# The role of small RNAs in caste determination and differentiation in the bumble bee, *Bombus terrestris*

David Henry Collins BSc (Hons)

Thesis submitted for the degree of Doctor of Philosophy

University of East Anglia

School of Biological Sciences

September 2014

## Thesis Abstract

Investigating the molecular basis of queen-worker caste determination and differentiation in eusocial insects allows researchers to address fundamental aspects of evolution and development. In this regard, little is known of the role of gene regulation by microRNAs (miRNAs), non-coding RNA molecules that regulate development in animals and plants. In this thesis I therefore investigated the role of miRNAs in caste determination and differentiation in the bumble bee *Bombus terrestris*. In **Chapter 2**, I used deep sequencing (miRNA-seq) and bioinformatics to annotate miRNAs in *B. terrestris* and a second *Bombus* species, *B. impatiens*. I found that *B. terrestris* yielded 131 miRNAs and *B. impatiens* yielded 114 miRNAs and that, of these, 17 were new miRNAs that had not previously been sequenced in any species. In **Chapter 3**, using miRNA-seq and Northern blot analysis of female *B. terrestris* larvae, I isolated a miRNA (Bte-miR-6001) that was more highly expressed in queen- than in worker-destined larvae. This miRNA comprised an entire intron of the gene *Very high density lipoprotein* (*Vhdl*), which is homologous to the gene for the key storage protein vitellogenin. In **Chapter 4**, using miRNA-seq and Northern blot analysis of adult females, I isolated some miRNAs (e.g. Bte-miR-279b, Bte-miR-279c) that were more highly expressed in queens and reproductive workers and others (e.g. Bte-miR-184, Bte-miR-133) that were highly expressed in non-reproductive workers. Finally, in **Chapter 5**, using RNAi (RNA interference), I tested whether expression at the gene locus *foraging* (*for*) is associated with locomotory behaviour of *B. terrestris* foundress queens. Although RNAi did not induce gene knockdown, there was a positive association between *for* expression and queen reproduction. Overall, by isolating miRNAs associated with caste and reproduction in *B. terrestris*, the work reported in this thesis contributes substantially to our understanding of the molecular basis of caste in eusocial in insects.

## Chapter contributions and publication

Each chapter of this thesis has been written by David Collins, in consultation with Andrew Bourke, Tamas Dalmay and Matthew Beckers (and with Jacob Holland for **Chapter 5**). I have provided estimates of the approximate percentage contributions to the initial concepts, planning, conducting (data collection) and analyses of the experiments

***Chapter 2:***

Work from this chapter contributed to the following publication: Sadd BM, Barribeau SM, Bloch G, de Graff DC, Dearden P, Elsik CG, Gadau J, Grimmelikhuijzen CJP, Hasselmann M, Lozier JD et al. 2015. The genomes of two key bumblebee species with primitive eusocial organization. *Genome Biol* **16**:76.

- Concept: **David Collins 70%**; Andrew Bourke 20%; Tamas Dalmay 10%
- Planning: **David Collins 70%**; Matthew Beckers 10%; Andrew Bourke 10%; Tamas Dalmay 10%
- Conducting: **David Collins 100%** (Sanger sequencing service provided by the Genome Analysis Centre (TGAC), Norwich Research Park, Norwich, UK; miRNA-seq provided by BaseClear B.V, Leiden, The Netherlands)
- Analysis: **David Collins 60%**; Matthew Beckers 40%

***Chapter 3:***

Not yet published

- Concept: **David Collins 40%**; Andrew Bourke 30%; Tamas Dalmay 30%
- Planning: **David Collins 50%**; Andrew Bourke 30% Tamas Dalmay 20%
- Conducting: **David Collins 100%** (Sanger sequencing service provided by TGAC; miRNA-seq service provided by BaseClear B.V.)
- Analysis: **David Collins 50%**; Matthew Beckers 50%

***Chapter 4:***

Not yet published

- Concept: **David Collins 60%**; Andrew Bourke 30%; Tamas Dalmay 10%
- Planning: **David Collins 70%**; Andrew Bourke 30%
- Conducting: **David Collins 100%** (Sanger sequencing service provided by Eurofins Genomics, Ebersberg, Germany; miRNA-seq service provided by TGAC)
- Analysis: **David Collins 50%**; Matthew Beckers 50%

***Chapter 5:***

Not yet published

- Concept: Jacob Holland 40%; Andrew Bourke 40%; **David Collins 20%**
- Planning: **David Collins 40%**; Jacob Holland 40%; Andrew Bourke 20%
- Conducting: **David Collins 40%**; Jacob Holland 40%; Faaziah Chishty 20% (qPCR service provided by qStandard, UCL, London, UK)
- Analysis: Jacob Holland 80%; **David Collins 20%**

Table of contents

List of Figures and Tables

## Acknowledgments

I would like to thank my supervisor, Andrew Bourke, who has provided advice throughout my study at UEA, and for making suggestions that greatly improved this thesis. I would also like to thank my co-supervisor, Tamas Dalmay, who provided me with bench space, most of the reagents and equipment for the molecular work, and for his advice during numerous lab meetings throughout the course of my PhD. Both of these men have extensive knowledge and experience in their respective areas of expertise (eusocial insects and miRNAs), which have really helped improve the quality of this work.

I would like to thank the members of the Bourke and Dalmay labs. Karim Sorefan deserves special mention for being like a third supervisor, and for sharing his interesting perspectives on the philosophy of science. In the lab I am very grateful to Adam Hall, Carly Turnball, Guy Wheeler, and Ping Xu for all their help. I thank Matthew Beckers for his bioinformatics contributions, which helped uncover the majority of the results. Edd Almond, Jacob Holland, and Tim Huggins taught me how to handle bees, and have all been great friends.

I am privileged to work at a University with so many wonderful people. Several have helped me with this thesis and all have helped me to refine my understanding of biology and the philosophy of science in general. They include Liam Crowther, Matt Dickinson, 'Ents' Fairfield, Becca Franklin, Conrad Gillett, Danni Gilroy, Neil Holmes, Sarah Marburger, Osgur McDermott Long, Will Nash, Cock Van Oosterhaut, Karl Philips, Amy Romans, Damian Smith, Martin Sullivan, and many others to whom I will remain eternally grateful.

I would like to thank the University of East Anglia, which provided the facility and the majority of the funding for this PhD. I further thank the Pamela Salter Trust which provided me with funding for the sequencing in Chapter 4, and, together with the International Union for the Study of Social Insects, provided funding to travel to various National and International conferences with the social insect community. I would like to thank Ben Sadd, and the *Bombus* genome consortium for granting me access to the genome resources.

Finally I would like to thank Jenny Donelan, who has been much more than a friend, whose support has been nothing short of fantastic during my work on this thesis, and who I have little doubt will be a brilliant naturalist one day.

# Chapter 1

## Thesis Introduction

*Abstract*

In this chapter, I will review the ultimate causes for the evolution of eusociality; this will include an introduction to inclusive fitness theory and a summary showing why recent critiques of it have failed to topple it as the leading theory of eusocial evolution. I will also review the literature on social genomics and show why an approach that considers all aspects of gene regulation, including epigenetics and post-transcriptional regulation, is important to the study of eusociality. Following this, I provide an introduction to miRNAs, their role in phenotypic plasticity in general and eusociality in particular. I will review the literature on miRNAs in social insects and outline five major outstanding areas of research that I had aimed to address in this thesis. Finally I outline the chapters presented in this thesis, focussing on how they show that miRNAs are important in caste differentiation in *B. terrestris* and hence in the evolution of eusociality.

### *The evolution of eusociality*

Eusocial insects exhibit a reproductive division of labour, whereby some individuals in a social colony have a reproducer phenotype and others have a non-reproducer phenotype (Wilson 1971; Bourke 2011a). The reproducers, usually termed queens, perform most but not necessarily all of the reproduction while the non-reproducers, which are usually termed workers, carry out tasks that aid reproducers, allowing them to rear larger numbers of offspring. Often this is to the cost of the non-reproducer itself, and sometimes it leads to complete sterility. Therefore eusociality represents an extreme form of altruism. Eusocial societies are sometimes referred to as 'super-organisms' because they are held to represent the analogue of organisms at the social level, with all of the individuals that make up a social colony being seen as analogous to cells in the body of a multicellular organism (Tautz and Heilmann 2008; Holldobler and Wilson 2009). Another form of eusociality among unicellular organisms is represented by social amoebae such as *Dictyostelium*. These aggregate into colonies containing both reproductive cells that continue to divide and non-reproductive cells that provide structure and support for the reproductive cells

(Schaap et al. 2006; Bourke 2011a). Such primitive multicellular societies are thought to represent the early basis for the evolution of complex multicellular species in the major transitions view of evolution (Maynard Smith and Szathmary 1995; Bourke 2011a). *Dictyostelium*, clonal multicellular organisms and eusocial animals are all examples of phenotypic plasticity, whereby multiple phenotypes can be produced from the same genotype depending on the environment (West-Eberhard 1989; 2003). Eusocial insect castes are also referred to as polyphenisms, which is an extreme form of phenotypic plasticity in which the phenotypes irreversibly develop along different pathways towards discreet phenotypes that share a common genotype (Nijhout 2003; DeWitt and Scheiner 2004; Moczek 2010). Environmental caste determination is the process where development of an individual is irreversibly fixed into one of the two social insect castes, and is hence an example of a polyphenism (Bortolotti et al. 2001). Although some eusocial insects have genetic caste determination (where different genotypes define different castes), environmental caste determination is more common, and is likely to be the ancestral state in most eusocial lineages (Hughes et al. 2008; Smith et al. 2008; Schwander et al. 2010). Explaining how multiple phenotypes can be produced from a single genotype, and how polyphenisms can evolve, is an important problem in evolutionary biology (West-Eberhard 2003). Therefore understanding the evolution of eusociality might help to explain the evolution of polyphenisms more generally.

Current estimates are that eusociality has evolved at least 24 times in extant animals, making it a striking example of convergent evolution (Bourke 2011a). Animal taxa with eusocial species include: all of the termites (Thorne 1997); many lineages of Hymenoptera (in which it is estimated that eusociality has evolved at least ten times (Bourke 2011a)); several lineages of aphids in the Pemphigidae and Hormaphididae (Stern and Foster 1996); two species of ambrosia beetle, *Austroplatypus incompertus* (Kent and Simpson 1992) and *Xyleborinus saxeseni* (Biedermann and Taborsky 2011); thrips, Phlaeothripinae (Crespi 1992); marine sponge dwelling shrimps, *Synalpheus* (Duffy 1996); two species of mole rats, *Fukomys damarensis* (Jarvis and Bennett 1993) and *Heterocephalus glaber* (Jarvis 1981); at least one social spider *Anelosimus eximius* (Vollrath 1986); and most recently a trematode flatworm (Hechinger et al. 2011).

Many of these lineages have a sterile worker caste, but workers remain capable of evolving novel adaptations, an apparent paradox that Darwin referred to as the 'special difficulty' posed by the eusocial insects for his theory of natural selection (Darwin 1859; Ratnieks et al. 2011). Darwin's solution to this difficulty was to use an analogy with domestication, in which individuals can continue to evolve through breeding from siblings showing the same traits, even when they do not directly contribute to the next generation themselves (Darwin 1859).

Following Darwin, researchers have speculated about how altruism (a social behaviour that benefits recipients of the behaviour, to the cost of the actor) could have evolved in the first place. Approaches for explaining the evolution of eusociality may be usefully divided into those invoking ultimate factors, which concern the evolutionary and specifically the selective reasons for the origin of a trait, and those invoking proximate factors, which concern the mechanisms underpinning how a trait develops and functions within an individual's lifetime (Tinbergen 1963). The modern reductionist theory of evolution uses the 'gene's-eye view', making the important distinction that genes are replicators and phenotypes are 'vehicles' for replication, an idea that was formulated by Williams (1966) and Dawkins (1976; 1982). Altruism, in which a gene that causes individuals to reproduce less might be expected to be selected against, represents an apparently even more challenging paradox for evolutionary reasoning than Darwin's special difficulty (Ratnieks et al. 2011). A solution to this paradox was provided by Hamilton's inclusive fitness theory (Hamilton 1964a; 1964b), which explained the evolution of worker sterility and therefore of altruism more generally. Inclusive fitness theory (IFT), also referred to as kin selection, invokes the gene's-eye view via the concept that a gene for altruism can undergo selection if the gene increases the overall fitness of its co-bearers (Hamilton 1964a; Hamilton 1972). In other words, a gene that reduces its bearer's reproductive output is not necessarily selected against if it codes for a behaviour that increases the reproductive output of co-bearers (usually kin). Likewise, a gene that increases its bearer's reproductive output is not necessarily selected for if it codes for a behaviour that decreases its co-bearers' reproductive output (Hamilton 1964b). Hamilton's rule, $rb>c$ is a simple inequality that formalises this concept, where $b$ is the benefit of altruism (often measured in terms of offspring gained) to relatives of the bearer, $c$ is the cost to the bearer (often measured in

terms of offspring lost), and *r* is the relatedness between them. Thus, if a gene coding for a behaviour provides a benefit that greatly increases the number of offspring that close relatives can produce (i.e. *b* and *r* are high), it will be favoured by selection even if the cost is high (e.g. worker sterility). Hamilton's rule inspired numerous lines of empirical research that have manipulated these variables and shown that IFT is highly predictive (Langer et al. 2004; Field and Cant 2009; Strassmann et al. 2011; Bourke 2014).

IFT is an ultimate explanation for the evolution of eusociality and has provided numerous insights to the field of social evolution (Hamilton 1972; Dawkins 1976; Abbot et al. 2011; Bourke 2011a; Bourke 2011b). One insight is that selection for a gene for a social behavior depends on the effects of that behavior on the overall fitness of the gene's co-bearers (Hamilton 1964a). Another insight is that all types of social interaction (co-operation, competition, spite, and altruism) can be considered in IFT terms (Hamilton 1964a; Bourke 2011a). For example, increasing *r* in a social interaction will increase the propensity of the actor to exhibit altruistic behaviour; meanwhile, selfishness is adaptive at any level of cost to non-relatives, but has higher costs as *r* increases (Frank 1995; Bourke 2011a). Finally, a third insight is the prediction of conflict in social groups (and even in intragenomic conflict) where individuals have unequal *r* within the same societies and therefore different IFT optima (Hamilton 1964a; Trivers 1974; Ratnieks and Reeve 1992).

*Critiques of inclusive fitness theory have not been successful*

Despite the success of IFT, it has remained controversial. Early arguments were often misconceptions and were catalogued and addressed by Dawkins (1979). Some of these were as follows. 1) Kin selection is a special, and largely irrelevant form of individual selection that only applies to rare cases of sibling-care behaviour (Charlesworth 1978; Grant 1978). This is a misunderstanding because sibling-care should not be considered fundamentally different to maternal-care; instead maternal-care is one type of altruistic behaviour that is predicted by IFT selection. If the benefits are high, individuals are selected to care for their closest relatives, whether those relative are indirectly related (as in the case of siblings, cousins, etc.) or directly related (as in the case of offspring). 2) IFT demands that individuals can perform advanced feats of reasoning for kin-selected traits to evolve (Sahlins 1977). This is a misunderstanding because altruism, like any behaviour, does not

require advanced mathematical calculations from the individual for it to be selected for. Selection is expected to increase the proportion of genes in a population that code for behaviours that benefit relatives. 3) The theory of kin selection predicted universal altruism because most individuals within a given species share the vast majority of genes, since most organisms are not altruistic then IFT must be false (Washburn 1978). This misunderstands the point that *r* is a relative concept; *r* must on average be higher between the provider and recipients of altruism than between the provider and a random member of the whole population for altruistic behaviour to be selected, and therefore cross-population (and cross-species) considerations of the proportion of shared genes across an entire genome are irrelevant to IFT.

Recently inclusive fitness theory has come under renewed criticism (Wilson 2005; Wilson and Holldobler 2005; Wilson and Wilson 2007; Nowak et al. 2010). These criticisms have come from authors who have argued that group selection represents a viable alternative to IFT (Wilson and Holldobler 2005; Wilson and Wilson 2007; Holldobler and Wilson 2009), when in fact there is no fundamental distinction between the two sets of models (Bourke and Franks 1995; Foster et al. 2006; Lehmann et al. 2007). Some have argued that ecological factors are routinely ignored by IFT (Wilson and Wilson 2007) and that relatedness is not necessary for the evolution of eusociality (Wilson 2005; Wilson and Holldobler 2005). However, it is not true that IFT ignores ecological factors (which are accounted for in the *b* and *c* terms of Hamilton's rule) and relatedness is essential for the evolution of eusociality. In all species with eusociality, positive *r* was present in the ancestral forms (Helantera and Bargum 2007; Hughes et al. 2008; Boomsma 2009), as predicted by IFT (Hamilton 1964a; Bourke 2011a). The only alternative to kinship as a mechanism for positive relatedness that leads to eusociality is the 'green beard effect' (where altruistic genes evolve independently but can recognise themselves in non-kin; Queller and Strassmann 1998). The evolution of 'green beards' are compatible with IFT but are unlikely in eusocial insects because a) the independent evolution of such a gene with the necessary pleitropic effects will be rare (Bourke 2011a) and b) green beards would be in intra-genomic conflict with other genes that do not benefit from altruism (Helantera and Bargum 2007). This is avoided by altruism evolving via kinship.

Recent critiques also confuse a perceived failure of the 'haplodiploidy hypothesis' (the prediction that a relatedness asymmetry between haplodiploid sisters and offspring favours the evolution of sibling- over maternal-care behaviour) with IFT (Wilson 2005; Wilson and Holldobler 2005; Hunt 2007; Wilson and Wilson 2007; Nowak et al. 2010). It seems likely that there is some association between haplodiploidy and eusociality, given the fact that eusociality has evolved at least ten times in haplodiploid Hymenoptera (Bourke 2011a), as well as in other haplodiploid lineages such as the ambrosia beetle *X. saxeseni* (Biedermann and Taborsky 2011) and some members of the gall-inducing thrips genus *Kladothrips* (Chapman et al. 2000). However, the haplodiploidy hypothesis is not sufficient or necessary as an explanation for altruism and IFT is not dependent on it (Bourke and Franks 1995; Strassmann et al. 2011). While the hypothesis is difficult to test, it is inaccurate to characterise it as disproven (Bourke 2011b). The haplodiploidy hypothesis has continued to be productive for the formulation of novel theories of social evolution (Herrera 2013). Furthermore, the critiques ignore the fact that in other eusocial lineages (the ambrosia beetle *A. incompertus*, eusocial aphids, polyembryonic wasps, social shrimps, eusocial molerats and eusocial flatworms) the social groups are clonal or live in inbred family groups exhibiting positive relatedness, as predicted by IFT (Bourke 2011a).

One recent high-profile paper purports to show a novel mathematical model showing that IFT is unnecessary for the evolution of eusociality, arguing instead that 'standard natural selection' is sufficient (Nowak et al. 2010). This study also provided a qualitative step-wise model for the evolution of eusociality, and offered a critique of a 'representative' sample of the IFT literature. The critique fails on numerous accounts. Firstly, the point that inclusive fitness and standard natural selection are different is wrong. IFT is an extension of modern natural selection, but one showing that genes in one body can have effects on that body's fitness via their interactions with the same genes in other bodies (Abbot et al. 2011; Bourke 2011b; Ferriere and Michod 2011; Herre and Wcislo 2011). Secondly, Nowak et al. (2010) also criticised the mathematical basis of IFT and claimed that it makes restrictive assumptions that complicate mathematical formulations of selection, including pairwise interactions, weak selection, linearity, additivity, and special population structures. However, modelling approaches have shown that such assumptions are not required (Queller 1992; Gardner et al. 2007; 2011; Herre and Wcislo 2011), that Hamilton did not

require them in his original formulation (Hamilton 1964a), and therefore that the mathematical basis of IFT is sound (Abbot et al. 2011; Gardner et al. 2011). Thirdly, the critique also willfully ignores some of the successes of IFT, claiming that it has provided no new insights into social evolution and characterising the field as an 'abstract enterprise' (Nowak et al. 2010). One widely-reported success is the ability of IFT to predict sex ratios in social species, showing that inclusive fitness theory can explain up to 96% of between-species sex ratio variation as well as the effect of factors such as inbreeding, dispersal, brood size, asymmetric larval competition, sibling mating, and presence of helpers in a range of social species (West 2009; Abbot et al. 2011). Fourthly, as with previous critiques, Nowak et al. (2010) confuse a perceived (but inaccurate) failure of the haplodiploidy hypothesis with a failure of IFT, arguing that the 'association between haplodiploidy and eusociality fell below statistical significance', without sourcing their statistical calculations.

Finally, both the empirical and mathematical models of Nowak et al. (2010) fail to make new predictions in addition to those already predicted by IFT (Bourke 2011b). Their mathematical model contains assumptions and makes predictions that are not supported by the literature (Bourke 2011b; Herre and Wcislo 2011). For example, the model allows social groups to form in multiple ways (e.g. random attraction), whereas the evidence is that, in most lineages, eusocial societies formed sub-socially within highly-related groups, as predicted by IFT (Boomsma et al. 2011; Bourke 2011b). Nowak et al. (2010) use their model to down-play the role of relatedness in eusocial evolution, but their model addresses the evolution of sterile workers in clonal or monogamous families only. Therefore *r* does not vary in their model, which means that the model cannot address the role of relatedness in eusocial evolution (Boomsma et al. 2011). The model also predicts that queens must be able to support a high increase in birth rate, and, because this appears a restrictive condition, the model predicts that the evolution of eusociality should be rare, when in fact it has evolved independently at least 24 times (and arguably many more; Bourke 2011a; Bourke 2011b). Overall, Hamilton's IFT has withstood criticism since its inception, and modern critiques have failed to topple it as the leading theory of eusocial evolution.

*Proximate approaches to the study of eusociality*

IFT explains the ultimate causes for the evolution of eusociality; however, the proximate causes are still under active investigation. For example, many studies have detailed the effects of hormones that induce specific morphometric changes across each caste pathway (Wyatt and Davey 1996; Bortolotti et al. 2001; Barchuk et al. 2002; Corona et al. 2007; Libbrecht et al. 2013). From this section I will refer to caste as the polyphenism between fully reproductive queens and less reproductive workers in eusocial Hymenoptera. Caste pathway is defined as the series of developmental steps that an individual undergoes once its caste has been determined. Recent advances in technology have greatly facilitated studies of hormones in eusocial insects; for example, in the honey bee *Apis mellifera*, a proximate explanation for the caste determining mechanism has recently been elucidated by the use of RNAi technology, which showed that it is dependent on the 57kDA protein, royalactin (Kamakura 2011).

Further advances in technology have also helped isolate the genes involved in caste differentiation (the process where an individual acquires a new phenotype and becomes more specialised towards a particular caste). One way to understand the proximate mechanism of caste differentiation is to identify the differentially expressed genes (DEGs) exhibited between castes, e.g. by comparing the transcriptomes of developing individuals of each caste in eusocial insects. Early studies on sociogenomics used microarrays to quantify DEGs between social phenotypes in several eusocial lineages (Evans and Wheeler 1999; Evans and Wheeler 2001b; Evans and Wheeler 2001a; Pereboom et al. 2005; Sumner et al. 2006; Barchuk et al. 2007; Cardoen et al. 2011). These have yielded several insights. The first insight is that caste differentiation is far more complicated than was originally anticipated. Caste differentiation involves large numbers of differentially expressed genes and the genes that are differentially expressed are often very highly expressed in their respective castes (Evans and Wheeler 1999; Evans and Wheeler 2001a). This stands in contrast to the naïve view that caste differentiation would involve a small number of dedicated caste 'switch' genes. In the bumble bee *Bombus terrestris*, one study (using suppression subtractive hybridization) showed that many DEGs found across queen- and worker-destined larvae are only differentially expressed at certain time points during development (Pereboom et al. 2005). It also showed that genes underlying caste

differences in adults are often different from the genes that are differentially expressed at the point that caste is decided in larvae, showing that caste task specialisation involves different genes to the genes that determine caste fate (Pereboom et al. 2005). Such studies further show that caste differentiation is highly dependent on appropriate regulation of the correct genes occurring at appropriate time points (Sumner et al. 2006). These conclusions are supported by a recent comparative analysis that shows that the regulatory features of eusocial insect genomes are those that are most highly conserved between eusocial lineages (Simola et al. 2013).

A second major insight from studies of transcriptomics is that many of the DEGs, and therefore many of the genes that are likely to be associated with caste differentiation in eusocial insects, have evolved *de novo*, while a small number of them are conserved between eusocial lineages (Ferreira et al. 2013; Feldmeyer et al. 2014; Sumner 2014). Previously it has been argued that caste systems evolved from a pre-existing 'ground plan', such that gene pathways associated with maternal-care have become co-opted in the evolution of sibling-care (West-Eberhard 1987; West-Eberhard 1996). At the molecular level, researchers have hypothesised a molecular 'toolkit' of genes that are supposedly essential for the evolution of eusociality (Toth and Robinson 2007; Toth and Robinson 2009). These are held to be ancient, conserved genes that maintain development processes and body plans in solitary lineages, but are then repeatedly co-opted in the multiple evolutions of eusociality. They then take on new roles as genes involved in caste differentiation. Microarray studies provided suggestive evidence that many of the DEGs were older conserved genes, and were being repeatedly co-opted in the evolution of sociality in different lineages. Examples of this included the gene *foraging* (*for*), which is involved in the cyclic GMP (cGMP) pathway and associated with feeding behaviour in solitary species such as *Drosophila* (Debelle and Sokolowski 1987; Pereira and Sokolowski 1993) and with novel social functions in *A. mellifera* and the ant *Pogonomyrmex barbatus* (Ben-Shahar et al. 2002; Ingram et al. 2005); and *vitellogenin* (*Vg*), a gene that retains its ancient conserved ancestral role as a storage protein in insects (Sappington and Raikhel 1998) but that has recently evolved functions linking it to sociality in honey bees (Toth and Robinson 2007). However, a problem with using microarray studies to identify DEGs is that, because the microarray must be constructed from the expressed genomes of existing

model species, the input sequences are more likely to be conserved genes (Ferreira et al. 2013). Therefore microarrays will bias their results towards genes that are conserved in evolutionary time, i.e. toolkit genes. Recently, advances in sequencing technology (next generation or deep sequencing) have enabled researchers to study a much larger number of DEGs in eusocial insects, including genes that are conserved and genes that have evolved *de novo* (Woodard et al. 2011; Chen et al. 2012; Ferreira et al. 2013; Feldmeyer et al. 2014). These advances have also allowed comparative analyses with previously discovered DEGs across different eusocial lineages (Johnson and Tsutsui 2011). Such analyses have shown that as many as 95% of the DEGs of one lineage are not shared with another lineage, even in species that share a common eusocial ancestor (Ferreira et al. 2013). They have also shown that only a minority of DEGs are conserved between social and solitary lineages, the majority of them being taxonomically-restricted genes that are only found in the specific lineages under study. This implies that many of the genes associated with caste differentiation evolved *de novo*. While this is a significant finding, it does not completely rule out the genomic toolkit or ground plan hypotheses. Firstly, the number of conserved caste-biased DEGs might be underestimated in experiments that utilise different methodologies (e.g. rearing conditions, RNA handling and extraction procedures, whether RNA was extracted from whole bodies or just individual tissues, and whether the developmental stages used to extract RNA were truly comparable). Secondly, it is still not clear what role most of the DEGs that are sequenced in eusocial insects play during caste differentiation, and therefore the relative importance of taxonomically restricted genes and conserved toolkit genes. Thus, there are two evolutionary scenarios where toolkit genes would still be important in the evolution of eusociality. In the first, taxonomically-restricted genes might make up the majority of DEGs but have evolved as a consequence of the evolution of castes (e.g. to adapt to the new functions required of a eusocial life style such as the evolution of novel communication systems), rather than being essential for the early evolution of eusociality. In the second, taxonomically-restricted genes and conserved genes might work in conjunction, with both being essential in the evolution of eusociality. Overall, further studies assessing the function of DEGs in caste fate are still needed to assess the relative importance of novel, taxonomically-restricted genes and conserved, toolkit genes during the molecular evolution of eusociality.

The ultimate aim of molecular studies in this field is to characterise the molecular nature of caste differentiation. Transcriptomic studies have helped to achieve this at least partially in model social species such as *A. mellifera*. Among the most important pathways identified in this insect are the *target of rapamycin* (TOR), *insulin/insulin-like growth factor signalling* (IIS), *Juvenile hormone* (JH), *vitellogenin* (Vg), and *epidermal growth factor* (Egfr) pathways (Corona et al. 2007; Patel et al. 2007; Kamakura 2011; Mutti et al. 2011; Wolschin et al. 2011; Chen et al. 2012). In *A. mellifera*, workers feed female larvae a secretion termed Royal Jelly, which contains the morphogenic protein *royalactin* and causes female larvae to develop as queens. Specifically, it induces changes in the *Egfr* pathway, which are directly responsible for queen-specific phenotypes such as increased size, increased longevity, increased fecundity, and decreased development time (Kamakura 2011). It also mediates the expression levels of *TOR*, which further acts to induce queen development (Patel et al. 2007; Kamakura 2011). Meanwhile changes in the larval diet induce changes in the *IIS* pathway, which, together with *Egfr* and *TOR* (which act in conjunction with each other), cause large *JH* peaks during the fourth and fifth instars of queen-destined larvae (de Azevedo and Hartfelder 2008; Kamakura 2011; Mutti et al. 2011). In *Drosophila*, *JH* is regulated by the release of the neuropeptides allatostatin and allatotropin from the corpus allatum (Tu et al. 2005) and it has been hypothesised that this happens during differential *JH* regulation in social species (Libbrecht et al. 2013). *JH* itself induces numerous developmental and endocrinal changes in eusocial insects, and peaks in *JH* are essential for caste development in *A. mellifera*, *B. terrestris*, and ants (Rachinsky et al. 1990; Bortolotti et al. 2001; Barchuk et al. 2002; Corona et al. 2007; Libbrecht et al. 2013). In *A. mellifera*, one of the most important effects of *JH* is on the production of *Vg*; accumulation of *Vg* happens through the action of ecdysteroids that are mutually responsive to *JH*, and this *Vg* build up is thought to be essential for queen development in honey bees (Engels et al. 1990; Barchuk et al. 2002). Ecdysteroids themselves are also important for some of the specific queen characteristics. For example, one ecdysteroid (*20-hydroxyecdysone*) is also induced by *Egfr* activity and is essential for the reduction in developmental time in queen-destined larvae (Kamakura 2011). While less is known about other species, at least some of these genes and gene pathways have been shown to be important for caste determination processes in bumble bees (Cnaani et al. 2000; Bortolotti et al. 2001), social wasps (Tibbetts

and Sheehan 2012), and ants such as *Pogonomyrmex* (Corona et al. 2013; Libbrecht et al. 2013).

While the study of these genes is promising in terms of developing an understanding of the proximate mechanisms of caste differentiation in eusocial insects, the genes still represent a relatively small number, while there are potentially thousands of DEGs (Johnson and Tsutsui 2011; Chen et al. 2012; Ferreira et al. 2013; Feldmeyer et al. 2014). There is still little information on the mechanism of action of these genes and on the regulatory mechanisms that program and control them. Previous studies have focussed mostly on differential regulation of protein-coding genes, but there are other mechanisms of regulation that work at different levels, and may prove more important in the regulation of caste gene expression. One mechanism of regulation is through epigenetics. For example, *A. mellifera* and other Hymenoptera have been demonstrated to have a functioning CpG methylation system (Wang et al. 2006), and some authors have shown that differential methylation is associated with reproductive division of labour in *A. mellifera* (Kucharski et al. 2008; Elango et al. 2009), *B. terrestris* (Amarasinghe et al. 2014), several ant species (Bonasio et al. 2012; Smith et al. 2012) and *Polistes* wasps (Weiner et al. 2013). These epigenetic systems of regulation may be regulated by the same pathways (*JH*, *Vg*, *IIS*) as those described above (Mutti et al. 2011). A comprehensive review of the potential regulatory role of epigenetics in caste differentiation has been provided elsewhere (Lyko and Maleszka 2011; Patalano et al. 2012). The rest of this thesis concerns a class of potentially important regulatory genes that have received little attention so far with regards to their potential for caste regulation at the post-transcriptional level (Dolezal and Toth 2014), the microRNAs (miRNAs).

*MicroRNAs regulate development and differentiation in polyphenisms*

MiRNAs are a class of small RNAs (sRNA, approximately 21-23 bp in length) that play a major role in gene regulation by binding to messenger RNAs (mRNA) and preventing them from being transcribed into protein by suppressing translation (Lagos-Quintana et al. 2001; Lau et al. 2001; Lee and Ambros 2001; Bartel 2004). The canonical biogenesis pathway of miRNAs starts when they are first transcribed as primary miRNA transcripts (Lee et al. 2002), that are then recognised and processed into precursors miRNAs (pre-miRNA) by the

RNase III enzyme *Drosha* in the cell nucleus (Lee et al. 2003). Pre-miRNAs are characterised by having a long hairpin loop secondary structure. They are exported into the cytoplasm by *exportin-5* (Lund et al. 2004), where their characteristic structure is recognised by a second RNase III enzyme called *Dicer*, which cleaves the loop into an approximately 21-23 imperfect nucleotide duplex with an approximately 2 nucleotide 3' overhang (Bartel 2004). One of the two strands (the mature miRNA) is organised into the RNA induced silencing complex (RISC) together with the *Argonaute* enzyme (Bartel 2004), where it recognises, and binds to an mRNA transcript which is complementary to the mature strand at 'seed sequence' (the sequence from 5' 2-7 nucleotides on the mature miRNA). The other strand of the duplex, sometimes termed the *strand, is usually degraded (Bartel 2004). Previous models of miRNA processing suggested that the mature strand is determined by the thermodynamic stability, and the structural properties of the miRNA duplex (Khvorova et al. 2003; Schwarz et al. 2003). However, some miRNAs can be processed from both arms and one study has shown that identical mature duplexes can produce different mature miRNAs in different species (Griffiths-Jones et al. 2011), and different mature sequences from the same precursor can target different genes (Marco et al. 2012). In such cases it is likely that the mature product is encoded by the pre-miRNA or primary miRNA transcript. Furthermore, the mature miRNA usually targets mRNA transcripts in the 3' UTR region (Bartel 2004), however it has also been shown that they can also cause translational inhibition by targeting the 5' UTR region (Lytle et al. 2007), and sometimes even the coding region of an mRNA transcript (Fang and Rajewsky 2011).

