. M., Helmkampf, M., Zimin, A., Yandell, M., Holt, C., Hu, H., Abouheif, E., Benton, R., Cash, E., Croset, V., Currie, C. R., Elhaik, E., Elsik, C. G., FavÃl, M.-J., Fernandes, V., Gibson, J. D., Graur, D., Gronenberg, W., Grubbs, K. J., Hagen, D. E., Viniegra, A. S. I., Johnson, B. R., Johnson, R. M., Khila, A., Kim, J. W., Mathis, K. A., Munoz-Torres, M. C., Murphy, M. C., Mustard, J. A., Nakamura, R., Niehuis, O., Nigam, S., Overson, R. P., Placek, J. E., Rajakumar, R., Reese, J. T., Suen, G., Tao, S., Torres, C. W., Tsutsui, N. D., Viljakainen, L., Wolschin, F., and Gadau, J. Draft genome of the red harvester ant *Pogonomyrmex barbatus*. *Proceedings of the National Academy of Sciences*, 108(14):5667–5672, January 2011a. ISSN 0027-8424, 1091-6490. doi:

10.1073/pnas.1007901108. URL http://www.pnas.org/content/early/2011/01/24/1007901108.

Smith, C. D., Zimin, A., Holt, C., Abouheif, E., Benton, R., Cash, E., Croset, V., Currie, C. R., Elhaik, E., Elsik, C. G., Fave, M.-J., Fernandes, V., Gadau, J., Gibson, J. D., Graur, D., Grubbs, K. J., Hagen, D. E., Helmkampf, M., Holley, J.-A., Hu, H., Viniegra, A. S. I., Johnson, B. R., Johnson, R. M., Khila, A., Kim, J. W., Laird, J., Mathis, K. A., Moeller, J. A., MuÃśoz-Torres, M. C., Murphy, M. C., Nakamura, R., Nigam, S., Overson, R. P., Placek, J. E., Rajakumar, R., Reese, J. T., Robertson, H. M., Smith, C. R., Suarez, A. V., Suen, G., Suhr, E. L., Tao, S., Torres, C. W., Wilgenburg, E. v., Viljakainen, L., Walden, K. K. O., Wild, A. L., Yandell, M., Yorke, J. A., and Tsutsui, N. D. Draft genome of the globally widespread and invasive Argentine ant (*Linepithema humile*). *Proceedings of the National Academy of Sciences*, 108(14):5673–5678, January 2011b. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1008617108. URL http://www.pnas.org/content/early/2011/01/26/1008617108.

Smith, Z. D. and Meissner, A. DNA methylation: roles in mammalian development. *Nature Reviews Genetics*, 14(3):204–220, March 2013. ISSN 1471-0056. doi: 10.1038/nrg3354. URL http://www.nature.com.ezproxy3.lib.le.ac.uk/nrg/journal/v14/n3/full/nrg3354.html.

Srinivasan, D. G. and Brisson, J. A. Aphids: A model for polyphenism and epigenetics. *Genetics Research International*, 2012, 2012. ISSN 2090-3154. doi: 10.1155/2012/431531. URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3335499/.

Steyaert, W. V. C., sandra. SNP-guided identification of monoallelic DNA-methylation events from enrichment-based sequencing data. *Nucleic acids research*, 2014. ISSN 0305-1048. doi: 10.1093/nar/gku847.

Strachecka, A., Paleolog, J., Borsuk, G., Olszewski, K., and Bajda, M. DNA methylation in the honey bee (*Apis mellifera*) and its importance for biological research. *Medycyna Weterynaryjna*, 68(7):391–396, 2012.