The first miRNAs, lin-14 and let-7 were discovered in 1993 and 2000 and respectively (Lee et al. 1993; Reinhart et al. 2000), and since this time tens of thousands have been discovered and recorded on *miRBase* (www.miRBase.org), the central repository for new miRNAs (Griffiths-Jones et al. 2008; Kozomara and Griffiths-Jones 2014). Most studies on caste-biased gene expression identify changes in the abundance of long mRNA transcripts (most of which are then translated into proteins). However, focusing on these genes alone ignores numerous changes that occur at the post-transcriptional levels of gene expression. Piwi-interacting small RNAs (piRNA) are a different class of sRNA that have been implicated in epigenetic mechanisms of inheritance in insects (Huang et al. 2013), and miRNAs are post-transcriptional regulators of gene expression (Bartel 2004). This means that

understanding the role of miRNAs and other sRNAs in general will be critical to gaining a full knowledge of the processes involved in caste determination and development.

MiRNAs have been found to be important in development in a variety of organisms, including plants, nematodes, *Drosophila*, other insects, and mammals (Bartel 2004; Baulcombe 2004; Alvarez-Garcia and Miska 2005; Biemar et al. 2005; Jones and Newbury 2010). One of their essential functions in higher eukaryotes is to fine-tune gene expression as part of larger gene regulatory networks (GRNs). GRNs contain specific patterns of interconnected molecules. To simplify the analysis of these networks, Shen-Orr et al. (2002) conceived of 'network motifs', which they defined as 'patterns of interconnections that recur in many different parts of a network at frequencies much higher than those found in randomized networks.' These motifs make up the basic building blocks of complex networks in model organisms such as *Escherichia coli*, *Saccharomyces cerevisiae*, and *Caenorhabditis elegans* (Milo et al. 2002; Shen-Orr et al. 2002). Mangan and Alon (2003) showed that one of the simplest, most common and most important type of network motif is the feed-forward loop, which is a three-gene pattern in which a pair of transcription factors act in concert to efficiently regulate a third target gene. Other authors (e.g. Hornstein and Shomron 2006; Peterson et al. 2009) have applied this model to post-transcriptional gene regulation by miRNAs, using the same classifications as Mangan and Alon (2003), in which feed-forward networks are defined as 'coherent' or 'incoherent'. A coherent feed-forward loop that contains miRNAs occurs when a miRNA is activated by a transcription factor that also switches off expression of the miRNA's target; the miRNA will therefore target the gene at the post-transcriptional level, while the transcription factor prevents the synthesis of more mRNA transcripts. This ensures that there are no residual levels of gene expression from either the genome itself or from mRNA transcripts that escaped being targeted, and thus ensures that the gene is switched off completely. Alternatively, an incoherent feed-forward loop occurs when a transcription factor activates both a miRNA and its respective target. In these systems, the target can still produce some protein, but because it is being targeted at the post-transcriptional level there is no risk of runaway gene expression which is likely to be costly, and even harmful, to the cell that produces it. By limiting transcription, such miRNA feed-forward loops reduce 'noise' (uncontrolled fluctuations in gene expression) in the systems that express them (Raser and

O'Shea 2005). This ensures that the expression of a particular protein will remain relatively constant (compared to systems that are not regulated at the post-transcriptional level). Constant gene expression of some proteins is optimal for the canalisation of development, where it is essential that a gene's expression is tightly controlled along very particular trajectories. This type of regulation by miRNAs is likely to have been important in the evolution of complex body plans in higher eukaryotes (Sempere et al. 2006; Peterson et al. 2009). Simple changes in the expression of miRNAs between related species have been shown to lead to significant changes in morphology (Arif et al. 2013). It has also been demonstrated that the phenotypic consequences of removing miRNAs can have major effects on development, often leading to organisms with more variable features, consistent with the amount of genetic noise in these individuals being higher (Li et al. 2006).

In species with phenotypic plasticity, it is especially important that development should occur along a precise trajectory, so that a sub-optimal, intermediate phenotype is not expressed in an inappropriate environment. Given that miRNA expression can have significant effects on morphology (Arif et al. 2013), it is not surprising that they have also been associated with regulating phenotypic changes in species with plastic phenotypes (Gutierrez et al. 2009; Legeai et al. 2010). For example, Gutierrez et al. (2009) showed that miRNAs play an important part in the regulatory process that causes adventitious rooting in *Arabidopsis* in response to variation in light levels. Within insects, a study that used deep sequencing of sRNAs (which I will refer to as miRNA-seq throughout this thesis) has shown that up to 17 miRNAs are differentially expressed between alternative phenotypes (winged and wingless) in the pea aphid *Acyrthosiphon pisum*, which arise in response to changes in population density (Legeai et al. 2010).

Recent studies have used deep sequencing approaches to characterise the role of miRNAs in social insect caste plasticity, primarily in the honey bee *A. mellifera* (Behura and Whitfield 2010; Chen et al. 2010; Greenberg et al. 2012; Liu et al. 2012; Guo et al. 2013; Nunes et al. 2013) but also in the ant *Camponotus floridanus* (Bonasio et al. 2010). A major example occurs in *A. mellifera* and concerns worker age polyethism (the process by which worker task repertoire shifts with age, with workers acting act as brood nurses when young and switching as they grow older to external foraging). This process has been associated with changes in miRNAs in the brains of transitioning workers (Behura and Whitfield 2010;

Greenberg et al. 2012; Liu et al. 2012). Among the most highly expressed and differentiated miRNAs were Ame-miR-184 and Ame-miR-133 which were more highly expressed in forager specialist workers compared to nurse specialist workers (Behura and Whitfield 2010; Greenberg et al. 2012; Liu et al. 2012), and Ame-miR-279b and Ame-miR-279c which were more highly expressed in nurse specialist workers compared to forager specialist workers (Liu et al. 2012). I will discuss the expression patterns of these miRNAs in further detail throughout this thesis. Likewise, in a comparison between the major and minor worker in *C. floridanus* Bonasio et al. (2010) showed that cfo-miR-64 was more highly expressed in minor workers compared to major workers, while cfo-miR-7 was more highly expressed in major workers compared to minor workers. Most recently, miRNAs have been shown to be involved in the response to the physiological changes induced by *Vg* in *A. mellifera* (Nunes et al. 2013), with *Vg* knockdown (reduction in gene expression using RNA interference artificially) changing the patterns of miRNA expression in foraging workers. Specifically Ame-let-7 and Ame-miR-281 were significantly less highly expressed in knockdown workers, and Ame-miR-3739, Ame-miR-3776, Ame-miR-3796, Ame-miR-316, and Ame-miR-3718a were significantly more highly expressed in knockdown workers compared to controls. As stated earlier, *Vg* itself has been associated with the regulation of social behaviour and caste biology in eusocial insects (Engels et al. 1990; Barchuk et al. 2002; Libbrecht et al. 2013), and the focal miRNAs had predicted targets in the *IIS*, *JH*, and *Ecdysteroid* pathways.

There has been little work to date on whether miRNAs are directly involved in regulating the reproductive division of labour in eusocial insects. This is an area of immense interest because, as earlier discussed: 1) gene regulation of different caste pathways, including the timing and locality of gene expression, is important for controlling queen-worker caste dimorphism (Pereboom et al. 2005; Sumner et al. 2006; Grozinger et al. 2007); and 2) miRNAs are important in other systems with phenotypic plasticity (Gutierrez et al. 2009; Legeai et al. 2010) and in the evolution and canalisation of development (Sempere et al. 2006; Peterson et al. 2009).

Two studies have characterised some of the miRNA expression differences occurring between the castes in *A. mellifera*. In the first study, Weaver et al. (2007) extracted RNA from the head, thorax, and abdomen of adult workers and queens of *A. mellifera*, and from

whole worker-destined pupae and queen-destined pupae. They used qRT-PCR to show that Ame-miR-2 and Ame-miR-9a were more highly expressed in queen versus worker abdomens, and Ame-miR-71 was more highly expressed in worker versus queen abdomens, and in queen-destined pupae compared to worker destined pupae. However, until recently, there have been no studies that specifically characterise the role of miRNAs at the point of caste determination in eusocial insects. In the second study, this was addressed by using miRNA-seq to sequence the miRNAs contained in Worker Jelly and Royal Jelly fed to female larvae, therefore examining the role of miRNAs in caste determination in *A. mellifera*. (Guo et al. 2013). The work showed that workers used mandibular gland secretions to enrich Worker Jelly with a greater diversity and higher proportion of miRNAs than Royal Jelly, with Ame-miR-184 being particularly highly expressed in Jelly fed to workers. Crucially, the study also showed that when Ame-miR-184 was added to Royal Jelly fed to queen-destined larvae, the treatment larvae developed with worker-like characteristics (Guo et al. 2013). This implies that miRNAs are causally involved in the caste determining processes of honey bees. These results, together with the studies that show differential expression of miRNAs across different behavioural sub-castes in *A. mellifera* (Behura and Whitfield 2010; Greenberg et al. 2012; Liu et al. 2012), imply that miRNAs have an important role for regulating polyphenisms and social behaviour in the larvae and adults of honey bees and possibly other eusocial insects.

Previous research has therefore shown that miRNAs are associated with eusocial insect caste polyphenisms, but there are still many unsolved problems remaining as regards the extent and nature of this association. Among the most important with respect to the role of miRNAs in caste differentiation are the following: 1) miRNAs are associated with caste and task differentiation in *A. mellifera* (Weaver et al. 2007; Greenberg et al. 2012) and possibly *C. floridanus* (Bonasio et al. 2010), but there has been no comparative work on miRNAs in other social species. Hence it remains unclear whether miRNAs have a general role in regulating eusocial insect caste differentiation across several lineages, or whether the roles identified so far are specific to advanced eusocial insects (eusocial insects with morphologically, and physiologically distinct castes) or even to *A. mellifera* alone, rather than primitively eusocial insects (eusocial insects where the caste differences are behavioural, and therefore their sociality more closely resembles an ancestral state to the

evolution of caste divisions). 2) The two studies that have addressed this have focussed only on general miRNA expression differences between the two castes (Weaver et al. 2007; Guo et al. 2013). No studies of miRNAs in eusocial insects have characterised the expression of miRNAs in developing larvae at the point when the castes first begin to differentiate. This means that it is unclear whether miRNAs have a direct causal role in caste determination, or whether they regulate downstream processes and their differential expression is a consequence rather than a cause of caste differentiation. 3) There has been no attempt to show whether miRNAs are implicated in more general systems of the reproductive division of labour in social insects. Two studies (Weaver et al. 2007; Guo et al. 2013) have focussed on caste differences between workers and queens, but there has been no research on the association of miRNAs with the reproductive division of labour that exists within each caste, specifically between gynes (young, unmated queens with inactive ovaries) and mature queens, or between inactive-ovary workers and active-ovary workers (which produce unfertilised eggs developing into males). 4) Many of the results that show differential expression of miRNAs in *A. mellifera* have been poorly validated and replicated (Chen et al. 2010; Liu et al. 2012; Guo et al. 2013), sometimes relying entirely on sequencing evidence to show that new miRNAs are involved in social regulation and often only using one library per phenotype. Previous studies have shown that results from indirect approaches to establishing gene expression changes, such as deep sequencing, are not reproduced when they are checked using more direct approaches such as Northern blots and qRT-PCR (Baker 2010; Greenberg et al. 2012). In addition, gene expression is often highly variable, which means that studies that only use one library per phenotype are likely to have a high number of type-I errors. Therefore it is unclear whether associations of miRNAs with variation across social phenotypes that have not been validated or well replicated will show the same results when more direct measures of gene expression are used on the same samples with higher numbers of replicates. 5) Finally, very few previous studies have characterised the function of miRNAs in caste development. For example, no studies on eusocial insects have investigated whether miRNAs are differentially expressed between castes in specific tissues alone or at a whole-organism level. No studies have shown whether the timing of miRNA expression is important as it is for other genes. This is important because, similarly to protein coding gene, miRNAs are often mediated by regulatory factors that cause them to be expressed in a time-specific and tissue specific

manner (Biemar et al. 2005). In addition, only one study has predicted the genes that differentially expressed miRNAs target, and linked those genes to the caste differentiation processes (Nunes et al. 2013). In this thesis, I aimed to address all these questions, specifically by characterising the role of miRNAs in the social processes that affect reproductive division of labour in the bumble bee *B. terrestris*.

*Bumble bees as model organisms for the study of the evolution of eusociality*

Bumble bees (*Bombus* species), and *B. terrestris* in particular, are an ideal taxon for understanding the proximate mechanisms of insect sociality. First, they are among the few social insects to have had their genomes sequenced (Sadd et al. 2015). Two species of *Bombus* (*B. terrestris* and *B. impatiens*; genomes available at http://www.ncbi.nlm.nih.gov/genbank/ under the accession numbers AELG00000000.1 and AEQM0000000.2 respectively), in different sub-genera, have had their genomes sequenced and this will allow for extensive genomic comparisons across the genus, and between *Bombus* and their nearest eusocial relative with a published genome, *A. mellifera* (Weinstock et al. 2006; Elsik et al. 2014). Second, there are extensive additional molecular resources available for *B. terrestris* (Pereboom et al. 2005; Riddell et al. 2009; Colgan et al. 2011; Stolle et al. 2011; Woodard et al. 2013; Amarasinghe et al. 2014; Deshwal and Mallon 2014, Harrison et al. accepted), which are not yet available for many other eusocial insects. Linking the DEGs isolated in some of these studies with post-transcriptional mechanisms of gene regulation will be an ongoing task in gaining an understanding of the proximate mechanisms underpinning eusociality in these species. Third, the caste systems, social organisation, life histories, and ecology of *B. terrestris* have all been studied extensively, which allows molecular studies to be placed in the relevant context when addressing the proximate mechanisms for the evolution of eusociality (Cnaani et al. 2000; Goulson 2003; Cameron et al. 2007; Williams et al. 2007; Zanette et al. 2012; Goulson 2013; Holland et al. 2013). Fourth, bumble bees have an annual colony cycle and relatively small colony sizes, and their colonies are available commercially, and so are ideal species for rearing under laboratory conditions (Alford 1975; Velthuis 2002; Goulson 2003). Fifth, bumble bees in general are globally important pollinators of wild flowers and commercial crops, and over the course of decades have been in global decline (Goulson 2003; Goulson 2013). However, *B. terrestris* has been robust against declines that have affected other bumble bees across

its native range (Dupont et al. 2011), and it is a pest species in areas where it has been imported (Schmid-Hempel et al. 2007; Goulson 2013; Schmid-Hempel et al. 2014). Therefore a molecular understanding of *B. terrestris* and other bumble bee species will be an invaluable tool for devising and implementing management strategies for these ecologically and economically important pollinators.

*Chapter aims*

Over the course of this thesis we aimed to identify and characterise the role of miRNAs in *B. terrestris*. In **Chapter 2** we described some of the findings of the ongoing *Bombus* genome project, and a miRNA-seq experiment that we used to annotate the miRNAs of *B. terrestris* to the genome. We also compared the miRNA profiles of *B. terrestris* with a second *Bombus* species, *B. impatiens*, and then compared the profiles of both species to that of *A. mellifera*, a species sharing a common eusocial ancestor with *Bombus* (Weinstock et al. 2006; Cardinal and Danforth 2011). We found that *B. terrestris* yielded 131 miRNAs and *B. impatiens* yielded 114 miRNAs and that, of these, 17 were new miRNAs that had not previously been sequenced in any species. These experiments showed that, of 217 miRNAs in *A. mellifera*, as many as 103 were not found in either of the genomes of the two *Bombus* species. This could suggest that the miRNAs of *A. mellifera* were strongly diverged from those found in *Bombus*, representing an example of rapid miRNA evolution during the transition from primitive to advanced eusociality. This explanation was supported by the findings of a previous study that miRNAs evolve rapidly in insects (Marco et al. 2010). However, we also considered the possibility that this arises from systematic misidentification of miRNAs in the eusocial insect literature. We concluded that the high number of lineage specific miRNAs in *A. mellifera* was likely to be a combination of both of these reasons.

Having characterised miRNA sequences in *B. terrestris*, we set out to establish whether they were important in social behavioural processes such as caste determination and reproductive division of labour. In **Chapter 3** we used miRNA-seq and Northern blot approaches to test whether there was an association between miRNA expression and *B. terrestris* queen-worker caste determination and differentiation in developing female larvae. We showed that at least one miRNA (Bte-miR-6001) was associated with caste

development, being more strongly expressed in queen-destined larvae compared to worker-destined larvae. We also found that this miRNA exhibited stage- and tissue-specificity in its expression profile, and was most highly expressed in the outer cuticle of developing queen-destined larvae, but only after the point that their caste fate became fixed, while expression levels declined in the pupae. We also showed that Bte-miR-6001 is a mirtron (a miRNA that is spliced from an intron), expressed from the gene *vhdl*, a homolog of *Vg* that itself has numerous links to caste differentiation in eusocial insects (Engels et al. 1990; Corona et al. 2007; Corona et al. 2013; Libbrecht et al. 2013). Due to the genomic location of mirtrons and their mechanism of expression it is likely that the two genes (*vhdl* and *mir-6001*) have linked expression patterns during caste differentiation.

Whilst caste is determined in the larval stages of *B. terrestris* development, the physiological and behavioural differences that distinguish the castes only become manifest once they reach adulthood. This culminates in the reproductive division of labour, by which queens perform most of the reproduction and workers perform most of the colony maintenance tasks (Wilson 1971). However, later in the colony cycle of *B. terrestris*, the queen switches to producing haploid eggs (Duchateau and Velthuis 1988) and shortly afterwards some of the workers activate their own ovaries and start to lay eggs (Duchateau and Velthuis 1989; Lopez-Vaamonde et al. 2007). Therefore a second form of reproductive division of labour occurs between active-ovary workers, which become reproducers along with the queen, and inactive-ovary workers, which continue to perform colony-associated tasks (Duchateau and Velthuis 1988; Duchateau and Velthuis 1989). As an extension of our work in Chapter 3, in **Chapter 4** we used miRNA-seq and Northern blots to test whether there was an association between miRNAs and caste differentiation in *B. terrestris* adult females. Our aim was similar to that of an early study that compared differential gene expression between larvae and adult *B. terrestris* (Pereboom et al. 2005), except that this study, being based on suppressive subtraction hybridization, investigated far fewer genes and did not profile miRNAs. We also aimed to test whether miRNAs were associated with the reproductive division of labour between inactive-ovary workers and active-ovary workers, and whether there was a link between queen-worker differentiation, and differentiation of adult workers into active-ovary and inactive-ovary females.

Overall we found no evidence for a link between differentially expressed miRNAs in female larvae and adults. However, this study revealed an association between miRNA expression and the reproductive division of labour in adults. We found that miRNAs associated with ovary activation in *B. terrestris* queens, have the same expression profiles as miRNAs associated with ovary activation in active-ovary workers. Of particular interest was Bte-miR-184, which was differentially expressed between queens and workers, and was also differentially expressed between active-ovary and inactive-ovary workers.

In this chapter we also found evidence for the ground plan hypotheses of West-Eberhard (1987) and Amdam et al. (2004), which predict that genes associated with reproduction should also be more highly expressed in individuals that exhibit maternal care behaviour (e.g. nursing workers), and genes associated with maintaining a reproductively inactive state should also be more highly expressed in individuals that exhibit behaviour not directly related to maternal care (e.g. foraging workers). While our study was not an exclusive test of these hypotheses, the miRNA expression patterns that we found were consistent with their key predictions. Specifically, we showed that miRNAs associated with foraging behaviour in *A. mellifera*, such as miR-184 and miR-133 (Greenberg et al. 2012; Liu et al. 2012), were much more highly expressed in the ovaries of inactive-ovary *B. terrestris* workers, while miRNAs associated with nursing behaviour in *A. mellifera*, such as miR-279b and miR-279c (Liu et al. 2012), were more highly expressed in the ovaries of queens and active-ovary *B. terrestris* workers.

Chapters 2-4 showed that RNAi mechanisms in *B. terrestris* were associated with the evolution of sociality, with caste mechanisms, and, finally, with the reproductive division of labour. In our final study, detailed in **Chapter 5,** we aimed to make use of RNAi pathways to target and reduce gene expression in *Bombus* queens. Until very recently, no study has shown that dsRNA could be used to induce targeted 'knockdown' (reduction in gene expression due to treatment with dsRNA) in *Bombus* species (Deshwal and Mallon 2014). We targeted the gene *for*, a gene that has previously been linked to social processes in *B. terrestris* where it has been shown to be more highly expressed in asocial queens compared to social queens (Edd Almond, Personal communication), and has also been shown to be more highly expressed in forage specialist workers compared to nurse specialists (Tobback et al. 2011), although the opposite result was found in *B. ignites* (Kodaira et al. 2009). We

used RNA knockdown and qPCR to test for an association between *B. terrestris* foundress queens cessation of foraging, and expression of the gene *for*. We generated dsRNA to target *for*, so that we could monitor the behaviour of foundress queens when *for* expression was silenced. We provided the dsRNA in sugar-syrup solution and then treated laying- and non-laying foundress queens by oral ingestion of the dsRNA solution. We then monitored the activity levels and circadian rhythmicity of the queens using tracking software before dissecting the queens and measuring their *for* expression levels using qRT-PCR.

Overall, we were unsuccessful in our aim to knock down *for*, and we found no link between *for* expression and queen activity levels. However, we showed that laying foundress queens expressed *for* at a higher rate than non-laying foundress queens, and they were also more likely to be arrhythmic (when food was restricted) in comparison with the non-laying queens. We concluded that there was a complex relationship between the effects of *for*, food availability and activity levels in *Bombus* queens.

In conclusion, this thesis makes novel contributions to our understanding of the molecular basis of eusocial evolution in several ways. Firstly, we used miRNA-seq and bioinformatics to identify the miRNAs in *B. terrestris*, and showed that miRNAs are strongly diverged between *B. terrestris* and *A. mellifera* (**Chapter 2**). Secondly, we used miRNA-seq and Northern blot validation to show for the first time that a miRNA, Bte-miR-6001-5p, is associated with the caste determination mechanisms of *B. terrestris* larvae (**Chapter 3**). Thirdly, we also used miRNA-seq and Northern blot validation to show for the first time that miRNAs (including Bte-miR-184, Bte-miR-279b, Bte-miR-279c, Bte-miR-133, Bte-miR-92b, and Bte-miR-9a) are associated with the reproductive division of labour in adult *B. terrestris* (**Chapter 4**). In this work, we showed that Bte-miR-184 was more highly expressed in worker adults than in queens, and that Bte-miR-184 and Bte-miR-133 were more highly expressed in inactive-ovary workers compared to active-ovary workers and queens, while Bte-miR-279b, Bte-miR-279c, Bte-miR-92b, and Bte-miR-9a were more highly expressed in active-ovary workers and queens compared to inactive-ovary workers. Finally, we show for the first time that the gene *for* was associated with *B. terrestris* queen egg laying behaviour, as the gene was significantly more highly expressed in laying *B. terrestris* queens compared to non-laying queens, and that the effect occurred independently of age (**Chapter 5**).

*References*

Abbot P Abe J Alcock J Alizon S Alpedrinha JAC Andersson M Andre JB van Baalen M Balloux F Balshine S et al. 2011. Inclusive fitness theory and eusociality. *Nature* **471**: E1-E4.

Alford DV. 1975. *Bumblebees*. Davis-Poynter, London, UK.

Alvarez-Garcia I, Miska EA. 2005. MicroRNA functions in animal development and human disease. *Development* **132**: 4653-4662.

Amarasinghe HE, Clayton CI, Mallon EB. 2014. Methylation and worker reproduction in the bumble-bee (*Bombus terrestris*). *Proc R Soc B* **281**: 20132502.

Amdam GV, Norberg K, Fondrk MK, Page RE. 2004. Reproductive ground plan may mediate colony-level selection effects on individual foraging behavior in honey bees. *Proc Natl Acad Sci USA* **101**: 11350-11355.

Arif S, Murat S, Almudi I, Nunes MDS, Bortolamiol-Becet D, McGregor NS, Currie JMS, Hughes H, Ronshaugen M, Sucena E et al. 2013. Evolution of *mir-92a* underlies natural morphological variation in *Drosophila melanogaster*. *Curr Biol* **23**: 523-528.

Baker M. 2010. MicroRNA profiling: separating signal from noise. *Nat Methods* **7**: 687-692.

Barchuk AR, Bitondi MM, Simoes ZL. 2002. Effects of juvenile hormone and ecdysone on the timing of vitellogenin appearance in hemolymph of queen and worker pupae of *Apis mellifera*. *J Insect Sci* **2**:1.

Barchuk AR, Cristino AS, Kucharski R, Costa LF, Simoes ZLP, Maleszka R. 2007. Molecular determinants of caste differentiation in the highly eusocial honeybee *Apis mellifera*. *Bmc Dev Biol* **7**:70.

Bartel DP. 2004. MicroRNAs: Genomics, biogenesis, mechanism, and function. *Cell* **116**: 281-297.

Baulcombe D. 2004. RNA silencing in plants. *Nature* **431**: 356-363.

Behura SK, Whitfield CW. 2010. Correlated expression patterns of microRNA genes with age-dependent behavioural changes in honeybee. *Insect Mol Biol* **19**: 431-439.

Ben-Shahar Y, Robichon A, Sokolowski MB, Robinson GE. 2002. Influence of gene action across different time scales on behavior. *Science* **296**: 741-744.

Biedermann PHW, Taborsky M. 2011. Larval helpers and age polyethism in ambrosia beetles. *Proc Natl Acad Sci USA* **108**: 17064-17069.

Biemar F, Zinzen R, Ronshaugen M, Sementchenko V, Manak JR, Levine MS. 2005. Spatial regulation of microRNA gene expression in the *Drosophila* embryo. *Proc Natl Acad Sci USA* **102**: 15907-15911.

Bonasio R, Li QY, Lian JM, Mutti NS, Jin LJ, Zhao HM, Zhang P, Wen P, Xiang H, Ding Y et al. 2012. Genome-wide and caste-specific DNA methylomes of the ants *Camponotus floridanus* and *Harpegnathos saltator*. *Curr Biol* **22**: 1755-1764.

Bonasio R, Zhang G, Ye C, Mutti NS, Fang X, Qin N, Donahue G, Yang P, Li Q, Li C et al. 2010. Genomic comparison of the ants *Camponotus floridanus* and *Harpegnathos saltator*. *Science* **329**: 1068-1071.

Boomsma JJ. 2009. Lifetime monogamy and the evolution of eusociality. *Philos T R Soc B* **364**: 3191-3207.

Boomsma JJ, Beekman M, Cornwallis CK, Griffin AS, Holman L, Hughes WOH, Keller L, Oldroyd BP, Ratnieks FLW. 2011. Only full-sibling families evolved eusociality. *Nature* **471**: E4-E5.

Bortolotti L, Duchateau MJ, Sbrenna G. 2001. Effect of juvenile hormone on caste determination and colony processes in the bumblebee *Bombus terrestris*. *Entomol Exp Appl* **101**: 143-158.

Bourke AFG. 2011a. *Principles of Social Evolution*. Oxford University Press, Oxford, UK.

-. 2011b. The validity and value of inclusive fitness theory. *Proc R Soc B* **278**: 3313-3320.

-. 2014. Hamilton's rule and the causes of social evolution. *Phil Trans R Soc B* **369:** 20130362.

Bourke AFG, Franks NR. 1995. *Social Evolution in Ants*. Princeton University Press, Princeton, New Jersey, USA.

Cameron SA, Hines HM, Williams PH. 2007. A comprehensive phylogeny of the bumble bees (*Bombus*). *Biol J Linn Soc* **91**: 161-188.

Cardinal S, Danforth BN. 2011. The antiquity and evolutionary history of social behavior in bees. *PLOS One* **6**: e21086.

Cardoen D, Wenseleers T, Ernst UR, Danneels EL, Laget D, De Graaf DC, Schoofs L, Verleyen P. 2011. Genome-wide analysis of alternative reproductive phenotypes in honeybee workers. *Mol Ecol* **20**: 4070-4084.

Chapman TW, Crespi BJ, Kranz BD, Schwarz MP. 2000. High relatedness and inbreeding at the origin of eusociality in gall-inducing thrips. *Proc Natl Acad Sci USA* **97**: 1648-1650.

Charlesworth B. 1978. Some models of the evolution of altruistic behaviour between siblings. *J Theor Biol* **72**: 297-319.

Chen X, Hu Y, Zheng HQ, Cao LF, Niu DF, Yu DL, Sun YQ, Hu SN, Hu FL. 2012. Transcriptome comparison between honey bee queen- and worker-destined larvae. *Insect Biochem Mol Biol* **42**: 665-673.

Chen X, Yu X, Cai Y, Zheng H, Yu D, Liu G, Zhou Q, Hu S, Hu F. 2010. Next-generation small RNA sequencing for microRNAs profiling in the honey bee *Apis mellifera*. *Insect Mol Biol* **19**: 799-805.

Cnaani J, Robinson GE, Hefetz A. 2000. The critical period for caste determination in *Bombus terrestris* and its juvenile hormone correlates. *J Comp Physiol A* **186**: 1089-1094.

Colgan TJ, Carolan JC, Bridgett SJ, Sumner S, Blaxter ML, Brown MJF. 2011. Polyphenism in social insects: insights from a transcriptome-wide analysis of gene expression in the life stages of the key pollinator, *Bombus terrestris*. *BMC Genomics* **12**: 623.

Corona M, Libbrecht R, Wurm Y, Riba-Grognuz O, Studer RA, Keller L. 2013. Vitellogenin underwent subfunctionalization to acquire caste and behavioral specific expression in the harvester ant *Pogonomyrmex barbatus*. PLOS Genet **9**: e1003730

Corona M, Velarde RA, Remolina S, Moran-Lauter A, Wang Y, Hughes KA, Robinson GE. 2007. Vitellogenin, juvenile hormone, insulin signaling, and queen honey bee longevity. *Proc Natl Acad Sci USA* **104**: 7128-7133.

Crespi BJ. 1992. Eusociality in Australian gall thrips. *Nature* **359**: 724-726.

Darwin C. 1859. *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. John Murray, London, UK.

Dawkins R. 1976. *The Selfish Gene*. Oxford University Press, Oxford, UK.

-. 1979. 12 Misunderstandings of Kin Selection. *Z Tierpsychol* **51**: 184-200.

-. 1982. *The Extended Phenotype*. Oxford University Press, Oxford, UK.

de Azevedo SV, Hartfelder K. 2008. The insulin signaling pathway in honey bee (*Apis mellifera*) caste development - differential expression of insulin-like peptides and insulin receptors in queen and worker larvae. *J Insect Physiol* **54**: 1064-1071.

Debelle JS, Sokolowski MB. 1987. Heredity of rover sitter - alternative foraging strategies of *Drosophila melanogaster* larvae. *Heredity* **59**: 73-83.

Deshwal S, Mallon EB. 2014. Antimicrobial peptides play a functional role in bumblebee anti-trypanosome defense. *Dev Comp Immunol* **42**: 240-243.

DeWitt T, Scheiner S. 2004. *Phenotypic plasticity: Functional and conceptual approaches*. Oxford University Press, Oxford, UK.

Dolezal AG, Toth AL. 2014. Honey bee sociogenomics: a genome-scale perspective on bee social behavior and health. *Apidologie* **45**: 375-395.

Duchateau MJ, Velthuis HHW. 1988. Development and reproductive strategies in *Bombus* colonies. *Behaviour* **107**: 186-207.

-. 1989. Ovarian development and egg-Laying in workers of *Bombus terrestris*. *Entomol Exp Appl* **51**: 199-213.

Duffy JE. 1996. Eusociality in a coral-reef shrimp. *Nature* **381**: 512-514.

Dupont YL, Damgaard C, Simonsen V. 2011. Quantitative historical change in bumblebee (*Bombus* spp.) assemblages of red clover fields. *PLOS One* **6**: e25172

Elango N, Hunt BG, Goodisman MAD, Yi SV. 2009. DNA methylation is widespread and associated with differential gene expression in castes of the honeybee, *Apis mellifera*. *Proc Natl Acad Sci USA* **106**: 11206-11211.

Elsik CG, Worley KC, Bennett AK, Beye M, Camara F, Childers CP, de Graaf DC, Debyser G, Deng JX, Devreese B et al. 2014. Finding the missing honey bee genes: lessons learned from a genome upgrade. *BMC Genomics* **15**: 86

Engels W, Kaatz H, Zillikens A, Simoes ZLP, Trube A, Braun R, Dittrich F. 1990. Honey bee reproduction - *Vitellogenin* and caste-specific regulation of fertility. In *Advances in Invertebrate Reproduction 5*, (ed. M Hoshi, O Yamashita), pp. 495-502. Elsevier, Amsterdam, The Netherlands.

Evans JD, Wheeler DE. 1999. Differential gene expression between developing queens and workers in the honey bee, *Apis mellifera*. *Proc Natl Acad Sci USA* **96**: 5575-5580.

-. 2001a. Expression profiles during honeybee caste determination. *Genome Biol* **2**: 1.

-. 2001b. Gene expression and the evolution of insect polyphenisms. *Bioessays* **23**: 62-68.

Fang Z, Rajewsky N. 2011. The impact of miRNA target sites in coding sequences and in 3'UTRs. *PLOS One* **6**: e18067.

Feldmeyer B, Elsner D, Foitzik S. 2014. Gene expression patterns associated with caste and reproductive status in ants: worker-specific genes are more derived than queen-specific ones. *Mol Ecol* **23**: 151-161.

Ferreira PG, Patalano S, Chauhan R, Ffrench-Constant R, Gabaldon T, Guigo R, Sumner S. 2013. Transcriptome analyses of primitively eusocial wasps reveal novel insights into the evolution of sociality and the origin of alternative phenotypes. *Genom Biol* **14**: R20.

Ferriere R, Michod RE. 2011. Inclusive fitness in evolution. *Nature* **471**: E6-E8.

Field J, Cant MA. 2009. Social stability and helping in small animal societies. *Phil Trans R Soc B* **364**: 3181-3189.

Foster KR, Wenseleers T, Ratnieks FLW. 2006. Kin selection is the key to altruism. *Trends Ecol Evol* **21**: 57-60.

Frank SA. 1995. Mutual policing and repression of competition in the evolution of cooperative groups. *Nature* **377**: 520-522.

Gardner A, West SA, Barton NH. 2007. The relation between multilocus population genetics and social evolution theory. *Am Nat* **169**: 207-226.

Gardner A, West SA, Wild G. 2011. The genetical theory of kin selection. *J Evolution Biol* **24**: 1020-1043.

Goulson D. 2003. *Bumblebees: Behaviour, ecology and conservation*. Oxford University Press, Oxford, UK.

Goulson D. 2013. *A Sting in the Tale*. Jonathan Cape, London, UK.

Grant V. 1978. Kin Selection: A critique. *Biol Zbl* **97**: 385-392.

Greenberg JK, Xia J, Zhou X, Thatcher SR, Gu X, Ament SA, Newman TC, Green PJ, Zhang W, Robinson GE et al. 2012. Behavioral plasticity in honey bees is associated with differences in brain microRNA transcriptome. *Genes Brain Behav* **11**: 660-670.

Griffiths-Jones S, Hui JHL, Marco A, Ronshaugen M. 2011. MicroRNA evolution by arm switching. *Embo Rep* **12**: 172-177.

Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ. 2008. *miRBase*: tools for microRNA genomics. *Nucleic Acids Res* **36**: D154-D158.

Grozinger CM, Fan YL, Hoover SER, Winston ML. 2007. Genome-wide analysis reveals differences in brain gene expression patterns associated with caste and reproductive status in honey bees (*Apis mellifera*). *Mol Ecol* **16**: 4837-4848.

Guo XQ, Su SK, Skogerboe G, Dai SAJ, Li WF, Li ZG, Liu F, Ni RF, Guo Y, Chen SL et al. 2013. Recipe for a busy bee: microRNAs in honey bee caste determination. *PLOS One* **8**: e81661.

Gutierrez L, Bussell JD, Pacurar DI, Schwambach J, Pacurar M, Bellini C. 2009. Phenotypic plasticity of adventitious rooting in *Arabidopsis* is controlled by complex regulation of *auxin* response factor transcripts and microRNA abundance. *Plant Cell* **21**: 3119-3132.

Hamilton WD. 1964a. The genetical evolution of social behaviour. I. *J Theor Biol* **7**: 1-16.

-. 1964b. The genetical evolution of social behaviour. II. *J Theor Biol* **7**: 17-52.

-. 1972. Altruism and related phenomena, mainly in social insects. *Annu Rev Ecol Syst* **3**: 193-232.

Harrison MC, Hammond RL, Mallon EB. (accepted). Reproductive workers show queen-like gene expression in an intermediately eusocial insect, the buff-tailed bumble bee Bombus terrestris. *Mol Ecol*.

Hechinger RF, Wood AC, Kuris AM. 2011. Social organization in a flatworm: trematode parasites form soldier and reproductive castes. *Proc R Soc B* **278**: 656-665.

Helantera H, Bargum K. 2007. Pedigree relatedness, not greenbeard genes, explains eusociality. *Oikos* **116**: 217-220.

Herre EA, Wcislo WT. 2011. In defence of inclusive fitness theory. *Nature* **471**: E8-E9.

Herrera EA. 2013. Ecological and genetic distribution of eusociality: the case for kin selection. *Behav Ecol* **24**: 331-333.

Holland JG, Guidat FS, Bourke AFG. 2013. Queen control of a key life-history event in a eusocial insect. *Biology Letters* **9**: 20130056

Holldobler B, Wilson EO. 2009. *The Superorganism: the beauty, elegance, and strangeness of insect societies*. W.W. Norton, New York, USA.

Hornstein E, Shomron N. 2006. Canalization of development by microRNAs. *Nat Genet* **38**: S20-S24.

Huang XA, Yin H, Sweeney S, Raha D, Snyder M, Lin HF. 2013. A major epigenetic programming mechanism guided by piRMAs. *Dev Cell* **24**: 502-516.

Hughes WOH, Oldroyd BP, Beekman M, Ratnieks FLW. 2008. Ancestral monogamy shows kin selection is key to the evolution of eusociality. *Science* **320**: 1213-1216.

Hunt JH. 2007. *The Evolution of Social Wasps*. Oxford University Press, New York, USA.

Ingram KK, Oefner P, Gordon DM. 2005. Task-specific expression of the foraging gene in harvester ants. *Mol Ecol* **14**: 813-818.

Jarvis JUM. 1981. Eusociality in a mammal - cooperative breeding in naked mole-rat colonies. *Science* **212**: 571-573.

Jarvis JUM, Bennett NC. 1993. Eusociality has evolved independently in 2 genera of bathyergid mole-rats - but occurs in no other subterranean mammal. *Behav Ecol Sociobiol* **33**: 253-260.

Johnson BR, Tsutsui ND. 2011. Taxonomically restricted genes are associated with the evolution of sociality in the honey bee. *BMC Genomics* **12**: 164.

Jones CI, Newbury SF. 2010. Functions of microRNAs in *Drosophila* development. *Biochem Soc Trans* **38**: 1137-1143.

Kamakura M. 2011. *Royalactin* induces queen differentiation in honeybees. *Nature* **473**: 478-483.

Kent DS, Simpson JA. 1992. Eusociality in the beetle *Austroplatypus incompertus* (Coleoptera, Curculionidae). *Naturwissenschaften* **79**: 86-87.

Khvorova A, Reynolds A, Jayasena SD. 2003. Functional siRNAs and miRNAs exhibit strand bias. *Cell* **115**: 209-216.

Kodaira Y, Ohtsuki H, Yokoyama J, Kawata M. 2009. Size-dependent *foraging* gene expression and behavioral caste differentiation in *Bombus ignitus*. *BMC Res Notes* **2**: 184-188.

Kozomara A, Griffiths-Jones S. 2014. *miRBase*: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res* **42**: D68-D73.

Kucharski R, Maleszka J, Foret S, Maleszka R. 2008. Nutritional control of reproductive status in honeybees via DNA methylation. *Science* **319**: 1827-1830.

Lagos-Quintana M, Rauhut R, Lendeckel W, Tuschl T. 2001. Identification of novel genes coding for small expressed RNAs. *Science* **294**: 853-858.

Langer P, Hogendoorn K, Keller L. 2004. Tug-of-war over reproduction in a social bee. *Nature* **428**: 844-847.

Lau NC, Lim LP, Weinstein EG, Bartel DP. 2001. An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science* **294**: 858-862.

Lee RC, Ambros V. 2001. An extensive class of small RNAs in *Caenorhabditis elegans*. *Science* **294**: 862-864.

Lee RC, Feinbaum RL, Ambros V. 1993. The C-Elegans heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell* **75**: 843-854.

Lee Y, Ahn C, Han JJ, Choi H, Kim J, Yim J, Lee J, Provost P, Radmark O, Kim S et al. 2003. The nuclear RNase III *Drosha* initiates microRNA processing. *Nature* **425**: 415-419.

Lee Y, Jeon K, Lee JT, Kim S, Kim VN. 2002. MicroRNA maturation: stepwise processing and subcellular localization. *Embo J* **21**: 4663-4670.

Legeai F, Rizk G, Walsh T, Edwards O, Gordon K, Lavenier D, Leterme N, Mereau A, Nicolas J, Tagu D et al. 2010. Bioinformatic prediction, deep sequencing of microRNAs and expression analysis during phenotypic plasticity in the pea aphid, *Acyrthosiphon pisum*. *BMC Genomics* **11**: 281-289.

Lehmann L, Keller L, West S, Roze D. 2007. Group selection and kin selection: two concepts but one process. *Proc Natl Acad Sci USA* **104**: 6736-6739.

Li Y, Wang F, Lee JA, Gao FB. 2006. MicroRNA-9a ensures the precise specification of sensory organ precursors in *Drosophila*. *Gene Dev* **20**: 2793-2805.

Libbrecht R, Corona M, Wende F, Azevedo DO, Serrao JE, Keller L. 2013. Interplay between *insulin signaling*, *juvenile hormone*, and *vitellogenin* regulates maternal effects on polyphenism in ants. *Proc Natl Acad Sci USA* **110**: 11050-11055.

Liu F, Peng W, Li Z, Li W, Li L, Pan J, Zhang S, Miao Y, Chen S, Su S. 2012. Next-generation small RNA sequencing for microRNAs profiling in *Apis mellifera*: comparison between nurses and foragers. *Insect Mol Biol* **21**: 297-303.

Lopez-Vaamonde C, Brown RM, Lucas ER, Pereboom JJM, Jordan WC, Bourke AFG. 2007. Effect of the queen on worker reproduction and new queen production in the bumble bee *Bombus terrestris*. *Apidologie* **38**: 171-180.

Lund E, Guttinger S, Calado A, Dahlberg JE, Kutay U. 2004. Nuclear export of microRNA precursors. *Science* **303**: 95-98.

Lyko F, Maleszka R. 2011. Insects as innovative models for functional studies of DNA methylation. *Trends Genet* **27**: 127-131.

Lytle JR, Yario TA, Steitz JA. 2007. Target mRNAs are repressed as efficiently by microRNA-binding sites in the 5 ' UTR as in the 3 ' UTR. *Proc Natl Acad Sci USA* **104**: 9667-9672.

Mangan S, Alon U. 2003. Structure and function of the feed-forward loop network motif. *Proc Natl Acad Sci USA* **100**: 11980-11985.

Marco A, Hui JHL, Ronshaugen M, Griffiths-Jones S. 2010. Functional shifts in insect microRNA evolution. *Genome Biol Evol* **2**: 686-696.

Marco A, Macpherson JI, Ronshaugen M, Griffiths-Jones S. 2012. MicroRNAs from the same precursor have different targeting properties. *Silence* **3**: 8.

Maynard Smith J, Szathmary E. 1995. *The Major Transitions in Evolution*. W. H. Freeman, Oxford, UK.

Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D, Alon U. 2002. Network motifs: Simple building blocks of complex networks. *Science* **298**: 824-827.

Moczek AP. 2010. Phenotypic plasticity and diversity in insects. *Phil Trans R Soc B* **365**: 593-603.

Mutti NS, Dolezal AG, Wolschin F, Mutti JS, Gill KS, Amdam GV. 2011. *IRS* and *TOR* nutrient-signaling pathways act via juvenile hormone to influence honey bee caste fate. *J Exp Biol* **214**: 3977-3984.

Nijhout HF. 2003. Development and evolution of adaptive polyphenisms. *Evol Dev* **5**: 9-18.

Nowak MA, Tarnita CE, Wilson EO. 2010. The evolution of eusociality. *Nature* **466**: 1057-1062.

Nunes FM, Ihle K, Mutti N, Simões Z, Amdam G. 2013. The gene *vitellogenin* affects microRNA regulation in honey bee (*Apis mellifera*) fat body and brain. *J Exp Biol* **216**: 3724-3732.

Patalano S, Hore TA, Reik W, Sumner S. 2012. Shifting behaviour: epigenetic reprogramming in eusocial insects. *Curr Opin Cell Biol* **24**: 367-373.

Patel A, Fondrk MK, Kaftanoglu O, Emore C, Hunt G, Frederick K, Amdam GV. 2007. The making of a queen: *TOR* pathway is a key player in diphenic caste development. *PLOS One* **2**: e509.

Pereboom JJM, Jordan WC, Sumner S, Hammond RL, Bourke AFG. 2005. Differential gene expression in queen-worker caste determination in bumble-bees. *Proc R Soc B* **272**: 1145-1152.

Pereira HS, Sokolowski MB. 1993. Mutations in the larval *foraging* gene affect adult locomotory behavior after feeding in *Drosophila melanogaster*. *Proc Natl Acad Sci USA* **90**: 5044-5046.

Peterson KJ, Dietrich MR, McPeek MA. 2009. MicroRNAs and metazoan macroevolution: insights into canalization, complexity, and the Cambrian explosion. *Bioessays* **31**: 736-747.

Queller DC. 1992. A general-model for kin selection. *Evolution* **46**: 376-380.

Queller DC, Strassmann JE. 1998. Kin selection and social insects. *Bioscience* **48**: 165-175.

Rachinsky A, Strambi C, Strambi A, Hartfelder K. 1990. Caste and metamorphosis - hemolymph titers of *juvenile-hormone* and *ecdysteroids* in last instar honeybee larvae. *Gen Comp Endocr* **79**: 31-38.

Raser JM, O'Shea EK. 2005. Noise in gene expression: Origins, consequences, and control. *Science* **309**: 2010-2013.

Ratnieks FLW, Foster KR, Wenseleers T. 2011. Darwin's special difficulty: the evolution of "neuter insects" and current theory. *Behav Ecol Sociobiol* **65**: 481-492.

Ratnieks FLW, Reeve HK. 1992. Conflict in single-queen Hymenopteran societies - the structure of conflict and processes that reduce conflict in advanced eusocial species. *J Theor Biol* **158**: 33-65.

Reinhart BJ, Slack FJ, Basson M, Pasquinelli AE, Bettinger JC, Rougvie AE, Horvitz HR, Ruvkun G. 2000. The 21-nucleotide let-7 RNA regulates developmental timing in *Caenorhabditis elegans*. *Nature* **403**: 901-906.

Riddell C, Adams S, Schmid-Hempel P, Mallon EB. 2009. Differential expression of immune defences is associated with specific host-Parasite interactions in insects. *PLOS One* **4**: e7621.

Sadd BM, Barribeau SM, Bloch G, de Graff DC, Dearden P, Elsik CG, Gadau J, Grimmelikhuijzen CJP, Hasselmann M, Lozier JD et al. 2015. The genomes of two key bumblebee species with primitive eusocial organization. *Genome Biol* **16**:76.

Sahlins M. 1977. *The Use and Abuse of Biology*. University of Michigan Press, Ann Arbor, Michigan, USA.

Sappington TW, Raikhel AS. 1998. Molecular characteristics of insect *vitellogenins* and *vitellogenin* receptors. *Insect Biochem Molec* **28**: 277-300.

Schaap P, Winckler T, Nelson M, Alvarez-Curto E, Elgie B, Hagiwara H, Cavender J, Milano-Curto A, Rozen DE, Dingermann T et al. 2006. Molecular phylogeny and evolution of morphology in the social amoebas. *Science* **314**: 661-663.

Schmid-Hempel P, Schmid-Hempel R, Brunner PC, Seeman OD, Allen GR. 2007. Invasion success of the bumblebee, *Bombus terrestris*, despite a drastic genetic bottleneck. *Heredity* **99**: 414-422.

Schmid-Hempel R, Eckhardt M, Goulson D, Heinzmann D, Lange C, Plischuk S, Escudero LR, Salathe R, Scriven JJ, Schmid-Hempel P. 2014. The invasion of southern South America by imported bumblebees and associated parasites. *J Anim Ecol* **83**: 823-837.

Schwander T, Lo N, Beekman M, Oldroyd BP, Keller L. 2010. Nature versus nurture in social insect caste differentiation. *Trends Ecol Evol* **25**: 275-282.

Schwarz DS, Hutvagner G, Du T, Xu ZS, Aronin N, Zamore PD. 2003. Asymmetry in the assembly of the RNAi enzyme complex. *Cell* **115**: 199-208.

Sempere LF, Cole CN, McPeek MA, Peterson KJ. 2006. The phylogenetic distribution of metazoan microRNAs: Insights into evolutionary complexity and constraint. *J Exp Zool B* **306B**: 575-588.

Shen-Orr SS, Milo R, Mangan S, Alon U. 2002. Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nat Genet* **31**: 64-68.

Simola DF, Wissler L, Donahue G, Waterhouse RM, Helmkampf M, Roux J, Nygaard S, Glastad KM, Hagen DE, Viljakainen L et al. 2013. Social insect genomes exhibit dramatic evolution in gene composition and regulation while preserving regulatory features linked to sociality. *Genome Res* **23**: 1235-1247.

Smith CR, Mutti NS, Jasper WC, Naidu A, Smith CD, Gadau J. 2012. Patterns of DNA methylation in development, division of labor and hybridization in an ant with genetic caste determination. *PLOS One* **7**: 42433.

Smith CR, Toth AL, Suarez AV, Robinson GE. 2008. Genetic and genomic analyses of the division of labour in insect societies. *Nat Rev Genet* **9**: 735-748.

Stern DL, Foster WA. 1996. The evolution of soldiers in aphids. *Biol Rev* **71**: 27-79.

Stolle E, Wilfert L, Schmid-Hempel R, Schmid-Hempel P, Kube M, Reinhardt R, Moritz RFA. 2011. A second generation genetic map of the bumblebee *Bombus terrestris* (Linnaeus, 1758) reveals slow genome and chromosome evolution in the Apidae. *Bmc Genomics* **12**: 48.

Strassmann JE, Page RE, Robinson GE, Seeley TD. 2011. Kin selection and eusociality. *Nature* **471**: E5-E6.

Sumner S. 2014. The importance of genomic novelty in social evolution. *Mol Ecol* **23**: 26-28.

Sumner S, Pereboom JJM, Jordan WC. 2006. Differential gene expression and phenotypic plasticity in behavioural castes of the primitively eusocial wasp, *Polistes canadensis*. *Proc R Soc B* **273**: 19-26.

Tautz J, Heilmann HR. 2008. *The Buzz about Bees - Biology of a Superorganism*. Springer-Verlag, Berlin, Germany.

Thorne BL. 1997. Evolution of eusociality in termites. *Annu Rev Ecol Syst* **28**: 27-54.

Tibbetts EA, Sheehan MJ. 2012. The effect of juvenile hormone on *Polistes* wasp fertility varies with cooperative behavior. *Horm Behav* **61**: 559-564.

Tinbergen N. 1963. On aims and methods of ethology. *Zeitschrift fur Tierpsychologie* **20**: 410-433.

Tobback J, Mommaerts V, Vandersmissen HP, Smagghe G, Huybrechts R. 2011. Age- and task-dependent *foraging* gene expression in the bumble bee *Bombus terrestris*. *Arch Insect Biochem* **76**: 30-42.

Toth AL, Robinson GE. 2007. Evo-devo and the evolution of social behavior. *Trends Genet* **23**: 334-341.

-. 2009. Evo-devo and the evolution of social behavior: Brain gene expression analyses in social insects. *Cold Spring Harb Symp Quant Biol* **74**: 419-426.

Trivers RL. 1974. Parent-Offspring Conflict. *Am Zool* **14**: 249-264.

Tu MP, Yin CM, Tatar M. 2005. Mutations in *insulin signaling* pathway alter *juvenile hormone* synthesis in *Drosophila melanogaster*. *Gen Comp Endocr* **142**: 347-356.

Velthuis HHW. 2002. The Historical background of the domestication of the bumble-bee, *Bombus terrestris*, and its introduction into agriculture. In *Pollinating Bees - The Conservation Link Between Agriculture and Nature*, (ed. P Kevan, V Imperatriz Fonseca). Ministry of the Environment, Brasilia, Brazil.

Vollrath F. 1986. Eusociality and Extraordinary Sex-Ratios in the Spider *Anelosimus-Eximius* (Araneae, Theridiidae). *Behav Ecol Sociobiol* **18**: 283-287.

Wang Y, Jorda M, Jones PL, Maleszka R, Ling X, Robertson HM, Mizzen CA, Peinado MA, Robinson GE. 2006. Functional CpG methylation system in a social insect. *Science* **314**: 645-647.

Washburn SL. 1978. Human behavior and the behavior of other animals. *American Psychol* **33**: 405-418.

Weaver DB, Anzola JM, Evans JD, Reid JG, Reese JT, Childs KL, Zdobnov EM, Samanta MP, Miller J, Elsik CG. 2007. Computational and transcriptional evidence for microRNAs in the honey bee genome. *Genome Biol* **8**: R97.

Weiner SA, Galbraith DA, Adams DC, Valenzuela N, Noll FB, Grozinger CM, Toth AL. 2013. A survey of DNA methylation across social insect species, life stages, and castes reveals abundant and caste-associated methylation in a primitively social wasp. *Naturwissenschaften* **100**: 795-799.

Weinstock GM Robinson GE Gibbs RA Worley KC Evans JD Maleszka R Robertson HM Weaver DB Beye M Bork P et al. 2006. Insights into social insects from the genome of the honeybee *Apis mellifera*. *Nature* **443**: 931-949.

West-Eberhard MJ. 1987. Flexible strategy and social evolution. In *Animal societies: theories and facts*, (ed. Y Ito, J Brown, J Kikkawa), pp. 35-51. Scientific Society Press, Tokyo, Japan.

-. 1989. Phenotypic plasticicty and the origins of diversity. *Annual Review of Ecology and Systematics* **20**: 249-278.

-. 1996. Wasp societies as microcosms for the study of development and evolution., pp. 290–317. In Natural history and evolution of paper wasps. (editors, West-Eberhard, M-J. & S. Turillazzi) Oxford University Press, Oxford.

-. 2003. *Developmental plasticity and evolution.* Oxford University Press, New York, USA.

West SA. 2009. *Sex Allocation*. Princton University Press, Princeton, New Jersey.

Williams GC. 1966. *Adaptation and Natural Selection: A Critique of Some Current Evolutionary Thought* Princeton University Press, Princeton, New Jersey.

Williams PH, Araujo MB, Rasmont P. 2007. Can vulnerability among British bumblebee (*Bombus*) species be explained by niche position and breadth? *Biol Conserv* **138**: 493-505.

Wilson DS, Wilson EO. 2007. Rethinking the theoretical foundation of sociobiology. *Q Rev Biol* **82**: 327-348.

Wilson EO. 1971. *The Insect Societies*. Belknap Press of Harvard University Press, Cambridge, Massachusetts, USA.

Wilson EO. 2005. Kin selection as the key to altruism: Its rise and fall. *Soc Res* **72**: 159-166.

Wilson EO, Holldobler B. 2005. Eusociality: Origin and consequences. *Proc Natl Acad Sci USA* **102**: 13367-13371.

Wolschin F, Mutti NS, Amdam GV. 2011. *Insulin receptor* substrate influences female caste development in honeybees. *Biol Letters* **7**: 112-115.

Woodard SH, Bloch G, Band MR, Robinson GE. 2013. Social regulation of maternal traits in nest-founding bumble bee (*Bombus terrestris*) queens. *J Exp Biol* **216**: 3474-3482.

Woodard SH, Fischman BJ, Venkat A, Hudson ME, Varala K, Cameron SA, Clark AG, Robinson GE. 2011. Genes involved in convergent evolution of eusociality in bees. *Proc Natl Acad Sci USA* **108**: 7472-7477.

Wyatt GR, Davey KG. 1996. Cellular and molecular actions of *juvenile hormone* .2. Roles of juvenile hormone in adult insects. *Adv Insect Physiol* **26**: 1-155.

Zanette LRS, Miller SDL, Faria CMA, Almond EJ, Huggins TJ, Jordan WC, Bourke AFG. 2012. Reproductive conflict in bumblebees and the evolution of worker policing. *Evolution* **66**: 3765-3777.

# Chapter 2

## MiRNAs in the genomes of *Bombus terrestris* and *B. impatiens*

*Abstract*

Sequencing genomes is important for understanding the structure, function, and evolution of genes; genome sequences also act as maps for sequences derived from transcriptomic RNA-seq experiments. To understand the evolution of eusociality (sociality with a reproductive division of labour) in bee species, the genomes of *Bombus impatiens* and *B. terrestris* have been sequenced, allowing comparisons with the previously sequenced genome of the honey bee *Apis mellifera*. One class of genes that have previously been suggested to be important in eusocial insects are the microRNAs (miRNAs); the sequences of these small RNAs have not yet been annotated in either of the newly-sequenced bee genomes. We therefore investigated the miRNAs present in the genome of *Bombus terrestris* and *B. impatiens*. Our aims in this study were: 1) to identify the miRNAs of *B. terrestris* and *B. impatiens*; 2) to compare the total numbers of miRNAs between both bumble bee species and between them and *A. mellifera*; and 3) to test whether the sequences and expression patterns of miRNAs were conserved between species. To investigate each of these aims we used miRNA-seq (deep sequencing of miRNA-enriched libraries) and bioinformatics prediction programs to identify miRNAs in both *Bombus* species. We identified 131 miRNAs in *B. terrestris*, and 114 in *B. impatiens*; of these, 17 were new miRNAs that had not previously been sequenced in any species. We found a striking level of difference in the miRNAs present between *Bombus* and *A. mellifera*, with 103 miRNAs in *A. mellifera* not being present in the genomes of the two bumble bees. Possible reasons for this include: 1) a large degree of divergence in miRNAs between *Bombus* and *A. mellifera*; and 2) a high level of miRNA misidentification in previous studies on the miRNAs of *A. mellifera*. A combination of these two reasons is likely because insect miRNAs have previously been shown to evolve and diverge very quickly, and most studies of miRNAs in *A. mellifera* have been overly reliant on deep sequencing methods to determine new miRNAs. This work will be published as part of the *Bombus* genome project (Sadd et al. 2015).

## *2.1 Introduction*

### *Eusocial insects are model organisms for studying within-species diversification of phenotypes*

Eusocial insects are species with a reproductive division of labour; they have phenotypically distinct castes that include a 'queen' or 'reproducer' caste, which specialises in reproduction, and a 'worker' caste that specialises in other tasks such as foraging and nursing (Wilson 1971). In some species there is evidence of a genetic influence on caste, but the majority of species have environmental caste determination (Smith et al. 2008; Schwander et al. 2010). This makes caste in eusocial insects an example of polyphenism, which is defined as the coexistence of two or more alternative and irreversible morphological or physiological states that are induced by an environmental change (Simpson et al. 2011). Understanding the factors that can cause polyphenisms is an important challenge for biologists, and, because polyphenisms produce variation under different environmental conditions, it will help to explain how particular phenotypes are produced from their corresponding genotypes. This makes eusocial insects an important group of model organisms for the study of phenotypic variation in general.

### *The importance of studying genomes in biology*

A genome can be defined as all of the genes present in the DNA of an organism (Ridley 1999); this includes protein-coding genes, upstream and downstream regulatory elements, and non-coding RNAs (genes that are not translated into proteins but have a regulatory role in organism development and physiology). Advances in sequencing technology have made it possible to sequence whole genomes (Roach et al. 1995). The first research groups to do this produced an almost complete genome sequence for the gram-positive bacteria strain *Haemophilus influenzae* using a modified version of Sanger cloning called shotgun sequencing (Fleischmann et al. 1995). By sequencing the first free-living organism, the group was able to generate insights into the bacterium's disease pathology, including the discovery of novel virulence factors. This demonstrated both the feasibility of whole organism genome sequencing and the value of genome sequencing to biology. A similar approach was used to sequence the first insect genome, that of *Drosophila melanogaster* (Adams et al. 2000), and subsequently the human genome (Venter et al. 2001), which was

sequenced using both shotgun sequencing and a BAC-based hierarchical sequencing approach (Lander et al. 2001). Obtaining the *Drosophila* genome was an important step, because *Drosophila* is a model organism for the study of genetics and as the first insect to have its genome sequenced it is still occasionally used as a solitary comparison species to predict gene sets in eusocial insects (Elsik et al. 2014).

An important step for the study of social genomics was the publication of the first genome of a eusocial insect, that of the honey bee *Apis mellifera* (Weinstock et al. 2006; Elsik et al. 2014). The *A. mellifera* genome provided a reference map for subsequent studies in the honey bee that aimed to investigate the molecular basis of social behaviour by identifying differentially expressed genes in different castes and in subgroups of workers specialising on different tasks (Thompson et al. 2008; Begna et al. 2012; Chen et al. 2012; Zayed and Robinson 2012). In addition to seeking to understand the molecular basis of social behaviour, some research groups have attempted to use the honey bee genome to investigate world-wide declines in honey bee populations (Alaux et al. 2011; Johnson et al. 2012) and specifically their vulnerabilities to disease (Cornman et al. 2012). Therefore the first social insect genome sequence has been shown have important conservation applications.

In recent years whole genome shotgun sequencing has been largely replaced with 'next generation' sequencing, which parallelises the sequencing process to make it faster, more accurate and cheaper. The impact of these technologies across a wide range of subject areas within modern biology has been immense. Within the field of sociogenomics, advances in sequencing technologies have led to the publication of eight more eusocial insect genomes including those of the ants *Camponotus floridanus* and *Harpegnathos saltator* (Bonasio et al. 2010), *Acromyrmex exhinatior* (Nygaard et al. 2011), *Atta cephalotes* (Suen et al. 2011), *Linepithema humile* (Smith et al. 2011a), *Pogonomyrmex barbatus* (Smith et al. 2011b) and *Solenopsis invicta* (Wurm et al. 2011). More recently, the whole genome of a facultatively eusocial sweat bee, *Lasioglossum albipes*, has also been sequenced (Kocher et al. 2013). These species have often had their genomes compared with that of the parasitoid *Nasonia vitripennis*, which has been used to serve as a non-social Hymenopteran outgroup (Werren et al. 2010). A wide-ranging comparison of these species has shown that many of the social species exhibit a large, rapidly evolving set of newly-

acquired lineage-specific genes, but also that there are some regulatory features conserved across the social lineages, which implies that regulation of gene expression, and a relatively small number of key genes, are important in the evolution of sociality (Simola et al. 2013). Therefore genomic approaches are already starting to yield insights into the molecular basis of eusocial evolution from an ancestral solitary state.

*The genomes of the bumble bees, Bombus terrestris and B. impatiens*

Recently the genomes of another two social insect species (the bumble bees *Bombus terrestris* and *B. impatiens*) have been sequenced and are being prepared for publication (Sadd et al. 2015; www.hymenopteragenome.org). Phylogenetic evidence suggests that bumble bees share a primitively eusocial common social ancestor with *A. mellifera* (Cardinal and Danforth 2011), but otherwise their social behaviour, caste determination mechanisms, and means of reproduction are very different. Current estimates put the divergence time between the two groups at approximately 77-95 million years ago (Cardinal and Danforth 2011). Bumble bees are increasingly being used as models for the evolution of sociality. In addition, their colonies are commercially available and are much smaller and, in terms of kin structure, simpler than those of honey bees. Despite this, bumble bees still have morphologically distinguishable queen and worker castes and mechanisms for the irreversible determination of these castes (Goulson 2003). *B. terrestris* in particular has been investigated extensively in fields such as social biology and evolution (Duchateau and Velthuis 1988; Bloch et al. 2000; Cnaani et al. 2000; Mallon and Schmid-Hempel 2004; Zanette et al. 2012; Holland et al. 2013), socio-genomics (Pereboom et al. 2005; Colgan et al. 2011; Stolle et al. 2011; Amarasinghe et al. 2014), learning (Ings et al. 2009), host-parasite interactions and immunology (Schmid-Hempel 2001; Riddell et al. 2009; Deshwal and Mallon 2014), and plant-pollinator interactions (Gegear and Burns 2007).

The *B. terrestris* genome was sequenced *de novo* using DNA prepared from haploid males, and assembled into a draft genome containing 236 Mb of sequence. The *B. impatiens* genome was sequenced and assembled using a whole genome shotgun strategy with Illumina GAIIx sequence data, and assembled into a draft genome containing 242.6 Mb of sequence. The two genomes, identified as Bter_1.0 for *B. terrestris* and Bimp_2.0 for *B.*

*impatiens*, were uploaded to *Genbank* (http://www.ncbi.nlm.nih.gov/genbank/) under the accession numbers AELG00000000.1 and AEQM0000000.2, respectively. They were annotated using *Apollo* software by an international team of 130 researchers from several institutions around the world. The genomes show several similarities with the *A. mellifera* genome (Weinstock et al. 2006; Sadd et al. 2015). For example, the methylome and methylating mechanism are well conserved across the *Bombus* species and *A. mellifera*, while several of the methylating genes are absent in distantly related species such as *Drosophila* (Wang et al. 2006; Weinstock et al. 2006). Conversely, several important developmental genes that are responsible for early developmental patterning in *Drosophila* (e.g. *torso, trunk, gurken, bicoid, swallow* and *oskar*) are absent in the two *Bombus* species and *A. mellifera*, possibly because these species have caste determination mechanisms that drastically alter early development patterning along alternative caste pathways (Weinstock et al. 2006; Sadd et al. 2015).

The *Bombus* genome project has also highlighted some important differences in the genomes of honey bees and bumble bees. For example, one interesting finding from the honey bee genome was the discovery of a wide expansion in odorant receptor genes, which are important in chemosensation and have been linked to social behaviour (Weinstock et al. 2006). The recently published ant and *Nasonia* genomes (Werren et al. 2010) exhibit the same expansion in chemosensory genes but there has been less evidence of this in *Bombus* (Sadd et al. 2015). Instead the main change as regards chemosensory genes has been in the gustatory receptors, which have undergone an expansion compared to what is found in other insects with sequenced genomes.

In addition, *Bombus* and *A. mellifera* have highly differentiated sets of the neuropeptides that control physiology and behaviour in insects in general (Weinstock et al. 2006; Sadd et al. 2015). Each lineage exhibits a unique suite of neuropeptides, with *sulfakinin* being found only in *A. mellifera* and *trissin* being found only in *Bombus*. It is possible that these unique gene sets contribute to the specific differences in social behaviour found between the two groups.

Overall, the *Bombus* genome is already starting to be important in uncovering some of the molecular processes involved in social evolution. Many of the genes that are thought to be

important in social evolution in the Hymenoptera are genes that are involved in regulation of gene expression (Simola et al. 2013). A particularly important set of genes that regulate gene expression are microRNAs (miRNAs). MiRNAs have already been implicated in the regulation of social behaviour between the genomes of different species of eusocial insects (Bonasio et al. 2010), and it is therefore important that they are characterised and annotated in the recently sequenced *Bombus* genomes (Sadd et al. 2015).