Suen, G., Teiling, C., Li, L., Holt, C., Abouheif, E., Bornberg-Bauer, E., Bouffard, P., Caldera, E. J., Cash, E., Cavanaugh, A., Denas, O., Elhaik, E., Fave, M. J.,

Gadau, J., Gibson, J. D., Graur, D., Grubbs, K. J., Hagen, D. E., Harkins, T. T., Helmkampf, M., Hu, H., Johnson, B. R., Kim, J., Marsh, S. E., Moeller, J. A., Munoz-Torres, M. C., Murphy, M. C., Naughton, M. C., Nigam, S., Overson, R., Rajakumar, R., Reese, J. T., Scott, J. J., Smith, C. R., Tao, S., Tsutsui, N. D., Viljakainen, L., Wissler, L., Yandell, M. D., Zimmer, F., Taylor, J., Slater, S. C., Clifton, S. W., Warren, W. C., Elsik, C. G., Smith, C. D., Weinstock, G. M., Gerardo, N. M., and Currie, C. R. The genome sequence of the Leaf-cutter ant *Atta cephalotes* reveals insights into its obligate symbiotic lifestyle. *PLoS Genet*, 7(2):e1002007, February 2011. doi: 10.1371/journal.pgen.1002007. URL http://dx.doi.org/10.1371/journal.pgen.1002007.

Suzuki, M. M. and Bird, A. DNA methylation landscapes: provocative insights from epigenomics. *Nat Rev Genet*, 9(6):465–476, June 2008. ISSN 1471-0056. doi: 10.1038/nrg2341. URL http://dx.doi.org/10.1038/nrg2341.

Szyf, M., McGowan, P., and Meaney, M. J. The social environment and the epigenome. *Environmental and Molecular Mutagenesis*, 49(1):46–60, January 2008. ISSN 0893-6692. doi: 10.1002/em.20357.

Takai, D. and Jones, P. A. Comprehensive analysis of CpG islands in human chromosomes 21 and 22. *Proceedings of the National Academy of Sciences*, 99(6):3740–3745, March 2002. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.052410099. URL http://www.pnas.org/content/99/6/3740.

Takeuchi, H., Paul, R. K., Matsuzaka, E., and Kubo, T. EcR- A expression in the brain and ovary of the honeybee (*Apis mellifera L.*). *Zoological Science*, 24(6): 596–603, June 2007. ISSN 0289-0003. doi: 10.2108/zsj.24.596.

Thompson, G. J., Yockey, H., Lim, J., and Oldroyd, B. P. Experimental manipulation of ovary activation and gene expression in honey bee (*Apis mellifera*) queens and workers: testing hypotheses of reproductive regulation. *Journal of Experimental Zoology Part A: Ecological Genetics and Physiology*, 307A (10):600–610, October 2007. ISSN 1932-5231. doi: 10.1002/jez.415. URL http://onlinelibrary.wiley.com/doi/10.1002/jez.415/abstract.

Toth, A. L., Tooker, J. F., Radhakrishnan, S., Minard, R., Henshaw, M. T., and Grozinger, C. M. Shared genes related to aggression, rather than chemical communication, are associated with reproductive dominance in paper wasps (*Polistes metricus*). *BMC Genomics*, 15(1):75, January 2014. ISSN 1471-2164. doi: 10.1186/1471-2164-15-75. URL http://www.biomedcentral.com/1471-2164/15/75/abstract.

Trivers, R. L. and Hare, H. Haploidiploidy and the evolution of the social insect. *Science (New York, N.Y.)*, 191(4224):249–263, January 1976. ISSN 0036-8075.

Trivers, R. Parent-offspring conflict. *American Zoologist*, 14(1):249–264, 1974. ISSN 0003-1569. WOS:A1974T008500016.

Uvarov, B. *Grasshoppers and Locusts*, volume 1 of *A Handbook of General Acridology*. Cambridge University Press, Cambridge, 1966.

van Honk, C. G. J., Roseler, P., Velthuis, H. H. W., and Hoogeveen, J. C. Factors influencing the egg laying of workers in a captive *Bombus terrestris* colony. *Behavioral Ecology and Sociobiology*, 9(1):9–14, August 1981. ISSN 0340-5443, 1432-0762. doi: 10.1007/BF00299847. URL http://link.springer.com.ezproxy4.lib.le.ac.uk/article/10.1007/BF00299847.

van Honk, C. and Hogeweg, P. The ontogeny of the social structure in a captive *Bombus terrestris* colony. *Behavioral Ecology and Sociobiology*, 9(2):111–119, 1981. ISSN 0340-5443. doi: 10.1007/BF00293582. URL http://dx.doi.org/10.1007/BF00293582.