*The role of miRNAs in sociogenomics*

MiRNAs are a class of regulatory non-coding RNAs that have short sequence lengths (approximately 21-24bp). These are components of the RNA interference pathway in the majority of eukaryotes. The *Bombus* genome project have shown that the main genes involved in this pathway (i.e. *Drosha*, *pasha*, *dicer*, *argonaute*) are conserved in *Bombus* species (Sadd et al. 2015). These genes work by binding to the 3'UTR region of a target mRNA transcript by complementary base pairing. Together with *argonaute*, they are recruited into the RNA induced silencing complex (RISC), which then binds to the target and prevents it from being expressed, either by causing the target transcript to be degraded by *argonaute*, or by simply preventing the transcript from binding to the ribosomes prior to translation (Bartel 2004). In the canonical miRNA processing pathway, miRNAs are first expressed from the genome as long RNA transcripts (which often contain several miRNAs). They are then cleaved into miRNA precursor sequences by *Drosha-Pasha*, and these miRNA precursors can then be recognised by their characteristic hairpin-loop secondary structures. The precursors are recognised by the gene *Dicer*, which then cleaves them and processes them into a double-stranded miRNA duplex. One of the arms of this duplex (the pilot strand) is then further processed into the ~22 bp mature miRNA, which has biological function when combined with the *RISC* (Kim et al. 2009).

Previous experiments have shown that miRNAs play an important role in regulating phenotypic plasticity (where multiple phenotypes are expressed from the same genotype) in plants and insects (Gutierrez et al. 2009; Legeai et al. 2010). Their expression patterns have also been shown to correlate with several processes in *A. mellifera*, including development (Weaver et al. 2007; Chen et al. 2010), queen-worker caste dimorphism (Weaver et al. 2007; Guo et al. 2013), the temporal polyethisms that define task

specialisation in honey bee workers (Behura and Whitfield 2010; Greenberg et al. 2012; Liu et al. 2012), and *vitellogenin*-associated effects on caste physiology (Nunes et al. 2013). The fact that miRNAs are known to affect phenotypic plasticity, caste biology and social processes, and are important in numerous developmental processes, means that identifying them and annotating them in the sequenced social insect genomes should be a high priority in the study of their potential effects on the evolution of eusociality. It is also important to use a comparative approach to study the impact of miRNAs on eusocial evolution. A similar approach has previously been used to compare the miRNAs of the solitary insect species *Drosophila melanogaster* and *Tribolium castaneum* (Marco et al. 2010). This showed that new miRNAs can evolve rapidly in insect lineages, and that the function and expression patterns of miRNAs can sometimes undergo large shifts, even when the sequences are conserved between species (Marco et al. 2010).

*Aims*

This study used Illumina miRNA-seq and bioinformatics to identify the miRNAs of *B. terrestris* and *B. impatiens*, the genomes of which have both recently been sequenced. This study makes up part of the *Bombus* genome project, which sequenced and annotated the genomes of the two species in order to compare and contrast the evolution of eusociality across *Bombus* and *Apis* (Sadd et al. 2015). The broad aims of the present study were: First, to sequence and characterise the miRNAs of *B. terrestris* and to infer the miRNAs that were not sequenced but nevertheless still present in the genomes of *B. terrestris* and *B. impatiens* using computational predictions; second, to compare the total numbers of miRNAs between both bumble bee species and *A. mellifera*; and third, to test whether the sequences of miRNAs were conserved between these species and also to test whether the expression patterns were the same between them. To identify the miRNAs, we sequenced RNA fragments extracted from *B. terrestris* female larvae that were between 16-32 bp in length, mapped them to the *B. terrestris* genome, and then compared them between the two *Bombus* species and *A. mellifera*, previously the only other eusocial insect with a well-characterised set of miRNAs (Weaver et al. 2007; Behura and Whitfield 2010; Greenberg et al. 2012). We also used prediction software to determine the identity of putative miRNAs that were not expressed during larval development and of the non-sequenced miRNAs from *B. impatiens*.

## *2.2 Methods*

### *Bumble bee rearing and husbandry*

We received four solitary queens of *B. terrestris* with a central brood mass containing eggs, larvae, and pupae, but no workers from Syngenta Bioline (Weert, Netherlands). By ordering solitary queens we were able to record the first worker eclosion (when an adult bee emerges from its pupal casing), which allowed us to estimate when the queen stopped producing diploid eggs (which develop into females), and instead produced haploid eggs (which develop into males). This point is the 'switch point' (Duchateau and Velthuis 1988) and it usually occurs 14-20 days after the first workers eclose, we collected larvae before this point to ensure that all of them were workers (and therefore more easily comparable between replicates). We housed the colonies in boxes kept at 28°C and 60% humidity under red light. We supplied the colony with sugar-syrup (Syngenta Bioline Bees B.V, Weert, The Netherlands) and freeze dried pollen (Koppert) *ad libitum*.

Bumble bee larvae undergo four instars (growth phases between moults) during their larval development. To sample the larvae we collected 164 first to second instar (early-instar) larvae and 85 fourth instar (late-instar) larvae from across the four colonies, keeping each colony and each age cohort separate. This produced eight larval samples in total. In addition, we let 148 larvae across the colonies develop into adults to ensure that all of the sampled individuals were from the same sex and caste (female, workers). Further details of the larval collection methods are provided in **Chapter 3.2** '*Sampling of queen- and worker-destined larvae from bumble bee colonies*' in this thesis. The four colonies that were used to produce RNA libraries in this chapter are labelled QR-2, QR-17, QR-15 and QR-17 in **Table 3.1** in **Chapter 3**. The larvae that were removed from each colony were pooled together into eight larval homogenate mixes (two from each colony: one representing early-instar larvae, one representing late-instar larvae). Each larval homogenate was used to make a cDNA library, the eight cDNA libraries are labelled EW1-4 and LW1-4 in **Table 3.3** in **Chapter 3**.

*RNA extractions*

Below is a summary of the methods for extracting RNA and making miRNA-seq libraries. Detailed methods for these procedures are provided in **Appendix 1**. We homogenised the larval pooled tissue with liquid nitrogen and Trizol reagent (Ambion, Foster City, California, USA) using a mortar and pestle. We then extracted total RNA according to the Trizol manufacturer's instructions.

To ensure that the RNA was not degraded, we separated each sample on a 1.2% agarose gel. When insect RNA is heat-treated, the large ribosomal sub-unit splits into two equally-sized bands that both separate at the same size fraction as the small ribosomal sub-unit (Winnebeck et al. 2010). As a result, only one band that comprises both the small ribosomal sub-unit and the two fragments of the large ribosomal sub-unit is clearly visible instead of the usual two. We used the presence of this single band as an indicator of the amount of RNA degradation. None of the RNA extracted in this way showed a ribosomal band that appeared faint or degraded. Therefore the RNA was of sufficient quality to be used for miRNA-seq library preparation.

*MiRNA-seq library preparation*

Libraries were prepared with the Illumina TruSeq RNA sample preparation kit (version 1.5; Epicentre Technologies, Madison, Wisconsin, USA). MiRNA-seq has been shown to produce a bias in its read distributions, such that certain sequences become over- or under-represented relative to their actual abundances (Hafner et al. 2011; Jayaprakash et al. 2011; Sun et al. 2011; Sorefan et al. 2012). To adjust for this bias, we used a custom-designed 3' adaptor sequence with a 4 bp degenerate nucleotide end (NNNN) that has been shown to reduce miRNA-seq bias (Jayaprakash et al. 2011; Sorefan et al. 2012). Previous experiments have shown that the bias comes from the stability of the secondary structure formed when the sRNA sequence ligates to both the 5' and 3' adaptors. However, in the present experiments, the ligation protocol did not produce sufficient quantities of the adaptor ligated product with the customised adaptors to be useful for cDNA library synthesis; we found that the library preparations only produced a clear product when one adaptor was modified in this way (data not shown). Much of the sequencing bias comes from the use of the truncated ligase 2 during the first adaptor ligation when the pre-

adenylated 3' adaptor is ligated to the sequences of interest. Less bias comes from *ligase 1* which ligates the non-adenylated 5' adaptor to the 5' end of the sequences of interest (Jayaprakash et al. 2011). Therefore we modified the 3' adaptor with the 4 bp degenerate nucleotide end but not the 5' adaptor, this method allowed us to minimise the sequencing bias inherent in miRNA-seq (Sorefan et al. 2012).

We ligated the adaptors to the RNA sequences of interest using a modified version of the TruSeq RNA sample preparation kit manufacturer's instructions. In the modified version, the custom-made 3' adaptor described above was used instead of the one provided with the kit and all of the steps were followed using half reactions. Following preparation of the cDNA, we amplified the DNA with primers complementary to the adaptor sequences. We separated the PCR products on an 8% polyacrylamide gel, and cut out the 21-24mer band that contained miRNAs and miRNA-length sequences. We separated the nucleic acid products from the gel fragments by shaking them overnight in New England Bio labs (NEB) buffer 2. We then precipitated the products in 75% ethanol and dissolved them in water. This protocol can sometimes produce adaptor-adaptor products where the 5' adaptor ligates to surplus 3' adaptor from the first ligation reaction. These reactions can leave a large product that separates just below the band containing miRNAs, causing contamination of the extracted bands. This ultimately reduces the number of miRNAs that are sequenced because they compete with the adaptor-adaptor products in the sequencing process. To minimise the risk of adaptor-adaptor products contaminating the miRNA fragment in this way, we separated the products on a second gel. We then extracted the nucleic acid as before. We separated one sixth of the product on a third polyacrylamide gel to ensure the same nucleic acid fraction was still present.

*Sanger sequencing and Illumina deep sequencing*

Having prepared the cDNA library samples, we used Sanger chain termination sequencing to ensure the libraries contained products of length 21-24 bp, which corresponds with the expected length of miRNAs, therefore ensuring that the libraries contained miRNAs. We used a fifth of the cDNA library and cloned the products into *Escherichia coli* using pGEMTeasy (Clontech, Saint-Germain-en-Laye, France) as a vector. We transformed the recombinant plasmids into DH5α super-competent *E. coli* using blue-white colony staining.

We removed ten white colonies from the library cultures and extracted the recombinant plasmids using a Qiaprep Spin Mini-prep kit (Qiagen, Manchester, UK) according to the manufacturer's instructions. We sent the samples to The Genome Analysis Centre (TGAC, Norwich, UK) for Sanger chain termination sequencing.

We analysed the returned reads (one read is a single sequence of RNA that has been returned by miRNA-seq) using the sequence analysis software *FinchTV* (http://www.geospiza.com/Products/finchtv.shtml). This confirmed that each of the ten sequenced plasmids from the sequenced library contained miRNAs or miRNA-length sequences (data not shown). The method showed that miRNAs that map to the *B. terrestris* genome were present in the libraries.

Having verified that the Sanger libraries contained miRNAs we then sent the libraries to BaseClear B.V (Leiden, The Netherlands) for Illumina deep sequencing on a HiSeq2000.

*Bioinformatics analysis*

The Illumina sequencing returned approximately 29 million reads across all eight libraries. We stripped the returned reads of their adaptor sequences, matching the 8 nucleotide sequence at the 5' end of the 3' adaptor and then removing it together with the 4 nucleotide overhang that made up the degenerate nucleotide sequence of the 3' adaptor. We then mapped sequences to the *B. terrestris* genome with *PatMaN* and excluded any sequences with mismatches. This lead to 53.5% of the sequenced reads being mapped to the *B. terrestris* genome. We predicted new *Bombus* miRNAs using *MapMi* to align all mature sequences from all species on *miRBase* (Kozomara and Griffiths-Jones 2011). We also aligned the sequences to a list of new *A. mellifera* miRNAs that had not yet been added to *miRBase* (Greenberg et al. 2012).

To identify the mature miRNAs, we first identified the precursor sequences from their secondary structures. To do this, we used the miRNA prediction software, *miRCat* (Moxon et al. 2008), which employs both miRNA-seq data and the genomic context of the mapped sequences to identify the putative precursor sequences from the genome of the focal organism (here, *B. terrestris*). This software uses a variety of parameters such as precursor length, number of paired nucleotides, and the minimum free energy value of the secondary structure (Moxon et al. 2008). It is thus conservative in its ability to predict miRNAs. Using

*miRCat*, we predicted 67 miRNAs in *B. terrestris* and we compared these to the known miRNAs of *A. mellifera* published in *miRBase* (Kozomara and Griffiths-Jones 2011). This procedure revealed 14 miRNAs in *B. terrestris* that had not previously been identified in *A. mellifera*. We used *Basic Local Alignment Search Tool* (BLAST; v.2.2.15) to analyse these sequences against the *A. mellifera* and *B. impatiens* genomes. We identified homologues as sequences that had >85% homology, and that still formed hairpin secondary structures when input into *RNAfold* (http://srna-tools.cmp.uea.ac.uk/animal/cgi-bin/srna-tools.cgi).

To identify more miRNAs, including those in *B. impatiens*, we downloaded the precursor sequences of all the published miRNAs for *A. mellifera* from *miRBase* (Kozomara and Griffiths-Jones 2011) and performed a search in *BLAST* to identify the homologues of these sequences in the *B. terrestris* and *B. impatiens* genomes. For all of the miRNAs that had homologues, we then used the miRNA prediction tool *miR-abela* (Sewer et al. 2005) to identify hairpin-loop structures in the 500 bp regions around each of the sequences identified from the *BLAST* search. Unlike *miRCat*, *miR-abela* requires only the genome information to make precursor predictions.

We assessed all the *Bombus* orthologues of the miRNAs that had been published in *A. mellifera* (Chen et al. 2010; Liu et al. 2012) but were not predicted by either *miRCat* or *miR-abela*. We assessed which ones (a) showed a high mature sequence similarity to *A. mellifera* (>85%), (b) showed a clear hairpin secondary structure in their putative precursor sequences and (c), in the case of *B. terrestris*, had a read copy number of more than 100 in at least one of the B. terrestris miRNA-seq libraries. The thresholds in these criteria were selected to reduce the numbers of putative miRNAs that were false positives when identified from the *BLAST* searches. We separated the putative miRNAs from all three prediction methods into three tables, i.e. sequenced miRNAs in *B. terrestris* (containing all the miRNAs that had a read copy number of more than 100 times, **Appendix 2**), non-sequenced miRNAs in *B. terrestris* (containing all of the miRNAs that had a read copy number of less than 100 times, **Appendix 3**), and non-sequenced miRNAs in *B. impatiens* (**Appendix 4**).

Having identified all of the miRNAs we assessed their genomic locations using *GBrowse* (http://hymenopteragenome.org/cgi-bin/gb2/gbrowse/bter_v1_0_v1/). We classified

miRNAs that occurred in the introns of protein coding genes as intronic miRNAs, and miRNAs that occurred within 1 Kb of other miRNAs as clustered miRNAs. We found that 34 *Bombus* miRNAs (26%) were located in the intronic regions and 38 miRNAs (31%) were located in clusters of two or more miRNAs close together. It is common for miRNAs to be clustered in this way and the numbers of miRNAs in intronic regions or clusters approximate to the numbers reported in other animals (Rodriguez et al. 2004), including insects (Marco et al. 2010).

We used *BLAST* to analyse the amount of miRNA sequence conservation between species using only the homologues that were found in all three taxa. We calculated the Levenshtein distance between each miRNA across all three species (Kruskall 1999). We excluded miRNAs that had undergone 'arm switching' (when the mature miRNA switches arms on the miRNA precursor sequence; see below) as the mature products of these miRNAs were no longer comparable. We also excluded miRNAs that had nucleotides added at the start and end of the mature sequence in the *Bombus* species compared to *A. mellifera* because many of these miRNAs are likely to be different due to different prediction methods between studies.

Finally, we looked for evidence of mature miRNA arm switching between *B. terrestris* and *A. mellifera*. Historically, the most highly expressed sequence has been classified as the mature miRNA, since this sequence has a regulator function while the other sequence (historically termed the '*sequence') is lost or degraded (Gregory et al. 2005). It is becoming increasingly clear that sometimes these sequences switch between species, so the *sequence in one species might not be the same in a related species (Chen and Rajewsky 2007; Marco et al. 2010; Griffiths-Jones et al. 2011). In addition, it is clear that for some miRNAs both arms of the duplex have biological activity and the most highly expressed will vary in different tissues or developmental stages (Ro et al. 2007; Glazov et al. 2008). A previous study has used comparative approaches to identify arm switches between species (Marco et al. 2010). In our analyses we defined miRNA expression as the number of reads in a library for each miRNA. Every miRNA produced two complementary groups of reads which make up the sequences of the miRNA duplex, we classified the mature miRNA sequence as the most abundantly expressed of the two sequences. We then compared these to the miRNAs that were classified as the mature sequence in *A.*

*mellifera* to determine whether any of the miRNAs had undergone arm-switching between species.

## *2.3 Results*

*New miRNAs in B. terrestris and B. impatiens*

Our first aim was to identify new and conserved miRNAs in *B. terrestris* and *B. impatiens*. In our analysis, *miRCat* revealed 67 miRNAs in *B. terrestris* (11 of which were new miRNAs that were conserved in *B. impatiens*); *miR-abela* revealed 28 miRNAs in *B. terrestris* and 60 miRNAs in *B. impatiens*; and the homology searches revealed 35 miRNAs in *B. terrestris* and 55 miRNAs in *B. impatiens*. Overall, using the combined approaches, we predicted 130 miRNAs in the genome of *B. terrestris* (divided into miRNAs that were sequenced more than 100 times and miRNAs that were sequenced less than 100 times; **Appendix 2** and **Appendix 3**, respectively) and 126 miRNAs in the genome of *B. impatiens* (**Appendix 4**).

The miRNA-seq revealed 16 new miRNAs that have not been sequenced in other species. Of the new miRNAs, two (MC732 and MC762 in **Appendix 2**) were conserved across all three bee species. Nine of the new miRNAs occurred in the two bumble bee species and five of them were found in the genome of *B. terrestris* only. In addition, the *miR-abela* prediction program revealed a miRNA duplicate in *B. impatiens* that was not present in either *B. terrestris* or *A. mellifera*.

In addition to isolating the above confirmed miRNAs, we compared the list of 18 *A. mellifera* miRNAs that was published by Greenberg et al. (2012). This showed 9 putative miRNAs that were conserved in the genomes of the two *Bombus* species (**Appendix 5**).

*Cross-species comparisons of miRNAs*

Our second aim was to compare the miRNAs present in the genome in the two bumble bee species with the miRNAs that have previously been published in *A. mellifera* (**Figure 2.1**). One striking result from these analyses was the amount of differentiation in the numbers of miRNAs in the *A. mellifera* genome compared to the two *Bombus* species. *Apis* had 217 miRNAs identified in *miRBase* (Kozomara and Griffiths-Jones 2011) and for at least 100 of

these there was very little or no evidence that they were present in the genomes of the two bumble bee species, and neither was there evidence (according to *miRBase*) that they were present in other species with sequenced miRNAs. Furthermore, three *A. mellifera* miRNA genes (Bte-miR-1, Bte-miR-87, and Bte-miR-92b) were duplicated in the *A. mellifera* genome but have only one copy in the *Bombus* genomes, meaning there were 103 miRNA genes that appeared exclusively in the *A. mellifera* genome, and in no other species with a sequenced genome (**Figure 2.1**). Our sequencing results revealed two new miRNAs (MC732 and MC762 in **Appendix 2**) that are conserved in the genome of *A. mellifera*, which means that the total number of currently known putative miRNAs in *A. mellifera* is 219 (not including the newly sequenced miRNAs that were published by Greenberg et al. 2012).

Compared to the number in *A. mellifera*, the number of taxonomically-restricted miRNAs in the two bumble bee species was very small, comprising 15 miRNAs that were not observed in *A. mellifera*, of which there were only six that were restricted to *B. terrestris* alone (**Figure 2.1**). There was only one miRNA that was found exclusively in the *B. impatiens* genome (**Figure 2.1**); however, this species has had no sequencing experiments performed on it.

*MiRNA sequence conservation and arm switching*

Our third aim was to compare the mature miRNA sequences and expression patterns of the conserved miRNAs that were found in all three species. The miRNAs that had homologues in *A. mellifera* and the two bumble bees were highly conserved, with all of the identified sequences being identical or nearly identical. For example, between *B. terrestris* and *B. impatiens* there were only five miRNAs (out of 125 that were conserved between species) with two or more nucleotide substitutions (Bte-miR-750, Bte-miR-3791, Bte-miR-6000b, Bte-miR-6040, Bte-miR-6046), and between *B. terrestris* and *A. mellifera* there were 11 miRNAs (out of 116 that were conserved between species) with two or more nucleotide substitutions (Bte-miR-9c, Bte-miR-184, Bte-miR-190, Bte-miR-263a, Bte-miR-282, Bte-miR-3736, Bte-miR-3777, Bte-miR-3788, Bte-miR-3791, Bte-miR-6000, Bte-miR-6040, Bte-miR-6046; **Appendix 2, 3**, and **4**). Among all of the miRNAs compared across all three species, only two miRNAs (Bte-miR-263a and Bte-miR-3736) showed substitutions in bases 2-8 of their mature sequences. This is the 'seed region', which is reverse-complementary

to the miRNA target site (Bartel 2004), such that substitutions in this region are most likely to have functional effects by changing the genes that the miRNA targets.

We identified ten miRNAs that had switched arms between *B. terrestris* and *A. mellifera* (including Bte-miR-13a, Bte-miR-14, Bte-miR-33, Bte-miR-79, Bte-miR-278, Bte-miR-281, Bte-miR-316, Bte-miR-1175, Bte-miR-3715, and Bte-miR-3786; **Table 2.1**). For these ten miRNAs, the miRNA-seq showed that the most abundantly expressed arm in *B. terrestris* was the opposite of that in *A. mellifera*. This suggests that, even though the sequences of the miRNAs were conserved, they could have still undergone functional changes via arm switching, and could therefore also have changed their target mRNA sequences. We could not verify whether the same miRNAs exhibited arm switching in *B. impatiens* because of the absence of sequencing data for that species.

## *2.4 Discussion*

*New miRNAs identified in the genomes of Bombus species*

In this study our first aim was to sequence the miRNAs of *B. terrestris*, and isolate non-sequenced miRNAs that were conserved across the genomes of *B. terrestris* and *B. impatiens*. Overall, by using three methods, one experimental and two computational, we identified 130 miRNAs in *B. terrestris* and 126 miRNAs in *B. impatiens*. However, of these miRNAs, only 101 miRNAs were identified by deep sequencing in *B. terrestris* and hence have experimental evidence to support them (**Appendix 2**). The other miRNAs, including all of the miRNAs in *B. impatiens*, are predictions, and future studies should aim to verify whether they are expressed in the two bumble bee species by sequencing from other developmental stages.

It should be noted that, even for the 101 sequenced miRNAs in *B. terrestris*, none of the sequencing results presented here have been validated using qRT-PCR and Northern blots in the present study. These methods are used to give a more accurate indication of the expression differences between experimental phenotypes, and to confirm that the sequences are miRNAs. In future chapters we will address this by looking at miRNA expression differences between different castes and developmental stages. In **Chapter 3**

we used Northern blots to validate the presence, and expression pattern of ten miRNAs in *B. terrestris* female larvae (**Chapter 3, Figure 3.5**), and in **Chapter 4** we use Northern blots to validate the presence and expression pattern of eight miRNAs in the ovaries (**Chapter 4**, **Figure 4.5**), and three miRNAs in the brains (**Chapter 4**, **Figure 4.7**) in *B. terrestris* adult females. However this still leaves several *B. terrestris* miRNAs, including all of the new miRNAs presented in this study (miRNAs beginning 'MC' in **Appendix 2**), that have not been validated in any study, which should be addressed in future studies on the function of *B. terrestris* miRNAs (e.g. See **Appendix 10**).

We also found evidence in the genome of *B. terrestris* of nine other miRNAs that were recently sequenced by Greenberg et al. (2012) in *A. mellifera* but the sequences of which have not yet been uploaded to *miRBase*. These putative miRNAs are well conserved between *B. terrestris* and *A. mellifera* and show features of non-coding regulatory RNAs (such as the hairpin loop structure of the precursor sequence which characterises sequences that are processed by *Dicer*) but as they have not yet been formally classified as miRNAs and were not sequenced in high frequencies in our libraries, we did not investigate them further.

*Diversification of miRNAs between three bee species*

Our second aim was to compare the total numbers of miRNAs between both bumble bee species and *A. mellifera*. One striking result from this study is the large differences in the numbers of taxonomically-restricted miRNAs between *A. mellifera* and the two *Bombus* species. This is striking because such large differences would not be expected to exist between lineages that diverged only 77-95 million years ago (Cardinal and Danforth 2011) and that show strong similarities in most other parts of their genomes (Hymenopteragenome.org; unpublished data). The *B. terrestris* genome itself has evolved slowly and is not very strongly diverged with *A. mellifera* (Stolle et al. 2011) which makes these results even more surprising.

To an extent, it might be expected that there would be several recently-evolved miRNAs in any given species. One model for miRNA evolution is the 'transcriptional control model', which posits that young, recently-evolved miRNAs are first expressed only at low levels, are tissue- or stage-specific, and have little effect on the overall phenotype (Chen and Rajewsky

2007). With time, most of these are lost and selection acts on an increasingly narrow number that then come to have more extensive effects on the whole organism. This model is supported by evidence in a range of organisms showing that new miRNAs of recent origin tend to be under relaxed selection, and are not as highly or broadly expressed as more ancient miRNAs (Roux et al. 2012). However, the large number of miRNAs unique to *A. mellifera* is still surprising because the rate at which new miRNAs evolve and are then maintained is thought to be quite low (approximately 0.3 long lived miRNAs per million years) in most organisms (Wheeler et al. 2009). Hence related species (up to 60 million years diverged, which relatively speaking is not very different to the estimate of approximately 77-95 million years divergence between *Bombus* and *Apis*) are usually very similar in their miRNA profiles (Lu et al. 2008; de Wit et al. 2009).

One explanation for the apparently high number of miRNAs unique to *A. mellifera* is that insects may have a particularly high rate of miRNA generation. This seems to be the case in *Drosophila* (Lu et al. 2008), and a recent study that characterised the miRNAs of *Tribolium castaneum* showed that only a third of the miRNAs were conserved with *Drosophila melanogaster* (Marco et al. 2010). Therefore it is possible that other species of insect have higher rates of miRNA evolution too, though data confirming this in eusocial insects are lacking since miRNAs have only been sequenced in *A. mellifera* (Chen et al. 2010), *Camponotus floridanus* (Bonasio et al. 2010), and now *B. terrestris* (this study) among the eusocial insects. Another possible explanation is that many of the apparently unique *A. mellifera* 'miRNAs' superficially show the characteristic secondary structure that identifies their precursors but are otherwise biologically non-functional. In support of this explanation, many of the identified sequences on *miRBase* yielded only very low read counts when they were subjected to high-throughput sequencing analysis (Chen et al. 2010; Liu et al. 2012). Furthermore the mature sequence predictions of some miRNAs are likely to be erroneous, for example the *miRBase* predictions for Ame-miR-317, Ame-miR-190, and Ame-miR-282 have mature sequence lengths of 25, 26, and 28 nucleotides respectively, which is unlikely given that miRNAs are normally only 21-23 bp in length (Bartel 2004). One potential problem for many miRNA studies is that they rely primarily on high-throughput sequencing to characterise the miRNAs, but they do not always adequately control for the fact that many of their identified sequences will be RNA

fragments of the same size as miRNAs with immediate genomic contexts presenting a superficially precursor-like general secondary structure (Kozomara and Griffiths-Jones 2011). In this study we reduced the number of false positives arising from this cause by using comparative approaches to ensure that most of the identified miRNAs were conserved between bee species. For the few miRNAs that were not conserved, we only used predictions of precursors from *miRCat* that took account of a larger number of parameters than other prediction programs (e.g. precursor length, number of paired nucleotides, and minimum free energy value of the secondary structure) and hence were more robust in omitting false positives (Moxon et al. 2008).

Compared to the potential number in *A. mellifera*, the number of taxonomically restricted miRNAs in the *Bombus* species was very small. The contrast between these results and the much greater number of miRNAs unique to *A. mellifera* implies that, within *Bombus*, there have been limited changes in the miRNAs since the genus split from the other members of Apidae. This is consistent with patterns shown by other genes in the *Bombus* genome (Sadd et al. 2015). To test this, further sequencing studies from a wider variety of tissues and developmental stages will be needed, given that the *A. mellifera* miRNAs come from sequencing studies mainly of adults (Behura and Whitfield 2010; Chen et al. 2010; Greenberg et al. 2012; Liu et al. 2012) whereas data in *B. terrestris* used in the present study derive from sequencing studies of larvae only.

*MiRNA sequence conservation*

Our third aim was to test whether sequences of miRNAs are conserved between species and to test whether the miRNA expression patterns were the same between species. We found that there was a high degree of sequence conservation in the miRNAs that were shared between species, with only miR-263a and miR-3736 showing substitutions in their 'seed regions' that would be likely to affect the identity of the targets of these miRNAs. Such a high degree of sequence conservation can be expected, because miRNAs often have very slow rates of evolution (Wheeler et al. 2009). In the 'transcriptional control' model mentioned above, miRNAs that are older and have broader functions and a higher degree of expression are predicted to be very highly conserved across a broad range of species (Chen and Rajewsky 2007).

Finally, we looked for evidence of mature miRNA arm switching between *B. terrestris* and *A. mellifera.* Previous studies have shown that miRNAs switch arms between species, and the phenomenon is likely to be common (Marco et al. 2010; Griffiths-Jones et al. 2011). For example a recent study showed that miR-10 has an identical mature miRNA duplex between *D. melanogaster* and *T. castaneum* even though the dominant sequence is on the opposite arm in each insect. In these species, the dominant arm was processed through information encoded in the primary miRNA transcript (Griffiths-Jones et al. 2011). In the present study, we found ten miRNAs that had switched arms (**Table 2.1**), with the most abundantly expressed arm in *B. terrestris* being the opposite of that in *A. mellifera*. Because, opposite arms have different sequences, they are likely to have different target sites. A recent study showed that two mature miRNAs from the same precursor target the 3' UTR regions of completely different mRNA transcripts, including genes that were not functionally related (Marco et al. 2012). This implies that even though the sequences of the miRNAs in the present study were conserved, there were still numerous functional changes in miRNA expression caused by arm switching in the mature miRNAs. Further work would need to focus on identifying the functions of these arm-switched miRNAs, and determining whether they target different genes in honey bees and bumble bees.

## *2.5 References*

Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, Scherer SE, Li PW Hoskins RA, Galle RF et al. 2000. The genome sequence of *Drosophila melanogaster*. *Science* **287**: 2185-2195.

Alaux C, Dantec C, Parrinello H, Le Conte Y. 2011. Nutrigenomics in honey bees: digital gene expression analysis of pollen's nutritive effects on healthy and *varroa*-parasitized bees. *BMC Genomics* **12**: 496.

Amarasinghe HE, Clayton CI, Mallon EB. 2014. Methylation and worker reproduction in the bumble-bee (*Bombus terrestris*). *Proc R Soc B* **281**: 20132502.

Bartel DP. 2004. MicroRNAs: Genomics, biogenesis, mechanism, and function. *Cell* **116**: 281-297.

Begna D, Han B, Feng M, Fang Y, Li JK. 2012. Differential expressions of nuclear proteomes between honeybee (*Apis mellifera* L.) queen and worker larvae: A deep insight into caste pathway decisions. *J Proteome Res* **11**: 1317-1329.

Behura SK, Whitfield CW. 2010. Correlated expression patterns of microRNA genes with age-dependent behavioural changes in honeybee. *Insect Mol Biol* **19**: 431-439.

Bloch G, Hefetz A, Hartfelder K. 2000. Ecdysteroid titer, ovary status, and dominance in adult worker and queen bumble bees (*Bombus terrestris*). *J Insect Physiol* **46**: 1033-1040.

Bonasio R, Zhang G, Ye C, Mutti NS, Fang X, Qin N, Donahue G, Yang P, Li Q, Li C et al. 2010. Genomic comparison of the ants *Camponotus floridanus* and *Harpegnathos saltator*. *Science* **329**: 1068-1071.

Cardinal S, Danforth BN. 2011. The antiquity and evolutionary history of social behavior in bees. *PLOS One* **6**: e21086

Chen K, Rajewsky N. 2007. The evolution of gene regulation by transcription factors and microRNAs. *Nat Rev Genet* **8**:93-103.

Chen X, Hu Y, Zheng HQ, Cao LF, Niu DF, Yu DL, Sun YQ, Hu SN, Hu FL. 2012. Transcriptome comparison between honey bee queen- and worker-destined larvae. *Insect Biochem Mol Biol* **42**: 665-673.

Chen X, Yu X, Cai Y, Zheng H, Yu D, Liu G, Zhou Q, Hu S, Hu F. 2010. Next-generation small RNA sequencing for microRNAs profiling in the honey bee *Apis mellifera*. *Insect Mol Biol* **19**: 799-805.

Cnaani J, Robinson GE, Bloch G, Borst D, Hefetz A. 2000. The effect of queen-worker conflict on caste determination in the bumblebee *Bombus terrestris*. *Behav Ecol Sociobiol* **47**: 346-352.

Colgan TJ, Carolan JC, Bridgett SJ, Sumner S, Blaxter ML, Brown MJF. 2011. Polyphenism in social insects: insights from a transcriptome-wide analysis of gene expression in the life stages of the key pollinator, *Bombus terrestris*. *BMC Genomics* **12**: 623.

Cornman RS, Tarpy DR, Chen YP, Jeffreys L, Lopez D, Pettis JS, van Engelsdorp D, Evans D. 2012. Pathogen webs in collapsing honey bee colonies. *PLOS One* **7**: E43562

de Wit E, Linsen SEV, Cuppen E, Berezikov E. 2009. Repertoire and evolution of miRNA genes in four divergent nematode species. *Genome Res* **19**: 2064-2074.

Deshwal S, Mallon EB. 2014. Antimicrobial peptides play a functional role in bumblebee anti-trypanosome defense. *Dev Comp Immunol* **42**: 240-243.

Duchateau MJ, Velthuis HHW. 1988. Development and reproductive strategies in *Bombus* colonies. *Behaviour* **107**: 186-207.

Elsik CG, Worley KC, Bennett AK, Beye M, Camara F, Childers CP, de Graaf DC, Debyser G, Deng JX, Devreese B et al. 2014. Finding the missing honey bee genes: lessons learned from a genome upgrade. *BMC Genomics* **15**: 86.

Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, Kerlavage AR, Bult CJ, Tomb JF, Dougherty BA, Merrick JM et al. 1995. Whole-genome random sequencing and assembly of *Haemophilus-Influenzae Rd*. *Science* **269**: 496-512.

Gegear RJ, Burns JG. 2007. The birds, the bees, and the virtual flowers: Can pollinator behavior drive ecological speciation in flowering plants? *Am Nat* **170**: 551-566.

Glazov EA, Cottee PA, Barris WC, Moore RJ, Dalrymple BP, Tizard ML. 2008. A microRNA catalog of the developing chicken embryo identified by a deep sequencing approach. *Genome Res* **18**: 957-964.

Goulson D. 2003. *Bumblebees: Behaviour, ecology and conservation*. Oxford University Press, Oxford, UK.

Greenberg JK, Xia J, Zhou X, Thatcher SR, Gu X, Ament SA, Newman TC, Green PJ, Zhang W, Robinson GE et al. 2012. Behavioral plasticity in honey bees is associated with differences in brain microRNA transcriptome. *Genes Brain Behav* **11**: 660-670.

Gregory RI, Chendrimada TP, Cooch N, Shiekhattar R. 2005. Human RISC couples microRNA biogenesis and posttranscriptional gene silencing. *Cell* **123**: 631-640.

Griffiths-Jones S, Hui JHL, Marco A, Ronshaugen M. 2011. MicroRNA evolution by arm switching. *Embo Reports* **12**: 172-177.

Guo XQ, Su SK, Skogerboe G, Dai SAJ, Li WF, Li ZG, Liu F, Ni RF, Guo Y, Chen SL et al. 2013. Recipe for a busy bee: microRNAs in honey bee caste determination. *PLOS One* **8**: e81661.

Gutierrez L, Bussell JD, Pacurar DI, Schwambach J, Pacurar M, Bellini C. 2009. Phenotypic plasticity of adventitious rooting in *Arabidopsis* is controlled by complex regulation of *auxin* response factor transcripts and microRNA abundance. *Plant Cell* **21**: 3119-3132.

Hafner M, Renwick N, Brown M, Mihailovic A, Holoch D, Lin C, Pena JTG, Nusbaum JD, Morozov P, Ludwig J et al. 2011. RNA-ligase-dependent biases in miRNA representation in deep-sequenced small RNA cDNA libraries. *RNA* **17**: 1697-1712.

Holland JG, Guidat FS, Bourke AFG. 2013. Queen control of a key life-history event in a eusocial insect. *Biology Letters* **9**: 20130056.

Ings TC, Raine NE, Chittka L. 2009. A population comparison of the strength and persistence of innate colour preference and learning speed in the bumblebee *Bombus terrestris*. *Behav Ecol Sociobiol* **63**: 1207-1218.

Jayaprakash AD, Jabado O, Brown BD, Sachidanandam R. 2011. Identification and remediation of biases in the activity of RNA ligases in small-RNA deep sequencing. *Nucleic Acids Res* **39**: e141.

Johnson RM, Mao WF, Pollock HS, Niu GD, Schuler MA, Berenbaum MR. 2012. Ecologically appropriate xenobiotics induce *cytochrome P450s* in *Apis mellifera*. *PLOS One* **7**: e31 051.

Kim VN, Han J, Siomi MC. 2009. Biogenesis of small RNAs in animals. *Nature Reviews Molecular Cell Biology* **10**: 126-139.

Kocher SD, Li C, Yang W, Tan H, Yi SV, Yang XY, Hoekstra HE, Zhang GJ, Pierce NE, Yu DW. 2013. The draft genome of a socially polymorphic halictid bee, *Lasioglossum albipes*. *Genom Biol* **14**: R142.

Kozomara A, Griffiths-Jones S. 2011. miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res* **39**: D152-D157.

Kruskall JB. 1999. An overview of sequence comparison. In *Warps, String edits, and Macro molecules: The Theory and Practice of Sequence Comparison*, (ed. D Sankoff, JB Kruskall). CSLI puplications, Stanford, California, USA.

Lander ES Consortium IHGS Linton LM Birren B Nusbaum C Zody MC Baldwin J Devon K Dewar K Doyle M et al. 2001. Initial sequencing and analysis of the human genome. *Nature* **409**: 860-921.

Legeai F, Rizk G, Walsh T, Edwards O, Gordon K, Lavenier D, Leterme N, Mereau A, Nicolas J, Tagu D et al. 2010. Bioinformatic prediction, deep sequencing of microRNAs and expression analysis during phenotypic plasticity in the pea aphid, *Acyrthosiphon pisum*. *BMC Genomics* **11**: 281-289.

Liu F, Peng W, Li Z, Li W, Li L, Pan J, Zhang S, Miao Y, Chen S, Su S. 2012. Next-generation small RNA sequencing for microRNAs profiling in *Apis mellifera*: comparison between nurses and foragers. *Insect Mol Biol* **21**: 297-303.

Lu J, Shen Y, Wu Q, Kumar S, He B, Shi S, Carthew RW, Wang SM, Wu CI. 2008. The birth and death of microRNA genes in *Drosophila*. *Nat Genet* **40**: 351-355.

Mallon EB, Schmid-Hempel P. 2004. Behavioural interactions, kin and disease susceptibility in the bumblebee *Bombus terrestris*. *J Evolution Biol* **17**: 829-833.

Marco A, Hui JHL, Ronshaugen M, Griffiths-Jones S. 2010. Functional shifts in insect microRNA evolution. *Genome Biol Evol* **2**: 686-696.

Marco A, Macpherson JI, Ronshaugen M, Griffiths-Jones S. 2012. MicroRNAs from the same precursor have different targeting properties. *Silence* **3**: 8.

Moxon S, Schwach F, Dalmay T, MacLean D, Studholme DJ, Moulton V. 2008. A toolkit for analysing large-scale plant small RNA datasets. *Bioinformatics* **24**: 2252-2253.

Nunes FM, Ihle K, Mutti N, Simões Z, Amdam G. 2013. The gene *vitellogenin* affects microRNA regulation in honey bee (*Apis mellifera*) fat body and brain. *J Exp Biol* **216**: 3724-3732.

Nygaard S, Zhang GJ, Schiott M, Li C, Wurm Y, Hu HF, Zhou JJ, Ji L, Qiu F, Rasmussen M et al. 2011. The genome of the leaf-cutting ant *Acromyrmex echinatior* suggests key adaptations to advanced social life and fungus farming. *Genome Res* **21**: 1339-1348.

Pereboom JJM, Jordan WC, Sumner S, Hammond RL, Bourke AFG. 2005. Differential gene expression in queen-worker caste determination in bumble-bees. *Proc R Soc B* **272**: 1145-1152.

Riddell C, Adams S, Schmid-Hempel P, Mallon EB. 2009. Differential expression of immune defences is associated with specific host-parasite interactions in insects. *PLOS One* **4**: e7621.

Ridley M. 1999. *Genome*. HarperCollins, New York, USA.

Ro S, Park C, Young D, Sanders KM, Yan W. 2007. Tissue-dependent paired expression of miRNAs. *Nucleic Acids Res* **35**: 5944-5953.

Roach JC, Boysen C, Wang K, Hood L. 1995. Pairwise end sequencing - a unified approach to genomic mapping and sequencing. *Genomics* **26**: 345-353.

Rodriguez A, Griffiths-Jones S, Ashurst JL, Bradley A. 2004. Identification of mammalian microRNA host genes and transcription units. *Genome Res* **14**: 1902-1910.

Roux J, Gonzalez-Porta M, Robinson-Rechavi M. 2012. Comparative analysis of human and mouse expression data illuminates tissue-specific evolutionary patterns of miRNAs. *Nucleic Acids Res* **40**: 5890-5900.

Sadd BM, Barribeau SM, Bloch G, de Graff DC, Dearden P, Elsik CG, Gadau J, Grimmelikhuijzen CJP, Hasselmann M, Lozier JD et al. 2015. The genomes of two key bumblebee species with primitive eusocial organization. *Genome Biol* **16**:76.

Schmid-Hempel P. 2001. On the evolutionary ecology of host-parasite interactions: addressing the question with regard to bumblebees and their parasites. *Naturwissenschaften* **88**: 147-158.

Schwander T, Lo N, Beekman M, Oldroyd BP, Keller L. 2010. Nature versus nurture in social insect caste differentiation. *Trends Ecol Evol* **25**: 275-282.

Sewer A, Paul N, Landgraf P, Aravin A, Pfeffer S, Brownstein MJ, Tuschl T, van Nimwegen E, Zavolan M. 2005. Identification of clustered microRNAs using an ab initio prediction method. *BMC Bioinformatics* **6**: 267.

Simola DF, Wissler L, Donahue G, Waterhouse RM, Helmkampf M, Roux J, Nygaard S, Glastad KM, Hagen DE, Viljakainen L et al. 2013. Social insect genomes exhibit dramatic evolution in gene composition and regulation while preserving regulatory features linked to sociality. *Genome Res* **23**: 1235-1247.

Simpson SJ, Sword GA, Lo N. 2011. Polyphenism in insects. *Curr Biol* **21**: R738-R749.

Smith CD, Zimin A, Holt C, Abouheif E, Benton R, Cash E, Croset V, Currie CR, Elhaik E, Elsik CG et al. 2011a. Draft genome of the globally widespread and invasive Argentine ant (*Linepithema humile*). *Proc Natl Acad Sci USA* **108**: 5673-5678.

Smith CR, Smith CD, Robertson HM, Helmkampf M, Zimin A, Yandell M, Holt C, Hu H, Abouheif E, Benton R et al. 2011b. Draft genome of the red harvester ant *Pogonomyrmex barbatus*. *Proc Natl Acad Sci USA* **108**: 5667-5672.

Smith CR, Toth AL, Suarez AV, Robinson GE. 2008. Genetic and genomic analyses of the division of labour in insect societies. *Nat Rev Genet* **9**: 735-748.

Sorefan K, Pais H, Hall AE, Kozomara A, Griffiths-Jones S, Moulton V, Dalmay T. 2012. Reducing ligation bias of small RNAs in libraries for next generation sequencing. *Silence* **3**: 4.

Stolle E, Wilfert L, Schmid-Hempel R, Schmid-Hempel P, Kube M, Reinhardt R, Moritz RFA. 2011. A second generation genetic map of the bumblebee *Bombus terrestris*

(Linnaeus, 1758) reveals slow genome and chromosome evolution in the Apidae. *BMC Genomics* **12**: 48.

Suen G, Teiling C, Li L, Holt C, Abouheif E, Bornberg-Bauer E, Bouffard P, Caldera EJ, Cash E, Cavanaugh A et al. 2011. The genome sequence of the leaf-cutter ant *Atta cephalotes* reveals insights into its obligate symbiotic lifestyle. *PLOS Genet* **7**: e1002007.

Sun GH, Wu XW, Wang JH, Li HQ, Li XJ, Gao HL, Rossi J, Yen Y. 2011. A bias-reducing strategy in profiling small RNAs using Solexa. *RNA* **17**: 2256-2262.

Thompson GJ, Kucharski R, Maleszka R, Oldroyd BP. 2008. Genome-wide analysis of genes related to ovary activation in worker honey bees. *Insect Mol Biol* **17**: 657-665.

Venter JC Adams MD Myers EW Li PW Mural RJ Sutton GG Smith HO Yandell M Evans CA Holt RA et al. 2001. The sequence of the human genome. *Science* **291**: 1304-1351.

Wang Y, Jorda M, Jones PL, Maleszka R, Ling X, Robertson HM, Mizzen CA, Peinado MA, Robinson GE. 2006. Functional CpG methylation system in a social insect. *Science* **314**: 645-647.

Weaver DB, Anzola JM, Evans JD, Reid JG, Reese JT, Childs KL, Zdobnov EM, Samanta MP, Miller J, Elsik CG. 2007. Computational and transcriptional evidence for microRNAs in the honey bee genome. *Genom Biol* **8**: R97.

Weinstock GM Robinson GE Gibbs RA Worley KC Evans JD Maleszka R Robertson HM Weaver DB Beye M Bork P et al. 2006. Insights into social insects from the genome of the honeybee *Apis mellifera*. *Nature* **443**: 931-949.

Werren JH Richards S Desjardins CA Niehuis O Gadau J Colbourne JK Beukeboom LW Desplan C Elsik CG Grimmelikhuijzen CJP et al. 2010b. Functional and evolutionary insights from the genomes of three parasitoid *Nasonia* species. *Science* **327**: 343-348.

Wheeler BM, Heimberg AM, Moy VN, Sperling EA, Holstein TW, Heber S, Peterson KJ. 2009. The deep evolution of metazoan microRNAs. *Evol Dev* **11**: 50-68.

Wilson EO. 1971. *The Insect Societies*. Belknap Press of Harvard University Press, Cambridge, Massachusetts, USA.

Winnebeck EC, Millar CD, Warman GR. 2010. Why does insect RNA look degraded? *J Insect Sci* **10**: 159.

Wurm Y, Wang J, Riba-Grognuz O, Corona M, Nygaard S, Hunt BG, Ingram KK, Falquet L, Nipitwattanaphon M, Gotzek D et al. 2011. The genome of the fire ant *Solenopsis invicta*. *Proc Natl Acad Sci USA* **108**: 5679-5684.

Zanette LRS, Miller SDL, Faria CMA, Almond EJ, Huggins TJ, Jordan WC, Bourke AFG. 2012. Reproductive conflict in bumblebees and the evolution of worker policing. *Evolution* **66**: 3765-3777.

Zayed A, Robinson GE. 2012. Understanding the relationship between brain gene expression and social behavior: lessons from the honey bee. *Annu Rev Genet* **46**: 591-615.

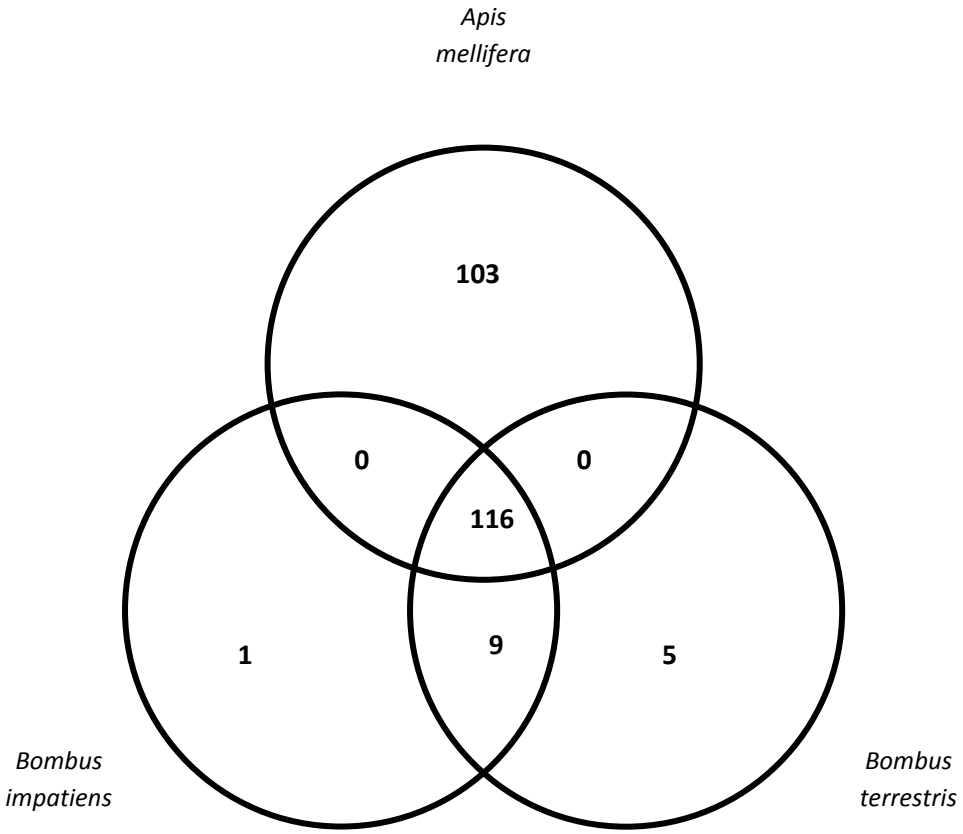


**Figure 2.1: Numbers of shared and unique miRNAs in a comparison of *Apis mellifera*, *Bombus impatiens* and *B. terrestris*.**

**Table 2.1: Arm switched miRNAs in a comparison between *Apis mellifera* and *Bombus terrestris*. For each miRNA the columns show the *Bombus* precursor sequence, and the mature sequence of each species. Brackets indicate the arm of the hairpin (5' or 3' arm) that each mature sequence is on in either species**

| MiRNA name | *B. terrestris* precursor sequence | *B. terrestris* mature sequence | *A. mellifera* mature sequence |
|---|---|---|---|
| miR-13a | ACCGAAAUGAAAAUACCUUUUGCG GUCUGAUACAUCAAAUUGGUUGUG GAAUGUUUCGAGUCAUAUCACAGC CAUUUUGUUGACUUUGGCCCGCAG AAUC | ACAUCAAAUUGGUUGUGGAAUG (5' arm) | UAUCACAGCCAUUUUGAUGAG (3' arm) |
| miR-14 | GCUAGGUCAGUGGGGGUGAGAAAC UGGCUUGGCUCUCUGUGCUACGAU AGUCAGUCUUUUUCUCUCUCCUAU CGACCUCGC | UCAGUCUUUUUCUCUCUCCUA (5' arm) | UCAGUCUUUUUCUCUCUCCUA (3' arm) |
| miR-33 | ACUUGUUACAAGUGUGCAUUGUAG UUGCAUUGCAUGUAAUAUAACUAU GCAAUACUUCUACAGUGCAACUCCU GUGGCAGGU | CAAUACUUCUACAGUGCAACUC (3' arm) | GUGCAUUGUAGUUGCAUUG (5' arm) |
| miR-79 | CUUUUGCCUUUACUUUGGUAAUAC AGCUCUAUGAUUUUAAUUAAGGAU CAUAAAGCUAGAUUACCAAAGCAAA GAGCACUGG | CUUUGGUAAUACAGCUCUAUGA (5' arm) | UAAAGCUAGAUUACCAAAGCA (3' arm) |
| miR-278 | UCACGAGCGGUGUGUCCGGAUGAG GUCUCCAUCGACCGUGAUUUAAUU UCAUAAGGUCGGUGGGACUUUCGU CCGUUUGCAAGACUCGA | CCGGAUGAGGUCUCCAUCGACC (5' arm) | UCGGUGGGACUUUCGUCCGUUU (3' arm) |
| miR-281 | GCGCUAUGAAGAGAGCUAUCCAUC GACAGUAUGGUGAUAAUAGACACU GUCAUGGAGUUGCUCUCUUUGUGG ACGC | AAGAGAGCUAUCCAUCGACAGU (5' arm) | UGUCAUGGAGUUGCUCUCUUUGU (3' arm) |
| miR-316 | GAGGGUGUGUCUUUUUCCGCUUU GCUGCCGCUGAAAUCGAGUCCGCCA GCAAAGGGGAACAGGCCGAcCCUC | CCAGCAAAGGGGAACAGGCCGA (3' arm) | UGUCUUUUUCCGCUUUGCUGCCG (5' arm) |
| miR-1175 | AGGGUUCAUGGUUCAAGUGGAGAA GUGGUCUCUACGCUUUGAAUUAAA GUGAGAUUCAACUCCUCCAACUUAA UCCUGAUCCCU | AAGUGGAGAAGUGGUCUCU (5' arm) | UGAGAUUCACUCCUCCAACUUAC (3' arm) |
| miR-3715 | GUGUGUCGUGCUAUCGGUAAGCAG AGUAUAAGACCUUUGAGAUGCGAA AUUUAAAGUAUUAUGCUCCGUUUA UCGUUGGUGCGCCAC | UAUUAUGCUCCGUUUAUCGUUG (3' arm) | UCGGUAAGCAGAGUAUAAGACCU (5' arm) |
| miR-3786 | ACAGGAACAAAACGUGUCCGACGUC UUGUCCUGGUUCAUGUAGGGCGGA AUGUUACGCAAACACUCUGUAUGG CUCAGGACGAUACGUCUGGCUCGU CCAACCA | UCUGUAUGGCUCAGGACGAUAC (3' arm) | CUUGUCCUGGUUCAUGUAGGGCG (5' arm) |

Chapter 3

# The role of microRNAs in caste determination and differentiation in the bumble bee, *Bombus terrestris*

*Abstract*

Eusocial insects are species with reproducer and non-reproducer phenotypes (usually termed queen castes and worker castes, respectively). Explaining the evolution and maintenance of castes represents a fundamental challenge in biology because the alternative phenotypes arise from the same genome but at some point in development they lose their totipotency (the ability to express both phenotypes). Previous studies have identified genes associated with caste determination and differentiation in some eusocial insects; however, one class of molecules that has been relatively ignored in these studies is the microRNAs (miRNA). MiRNAs are small non-coding RNAs that regulate gene expression at the post-transcriptional level. They have been implicated in a number of social processes in eusocial insects, but their role in caste determination has been under-investigated. In this study, we investigated the following research questions using the bumble bee, *Bombus terrestris*, a model species for the study of eusociality: 1) whether alternative queen caste and worker caste pathways are associated with miRNA expression changes; 2) whether caste-associated miRNAs induce whole-organism changes in gene expression or changes confined to specific tissues; and 3) the possible functions of caste-differentiated miRNAs in *B. terrestris*. We generated queen- and worker-destined larvae and used Illumina deep sequencing and Northern blots to profile the miRNAs of the larvae before and after the critical caste determination points. Using these methods, we show that Bte-miR-6001-5p was more highly expressed in queen-destined larvae following the loss of totipotency. Bte-miR-6001 is a mirtron expressed from the intron of a storage protein that has homology to vitellogenin, which itself has previously been associated with caste differentiation. This study describes the first well-validated miRNA associated with caste differentiation in a eusocial insect.

## *3.1 Introduction*

### *Polyphenisms and eusocial insect societies*

Phenotypic plasticity is defined as 'the ability of a single genotype to produce more than one alternative form of morphology, physiological state and/or behaviour in response to environmental conditions' (West-Eberhard 1989; Moczek 2010). The most extreme types of phenotypic plasticity are polyphenisms, which occur when an individual begins development in a state of totipotency (i.e. it is still capable of developing into more than one phenotype), and then develops via a series of irreversible changes to acquire a fixed phenotype determined by environmental conditions (Simpson et al. 2011). A fundamental problem in biology is to explain how multiple phenotypes can develop from the same genotype in this way. For example, in the development of multicellular organisms, cells specialise to perform particular tasks, even though they all have an identical starting genotype. Another example occurs in the eusocial insects, in which a single genome is able to produce multiple phenotypes that manifest as specialist castes. These are a reproducer caste (usually termed the queen caste) that specialises in reproduction and a non-reproducer caste (usually termed the worker caste) that specialises in rearing and nursing the developing larvae, foraging for pollen and nectar, and, when necessary, defending the colony against parasites and predators (Wilson 1971). In eusocial insects, queens and workers often vary between each other considerably in size, physiology, behaviour, and morphology. Thus understanding the mechanisms underlying the evolution of castes in social insects helps explain a much more fundamental biological problem.

The presence of castes in eusocial insects has long fascinated biologists. The fact that a sterile worker caste can continue to evolve and express unique traits that are not observed in the queen, even though they will never directly reproduce themselves, was considered a 'special difficulty' by Darwin (Darwin 1859). In addition, many biologists have been puzzled by the apparent paradox that a less reproductive or even completely sterile worker caste can exist at all (Dawkins 1976; Ratnieks et al. 2011). This paradox has been largely resolved by the concept of inclusive fitness, which showed that genes for altruism could undergo selection if they increased the fitness of their possessor's co-bearers (Hamilton 1964a; 1964b; Dawkins 1976). Thus Hamilton showed that a gene that is disadvantageous to its bearer is not necessarily selected against if it provides a large enough benefit to the

bearer's relatives. Conversely a gene that seems advantageous to its bearers will not necessarily be selected for if it bears a heavy cost to relatives of the bearer (Hamilton 1964b). The theory of inclusive fitness has been valuable and productive in explaining the ultimate causes of sociality (Bourke 2011a; 2011b), but the proximate mechanisms involved in the change from solitary to social insects and the development of specialised castes are far less well understood (Smith et al. 2008). Recently, transcriptomic approaches have been used to study how caste determination and differentiation work at the molecular level.

*Transcriptomics of caste differentiation in eusocial insects*

The terms caste determination, caste differentiation and caste fate may be defined in several ways. In this study, I use the following definitions: Caste determination is the point in larval development at which an individual loses its totipotency and starts down either caste pathway; caste differentiation describes the morphological, physiological and gene expression changes that occur as alternative caste pathways diverge during larval development following the loss of totipotency; and caste fate refers to the adult phenotype (queen or worker) that an individual will develop into once it has started along either caste pathway. In recent years the study of caste determination and differentiation has been greatly enhanced by the emerging field of sociogenomics, which broadly aims to identify the genes involved in social insect caste biology.

Sociogenomics uses large scale techniques such as microarrays, whole-genome sequencing and, in recent years, rapidly advancing deep sequencing technologies to isolate and identify the genes involved in caste determination and differentiation. Using these technologies permits investigators to uncover which genes are differentially expressed during caste differentiation, and therefore helps elucidate how caste systems evolved at a genomic level. Advances in this field were made possible by the publication of the honey bee (*Apis mellifera*) genome (Weinstock et al. 2006), and, accordingly, many of the early studies in sociogenomics focussed on isolating differentially regulated genes in tissues, castes and developmental stages in this eusocial insect (Evans and Wheeler 1999; Evans and Wheeler 2001b; Evans and Wheeler 2001a; Thompson et al. 2006; Grozinger et al. 2007; Thompson et al. 2007; Johnson and Tsutsui 2011; Chen et al. 2012; Niu et al. 2014).

Further advances have come with the publication of genomes of a number of ant species, including *Camponotus floridanus* and *Harpegnathos saltator* (Bonasio et al. 2010), *Acromyrmex exhinatior* (Nygaard et al. 2011), *Atta cephalotes* (Suen et al. 2011), *Linepithema humile* (Smith et al. 2011a), *Pogonomyrmex barbatus* (Smith et al. 2011b), and *Solenopsis Invicta* (Wurm et al. 2011). More recently the differential regulation of polyphenisms, including caste-associated differences in larvae and behavioural differences in adults, have begun to be addressed in other eusocial insects (Pereboom et al. 2005; Sumner et al. 2006; Colgan et al. 2011; Ferreira et al. 2013; Feldmeyer et al. 2014; Harrison et al. accepted).

Taken together, these studies have shown that: 1) caste fate is correlated with changes in expression in a wide variety of genes across both castes (Evans and Wheeler 1999; Colgan et al. 2011; Chen et al. 2012; Ferreira et al. 2013; Feldmeyer et al. 2014); and 2) that the timing and tissue-specificity of differential gene expression is often as important as the identity of the genes themselves in determining caste (Pereboom et al. 2005; Sumner 2006; Smith et al. 2008). In addition, studies on differential methylation patterns in eusocial insects have shown that many of the genes associated with caste differences are also important 'house-keeping genes' (Weiner and Toth 2012), i.e. genes that are constitutively expressed and maintain cell and tissue function, such that complete failure of expression would be lethal. These findings show that up-regulation and down-regulation of a wide variety of genes at different points in development cause larvae to develop towards alternative caste fates (Sumner 2006). This highlights the importance of regulatory genes in caste determination and differentiation. It has been said that increasing organismal complexity through evolution is (at the proximate level) a problem of regulation (Brenner 1999), and the above findings imply that this is also the case with social complexity.

Several hypotheses have been inspired by these large scale gene expression approaches to the study of eusocial insects. For example, it has been hypothesized that caste determination across eusocial lineages might involve a common genomic 'tool kit' such that the same genes or genetic pathways are co-opted in each independent origin of eusociality, and, as a corollary, that these genes are likely to be the most differentiated across different castes (Toth et al. 2007; Woodard et al. 2011). However, recent studies have used a variety of methods, including deep sequencing approaches, to show that many of the genes that

are differentially expressed between castes are taxonomically restricted, i.e. are novel genes unique to each independent eusocial lineage (Barchuk et al. 2007; Johnson and Tsutsui 2011; Ferreira et al. 2013; Feldmeyer et al. 2014). To address the relative importance of novel versus toolkit genes in caste differentiation, it is important to broaden the study of caste genes to more eusocial insects that have acquired their caste systems independently. Bumble bees represent a useful group for this purpose, because they and the other corbiculate bees (honey, stingless and orchid bees) share a common eusocial ancestor (Cardinal and Danforth 2011), suggesting, if the 'tool kit' concept is correct, that they should share genes associated with caste determination and differentiation with those found in *A. mellifera*.

*MicroRNAs and the mechanisms of phenotypic plasticity*

The study of phenotypic plasticity in general and the evolution of polyphenisms in particular has been greatly aided by the discovery of microRNAs (miRNA). These are post-transcriptional silencing molecules that make up part of the RNA interference (RNAi) pathway. The first miRNAs, lin-14 and let-7, were described in 1993 and 2000 respectively (Lee et al. 1993; Reinhart et al. 2000). MiRNAs are the best studied class of small RNA (sRNA; non-coding RNAs that regulate gene expression), and are approximately 21-23 bp in length. They regulate gene expression by being incorporated into the RNA induced silencing complex (RISC) and binding to complementary messenger RNAs (mRNA) at the 3'UTR region (Bartel 2004). This interaction leads to translational suppression and often RNA decay, thus preventing the synthesis of a particular protein. Therefore miRNAs are highly important post-transcriptional regulators of gene expression, with well understood roles in development (Kloosterman and Plasterk 2006), physiology (Chang and Mendell 2007), and evolution (Peterson et al. 2009; Axtell et al. 2011) across plants and animals. In recent years, variation in the expression of a single miRNA (miR-92a) between different strains of *Drosophila melanogaster* has been shown to induce significant changes in morphology (Arif et al. 2013), and miRNAs have also been associated with differential gene regulation in several examples of phenotypic plasticity (Weaver et al. 2007; Gutierrez et al. 2009; Legeai et al. 2010; Guo et al. 2013). For example, they have a key role in the regulation of adventitious rooting in *Arabidopsis*, a plastic phenotype that varies in response to sunlight (Gutierrez et al. 2009). Among insects, a recent study used deep

sequencing of miRNAs (henceforth, miRNA-seq) to show that up to 17 miRNAs are differentially expressed between whole body RNA extracts of winged and wingless pea aphids (*Acyrthosiphon pisum*), wing development being a plastic trait that is highly responsive to population size in aphids (Legeai et al. 2010).

There is good reason to suspect that miRNAs are important regulators of social insect caste differentiation. As described above, the field of sociogenomics has shown that caste differentiation is associated with differences in gene expression at different stages and tissues during development (Smith et al. 2008). Studies of methylation have shown that some of the genes involved in caste differentiation are 'house-keeping genes' that are always switched on but often fine-tuned to different levels at different stages of development (Elango et al. 2009). Furthermore many of the genes involved have been found to be non-coding RNAs, which often play an unappreciated role in the regulation of development (Ferreira et al. 2013; Humann et al. 2013). These results highlight the fact that gene regulation is of critical importance for understanding caste differentiation, but until recently gene regulation at the post-transcriptional level by miRNAs has been largely ignored in social insects, despite its apparent importance in other polyphenisms and in development in general.

Recent studies have addressed the role of miRNAs in regulating some aspects of social behaviour in *A. mellifera* (Weaver et al. 2007; Chen et al. 2010). For example, *A. mellifera* exhibits a complex age-related division of labour in which, soon after eclosing, young adult workers specialise as nurse bees and then gradually switch to foraging behaviour later on in their lives. This process is termed age polyethism and it is another example of phenotypic plasticity. In *A. mellifera*, miRNA-seq has been used to isolate miRNAs that are differentially expressed across phenotypes exhibiting such polyethism (Behura and Whitfield 2010; Greenberg et al. 2012; Liu et al. 2012). In addition, a study in *A. mellifera* used qRT-PCR to identify miRNA expression differences in adult workers and queens, and in worker-destined pupae and queen-destined pupae, showing that Ame-miR-2 and Ame-miR-9a were more highly expressed in adult queens compared to workers while Ame-miR-71 was more highly expressed in queen-destined pupae compared to worker-destined pupae (Weaver et al. 2007). Recently, Guo et al. (2013) showed the potential for miRNAs to be directly involved in regulating caste differentiation because they are enriched in Worker Jelly (nutritious jelly

that induces worker differentiation in *A. mellifera*) compared to Royal Jelly (nutritious jelly that induces queen differentiation in *A. mellifera*). This has led to the intriguing possibility that miRNAs silence gene expression when fed to the larvae, leading larvae to develop worker-like traits. To test this, the investigators enriched Royal Jelly with the most highly expressed miRNA in Worker Jelly (Ame-miR-184), which caused queen-destined larvae to develop worker-like traits such as a reduced body size (Guo et al. 2013). This demonstrated that, at least in principle, miRNAs have the ability to influence caste determination and differentiation.