Van Oystaeyen, A., Oliveira, R. C., Holman, L., van Zweden, J. S., Romero, C., Oi, C. A., d'Ettorre, P., Khalesi, M., Billen, J., WÃďckers, F., Millar, J. G., and Wenseleers, T. Conserved class of queen pheromones stops social insect workers from reproducing. *Science*, 343(6168):287–290, January 2014. doi: 10.1126/science.1244899. URL http://www.sciencemag.org/content/343/6168/287.abstract.

Vandegehuchte, M. B. and Janssen, C. R. Epigenetics and its implications for ecotoxicology. *Ecotoxicology (London, England)*, 20(3):607–624, May 2011. ISSN 1573-3017. doi: 10.1007/s10646-011-0634-0.

Varmuza, S. and Mann, M. Genomic imprinting - defusing the ovarian time bomb. *Trends in Genetics*, 10(4):118–123, April 1994. ISSN 0168-9525. doi: 10.1016/0168-9525(94)90212-7. URL http://www.sciencedirect.com/science/article/pii/0168952594902127.

Varriale, A. DNA methylation, epigenetics, and evolution in vertebrates: Facts and challenges. *International Journal of Evolutionary Biology*, 2014:e475981, January 2014. ISSN 2090-8032. doi: 10.1155/2014/475981. URL http://www.hindawi.com/journals/ijeb/2014/475981/abs/.

Vergoz, V., Lim, J., and Oldroyd, B. P. Biogenic amine receptor gene expression in the ovarian tissue of the honey bee *Apis mellifera*. *Insect Molecular Biology*, 21 (1):21–29, February 2012. ISSN 1365-2583. doi: 10.1111/j.1365-2583.2011.01106.x. URL http://onlinelibrary.wiley.com/doi/10.1111/j.1365-2583.2011.01106.x/abstract.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., van de Lee, T., Hornes, M., Frijters, A., Pot, J., Peleman, J., and Kuiper, M. AFLP: A new technique for DNA fingerprinting. *Nucleic Acids Research*, 23(21):4407–4414, November 1995. ISSN 0305-1048. URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC307397/.

Wang, S. Loss of imprinting of IGF2 correlates with hypomethylation of the H19 differentially methylated region in the tumor tissue of colorectal cancer patients. *Molecular Medicine Reports*, March 2012. ISSN 1791-2997, 1791-3004. doi: 10.3892/mmr.2012.833. URL http://www.spandidos-publications.com/mmr/5/6/1536.

Wang, W., Zhao, X., Pan, Y., Zhu, L., Fu, B., and Li, Z. DNA methylation changes detected by methylation-sensitive amplified polymorphism in two contrasting rice genotypes under salt stress. *Journal of Genetics and Genomics*, 38(9):419 – 424, 2011. ISSN 1673-8527. doi: http://dx.doi.org/10.1016/j.jgg.2011.07.006. URL http://www.sciencedirect.com/science/article/pii/S1673852711001305.

Wang, X. and Clark, A. G. Using next-generation RNA sequencing to identify imprinted genes. *Heredity*, 113(2):156–166, August 2014. ISSN 0018-067X. doi: 10.1038/hdy.2014.18. URL http://www.nature.com/hdy/journal/v113/n2/abs/hdy201418a.html.

Wang, X. and Kang, L. Molecular mechanisms of phase change in locusts. *Annual Review of Entomology*, 59(1):225–244, 2014. doi: 10.1146/annurev-ento-011613-162019. URL http://dx.doi.org/10.1146/annurev-ento-011613-162019.

Wang, X., Fang, X., Yang, P., Jiang, X., Jiang, F., Zhao, D., Li, B., Cui, F., Wei, J., Ma, C., Wang, Y., He, J., Luo, Y., Wang, Z., Guo, X., Guo, W., Wang, X., Zhang, Y., Yang, M., Hao, S., Chen, B., Ma, Z., Yu, D., Xiong, Z., Zhu, Y., Fan, D., Han, L., Wang, B., Chen, Y., Wang, J., Yang, L., Zhao, W., Feng, Y., Chen, G., Lian, J., Li, Q., Huang, Z., Yao, X., Lv, N., Zhang, G., Li, Y., Wang, J., Wang, J., Zhu, B., and Kang, L. The locust genome provides insight into swarm formation and long-distance flight. *Nature Communications*, 5, January 2014. doi: 10.1038/ncomms3957. URL http://www.nature.com/ncomms/2014/140114/ncomms3957/full/ncomms3957.html.