To date no study has looked at the effect of miRNAs on caste differentiation in eusocial insects other than *A. mellifera*. This is important because *A. mellifera* has 'advanced eusociality' (queens and workers are strongly morphologically differentiated, colonies are large, there is a strong worker division of labour), whereas to understand how caste systems originated it is important to study 'primitively eusocial' species, which have smaller differences between the castes. In addition, no study has investigated differences in miRNA profiles across development in larvae before and after the developmental stages at which they change from being totipotent to being fixed in their caste fate. Therefore no study has shown whether miRNAs are involved in regulating the loss of totipotency in the development of alternative caste pathways. Finally, most studies of miRNA expression in eusocial insects have focussed on sequencing miRNAs in the brain (Behura and Whitfield 2010; Greenberg et al. 2012; Liu et al. 2012; Nunes et al. 2013), and there are no studies of miRNA expression in other tissues. This is important because a study in *A. mellifera* has shown that different miRNAs are caste differentiated in different tissues (Weaver et al. 2007), and work in *Drosophila* larvae has shown that miRNAs are often under the regulation of cell-specific enhancers and are therefore commonly expressed in a tissue-specific manner (Biemar et al. 2005). Therefore understanding the functional significance of miRNAs will require identifying the tissues where they are most highly expressed, which could help elucidate which genes they are most likely to target.

*Caste determination in B. terrestris*

Bumble bees (*Bombus* spp.), like other eusocial insects, have specialised castes including a reproducer or 'queen' caste, and a non-reproducer or 'worker' caste. The queen and

workers are morphologically similar but they differ considerably in size, physiology and behaviour (Cumber 1949; Goulson 2003). In bumble bee colonies, most of the initial reproduction is done by the queen while the workers perform foraging and nursing duties.

The best studied bumble bee with respect to caste determination and differentiation is *B. terrestris*. The caste fate of any *B. terrestris* female is decided during two sensitive phases in the larval stages of development (**Figure 3.1**; Goulson 2003). When they first hatch, all of the diploid larvae are totipotent, i.e. capable of developing into either queens or workers (Free 1955; Cnaani et al. 2000b). They then undergo a series of endogenous changes that cause them to lose their totipotency. It is known that two of the key hormones involved in this process are *juvenile hormone* (JH) and *ecdysteroid* (Goulson 2003). *JH* tends to be higher in queen-destined larvae (Cnaani et al. 1997; Hartfelder et al. 2000), and queen development is associated with a number of specific peaks in *JH* titre (Cnaani et al. 2000b). Furthermore, when pure *JH* is added to early-instar larvae, they develop into queens (Bortolotti et al. 2001). Like *JH*, *ecdysteroid* occurs at higher levels in queen-destined larvae (Hartfelder et al. 2000), and is thought to act in synchrony with *JH* to regulate caste differentiation in *B. terrestris* larvae.

The main environmental influence that regulates caste in *B. terrestris* is thought to be an inhibitory non-volatile pheromone produced by the queen early in the colony life cycle that prevents the larvae from developing into new queens (**Figure 3.1**; Roseler 1970; Cnaani et al. 1997; Cnaani et al. 2000a; Lopez-Vaamonde et al. 2007). Exposure to this pheromone occurs during the first of the two sensitive phases of caste determination in *B. terrestris*. The larvae remain sensitive to this pheromone until between three to five days after they hatch (Cnaani et al. 2000b), since after this period in its presence their caste fate towards the worker phenotype becomes fixed. It is hypothesized that later on in the colony cycle the queen stops producing the pheromone, so allowing female larvae to develop along the queen pathway (Lopez-Vaamonde et al. 2007). This point in the colony cycle has been referred to as the onset of queen production (Cnaani et al. 2000a). Queen-destined larvae undergo a peak in JH production towards the end of their second instar that prevents them from moulting as early as workers. They also exhibit a second peak of JH in their third instar. The two JH peaks extend the length of time of each larval moult, thus making the total

larval development time longer. As a result the queen-destined larvae, and by extension the adult queens, are much larger than the workers (Cnaani et al. 2000b).

Even with these differences in JH production, a female larva that develops in the absence of the inhibitory queen pheromone may still revert to the worker-destined caste trajectory if, during a second sensitive phase of caste determination occurring in the later instars, it is not fed enough pollen (**Figure 3.1**). Thus the second sensitive phase of caste determination occurs later in queen-destined larval development than the first, and is dependent on the amount of food being fed to larvae (Pereboom et al. 2003). By contrast a worker-destined larva cannot change to the queen trajectory if it has already passed the second instar and has not been subjected to a peak in JH production, i.e. if it has passed the first sensitive phase in the queen's presence (Cnaani et al. 2000b). For this reason the end of the second instar is sometimes called the critical period of caste determination in *B. terrestris* (Cnaani et al. 2000b), even though the caste pathway of a queen-destined larvae in this species is still flexible. Despite sharing a common eusocial ancestor (Cardinal and Danforth 2011), the caste determining system in *B. terrestris* is substantially different from *A. mellifera*. In *B. terrestris* it is the presence of the queen and the amount of food that causes caste determination. In *A. mellifera*, the food content is more important because totipotent larvae must be fed royal jelly to develop into queens (Kamakura 2011).

*Transcriptomics in Bombus species*

To date, two studies have isolated differentially expressed genes that correlate with caste fate in *B. terrestris*. Pereboom et al. (2005) used suppression subtractive hybridization to isolate 12 genes differentially expressed between queen- and worker-destined larvae, having allowed larvae to develop as queens or workers by rearing them in queenright (with the colony queen present) or queenless conditions. These authors also showed that the timing of differential gene expression, as well as the identity of the genes involved, was likely to be important in influencing caste differentiation. Colgan et al. (2011) used 454 sequencing to profile mRNAs in larvae, pupae and adults in *B. terrestris* and thereby to characterize the gene expression changes occurring in development and between adult castes. Although these authors used only one replicate per phenotype, and did not test for caste differences in larvae, they did show that large numbers of genes were differentially

expressed between castes in adults, and that several of these genes were highly conserved genes essential for normal organismal development. Both these studies reinforce the conclusion that regulation of gene expression is of critical importance in caste differentiation in eusocial insects. Both studies have also shown that the genes involved in caste differentiation in *A. mellifera* and other eusocial lineages are not necessarily the same genes that affect caste differentiation in bumble bees. This means that addressing the molecular basis of caste determination will require comparative studies of many different species and caste systems.

The fact that caste differentiation involves the co-ordinated regulation of large numbers of important, conserved genes and is highly stage- and tissue-specific means that understanding how all caste-associated genes are regulated is critical to understanding the molecular basis of caste. To date there have been no studies in *Bombus* that have focussed on the regulatory aspects of caste determination. To approach this, research should be directed at understanding regulation of gene expression at all levels, ranging from the genome (Stolle et al. 2011), transcriptome (Pereboom et al. 2005; Colgan et al. 2011; Harrison et al. accepted), and epigenome (Amarasinghe et al. 2014), to the post-transcriptional level, including regulation by miRNAs (this study).

*Study aims*

In this study we used miRNA-seq and Northern blot validation of miRNAs to identify, first, whether queen- and worker-destined caste pathways in *B. terrestris* are associated with miRNA expression changes, i.e. whether miRNAs have a role in regulating caste determination and differentiation in bumble bees. We predicted expression differences to be more pronounced later on in development because this is the period when larvae have lost their totipotency, and can only develop along a single caste pathway. Second, we investigated whether caste-associated miRNAs are ubiquitously expressed and therefore likely to induce whole organism changes in gene expression or whether they are tissue- and developmental stage specific and thus more likely to be affecting changes in tissues that then become differentiated between the castes. Third, we used genome searches to identify the possible functions of caste-differentiated miRNAs.

## *3.2 Methods*

*Sampling of queen- and worker-destined larvae from bumble bee colonies*

Between May 2011 and April 2012, we obtained 60 *B. t. terrestris* colonies in two cohorts from Syngenta Bioline B.V. (Weert, The Netherlands), and in December 2013 we obtained three *B. t. terrestris* colonies from a third cohort from Koppert B.V. (Berkel en Rodenrijis, The Netherlands). *B. t. terrestris* is a common sub-species of *B. terrestris* found across much of Western Europe. To address our first aim (identifying caste differentiated miRNAs), we used cohort 1 (Summer 2011; 40 colonies) to generate queen-destined and worker-destined larvae for miRNA-seq and cohort 2 (Spring 2012; 20 colonies) to generate queen-destined and worker-destined larvae for sequence validation by Northern blot (**Table 3.1, Appendix 6**). To address our second aim (identifying the stage and tissue specificity of caste differentiated miRNAs), we used cohort 3 (Winter 2013; 3 colonies) to generate RNA from specific tissues and developmental stages in female pupae (**Table 3.2**).

Each colony in all three cohorts contained a single queen that had not yet produced any workers. This allowed us to record the date of first worker eclosion (when workers first emerge from their pupae), which in turn permitted us to estimate the date of the 'switch point'. In the colony cycle of *B. terrestris*, the switch point is defined as the point when the queen starts to produce male eggs; it usually occurs 2-4 weeks after the date of the first worker eclosion (Duchateau and Velthuis 1988; Holland et al. 2013). In the present study it was necessary to estimate the switch point because male larvae are not easily distinguishable from worker-destined larvae, and hence larvae present in post-switch point colonies could be male larvae. Accordingly, our sampling strategy involved sampling larvae only from pre-switch-point colonies. Our sampling strategy further involved the following steps: (a) sampling (by removal) approximately 50% of each batch of early-instar larvae produced per colony, and allowing the remaining 50% to develop into late-instar larvae; (b) of these late-instar larvae, sampling approximately 50% per colony, and allowing the remaining 50% to develop into adults. We then monitored the development of the unsampled late-instar larvae until eclosion as adults. This final step was essential in allowing us to verify the caste fate of sampled larvae from that of contemporarily-produced larvae (details below). We housed the bees in plastic colony boxes (measuring 200 mm × 200 mm × 150 mm) and kept them at constant environmental conditions (28°C and 60%

humidity, under red light). Colonies were supplied with sugar-syrup (Syngenta Bioline B.V.) and freeze dried pollen (Koppert B.V.) *ad libitum*.

For cohort 1, we randomly divided the colonies into two groups of 20, with one group being reared to produce queen-destined larvae and the other group being reared to produce worker-destined larvae (further details below). Once the experiment had started (when the first workers eclosed, i.e. emerged from pupation), we photographed each colony at 17:00 every day until the first males eclosed (28-93 days after receiving the colonies, with one notable exception discussed below). Using the photographs, we were able to determine the locations of the developing brood and any newly-laid eggs, so allowing us to follow the development of individual larvae after its first instar in each of the experimental colonies. Five of the queens (QR3, QR6, QR9, QR11, and QL26) died one or two days after the experiment began, so we excluded their colonies from further study. We also excluded colony QR20 (because it contained workers on arrival) and colony QR19 (because in this colony males eclosed on the same day as the first workers). This left 33 colonies in cohort 1. To induce female larvae to develop along the queen-destined pathway, we removed queens from 20 colonies 14 days after the first workers eclosed, leaving the remaining 13 colonies to develop with the queen still present (**Table 3.1**; **Appendix 6**). In *B. terrestris*, the queen produces an inhibitory pheromone that prevents female larvae developing as queens (see **3.1 Introduction**), so, following previous authors (Cnaani et al. 1997; Pereboom et al. 2005; Lopez-Vaamonde et al. 2007), by removing the queens from each colony we were able to generate queen-destined larvae. After two more days, we then removed up to half of the larvae that were in the 1$^{st}$ or 2$^{nd}$ instar (1-3 days old) from both queenless and queenright colonies. *B. terrestris* eggs hatch five days after they are produced so for the queenless colonies we monitored the colonies and removed half of the newly-hatched first or second instar larvae every two to three days, for six days after the queen was removed; from this point on the only eggs remaining will have been produced by workers. For the queenright colonies, we continued sampling half of the first or second instar larvae every two-three days until approximately 10-14 days after the first workers eclosed, after which further larvae were not sampled as colonies were likely to have passed their switch point (see above). We allowed all remaining larvae, i.e. larvae that had not been sampled, to develop to the fourth instar, which is beyond the point in individual

development when caste fate has been irreversibly determined (Cnaani et al. 2000b). We then sampled approximately half of the fourth instar larvae from both sets of colonies.

We used the photographs to follow the development of the larvae that had not been collected as they pupated and eclosed as adults. In eight of the 13 queenright colonies, all of the unsampled larvae eclosed as workers. These eight colonies were maintained for a further month after the focal larvae had all pupated and eclosed as workers, which allowed us to determine that no queens or males eclosed for at least two weeks after final larval sampling. This meant that all of the sampled larvae in these colonies would have been very likely to develop into workers; we selected the four of these colonies that had the highest sample sizes (early-instar and late-instar larvae) to produce libraries and we classified these colonies as worker-destined colonies (QR-2, QR-7, QR-15, and QR-17 in **Table 3.1**). All of the first and second instar larvae sampled from these colonies were classified as early-instar worker-destined larvae and all of the fourth instar larvae were classified as late-instar worker-destined larvae. We confirmed their instars by weighing each larva individually (Cnaani et al. 1997). The mean masses and sample sizes (mean ± SD mg, n = number of larvae) of each worker-destined larvae stage were as follows: early-instar worker-destined larvae (7.5 ± 2.3 mg, n = 164), late-instar worker-destined larvae (252.4 ± 68.8 mg, n = 85). The mean mass values confirmed that the early-instar larvae were in either the first instar or the early stages of the second instar, as second instar *B. terrestris* worker-destined larvae have previously been shown to have a maximum mass of 39.5 mg (Cnaani et al. 1997; Pereboom et al. 2005). Likewise the worker-destined late-instar larvae were confirmed to be in their fourth instar, which have previously been shown to have a size range of 95.8-375.3 mg (Cnaani et al. 1997; Pereboom et al. 2005).

Eight of the 19 queenless colonies produced queens. We used photos to show that of those eight, three of the colonies produced all queens from the sampled larvae, i.e. queen-destined larvae (QL-10, QL-13, and QL-19 in **Table 3.1**). In a fourth colony, more than 90% of the unsampled larvae developed into queens, while two of the larvae developed into workers (QL-14 in **Table 3.1**). As the proportion of larvae that developed into queens in this colony was still high, we treated the early-instar larvae from this fourth colony as queen-destined. As it is possible to tell late-instar worker-destined and queen-destined larvae apart by size alone (queen-destined larvae having a mass approximately four times greater;

Cnaani et al 1997), all of the late-instar larvae from all four colonies could be identified as queen-destined. The masses and sample sizes (mean ± SD mg, n = number of larvae) of queen-destined larvae were as follows: queen-destined early-instar larvae (9.3 ± 2.8 mg, n = 76), queen-destined late-instar (1168.7 ± 188.5 mg, n = 65). The mean mass values confirmed that the early-instar larvae were in either the first instar or the early stages of the second instar because second instar *B. terrestris* queen-destined larvae have previously been shown to have a maximum mass of 64.0 mg (Cnaani et al. 1997; Pereboom et al. 2005). Likewise the queen-destined late-instar larvae were confirmed to be in their fourth instar because queen-destined fourth-instar larvae have previously been shown to have a size range of 256.2-1291 mg (Cnaani et al. 1997; Pereboom et al. 2005).

In summary, using these methods on the colonies from cohort 1, in the four queenright colonies (QR-2, QR-7, QR-15, and QR-17 in **Table 3.1**), we monitored 397 larvae in total. Of these we sampled totals of 164 early-instar worker-destined larvae and 85 late-instar worker-destined larvae. Of the remaining 148 unsampled larvae, all 148 (100%) eclosed as workers. In the four queenless colonies (QL-10, QL-13, QL-14, and QL-19 in **Table 3.1**), we monitored 192 larvae in total. Of these we sampled totals of 76 early-instar queen-destined larvae and 65 late-instar larvae. Of the remaining 51 unsampled larvae 49 (96.1%) eclosed as queens, i.e. all except the two eclosing as workers described above (**Table 3.1**). This confirmed that larvae sampled from each colony type would have developed along the expected caste trajectory in the vast majority of cases.

Within each colony, larval stage and caste, we pooled sampled larvae and then snap-froze them in liquid nitrogen and crushed them into a mass of fine frozen particles. We dissolved the pooled mixture into an appropriate amount of Trizol reagent (Invitrogen, Carlsbad, California, USA), i.e. 1.5ml of Trizol for every 100 mg of tissue. We snap-froze the larvae-homogenate mixtures in liquid nitrogen and stored them at -80°C. Pooling was necessary because miRNA-seq requires up to 2000 ng of total RNA per phenotype and each Northern blot requires up to 10 µg of total RNA; 2000 ng was more RNA than the early-instar larvae yielded individually. Pooling also meant that we could average the miRNA read counts (the number of sequences returned from the miRNA-seq) between individuals, and therefore reflect differences between phenotypes rather than differences between individual bees (Feldmeyer et al. 2014). To ensure that the libraries were biological replicates that

represented colony variation, individuals from different colonies were kept separate. Pooled samples consisted of between 10-44 larvae per colony (and hence per library) for early-instar larvae and between 5-27 larvae per colony (per library) for late-instar larvae (**Table 3.3**). Overall there were 16 libraries representing four replicates of each phenotype. Each colony produced one early-instar and one late-instar library pool of the same caste fate (**Table 3.3**).

We used cohort 2 to produce larvae to generate RNA samples for Northern Blots (**Table 3.4**). We repeated the sampling procedures used for cohort 1, i.e. starting from colonies containing a queen, an unhatched brood mass, and no workers. We obtained 20 more *B. t. terrestris* colonies from Syngenta Bioline and kept them under the same conditions as the previous cohort. The queen died in two of these colonies (QL-21 and QL-26), so these were removed from analysis. Five of the remaining 18 colonies were kept queenright (so that they would produce worker-destined larvae) and the queen was removed from the final 13 colonies. We sampled larvae from all 18 colonies and determined which larvae were queen-destined and which were worker-destined as described for cohort 1. In the five queenright colonies (QR-21 – QR-25 in **Table 3.1**), we monitored 551 larvae in total. Of these we sampled totals of 223 early-instar worker-destined larvae and 92 late-instar worker-destined larvae. Of the remaining 236 larvae, 235 (99.6%) eclosed as workers (one of the eclosed bees was a male in QR-22). In five of the queenless colonies (QL-23, QL-27, QL-29, QL-32, QLR-33) we monitored 284 larvae in total. Of these we sampled 122 early-instar queen-destined larvae and 56 late-instar queen-destined larvae. Of the remaining 106 unsampled larvae, 103 (97.1%) eclosed as queens, i.e. all except for QR-23, QR-32, and QR-33 which produced one worker each. The 8 (out of 13) remaining queenless colonies from cohort 2 produced large numbers of workers (**Appendix 6**) and were therefore excluded from further analysis. Once again, because the majority of unsampled larvae from each colony developed into the expected phenotype, we treated all sampled larvae from queenright colonies as worker-destined and all sampled larvae from queenless colonies as queen-destined.

For cohort 3, we produced queen-destined larvae and pupae that could then be dissected to isolate the tissue and stage-specific patterns of gene expression in *B. t. terrestris* (**Table 3.2**). We obtained three *B. t. terrestris* colonies, comprising a single queen with no eclosed

workers from Koppert B.V. in November 2013. The queens from all three colonies were removed 14 days after the first worker eclosion. One of these colonies produced ten late-instar queen-destined larvae (**Table 3.2**). We homogenised two of the larvae in Trizol, then froze them in liquid nitrogen and stored them individually (i.e. not in sample pools) at -80°C. We dissected four of the larvae into head, digestive tract, and outer cuticle. These tissues were removed and homogenised in Trizol, frozen in liquid nitrogen, and stored individually at -80°C. The remaining four larvae were allowed to pupate in the colony. We sampled two of them two days after pupation and termed these 'early pupae' and we sampled the other two seven days after pupation and termed these 'late pupae.' We homogenised all four pupae in Trizol, froze them in liquid nitrogen, and stored them at -80°C. Late-instar queen-destined larvae and pupae are four times the size of the equivalent larval and pupal stages in workers and males (Cnaani et al. 1997), and therefore possible to tell apart by size alone. This meant that allowing unsampled larvae to develop in these colonies to verify caste fate was unnecessary.

*RNA extractions*

We extracted the larvae-homogenate mixtures separately for each library. We extracted total-RNA using Trizol according to the manufacturer's (Invitrogen) instructions with minor modifications (**Appendix 1**). Following RNA extraction, we quantified total RNA using a Nanodrop 8000 spectrophotometer (ThermoFisher Scientific, Loughborough, UK). We measured the integrity of the RNA on a 1.2% agarose gel. All of the RNA that we used for miRNA-seq and Northern blots was pure (260/280 ratio > 1.6) and there was no evidence of degradation (data not shown).

*Illumina cDNA library preparation*

To construct cDNA libraries, we used total RNA extracted from the queen- and worker-destined larvae from cohort 1 (**Table 3.1**). Following from the procedures described above, the 16 libraries consisted of four libraries of early-instar worker-destined larvae (EW1-4), four of late-instar worker-destined larvae (LW1-4), four of early-instar queen-destined larvae (EQ1-4) and four of late-instar queen-destined larvae (LQ1-4) (**Table 3.3**). We enriched the total RNA for small RNAs (sRNA) (i.e. enriching the fraction of total RNA that was less than 200 bp in length) using a mirVana miRNA isolation kit (Ambion, Foster City,

California, USA) according to the manufacturer's instructions. We then prepared the libraries using the TruSeq small RNA library preparation kit v.1.5 (Epicentre Technologies, Madison, Wisconsin, USA) with modifications to the 3' adaptor in order to reduce sequencing bias (Sorefan et al. 2012).

To ligate the adaptors to the sRNA sequences, we followed the protocol provided with the TruSeq 1.5 library preparation kit (Illumina) with some modifications (**Appendix 1**). Following preparation of the cDNA, we amplified each library with a unique index sequence using Illumina index primers (1-16). We tested the specificity of the PCRs by varying the cycle number for each library individually, choosing the cycle number that provided the most clear product when separated on an 8% polyacrylamide gel. We then used the selected cycle number to prepare four PCR reactions for each library. We separated the PCR products on an 8% polyacrylamide gel, and then scanned the gel using a Molecular imager FX pro plus and *Quantity One* software. We then identified the 21-23mer miRNA band on the gel, and cut out the gel section that contained it.

We extracted the nucleic acids from the gel fragments, then re-purified them using ethanol precipitation. To reduce the chances of adaptor-adaptor contamination, which is a common problem during library preparation (Yoshikawa et al. 1997), we separated the purified nucleic acid sample on a second 8% polyacrylamide gel and re-purified it by using ethanol precipitation and re-dissolving the pellet in ARG water. We took 1/6th of the purified product and separated it on a third gel to ensure that the miRNA band of interest was still present.

*Sanger sequencing and miRNA-seq*

To ensure that the libraries contained miRNAs, we cloned the construct from library EW-1 into a pGEMTeasy vector (Clontech, Saint-Germain-en-Laye, France) and then transformed into DH5α super-competent *Escherichia coli* using blue-white colony staining. We then extracted the vectors containing the RNA fragment of interest by culturing the *E. coli* overnight, picking the white colonies and making minipreps using the Qiagen mini-prep kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. We prepared ten samples using ready reactions (Life technologies, Paisley, UK) and then sent them to The

Genome Analysis Centre (TGAC, Norwich Research Park, UK) for Sanger chain-termination sequencing.

We used *FinchTV* sequence analysis software to identify each sequence. We were able to map six of the miRNA length sequences to the *B. terrestris* genome (Bter20110317) which is available at Genbank (http://www.ncbi.nlm.nih.gov/genbank/) under extension number AELG00000000.1. One of the six sequences was identified as Bte-miR-1 (We refer to *B. terrestris* miRNAs using the prefix Bte- and *A. mellifera* miRNAs using the prefix Ame- according to standard conventions detailed in *miRBase*; when referring to miRNAs in several species at once we use no prefix), confirming that the libraries contained miRNAs (data not shown).

We then sent the 16 prepared Illumina cDNA libraries to BaseClear B.V, Leiden, The Netherlands, on dry ice for miRNA-seq on a HiSeq2000 platform.

*Bioinformatics analysis*

The miRNA-seq returned 86 million reads across all 16 libraries. We stripped the returned reads of their adaptor sequences *in silico* by matching 8 nt of the 5' end of the 3' adaptor sequence, and then removing it together with a 4 bp overhang of the putative sRNA sequence. This 4 bp overhang accounts for the 4 bp degenerate nucleotide sequence that was added to the 5' end of the 3' adaptor sequence to reduce sequencing bias (Sorefan et al. 2012).

We then mapped all of the sequences to the *B. terrestris* genome (Bter20110317) with *PatMaN*, excluding reads with mismatches or gaps. Overall, we mapped 57.29% of the returned sequences to the genome (**Table 3.2**). We quality-checked the mapped reads using log-sequence count scatter plots, MA plots and a Jaccard index for the 500 most highly expressed sequences in each library (**Figure 3.2**). Log-sequence count scatter plots are the plots of the read counts of one replicate within a phenotype against every other replicate within a phenotype on a log scale. These plots can be visually assessed for deviations of each read from other replicates, and replicates that deviate appreciably can then be excluded from further analysis. MA plots were used to assess the log-offset fold-change ratio ($M$) of each sequence in two samples against the average abundance of each

sequence (*A*). For a sequence $i$ in samples $k$ and $k'$, the $M$ and $A$ values would be calculated as:

$$M_i = \log_2 \frac{k_i+O}{k_i'+O} \qquad \textbf{Equation 3.1}$$

$$A_i = \log_2 \frac{(k_i+O)(k_i'+O)}{2} \qquad \textbf{Equation 3.2}$$

Where $O$ is the offset, which was set at 20 in order to remove inherent biases towards fold-change differences in sequences with low read counts. MA plots naturally show a larger $M$ value spread at lower $A$ values, the distribution of which tightens towards zero from both directions at high $A$ values. A deviation away from zero that is not corrected by normalisation indicates a scaling bias towards one of the two samples, indicating that the data are not comparable between samples. The Jaccard index is a simple matrix that compares the proportion of reads for each sample with every other sample. A value of 1.00 indicates that 100% of the most highly expressed 500 reads in sample A are shared with the 500 most highly expressed reads in sample B. A value of 0.00 indicates that none of the 500 most highly expressed reads are shared between samples A and B.

Using these quality checks, we determined that none of the samples deviated strongly from any of the others in terms of the sequence reads on the log-sequence scatter plots (data not shown). The MA plots did not show strong deviation from the expected pattern of zero between samples except for EQ2 (data not shown), and, on the Jaccard indexes, 15/16 libraries shared 50% or more of the 500 most highly expressed sequences between replicates of the same phenotype (a rough proportion of 0.4 to 0.6 of the reads were shared across all of the libraries) except for LW4 which was an outlier (**Figure 3.2**). Libraries EQ2 and LW4 were removed from further analysis. Read length-abundance distribution plots showed that RNA sequences 21-22 bp in length were the most frequently expressed within each phenotype (**Figure 3.3**). This implies that miRNAs were one of the most abundant class of sequence in each of the libraries.

We normalised the remaining libraries using quantile normalisation (Bolstad et al. 2003). Prior to normalisation we boot-strapped the libraries by stripping the number of reads down to equal the library with the lowest read count (LW3, **Table 3.3**), reducing the size of the other libraries by random sampling of sequences. This method reduces the effect of smaller libraries having a disproportionate influence on differential expression between libraries (Beckers et al. In prep). We then predicted the new *Bombus* miRNAs using MapMi (Guerra-Assuncao and Enright 2010) to align all mature sequences from all species on the *miRBase* repository of miRNAs (Kozomara and Griffiths-Jones 2011) to the miRNA-seq reads. To address our first aim, we identified the reads that were differentially expressed between libraries. We calculated the $M$ between phenotypes for each miRNA to isolate caste-specific miRNAs (i.e. comparing EW vs EQ and LW vs LQ) and, within castes, developmentally differentiated miRNAs (i.e. comparing EQ vs LQ and EW vs LW). Following previous studies (Lopez-Gomollon et al. 2012; Ferreira et al. 2013; Feldmeyer et al. 2014), we defined sequences as differentially expressed between phenotypes if the log-offset fold difference between the average numbers of reads was greater than two (i.e. $M$ was > 1 or < -1, the more positive or more negative the $M$ value, the greater the degree of differential expression between phenotypes). To account for variation between samples we only classified them as differentially expressed if $M$ was > 1 or < -1 for more than half of the possible pairwise comparisons of each replicate between each phenotype (Beckers et al. In prep). We excluded samples where the highest read count in any one library was less than 100 reads because low read counts can inflate the calculated fold-change between samples and are less likely to be biologically meaningful.

To address our third aim, we identified the region where caste-differentiated miRNAs were expressed on the *B. terrestris* genome. We used *Basic Local Alignment Search Tool* (BLAST) to identify the region, and the program *GBrowse* to identify the nucleotide sequence (and therefore the neighbouring genes) around each miRNA.

*Northern Blotting*

To verify the presence of miRNAs in the libraries prepared from cohort 1, and to validate the expression patterns identified in them by the bioinformatics analysis, we used Northern blotting to probe for miRNAs in RNA samples extracted from ten colonies from cohort 2

(**Table 3.4)**. We designed probes that were reverse-complementary to ten miRNAs (**Table 3.5**). To investigate our first aim, we identified miRNAs that were caste differentiated and used probes to validate the expression of four differentially expressed miRNAs (Bte-miR-6001-5p, Bte-miR-6001-3p, Bte-miR-11, and Bte-miR-12) with large fold-changes between phenotypes. Bte-miR-6001-5p and Bte-miR-6001-3p are part of the same duplex, but as Bte-miR-6001-3p showed differential expression and had a read count > 100, we shall treat the two arms as separate miRNAs for the rest of this study. In addition to the four differentially expressed miRNAs we validated the expression of three miRNAs (Bte-miR-9a, Bte-miR-71, Bte-miR-184) that showed mild differences between phenotypes, and their homologues that have previously been associated with caste differentiation in *A. mellifera* (Weaver et al. 2007; Guo et al. 2013); two miRNAs (Bte-miR-275, Bte-miR-283) that showed mild differences between phenotypes, and their homologues that have previously been associated with polyethism in *A. mellifera* workers (Behura and Whitfield 2010; Greenberg et al. 2012; Liu et al. 2012); and one miRNA (Bte-miR-315) that showed mild differences between phenotypes, but does not have a homologue that has previously been associated with eusociality in eusocial bees. For each miRNA, we produced two Northern blots (one from each of two independent colony replicates) to compare the expression of the four phenotypes (early-instar and late-instar, queen-destined and worker-destined larvae) alongside each other. The expression pattern was considered to be validated if both Northern blots showed a pattern of differential gene expression in the same direction as that shown by the miRNA-seq results. To investigate our second aim, we used RNA extracted from cohort 3 to identify the tissue and pupae stage specificity of the two caste-differentiated miRNAs that were validated by the Northern blots (Bte-miR-6001-5p and Bte-miR-6001-3p).

In brief, to conduct the Northern blots (**Appendix 1**), we extracted total RNA from cohort 2 using the Trizol protocol described above. We then separated the RNA on a 15% denaturing polyacrylamide gel and transferred the RNA to a Hybond-NX nylon membrane (GE Healthcare, Amersham, UK) using Semi-dry membrane transfer apparatus (Scie-plas, Cambridge, UK). We bound the RNA to the membrane by chemical crosslinking.

We hybridized membranes overnight at 37°C in UltraHyb-Oligo hybridization buffer (Ambion) with probes that were reverse complementary to the miRNA of interest and were

labelled with 32P using T4 polynucleotide kinase and $\gamma^{-32}$ P ATP. We rinsed the membranes with a wash solution (0.2 x sodium chloride/sodium citrate, 0.1% (w/v) sodium dodecyl sulphate (SDS)), then exposed them on a blank phospho-imaging screen inside a radioactive cassette (Fujifilm, Billingham, UK). We then scanned the screens on a Molecular Imager FX Pro Plus using the software *Quantity One* to visualize the signal. To allow membrane re-use, we stripped the membrane of its radioactive signal by incubating it in a stripping solution (pH 8.5, 0.1% SDS, 5mM EDTA) at 95°C for 20-40 minutes. All membranes were re-probed with U6 which is a small nuclear RNA acting as a loading control (Lopez-Gomollon 2011).

## *3.3 Results*

### *Illumina library analysis*

Overall the miRNA-seq revealed 101 miRNAs in female larvae of *B. terrestris* (**Appendix 2**), of which 16 were new miRNAs (i.e. not sequenced in any other organism) and 85 were shared with *A. mellifera* (**Appendix 2**) and so appear conserved across the corbiculate bees. Of the 16 new miRNAs, 14 appear to be unique to *Bombus* (including *B. terrestris* and *B. impatiens*), and two miRNAs were conserved in the genome of *A. mellifera* (**Chapter 2**). In the current study, we focus on the miRNAs that were differentially expressed in the miRNA-seq and those that were subsequently validated by Northern blot analysis.

### *Differentially expressed miRNAs between B. terrestris castes*

Our first aim was to identify the miRNAs differentially expressed between queen- and worker-destined *B. terrestris* larvae. MiRNA-seq showed that four miRNAs were differentially expressed between late-instar queen-destined larvae and late-instar worker-destined larvae, these were Bte-miR-6001-5p, Bte-miR-6001-3p, Bte-miR-11, and Bte-miR-12 (**Figure 3.4a-d**, **Appendix 7**). Both the 5' and 3' arm of Bte-miR-6001 were more highly expressed in late-instar queen-destined larvae compared to worker-destined larvae of either instar. The pattern of expression was similar for both miRNA arms but, overall, Bte-miR-6001-5p was much more highly expressed than Bte-mir-6001-3p. Bte-miR-11 and Bte-miR-12 had more variability between colony replicates than Bte-miR-6001-5p and Bte-miR-

6001-3p (**Figure 3.4a-d**); however, Bte-miR-11 was more highly expressed in late-instar worker-destined larvae compared to the other phenotypes, while Bte-miR-12 was expressed at the lowest levels in late-instar queen-destined larvae compared to the other phenotypes. In addition all four miRNAs were differentially expressed between early and late larval instars in at least one of the castes.