Wang, Y., Jorda, M., Jones, P. L., Maleszka, R., Ling, X., Robertson, H. M., Mizzen, C. A., Peinado, M. A., and Robinson, G. E. Functional CpG Methylation System in a Social Insect. *Science*, 314(5799):645–647, October 2006. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1135213. URL http://www.sciencemag.org/content/314/5799/645.

Ward, K. N., Coleman, J. L., Clinnin, K., Fahrbach, S., and Rueppell, O. Age, caste, and behavior determine the replicative activity of intestinal stem cells in honeybees (*Apis mellifera L.*). *Experimental Gerontology*, 43(6):530 – 537, 2008. ISSN 0531-5565. doi: http://dx.doi.org/10.1016/j.exger.2008.03.012. URL http://www.sciencedirect.com/science/article/pii/S0531556508000922.

Weber, F., Fukino, K., Villalona-Calero, M., and Eng, C. Limitations of single-strand conformation polymorphism analysis as a high-throughput method for the detection of EGFR mutations in the clinical setting. *Journal of Clinical Oncology*, 23(24):5847–5848, August 2005. ISSN 0732-183X, 1527-7755. doi: 10.1200/JCO.2005.01.5222. URL http://jco.ascopubs.org/content/23/24/5847.

Weidner, C. I., Lin, Q., Koch, C. M., Eisele, L., Beier, F., Ziegler, P., Bauerschlag, D. O., Jockel, K.-H., Erbel, R., Muhleisen, T. W., Zenke, M., Brummendorf, T. H., and Wagner, W. Aging of blood can be tracked by DNA methylation changes at just three CpG sites. *Genome Biology*, 15(2):R24, February 2014.

ISSN 1465-6906. doi: 10.1186/gb-2014-15-2-r24. URL http://genomebiology.com/2014/15/2/R24/abstract.

Weiner, S. A. and Toth, A. L. Epigenetics in social insects: A new direction for understanding the evolution of castes. *Genetics Research International*, 2012: e609810, March 2012. ISSN 2090-3154. doi: 10.1155/2012/609810. URL http://www.hindawi.com/journals/gri/2012/609810/abs/.

Wenseleers, T. and Ratnieks, F. L. W. Comparative Analysis of Worker Reproduction and Policing in Eusocial Hymenoptera Supports Relatedness Theory. *The American Naturalist*, 168(6):E163–E179, December 2006a. ISSN 0003-0147. doi: 10.1086/508619. URL http://www.jstor.org/stable/10.1086/508619.

Wenseleers, T. and Ratnieks, F. L. W. Enforced altruism in insect societies. *Nature*, 444(7115):50–50, November 2006b. ISSN 0028-0836. doi: 10.1038/444050a. URL http://www.nature.com.ezproxy3.lib.le.ac.uk/nature/journal/v444/n7115/full/444050a.html.

Wenseleers, T., Helantera, H., Alves, D. A., Duenez-Guzman, E., and Pamilo, P. Towards greater realism in inclusive fitness models: the case of worker reproduction in insect societies. *Biology Letters*, 9(6):20130334, December 2013. ISSN 1744-9561, 1744-957X. doi: 10.1098/rsbl.2013.0334. URL http://rsbl.royalsocietypublishing.org/content/9/6/20130334.

Wilkins, J. F. and Haig, D. What good is genomic imprinting: The function of parent-specific gene expression. *Nature Reviews Genetics*, 4:359–368, May 2003. ISSN 1471-0056. URL ://000182664800016. 5.

Williamson, C. Mouse Imprinting Data and References (MRC Harwell, Oxfordshire), 2012. URL http://www.har.mrc.ac.uk/research/genomic_imprinting/.

Wilson, E. O. and Holldobler, B. Eusociality: Origin and consequences. *Proceedings of the National Academy of Sciences of the United States of America*, 102(38): 13367–13371, September 2005. doi: 10.1073/pnas.0505858102. URL http://www.pnas.org/content/102/38/13367.abstract.