Several miRNAs did not fulfil our criteria for differential expression (i.e. the miRNAs had a less than two-fold difference in log-offset fold changes and therefore $-1 < M < 1$ across all replicates), however they still showed consistent changes in miRNA expression between replicates. Bte-miR-9a was more highly expressed in early-instar worker-destined larvae compared to the other phenotypes (**Figure 3.4e**). Bte-miR-71 was expressed at a much lower level in late-instar worker-destined larvae (**Figure 3.4f**). Bte-miR-184 was more highly expressed in early-instar worker-destined larvae (**Figure 3.4g**). Bte-miR-275 was more highly expressed in late-instar queen-destined larvae (**Figure 3.4h**). Finally, Bte-miR-283 and Bte-miR-315 were both more highly expressed in early-instar worker-destined larvae compared to the other phenotypes (**Figure 3.4i-j**).

*Northern Blot library verification of caste differentiated miRNAs in B. terrestris larvae*

To further investigate the miRNAs that were differentially expressed between castes, we used Northern blots to validate the results of the miRNA-seq. Overall the pattern of regulation of Bte-mir-6001-5p observed in the miRNA-seq was confirmed by the Northern blots. Bte-miR-6001-5p was more highly expressed in late-instar larvae compared to early-instar larvae, and was also more highly expressed in the late-instar queen-destined larvae compared to the late-instar worker-destined larvae (**Figure 3.5a**). Likewise, Bte-miR-6001-3p showed a similar pattern of expression to that revealed by the miRNA-seq. That is, although it was present at a lower concentration than Bte-miR-6001-5p, it was more highly expressed in late-instar queen-destined larvae compared to worker-destined larvae, in which, unlike Bte-miR-6001-5p, it could not be detected at all (**Figure 3.5b**).

By contrast, and despite showing changes in expression between phenotypes in the miRNA-seq analysis, Bte-miR-11, Bte-miR-12, Bte-miR-315, Bte-miR-184, Bte-miR-9a, Bte-miR-71, Bte-miR-275, and Bte-miR-283, did not show differential expression patterns between castes in the Northern blot analysis (**Figure 3.5c-j**). Hence, of the four miRNAs showing

between-caste differential expression in larvae according to miRNA-seq, only two (Bte-mir-6001-5p and Bte-mir-6001-3p) were validated by Northern blot analysis.

*Stage- and tissue-specificity of miRNAs in queen-destined larvae*

Our second aim was to identify the stage-specific and tissue-specific expression of caste differentiated miRNAs. The probes for Bte-miR-6001-5p showed a much stronger signal than Bte-miR-6001-3p in the specific tissues (head, digestive tract and outer cuticle) of queen-destined larvae and in whole bodies of queen-destined pupae (**Figure 3.6**), suggesting that, as also suggested by the sequencing data, Bte-miR-6001-5p is more highly expressed than Bte-miR-6001-3p. Both miRNAs were much more strongly expressed in outer cuticle than in head, and both were expressed at the lowest levels in digestive tract (**Figure 3.6**). This implies a degree of tissue specificity, although both miRNAs were present in all of the larval tissues.

The two miRNAs (Bte-miR-6001-5p and Bte-miR-6001-3p) were also differentiated by developmental stage. MiRNA-seq and Northern blots on RNA from whole larvae showed that they were much more highly expressed in late-instar larvae than in early-instar larvae in both castes (**Figures 3.4a-b, 3.5a-b**). Consistent with this, tissue-specific Northern blots showed that that the two miRNAs were most highly expressed in the late-instar larvae and that the level of expression declined following pupation, with lower levels of both miRNAs being detected in the late pupae compared to early pupae and late-instar larvae (**Figure 3.6**).

*Caste differentiated miRNAs: possible targets*

The third aim of this study was to investigate the function of caste-differentiated miRNAs. A genome scan of the gene *bte-mir-6001* showed that it comprises the entire fourth intron of *Very High Density Lipoprotein* (VHDL, a gene with homology to *Vitellogenin*), signifying that it is highly likely to be a mirtron (Okamura et al. 2007). A *BLAST* search of this sequence against the *A. mellifera* genome revealed that the mirtron was conserved in the same intron of the same gene between both eusocial insects.

## *3.4 Discussion*

In this study we sequenced the miRNAs of worker-destined and queen-destined early-instar and late-instar larvae in the bumble bee *B. terrestris*. We then validated the miRNA-seq results using Northern blot analysis. The miRNA-seq revealed four miRNAs (Bte-miR-6001-5p, Bte-miR-6001-3p, Bte-miR-11, and Bte-miR-12) that were differentially expressed, and a further six miRNAs that had consistent changes in miRNA expression between queen-destined and worker-destined larvae (**Figure 3.4**); of these only Bte-miR-6001-5p and Bte-miR-6001-3p could be validated by Northern blot (**Figure 3.5a-b**). It is not unusual for Northern blots and miRNA-seq to show different results (Greenberg et al. 2012). Taking a conservative approach, we consider further only miRNAs that showed consistent patterns between miRNA-seq and Northern blots (Bte-miR-6001-5p and Bte-miR-6001-3p) and three miRNAs (Bte-miR-9a, Bte-miR-71, Bte-miR-184) that were not differentially expressed when validated by Northern blots, but have homologues that have previously been associated with caste determination in *A. mellifera* (Weaver et al. 2007; Guo et al. 2013).

*MiRNAs are differentially expressed between queen-destined and worker-destined larvae*

The first aim of this study was to identify whether queen-worker caste determination and differentiation in *B. terrestris* are associated with miRNA expression changes. Overall, we identified and validated two such miRNAs, which were more highly expressed in late-instar queen-destined larvae relative to late-instar worker-destined larvae, Bte-miR-6001-5p and Bte-miR-6001-3p (**Figure 3.4a-b** and **3.5a-b**). This finding suggests that these miRNAs are associated with caste determination, either because they influence caste determination or because they respond to factors associated with caste. One possibility is that one or both arms of the Bte-miR-6001 miRNA duplex targets a caste determining protein-coding gene that prevents larvae from developing along the queen-destined pathway, so that, by silencing the gene, the miRNA allows larvae to develop along the queen trajectory. As described in the introduction (**Section 3.1**), in *B. terrestris* a queen-produced pheromone stops female larvae from developing into queens (Roseler 1970; Cnaani et al. 2000b). The absence of this pheromone causes a JH peak that is likely to provoke a cascade of expression in other genes, so setting a larva along the queen-destined trajectory (Cnaani et al. 2000b), while genes that might be more closely associated with worker-determination show reduced gene expression (Pereboom et al. 2005). Given that natural variation in

miRNA expression has already been shown to induce changes in morphology in insects (Arif et al. 2013), we suggest that control of gene expression leading to changes in caste fate could be mediated by a peak in Bte-miR-6001, which might either respond directly to the JH peak or the absence of the inhibitory queen pheromone. A non-mutually exclusive possibility is that Bte-miR-6001-5p and Bte-miR-6001-3p are more strongly expressed in late-instar queen-destined larvae in response to the greater amount of food that such larvae receive (Cnaani et al. 1997).

As expected for miRNAs that are derived from the same precursor, Bte-miR-6001-5p and Bte-miR-6001-3p showed similar patterns of gene expression between the miRNA-seq and Northern blot results. However, the patterns were slightly different, as Bte-miR-6001-5p was detected in both castes in miRNA-seq and Northern blots whereas Bte-miR-6001-3p could only be detected in queen-destined larvae in Northern blots. This is probably a result of Bte-miR-6001-5p being much more highly expressed overall, relative to Bte-miR-6001-3p. One possible explanation for such higher expression could be that Bte-miR-6001-5p is the functional mature miRNA while Bte-miR-6001-3p may be the non-functional passenger strand of the Bte-miR-6001 duplex. The usual case for miRNAs is that one strand is functional and is incorporated into the *RISC* complex while the other is degraded (Bartel 2004). However, there is a possibility that Bte-miR-6001-3p is functional too, given that it was still highly expressed in the miRNA-seq and was detectable by Northern blot. In at least some miRNAs, the identity of the functional sequence (5p or 3p) differs in different organisms (Marco et al. 2010; Griffiths-Jones et al. 2011), and even occasionally within different tissues (Ro et al. 2007) and stages (Glazov et al. 2008) of the same organism. In such cases, the different, albeit complimentary, sequences of the two miRNAs mean that the genes they target are likely to be very different (Marco et al. 2010). For example, one recent study showed that the majority of miRNAs from the same precursor were likely to target different 3' UTR regions in genes that were functionally unrelated (Marco et al. 2012). Therefore it is possible that Bte-miR-6001-3p and Bte-miR-6001-5p have separate roles in caste differentiation.

We predicted that miRNAs would show a stronger pattern of differential expression between queen- and worker-destined larvae later in development. Interestingly, Bte-miR-6001 was only differentially expressed between the castes in late-instar queen-destined

larvae, i.e. there were no caste differences in gene expression in early-instar larvae. One explanation for this is that, as described in the introduction (**Section 3.1**), queen- and worker-destined larvae do not lose their totipotency until after the critical point in larval development which occurs towards the end of the second instar (Cnaani et al. 2000b). Therefore, if Bte-miR-6001 influences caste determination, it is likely that it would not show strong patterns of differential regulation until after the end of the second instar.

We used Northern blots to test the expression patterns of miRNAs that were not found to be differentially expressed between castes in *B. terrestris* larvae, but that have nevertheless been shown to be involved in caste differentiation in *A. mellifera*. One of the predictions of the 'genetic toolkit' hypothesis is that a common set of regulatory genes are co-opted in independently evolving eusocial lineages. If these toolkit genes include miRNAs then we would expect miRNAs that are differentially expressed in other social lineages to also be differentially expressed in bumble bees. Weaver et al. (2007) showed that Ame-miR-9a and Ame-miR-71 were both highly differentiated between *A. mellifera* queen- and worker-developmental trajectories. Meanwhile (Guo et al. 2013) showed that Ame-miR-184 could induce worker-like differentiation in queen-destined larvae. The homologues of these genes in *B. terrestris* were Bte-miR-9a, Bte-miR-71 and Bte-miR-184, respectively. The Northern blots showed that Bte-miR-184 and Bte-miR-9a were not differentially expressed between early and late-instars and they were not caste differentiated (**Figure 3.6f-g**). Bte-miR-71 was more highly expressed in early-instar larvae of both castes but showed no evidence for caste differentiation (**Figure 3.6h**). This shows that there is little evidence that caste-differentiated miRNAs in *A. mellifera* show similar patterns in *B. terrestris*. This suggests that, in line with the different systems of caste determination found in these two taxa (see **3.1 Introduction**), the taxa have diverged with respect to the role of miRNAs in caste determination and differentiation since the time they shared a common eusocial ancestor 77-95 million years ago (Cardinal and Danforth 2011).

*Bte-miR-6001 has stage- and tissue-specificity in queen-destined larvae*

Our second aim was to test for stage- tissue-specificity in caste-differentiated miRNAs, which may be informative for identifying their functions. With regard to stage-specificity, our results showed that, within each larval caste, both arms of Bte-miR-6001 are more

highly expressed in late-instar larvae relative to early-instar larvae (**Figure 3.5a-b**). The pattern of expression remained high in extracts from individual queen-destined larvae (**Figure3.6a-b**). In addition, both miRNAs were expressed at lower levels in the pupae than the larvae, and declined throughout pupation (**Figure 3.6a-b**), although this result was more clear for Bte-miR-6001-5p than Bte-miR-6001-3p (which was at the lower threshold for detection across all Northern blots). These results suggest that Bte-mir-6001-5p and Bte-miR-6001-3p could have an important role in caste differentiation during development, especially following the loss of totipotency in queen-destined larvae, but that it is less important in the caste differences between adult workers and queens. This possibility is investigated further in **Chapter 4**.

The results also suggest that Bte-miR-6001-5p and Bte-miR-6001-3p, in addition to having a role in caste determination and differentiation, are potentially important developmental genes. Development is tightly linked with caste determination and differentiation in eusocial insects such as *Bombus* and *Apis*, in which caste fate is decided in the larval stages. Transcriptomic studies in eusocial insects have shown that differences in social traits often involve large numbers of differentially expressed developmental genes, including some house-keeping genes (Evans and Wheeler 1999; Evans and Wheeler 2001b; Evans and Wheeler 2001a; Thompson et al. 2006; Grozinger et al. 2007; Thompson et al. 2007). Developmental genes are likely to interact with the environment in complex and synergistic ways; they will be expressed at different times of development, in different tissues, and at different levels (Pereboom et al. 2005; Sumner 2006; Ament et al. 2010). Expression differences in such genes would cause mostly subtle, but at times dramatic, changes in phenotype which are then continued into adulthood. If so, we would expect developmental genes to be important in caste determination and differentiation processes. Both arms of Bte-mir-6001 were expressed in both castes and both developmental stages that we investigated, albeit at different levels, and therefore it could be important in this more subtle system of regulation. These genes would not be expected to work by completely switching off or on any genes, because they are expressed at different levels in both castes; instead they could regulate developmental target genes to a specified level in both castes.

With regard to tissue-specificity, our results showed that Bte-miR-6001-5p and Bte-miR-6001-3p were tissue-specific to a certain extent. The miRNAs were expressed in all three of

the dissected larval tissues (head, digestive tract and outer cuticle), but they were most highly expressed in outer cuticle and only expressed at a comparatively low level in the digestive tract (**Figure 3.6)**. These findings could be informative because they suggest that Bte-miR-6001 may not target brain- and gut-specific genes, so reducing the list of possible targets for Bte-miR-6001. Of the possible target tissues, cuticle could be expected to be heavily involved in growth and in moulting between instars. Unlike the case in *A. mellifera*, the growth rate of queen-destined *B. terrestris* larvae is known not to differ from that of worker-destined larvae; however, the length of each instar is greater for queen-destined larvae, leading to their larger final body size (Cnaani et al. 1997). To speculate, it is possible that Bte-miR-6001 is highly expressed in the cuticle of queen-destined larvae because it targets genes that regulate moulting behaviour. Consistent with this, a transcriptional study in *B. terrestris* showed that genes involved in larval cuticular biogenesis are strongly elevated compared to other genes, and likely important for larval development (Colgan et al. 2011), however in that study there was no comparison of gene expression between worker-destined and queen-destined larvae.

*The function of caste differentiated miRNAs*

The third aim of this study was to identify the function of caste-differentiated miRNAs. To do this we did genome search of the locus for the *bte-mir-6001* gene, this approach showed that the precursor sequence of *bte-mir-6001* made up the entire fourth intron of the protein-coding gene *very high density lipoprotein* (vhdl). The fact that the entire intron is made up from a single miRNA precursor implies that the miRNA is a unique type of miRNA which is termed a mirtron (Okamura et al. 2007). Most miRNAs are expressed as long single-stranded pri-miRNAs which are then processed by the enzyme *Drosha* (Bartel 2004). In contrast, mirtrons are cleaved from the intron of a protein-coding gene by *lariat-debranching enzyme* following transcription of the mRNA (Okamura et al. 2007; Ruby et al. 2007). Once mirtrons have been spliced from the mRNA transcript of their host protein-coding genes, they are processed in the same manner as other miRNAs, i.e. they are recognised by *exportin 5* and transported into the cytoplasm where they are cleaved into a short 21-23 bp duplex by *Dicer* (Okamura et al. 2007). The trans-regulatory functions of mirtrons do not differ from those of regular miRNAs; both types of miRNA are incorporated into *RISC*, and they target mRNA transcripts for silencing in the same way (Westholm and

Lai 2011). It is less clear whether their different modes of biogenesis have any cis-regulatory consequences. It is not clear, for example, whether there is a relationship between mirtrons and their host protein-coding genes. One intriguing possibility is that the expression patterns of one might be associated with those of the other; for example, the miRNA might be regulated by the upstream regulatory sequences for the host protein-coding gene itself. Therefore the miRNA and the host gene will be co-expressed and affect the same pathways or phenotypes, in this case queen development.

Such a process has special relevance in the case of *bte-mir-6001* because its host gene *vhdl* is known in several eusocial insects and has sequence homology to vitellogenin (*vg*). *Vg*s are an interesting class of nutritive proteins that are induced by the action of JH and have a direct link to reproduction in numerous insects (Hagedorn and Kunkel 1979; Sappington and Raikhel 1998). In *A. mellifera* they are highly expressed in queens, and they seem to be more highly expressed in workers with active ovaries compared to workers with inactive ovaries (Koywiwattrakul et al. 2005; Koywiwattrakul and Sittipraneed 2009). Consistent with this, unpublished data have shown that *vg* genes are most highly expressed in ovary active *B. terrestris* workers compared to workers with inactive ovaries (A. F. G. Bourke, personal communication). They have also been suggested to have a role in caste determining processes in *A. mellifera* (Engels et al. 1990; Barchuk et al. 2002) and harvester ants (Libbrecht et al. 2013). In *B. terrestris*, vitellogenin has been less well studied, however it is expressed in the pupae of both castes and reaches a peak shortly after the adults emerge; the peaks are reached more quickly and at a higher level in queens compared to workers (Li et al. 2010). It is not clear whether *vhdl* itself has a direct link to caste determination in either *Apis* or *Bombus*, though it is thought to be an important storage protein in Hymenoptera (Wheeler and Buck 1995) and is highly expressed during larval and pupal development in *B. terrestris* (Harrison et al. 2015). Storage proteins play key roles in most holometabolous insects, since they accumulate in late-instar larvae and are used in the rapid synthesis of amino acids prior to metamorphosis. They are then depleted throughout the pupal stages and are absent in adulthood (Levenbook 1985; Hunt et al. 2003). Like other storage proteins, *vhdl* accumulates in late-instar larvae, and being a lipoprotein it is also likely to have a role in fatty acid metabolism and transport in larvae undergoing metamorphosis (Shipman et al. 1987). The gene may be of particular

importance in at least some social species; for example, in the ant *Crematogaster opuntiae* it continues to be accumulated and stored in newly-mated adult queens and is then used in the production of the first workers (Wheeler and Buck 1995). This means that *vhdl* has potential as a caste-differentiated gene in other eusocial insects and suggests that further investigations should focus on the potential link between the Bte-miR-6001 miRNA duplex expression and *vhdl*. A *BLAST* search revealed that the mirtron is conserved in the same intron of the corresponding gene in *A. mellifera*, so there is a broader potential for the two genes to be involved in social evolution in bees.

Another method to explore the function of miRNAs is to predict the genes that they target. Software such as *targetscan* can be used to scan new miRNAs against the 3'UTR regions of potential target genes in species for which the data are available. The 3'UTR region contains the sequence that is usually targeted by miRNAs (Bartel 2004) though there are numerous exceptions to this (Lytle et al. 2007; Fang and Rajewsky 2011). With the recent publication of the *B. terrestris* genome (Sadd et al. 2015), further annotation will make it possible to identify and scan the 3'UTR regions of *B. terrestris* (and *B. impatiens*).

Further studies will be needed to establish which genes are targeted by caste differentiated miRNAs in *B. terrestris*, first by using programs such as *targetscan*, then by using experimental methods such as luciferase assays (Akgül and Göktaş 2014) and RNA interference (**Chapter 5**). These methods will be essential to establish the possible functions of miRNAs such as miR-6001 that have been shown to be associated with caste determination in *B. terrestris* in this study.

## *3.5 References*

Akgül B, Göktaş C. 2014. Gene reporter assay to validate microRNA targets in *Drosophila* S2 cells. In *miRNomics: MicroRNA Biology and Computational Analysis Methods in Molecular Biology*, Vol 1107 (ed. Y M, A J), pp. 233-242. Humana Press, New York, USA.

Amarasinghe HE, Clayton CI, Mallon EB. 2014. Methylation and worker reproduction in the bumble-bee (*Bombus terrestris*). *Proc R Soc B* **281**: 20132502.

Ament SA, Wang Y, Robinson GE. 2010. Nutritional regulation of division of labor in honey bees: toward a systems biology perspective. *WIREs Syst Biol Med* **2**: 566-576.

Arif S, Murat S, Almudi I, Nunes MDS, Bortolamiol-Becet D, McGregor NS, Currie JMS, Hughes H, Ronshaugen M, Sucena E et al. 2013. Evolution of mir-92a Underlies Natural Morphological Variation in *Drosophila melanogaster*. *Curr Biol* **23**: 523-528.

Axtell MJ, Westholm JO, Lai EC. 2011. Vive la difference: biogenesis and evolution of microRNAs in plants and animals. *Genome Biol* **12**: 221.

Barchuk AR, Bitondi MM, Simoes ZL. 2002. Effects of juvenile hormone and ecdysone on the timing of vitellogenin appearance in hemolymph of queen and worker pupae of *Apis mellifera*. *J Insect Sci* **2**:1.

Barchuk AR, Cristino AS, Kucharski R, Costa LF, Simoes ZLP, Maleszka R. 2007. Molecular determinants of caste differentiation in the highly eusocial honeybee *Apis mellifera*. *BMC Dev Biol* **7**: 70.

Bartel DP. 2004. MicroRNAs: Genomics, biogenesis, mechanism, and function. *Cell* **116**: 281-297.

Behura SK, Whitfield CW. 2010. Correlated expression patterns of microRNA genes with age-dependent behavioural changes in honeybee. *Insect Mol Biol* **19**: 431-439.

Biemar F, Zinzen R, Ronshaugen M, Sementchenko V, Manak JR, Levine MS. 2005. Spatial regulation of microRNA gene expression in the *Drosophila* embryo. *Proc Natl Acad Sci USA* **102**: 15907-15911.

Bolstad BM, Irizarry RA, Astrand M, Speed TP. 2003. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* **19**: 185-193.

Bonasio R, Zhang G, Ye C, Mutti NS, Fang X, Qin N, Donahue G, Yang P, Li Q, Li C et al. 2010. Genomic comparison of the ants *Camponotus floridanus* and *Harpegnathos saltator*. *Science* **329**: 1068-1071.

Bortolotti L, Duchateau MJ, Sbrenna G. 2001. Effect of *juvenile hormone* on caste determination and colony processes in the bumblebee *Bombus terrestris*. *Entomol Exp Appl* **101**: 143-158.

Bourke AFG. 2011a. *Principles of Social Evolution*. Oxford University Press, Oxford, UK.

-. 2011b. The validity and value of inclusive fitness theory. *Proc R Soc B* **278**: 3313-3320.

Brenner S. 1999. Theoretical biology in the third millennium. *Phil Trans R Soc B* **354**: 1963-1965.

Cardinal S, Danforth BN. 2011. The antiquity and evolutionary history of social behavior in bees. *PLOS One* **6**: e21086.

Chang TC, Mendell JT. 2007. MicroRNAs in vertebrate physiology and human disease. *Annu Rev Genomics Hum Genet* **8**: 215-239.

Chen X, Hu Y, Zheng HQ, Cao LF, Niu DF, Yu DL, Sun YQ, Hu SN, Hu FL. 2012. Transcriptome comparison between honey bee queen- and worker-destined larvae. *Insect Biochem Mol Biol* **42**: 665-673.

Chen X, Yu X, Cai Y, Zheng H, Yu D, Liu G, Zhou Q, Hu S, Hu F. 2010. Next-generation small RNA sequencing for microRNAs profiling in the honey bee *Apis mellifera*. *Insect Mol Biol* **19**: 799-805.

Cnaani J, Borst DW, Huang ZY, Robinson GE, Hefetz A. 1997. Caste determination in *Bombus terrestris*: Differences in development and rates of JH biosynthesis between queen and worker larvae. *J Insect Physiol* **43**: 373-381.

Cnaani J, Robinson GE, Bloch G, Borst D, Hefetz A. 2000a. The effect of queen-worker conflict on caste determination in the bumblebee *Bombus terrestris*. *Behav Ecol Sociobiol* **47**: 346-352.

Cnaani J, Robinson GE, Hefetz A. 2000b. The critical period for caste determination in *Bombus terrestris* and its juvenile hormone correlates. *J Comp Physiol A* **186**: 1089-1094.

Colgan TJ, Carolan JC, Bridgett SJ, Sumner S, Blaxter ML, Brown MJF. 2011. Polyphenism in social insects: insights from a transcriptome-wide analysis of gene expression in the life stages of the key pollinator, *Bombus terrestris*. *BMC Genomics* **12**: 623.

Cumber RA. 1949. The biology of humble-bees, with special reference to the production of the worker caste. *Trans R Entomol Soc Lond* **100**: 1-45.

Daish TJ, Cakouros D, Kumar S. 2003. Distinct promoter regions regulate spatial and temporal expression of the *Drosophila caspase dronc*. *Cell Death Differ* **10**: 1348-1356.

Darwin C. 1859. *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. John Murray, London, UK.

Dawkins R. 1976. *The Selfish Gene*. Oxford University Press, Oxford, UK.

Duchateau MJ, Velthuis HHW. 1988. Development and reproductive strategies in *Bombus* colonies. *Behaviour* **107**: 186-207.

Elango N, Hunt BG, Goodisman MAD, Yi SV. 2009. DNA methylation is widespread and associated with differential gene expression in castes of the honeybee, *Apis mellifera*. *Proc Natl Acad Sci USA* **106**: 11206-11211.

Engels W, Kaatz H, Zillikens A, Simoes ZLP, Trube A, Braun R, Dittrich F. 1990. Honey bee reproduction - *Vitellogenin* and caste-specific regulation of fertility. In *Advances in Invertebrate Reproduction 5*, (ed. M Hoshi, O Yamashita), pp. 495-502. Elsevier, Amsterdam, The Netherlands.

Evans JD, Wheeler DE. 1999. Differential gene expression between developing queens and workers in the honey bee, *Apis mellifera*. *Proc Natl Acad Sci USA* **96**: 5575-5580.

-. 2001a. Expression profiles during honeybee caste determination. *Genome Biol* **2**: 1.

-. 2001b. Gene expression and the evolution of insect polyphenisms. *Bioessays* **23**: 62-68.

Fang Z, Rajewsky N. 2011. The impact of miRNA target sites in coding sequences and in 3'UTRs. PLOS One 6: e18067

Feldmeyer B, Elsner D, Foitzik S. 2014. Gene expression patterns associated with caste and reproductive status in ants: worker-specific genes are more derived than queen-specific ones. *Mol Ecol* **23**: 151-161.

Ferreira PG, Patalano S, Chauhan R, Ffrench-Constant R, Gabaldon T, Guigo R, Sumner S. 2013. Transcriptome analyses of primitively eusocial wasps reveal novel insights into the evolution of sociality and the origin of alternative phenotypes. *Genom Biol* **14**: R20.

Free JB. 1955. Queen production in colonies of bumble bees. *Proc R Entomol Soc A* **30**: 19-25.

Glazov EA, Cottee PA, Barris WC, Moore RJ, Dalrymple BP, Tizard ML. 2008. A microRNA catalog of the developing chicken embryo identified by a deep sequencing approach. *Genome Res* **18**: 957-964.

Goulson D. 2003. *Bumblebees: Behaviour, ecology and conservation*. Oxford University Press, Oxford, UK.

Greenberg JK, Xia J, Zhou X, Thatcher SR, Gu X, Ament SA, Newman TC, Green PJ, Zhang W, Robinson GE et al. 2012. Behavioral plasticity in honey bees is associated with differences in brain microRNA transcriptome. *Genes Brain Behav* **11**: 660-670.

Griffiths-Jones S, Hui JHL, Marco A, Ronshaugen M. 2011. MicroRNA evolution by arm switching. *Embo Reports* **12**: 172-177.

Grozinger CM, Fan YL, Hoover SER, Winston ML. 2007. Genome-wide analysis reveals differences in brain gene expression patterns associated with caste and reproductive status in honey bees (*Apis mellifera*). *Mol Ecol* **16**: 4837-4848.

Guerra-Assuncao JA, Enright AJ. 2010. *MapMi*: automated mapping of microRNA loci. *BMC Bioinformatics* **11**: 133.

Guo XQ, Su SK, Skogerboe G, Dai SAJ, Li WF, Li ZG, Liu F, Ni RF, Guo Y, Chen SL et al. 2013. Recipe for a busy bee: microRNAs in honey bee caste determination. *PLOS One* **8**: e81661.

Gutierrez L, Bussell JD, Pacurar DI, Schwambach J, Pacurar M, Bellini C. 2009. Phenotypic plasticity of adventitious rooting in arabidopsis is controlled by complex regulation of *auxin* response factor transcripts and microRNA abundance. *Plant Cell* **21**: 3119-3132.

Hagedorn HH, Kunkel JG. 1979. *Vitellogenin* and *vitellin* in insects. *Annu Rev Entomol* **24**: 475-505.

Hamilton WD. 1964a. The genetical evolution of social behaviour. I. *J Theor Biol* **7**: 1-16.

-. 1964b. The genetical evolution of social behaviour. II. *J Theor Biol* **7**: 17-52.

Harrison MC, Hammond RL, Mallon EB. (accepted). Reproductive workers show queen-like gene expression in an intermediately eusocial insect, the buff-tailed bumble bee Bombus terrestris. *Mol Ecol*.

Hartfelder K, Cnaani J, Hefetz A. 2000. Caste-specific differences in *ecdysteroid* titers in early larval stages of the bumblebee *Bombus terrestris*. *J Insect Physiol* **46**: 1433-1439.

Holland JG, Guidat FS, Bourke AFG. 2013. Queen control of a key life-history event in a eusocial insect. *Biol Letters* **9**: 20130056.

Humann FC, Tiberio GJ, Hartfelder K. 2013. Sequence and expression characteristics of long noncoding RNAs in honey bee caste development - potential novel regulators for transgressive ovary size. *PLOS One* **8**: e78915

Hunt JH, Buck NA, Wheeler DE. 2003. Storage proteins in vespid wasps: characterization, developmental pattern, and occurrence in adults. *J Insect Physiol* **49**: 785-794.

Johnson BR, Tsutsui ND. 2011. Taxonomically restricted genes are associated with the evolution of sociality in the honey bee. *BMC Genomics* **12**: 164.

Kamakura M. 2011. *Royalactin* induces queen differentiation in honeybees. *Nature* **473**: 478-483.

Kloosterman WP, Plasterk RHA. 2006. The diverse functions of microRNAs in animal development and disease. *Dev Cell* **11**: 441-450.

Koywiwattrakul P, Sittipraneed S. 2009. Expression of *vitellogenin* and *transferrin* in activated ovaries of worker honey bees, *Apis mellifera*. *Biochem Genet* **47**: 19-26.

Koywiwattrakul P, Thompson GJ, Sitthipraneed S, Oldroyd BP, Maleszka R. 2005. Effects of carbon dioxide narcosis on ovary activation and gene expression in worker honeybees, *Apis mellifera*. *J Insect Sci* **5**: 36.

Kozomara A, Griffiths-Jones S. 2011. *miRBase*: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res* **39**: D152-D157.

Lee RC, Feinbaum RL, Ambros V. 1993. The *C-Elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell* **75**: 843-854.

Legeai F, Rizk G, Walsh T, Edwards O, Gordon K, Lavenier D, Leterme N, Mereau A, Nicolas J, Tagu D et al. 2010. Bioinformatic prediction, deep sequencing of microRNAs and expression analysis during phenotypic plasticity in the pea aphid, *Acyrthosiphon pisum*. *BMC Genomics* **11**: 281-289.

Levenbook L. 1985. Insect storage proteins. In *Comprehensive insect physiology, biochemistry and pharmacology*, Vol 10 (ed. G Kerkut, L Gilbert), pp. 307-346. Pergamon Press, Oxford, UK.

Li JL, Huang JX, Cai WZ, Zhao ZW, Peng WJ, Wu J. 2010. The *vitellogenin* of the bumblebee, *Bombus hypocrita*: studies on structural analysis of the cDNA and expression of the mRNA. *J Comp Physiol B* **180**: 161-170.

Libbrecht R, Corona M, Wende F, Azevedo DO, Serrao JE, Keller L. 2013. Interplay between *insulin signaling*, *juvenile hormone*, and *vitellogenin* regulates maternal effects on polyphenism in ants. *Proc Natl Acad Sci USA* **110**: 11050-11055.

Liu F, Peng W, Li Z, Li W, Li L, Pan J, Zhang S, Miao Y, Chen S, Su S. 2012. Next-generation small RNA sequencing for microRNAs profiling in *Apis mellifera*: comparison between nurses and foragers. *Insect Mol Biol* **21**: 297-303.

Lopez-Gomollon S. 2011. Detecting sRNAs by Northern Blotting. In *MicroRNAs in Development: Methods and Protocols*, (ed. T Dalmay). Humana Press, New York, USA.

Lopez-Gomollon S, Mohorianu I, Szittya G, Moulton V, Dalmay T. 2012. Diverse correlation patterns between microRNAs and their targets during tomato fruit development indicates different modes of microRNA actions. *Planta* **236**: 1875-1887.

Lopez-Vaamonde C, Brown RM, Lucas ER, Pereboom JJM, Jordan WC, Bourke AFG. 2007. Effect of the queen on worker reproduction and new queen production in the bumble bee *Bombus terrestris*. *Apidologie* **38**: 171-180.

Lytle JR, Yario TA, Steitz JA. 2007. Target mRNAs are repressed as efficiently by microRNA-binding sites in the 5 ' UTR as in the 3 ' UTR. *Proc Natl Acad Sci USA* **104**: 9667-9672.

Marco A, Hui JHL, Ronshaugen M, Griffiths-Jones S. 2010. Functional shifts in insect microRNA evolution. *Genome Biol Evol* **2**: 686-696.

Marco A, Macpherson JI, Ronshaugen M, Griffiths-Jones S. 2012. MicroRNAs from the same precursor have different targeting properties. *Silence* **3**: 8.

Moczek AP. 2010. Phenotypic plasticity and diversity in insects. *Phil Trans R Soc B* **365**: 593-603.

Niu D, Zheng H, Corona M, Lu Y, Chen X, Cao L, Sohr A, Hu F. 2014. Transcriptome comparison between inactivated and activated ovaries of the honey bee *Apis mellifera* L. *Insect Mol Biol* **23**: 668-681.

Nunes FM, Ihle K, Mutti N, Simões Z, Amdam G. 2013. The gene *vitellogenin* affects microRNA regulation in honey bee (*Apis mellifera*) fat body and brain. *J Exp Biol* **216**: 3724-3732.