Woyke, J. Sex determination. *Bee Genetics and Breeding*, pages 91–119, 1986.

Wurm, Y., Wang, J., Riba-Grognuz, O., Corona, M., Nygaard, S., Hunt, B. G., Ingram, K. K., Falquet, L., Nipitwattanaphon, M., Gotzek, D., Dijkstra, M. B., Oettler, J., Comtesse, F., Shih, C.-J., Wu, W.-J., Yang, C.-C., Thomas, J., Beaudoing, E., Pradervand, S., Flegel, V., Cook, E. D., Fabbretti, R., Stockinger, H., Long, L., Farmerie, W. G., Oakey, J., Boomsma, J. J., Pamilo, P., Yi, S. V., Heinze, J., Goodisman, M. A. D., Farinelli, L., Harshman, K., Hulo, N., Cerutti, L., Xenarios, I., Shoemaker, D., and Keller, L. The genome of the fire ant *Solenopsis invicta*. *Proceedings of the National Academy of Sciences*, 108(14):5679–5684, January 2011. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1009690108. URL http://www.pnas.org/content/early/2011/01/24/1009690108.

Yagound, B., Blacher, P., Chameron, S., and Chaline, N. Social context and reproductive potential affect worker reproductive decisions in a Eusocial insect. *PLoS ONE*, 7(12):e52217, 2012. doi: 10.1371/journal.pone.0052217. URL http://dx.doi.org/10.1371%2Fjournal.pone.0052217.

Yan, H., Simola, D. F., Bonasio, R., Liebig, J., Berger, S. L., and Reinberg, D. Eusocial insects as emerging models for behavioural epigenetics. *Nat Rev Genet*, 15(10):677–688, October 2014. ISSN 1471-0056. URL http://dx.doi.org/10.1038/nrg3787.

Zemach, A., McDaniel, I. E., Silva, P., and Zilberman, D. Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science*, 328(5980):916–919, May 2010. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1186366. URL http://www.sciencemag.org/content/328/5980/916.

Zeng, J. and Yi, S. V. DNA methylation and Genome evolution in honeybee: Gene length, Expression, Functional enrichment covary with the evolutionary signature of DNA methylation. *Genome Biology and Evolution*, 2:770–780, 2010. ISSN 1759-6653. doi: 10.1093/gbe/evq060. URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2975444/.

Zhang, B., Zhou, Y., Lin, N., Lowdon, R. F., Hong, C., Nagarajan, R. P., Cheng, J. B., Li, D., Stevens, M., Lee, H. J., Xing, X., Zhou, J., Sundaram, V., Elliott, G., Gu, J., Shi, T., Gascard, P., Sigaroudinia, M., Tlsty, T. D., Kadlecek, T., Weiss, A., O'Geen, H., Farnham, P. J., Maire, C. L., Ligon, K. L., Madden, P. A. F., Tam, A., Moore, R., Hirst, M., Marra, M. A., Zhang, B., Costello, J. F.,

and Wang, T. Functional DNA methylation differences between tissues, cell types, and across individuals discovered using the M&M algorithm. *Genome Research*, 23(9):1522–1540, September 2013. ISSN 1549-5469. doi: 10.1101/gr.156539.113.

Zheng, D., Liu, B.-B., Liu, Y.-K., Kang, X.-N., Sun, L., Guo, K., Sun, R.-X., Chen, J., and Zhao, Y. Analysis of the expression of Slit/Robo genes and the methylation status of their promoters in the hepatocellular carcinoma cell lines. *Zhonghua Gan Zang Bing Za Zhi = Zhonghua Ganzangbing Zazhi = Chinese Journal of Hepatology*, 17(3):198–202, March 2009. ISSN 1007-3418.

Zwier, M. V., Verhulst, E. C., Zwahlen, R. D., Beukeboom, L. W., and van de Zande, L. DNA methylation plays a crucial role during early *Nasonia development. Insect Molecular Biology*, 21(1):129–138, February 2012. ISSN 1365-2583. doi: 10.1111/j.1365-2583.2011.01121.x. URL http://onlinelibrary.wiley.com.ezproxy4.lib.le.ac.uk/doi/10.1111/j.1365-2583.2011.01121.x/abstract.