Nygaard S, Zhang GJ, Schiott M, Li C, Wurm Y, Hu HF, Zhou JJ, Ji L, Qiu F, Rasmussen M et al. 2011. The genome of the leaf-cutting ant *Acromyrmex echinatior* suggests key adaptations to advanced social life and fungus farming. *Genome Res* **21**: 1339-1348.

Okamura K, Hagen JW, Duan H, Tyler DM, Lai EC. 2007. The mirtron pathway generates microRNA-class regulatory RNAs in *Drosophila*. *Cell* **130**: 89-100.

Pereboom JJM, Jordan WC, Sumner S, Hammond RL, Bourke AFG. 2005. Differential gene expression in queen-worker caste determination in bumble-bees. *Proc R Soc B* **272**: 1145-1152.

Pereboom JJM, Velthuis HHW, Duchateau MJ. 2003. The organisation of larval feeding in bumblebees (Hymenoptera, Apidae) and its significance to caste differentiation. *Insectes Soc* **50**: 127-133.

Peterson KJ, Dietrich MR, McPeek MA. 2009. MicroRNAs and metazoan macroevolution: insights into canalization, complexity, and the Cambrian explosion. *Bioessays* **31**: 736-747.

Ratnieks FLW, Foster KR, Wenseleers T. 2011. Darwin's special difficulty: the evolution of "neuter insects" and current theory. *Behav Ecol Sociobiol* **65**: 481-492.

Reinhart BJ, Slack FJ, Basson M, Pasquinelli AE, Bettinger JC, Rougvie AE, Horvitz HR, Ruvkun G. 2000. The 21-nucleotide let-7 RNA regulates developmental timing in *Caenorhabditis elegans*. *Nature* **403**: 901-906.

Ro S, Park C, Young D, Sanders KM, Yan W. 2007. Tissue-dependent paired expression of miRNAs. *Nucleic Acids Res* **35**: 5944-5953.

Roseler PF. 1970. Unterschiede in der kastendetermination zwischen den hummelarten *Bombus hypnorum* und *Bombus terrestris*. *Z Naturforsch* **25**: 543-548.

Ruby JG, Jan CH, Bartel DP. 2007. Intronic microRNA precursors that bypass *Drosha* processing. *Nature* **448**(7149): 83-86.

Sadd BM, Barribeau SM, Bloch G, de Graff DC, Dearden P, Elsik CG, Gadau J, Grimmelikhuijzen CJP, Hasselmann M, Lozier JD et al. 2015. The genomes of two key bumblebee species with primitive eusocial organization. *Genome Biol* **16**:76.

Sappington TW, Raikhel AS. 1998. Molecular characteristics of insect *vitellogenins* and *vitellogenin receptors*. *Insect Biochem Mol Biol* **28**: 277-300.

Shipman BA, Ryan RO, Schmidt JO, Law JH. 1987. Purification and properties of a very high-density-lipoprotein from the hemolymph of the honeybee *Apis mellifera*. *Biochemistry US* **26**: 1885-1889.

Simpson SJ, Sword GA, Lo N. 2011. Polyphenism in insects. *Curr Biol* **21**(18): R738-R749.

Smith CD, Zimin A, Holt C, Abouheif E, Benton R, Cash E, Croset V, Currie CR, Elhaik E, Elsik CG et al. 2011a. Draft genome of the globally widespread and invasive Argentine ant (*Linepithema humile*). *Proc Natl Acad Sci USA* **108**: 5673-5678.

Smith CR, Smith CD, Robertson HM, Helmkampf M, Zimin A, Yandell M, Holt C, Hu H, Abouheif E, Benton R et al. 2011b. Draft genome of the red harvester ant *Pogonomyrmex barbatus*. *Proc Natl Acad Sci USA* **108**: 5667-5672.

Smith CR, Toth AL, Suarez AV, Robinson GE. 2008. Genetic and genomic analyses of the division of labour in insect societies. *Nature Rev Genet* **9**: 735-748

Sorefan K, Pais H, Hall AE, Kozomara A, Griffiths-Jones S, Moulton V, Dalmay T. 2012. Reducing ligation bias of small RNAs in libraries for next generation sequencing. *Silence* **3**: 4.

Stolle E, Wilfert L, Schmid-Hempel R, Schmid-Hempel P, Kube M, Reinhardt R, Moritz RFA. 2011. A second generation genetic map of the bumblebee *Bombus terrestris* (Linnaeus, 1758) reveals slow genome and chromosome evolution in the Apidae. *BMC Genomics* **12**: 48

Suen G, Teiling C, Li L, Holt C, Abouheif E, Bornberg-Bauer E, Bouffard P, Caldera EJ, Cash E, Cavanaugh A et al. 2011. The genome sequence of the leaf-cutter ant *Atta cephalotes* reveals insights into its obligate symbiotic lifestyle. *PLOS Genet* **7**: e1002007.

Sumner S. 2006. Determining the molecular basis of sociality in insects: progress, prospects and potential in sociogenomics. *Ann Zool Fenn* **43**: 423-442.

Sumner S, Pereboom JJM, Jordan WC. 2006. Differential gene expression and phenotypic plasticity in behavioural castes of the primitively eusocial wasp, *Polistes canadensis*. *Proc R Soc B* **273**: 19-26.

Thompson GJ, Kucharski R, Maleszka R, Oldroyd BP. 2006. Towards a molecular definition of worker sterility: differential gene expression and reproductive plasticity in honey bees. *Insect Mol Biol* **15**: 637-644.

Thompson GJ, Yockey H, Lim J, Oldroyd BP. 2007. Experimental manipulation of ovary activation and gene expression in honey bee (*Apis mellifera*) queens and workers: Testing hypotheses of reproductive regulation. *J J Exp Zool Part A* **307A**: 600-610.

Toth AL, Varala K, Newman TC, Miguez FE, Hutchison SK, Willoughby DA, Simons JF, Egholm M, Hunt JH, Hudson ME et al. 2007. Wasp gene expression supports an evolutionary link between maternal behavior and eusociality. *Science* **318**(5849): 441-444.

Weaver DB, Anzola JM, Evans JD, Reid JG, Reese JT, Childs KL, Zdobnov EM, Samanta MP, Miller J, Elsik CG. 2007. Computational and transcriptional evidence for microRNAs in the honey bee genome. *Genom Biol* **8**: R97.

Weiner SA, Toth AL. 2012. Epigenetics in social insects: a new direction for understanding the evolution of castes. *Genet Res Int* **2012**: 609810.

Weinstock GM Robinson GE Gibbs RA Worley KC Evans JD Maleszka R Robertson HM Weaver DB Beye M Bork P et al. 2006. Insights into social insects from the genome of the honeybee *Apis mellifera*. *Nature* **443**: 931-949.

West-Eberhard MJ. 1989. Phenotypic plasticicty and the origins of diversity. *Annu Rev Ecol Syst* **20**: 249-278.

Westholm JO, Lai EC. 2011. Mirtrons: microRNA biogenesis via splicing. *Biochimie* **93**(11): 1897-1904.

Wheeler DE, Buck NA. 1995. Storage proteins in ants during development and colony founding. *J Insect Physiol* **41**: 885-894.

Wilson EO. 1971. *The Insect Societies*. Belknap Press of Harvard University Press, Cambridge, Massachusetts, USA.

Woodard SH, Fischman BJ, Venkat A, Hudson ME, Varala K, Cameron SA, Clark AG, Robinson GE. 2011. Genes involved in convergent evolution of eusociality in bees. *Proc Natl Acad Sci USA* **108**: 7472-7477.

Wurm Y, Wang J, Riba-Grognuz O, Corona M, Nygaard S, Hunt BG, Ingram KK, Falquet L, Nipitwattanaphon M, Gotzek D et al. 2011. The genome of the fire ant *Solenopsis invicta*. *Proc Natl Acad Sci USA* **108**: 5679-5684.

Yoshikawa T, Sanders AR, DeteraWadleigh SD. 1997. Contamination of sequence databases with adaptor sequences. *Am J Hum Gen* **60**: 463-466.

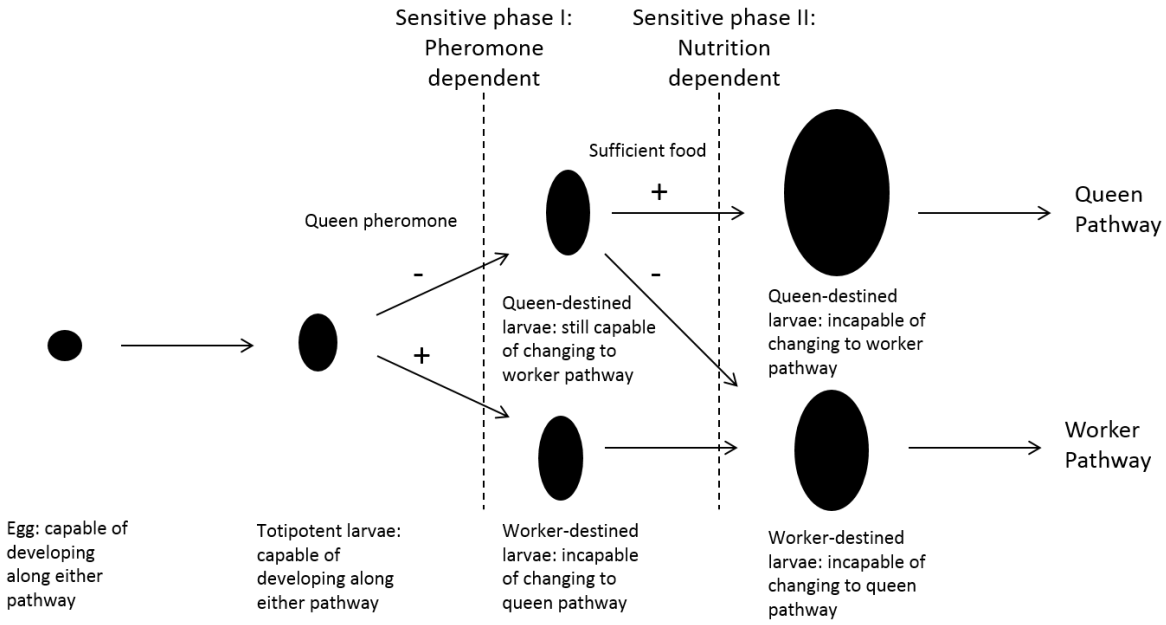


**Figure 3.1: Caste determination in *Bombus terrestris*. There are two sensitive phases in which female larvae change from being totipotent to being committed to a single caste fate. Queen-destined larvae can revert to the worker pathway during either of the sensitive phases while worker-destined larvae are irreversibly determined along the worker pathway once they have passed the first sensitive phase.**

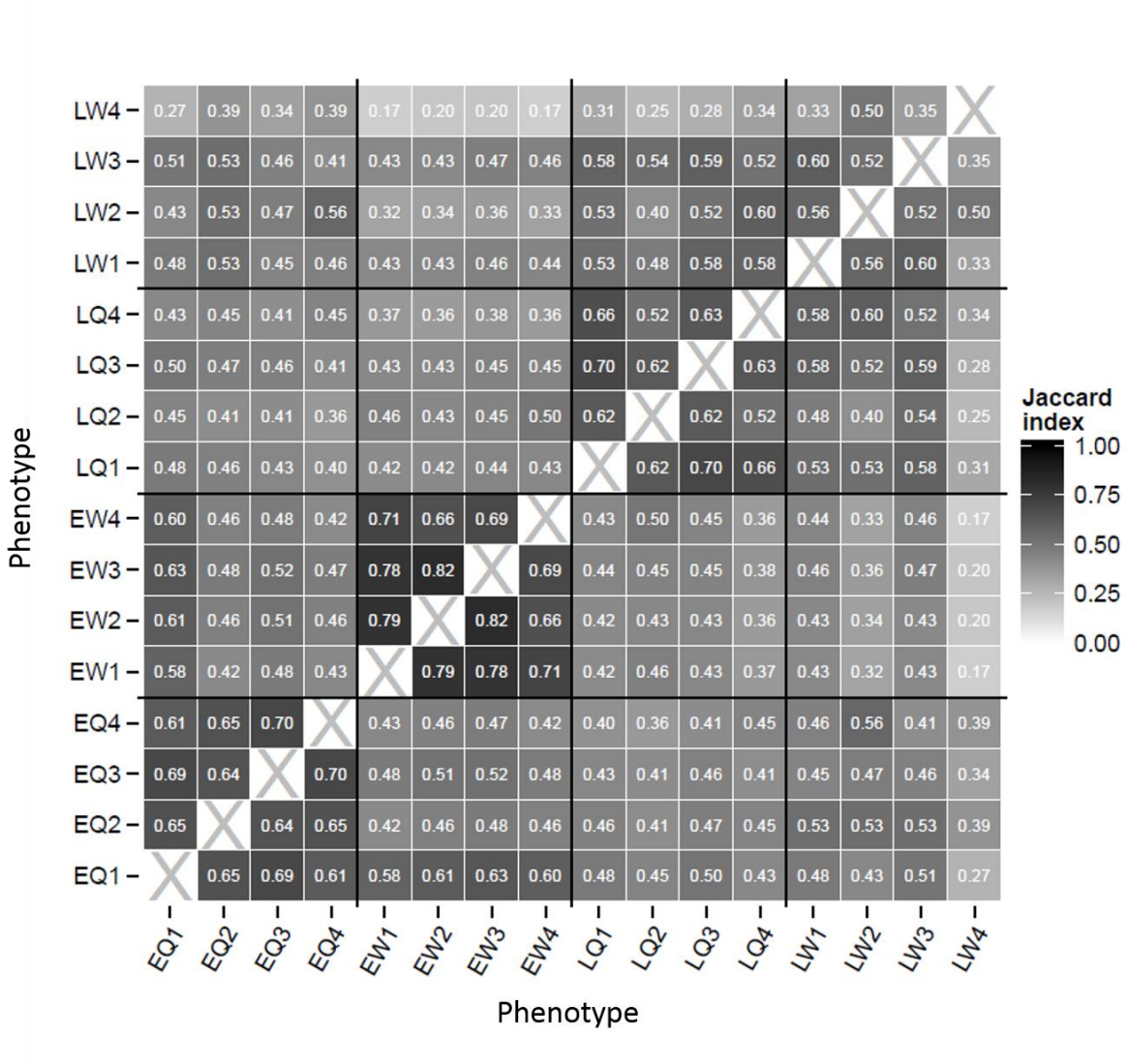


**Figure 3.2: Matrix of Jaccard indices comparing 16 miRNA-enriched libraries prepared from *Bombus terrestris* female larvae (4 replicates of each of 4 phenotypes, i.e. early- and late-instar worker-destined larvae and early- and late-instar queen-destined larvae). For the 500 most abundant sequences, the Jaccard index denotes the proportion of shared sequences between pairs of libraries. The index ranges from 0.00 to 1, with 0 representing libraries containing no shared sequences and 1.00 representing libraries containing exactly the same sequences. EQ = early-instar queen-destined larvae, EW = early-instar worker-destined larvae, LQ = late-instar queen-destined larvae, LW = late-instar worker-destined larvae.**

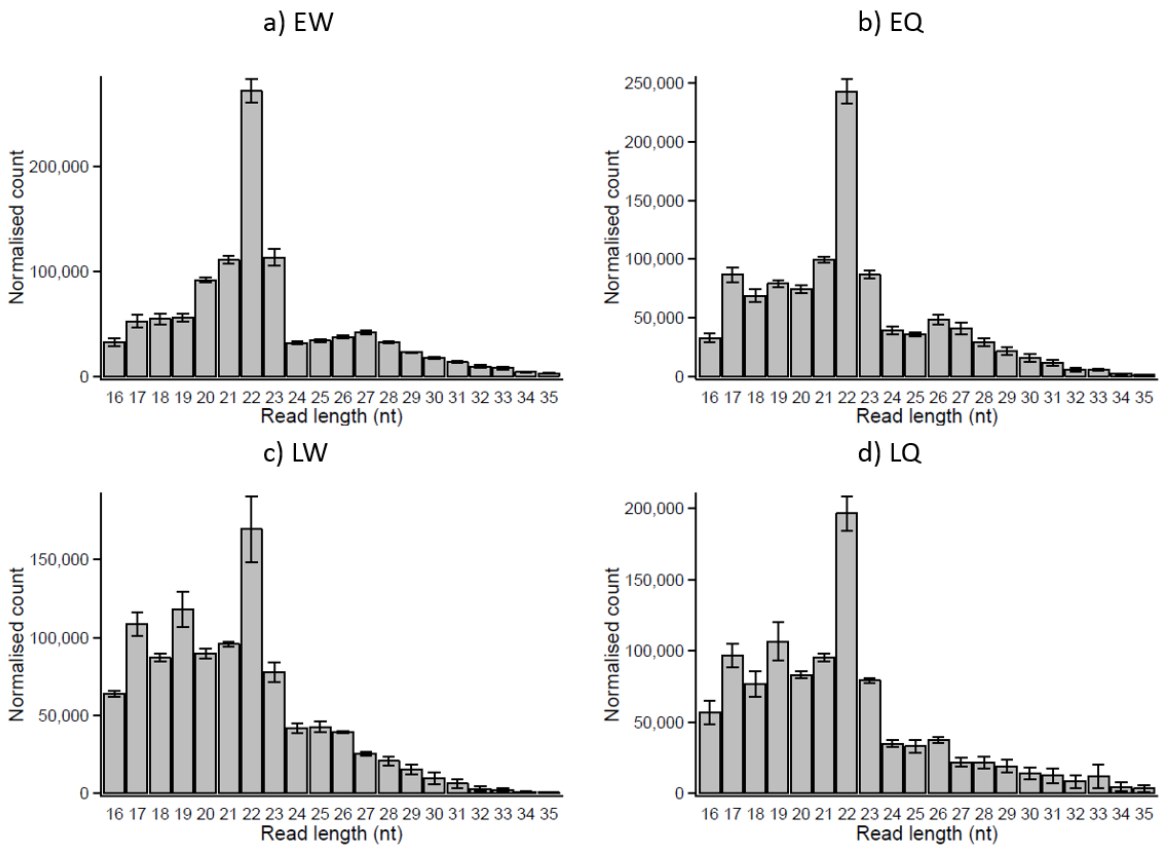


**Figure 3.3: Read length-abundance distributions (following quantile normalisation) of RNA sequences from the 14 miRNA-enriched libraries prepared from *Bombus terrestris* female larvae (four replicates of each of four phenotypes, excluding libraries EQ2 and LW4; phenotypes were early- and late-instar worker-destined larvae and early- and late-instar queen-destined larvae, with replicates averaged between phenotypes). Error bars represent the range of each read count across the four replicate libraries for each phenotype. The largest read count peaks in all libraries were at 21-22 bp. EQ = early-instar queen-destined larvae, EW = early-instar worker-destined larvae, LQ = late-instar queen-destined larvae, LW = late-instar worker-destined larvae.**

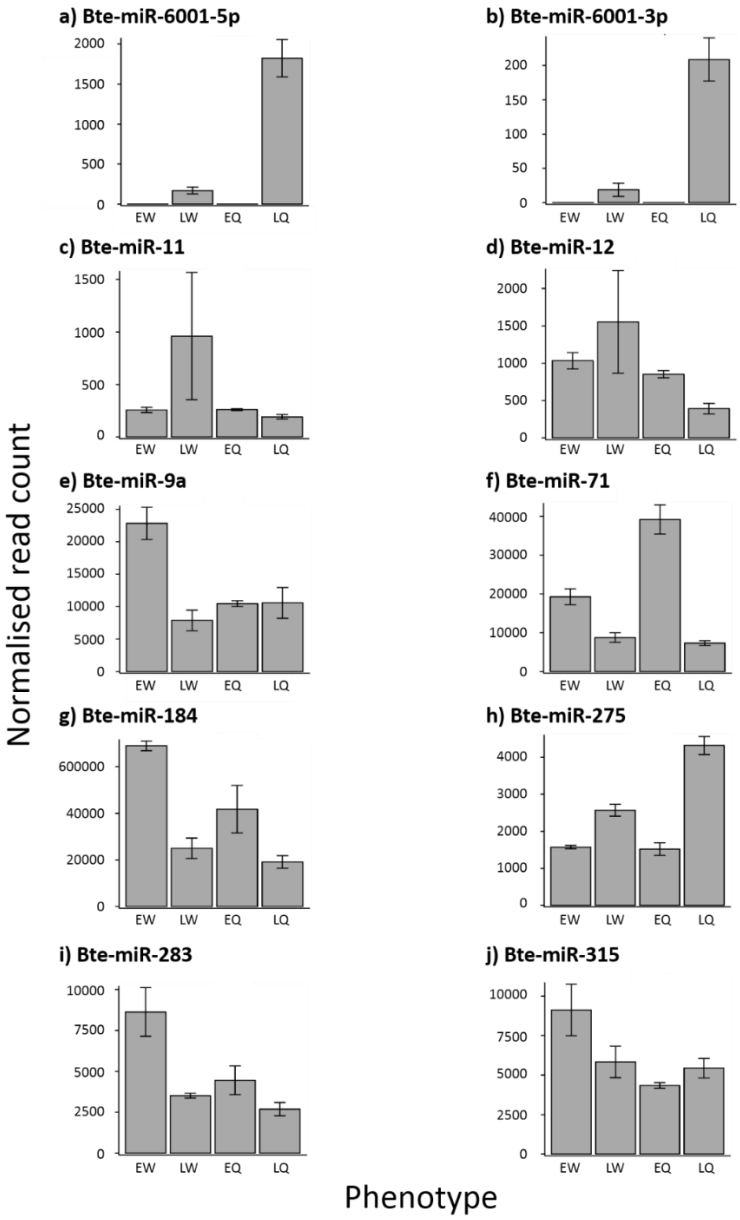


**Figure 3.4: Normalised read counts (reads per million) in the larvae of *Bombus terrestris* for (a-d) four miRNAs that were classified as differentially expressed between queen-destined larvae and worker-destined larvae in the miRNA-seq data, (e-g) three miRNAs that had large expression differences between castes that have been previously associated with caste differentiation in *Apis mellifera*, (h-j) three miRNAs that had large expression differences that have not previously been associated with caste in eusocial insects. Error bars represent the range of each read count across all the four replicate libraries for each phenotype (excluding libraries EQ2 and LW4). EW = early-instar worker-destined larvae, LW = late-instar worker-destined larvae, EQ = early-instar queen-destined larvae, LQ = late-instar queen-destined larvae.**

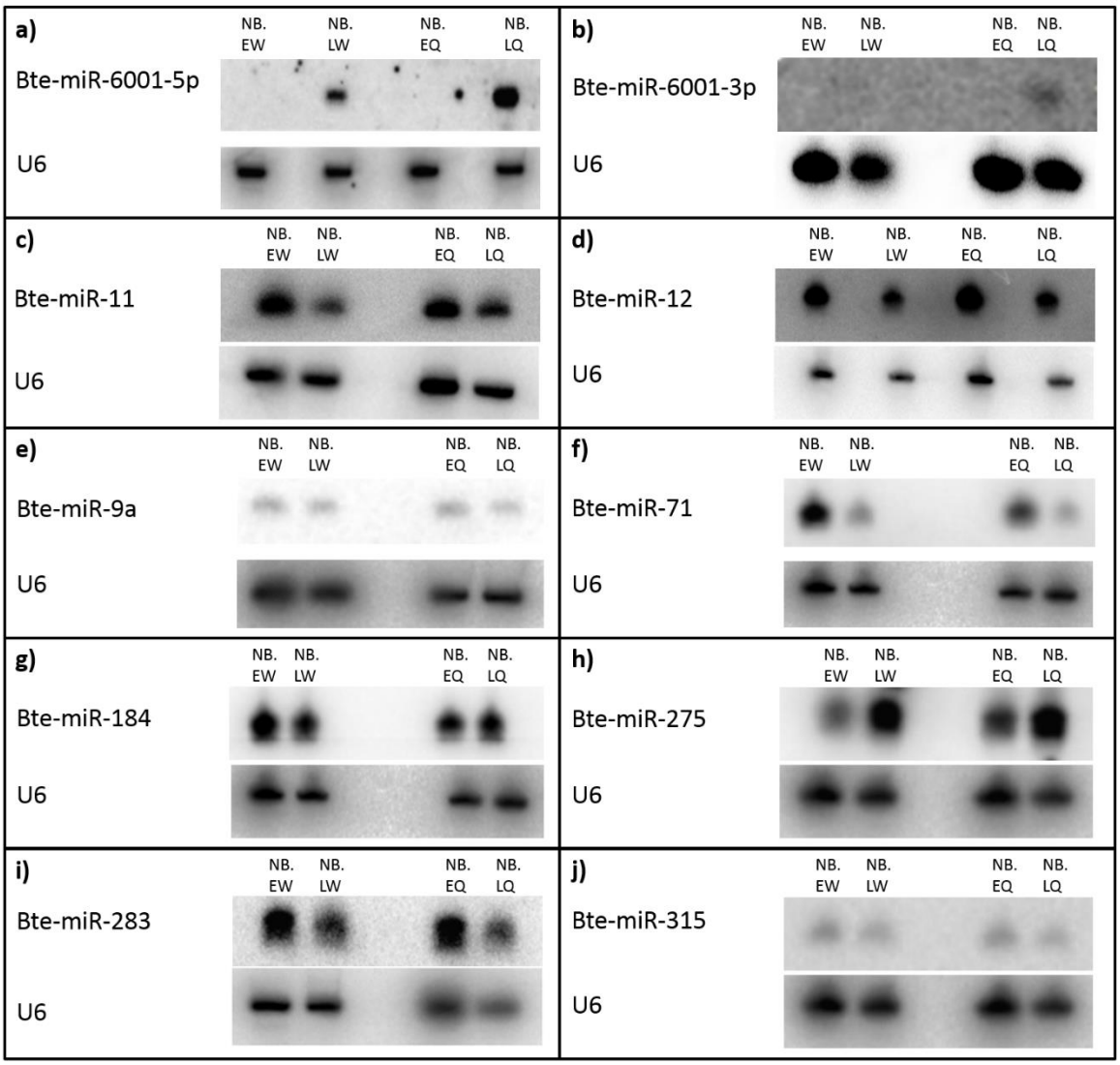


**Figure 3.5: Northern blot expression profiles in the larvae of *Bombus terrestris* for (a-d) four miRNAs that were classified as differentially expressed between queen-destined larvae and worker-destined larvae in the miRNA-seq data, (e-g) three miRNAs that had large expression differences between castes that have been associated with caste differentiation in *Apis mellifera*, (h-j) three miRNAs that had large expression differences that have not previously been associated with caste in eusocial insects. The U6 panels (control) demonstrate equal loading for each membrane. EW = early-instar worker-destined larvae, LW = late-instar worker-destined larvae, EQ = early-instar queen-destined larvae, LQ = late-instar queen-destined larvae. 'NB.' prefix denotes sample for Northern blot (Table 3.1).**

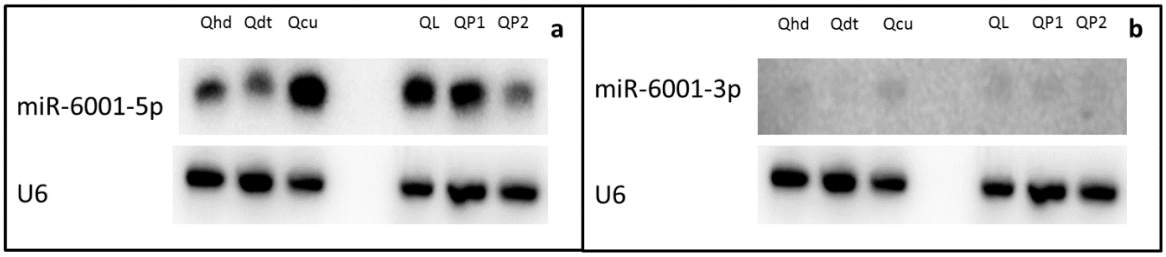


**Figure 3.6: Northern blot expression profiles of *Bombus terrestris* queen-destined fourth instar larvae and queen-destined pupae showing stage-specific expression of two caste differentiated miRNAs. The U6 panels (control) demonstrate equal loading for each membrane. Qhd= queen-destined larva head, Qdt = queen-destined larva digestive tract, Qcu = queen-destined larva cuticle, QL = queen-destined late-instar larva whole body preparation, QP1 = queen-destined early pupa whole body preparation, QP2= queen-destined late pupa whole body preparation.**

**Table 3.1: Details of samples of *Bombus terrestris* larvae used for miRNA-seq and Northern blot analysis. Sixteen colonies were sampled, divided into two cohorts of 8 colonies each (cohorts 1 and 2). Colonies were sampled for larvae either after queen removal (queenless) or with the queen still present (queenright). Columns of numerical data represent: total number of larvae that hatched during experiment days (the period when brood development was being monitored by photographing colonies); of these, the number of early-instar larvae sampled (removed) for RNA extraction and the number of early-instar larvae left within colony to complete development to adulthood; of the last-mentioned category (following further development), the number of late-instar larvae sampled for RNA extraction and the number of late-instar larvae left within colony to complete development to adulthood; and the adult caste of the remaining unsampled larvae, i.e. numbers of adult workers and queens developing from the remaining unsampled larvae. The far right column shows the sequencing library or Northern blot sample to which each colony contributed.**



| Cohort | Colony | Queen status | Total number of larvae | Early-instar larvae sampled | Early-instar larvae left | Late-instar larvae sampled | Late-instar larvae left | Adult caste (unsampled larvae) | Experiment |
|---|---|---|---|---|---|---|---|---|---|
| 1 (Summer, 2011) | QR-2 | Queenright | 85 | 41 | 44 | 19 | 25 | 25 workers<br>0 queens | Sequencing library EW1, LW1 |
| | QR-7 | Queenright | 105 | 44 | 61 | 21 | 40 | 40 workers<br>0 queens | Sequencing library EW2, LW2 |
| | QR-15 | Queenright | 111 | 43 | 68 | 23 | 45 | 45 workers<br>0 queens | Sequencing library EW3, LW3 |
| | QR-17 | Queenright | 96 | 36 | 60 | 22 | 38 | 38 workers<br>0 queens | Sequencing library EW3, LW4 |
| | **Sub-total** | **Queenright** | **397** | **164** | **237** | **85** | **148** | **148 workers<br>0 queens** | **4 Sequencing worker libraries** |
| | QL-10 | Queenless | 69 | 23 | 46 | 22 | 24 | 0 workers<br>24 queens | Sequencing library EQ1, LQ1 |
| | QL-13 | Queenless | 20 | 10 | 10 | 5 | 5 | 0 workers<br>5 queens | Sequencing library EQ2, LQ2 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | QL-14 | Queenless | 65 | 25 | 33 | 27 | 13 | 2 workers<br>11 queens | Sequencing library EQ3, LQ3 |
| | QL-19 | Queenless | 38 | 18 | 20 | 11 | 9 | 0 workers<br>9 queens | Sequencing library EQ4, LQ4 |
| | **Sub-total** | **Queenless** | **192** | **76** | **109** | **65** | **51** | **2 workers<br>49 queens** | **4 Sequencing queen libraries** |
| 2(Spring, 2012) | QR-21 | Queenright | 137 | 59 | 78 | 20 | 58 | 58 workers<br>0 queens | Northern Blot NB.EW1, NB.LW1 |
| | QR-22 | Queenright | 95 | 43 | 52 | 6 | 46 | 45 workers<br>0 queens<br>1 male | Northern Blot NB.EW2, NB.LW2 |
| | QR-23 | Queenright | 104 | 40 | 64 | 22 | 42 | 42 workers<br>0 queens | Northern Blot NB.EW3, NB.LW3 |
| | QR-24 | Queenright | 125 | 47 | 78 | 25 | 53 | 53 workers<br>0 queens | Northern Blot NB.EW4, NB.LW4 |
| | QR-25 | Queenright | 90 | 34 | 56 | 19 | 37 | 37 workers<br>0 queens | Northern Blot NB.EW5, NB.LW5 |
| | **Sub-total** | **Queenright** | **551** | **223** | **328** | **92** | **236** | **235 workers<br>0 queens<br>1 male** | **5 Northern Blot worker samples** |
| | QR-23 | Queenless | 81 | 33 | 48 | 17 | 31 | 1 worker<br>30 queens | Northern Blot NB.EQ1, NB.LQ1 |
| | QR-27 | Queenless | 63 | 31 | 32 | 11 | 21 | 0 workers<br>21 queens | Northern Blot NB.EQ2, NB.LQ2 |
| | QR-29 | Queenless | 79 | 30 | 49 | 10 | 39 | 0 workers<br>39 queens | Northern Blot NB.EQ3, NB.LQ3 |
| | QR-32 | Queenless | 26 | 11 | 15 | 10 | 5 | 1 worker<br>4 queens | Northern Blot NB.EQ4, NB.LQ4 |
| | QR-33 | Queenless | 35 | 17 | 18 | 8 | 10 | 1 worker<br>9 queens | Northern Blot NB.EQ5, NB.LQ5 |
| | **Sub-total** | **Queenless** | **284** | **122** | **162** | **56** | **106** | **3 workers<br>103 queens** | **5 Northern Blot queen samples** |

**Table 3.2: Late-instar queen-destined larvae and queen-destined pupae sampled from three queenless colonies of *Bombus terrestris* from cohort 3 for the analysis of stage-specific miRNA expression. Columns of numerical data show the total number of sampled individuals and, of these, the number of late-instar queen-destined larvae, early-stage queen-destined pupae and late-stage queen-destined pupae sampled. Only one of the three colonies produced queen-destined larvae in this cohort.**

| Cohort | Colony | Total number of sampled individuals | Late-instar queen- larvae | Early-stage queen-destined pupae | Late-stage queen-destined pupae |
|---|---|---|---|---|---|
| 3 (Winter 2013) | QL-36 | 10 | 6 | 2 | 2 |

**Table 3.3: Details of libraries prepared from queen- and worker-destined *Bombus terrestris* larvae and used for miRNA-seq (see also Table 3.1). Columns show: the colony that each library was constructed from; the number of larvae pooled to make each library; read count (the total number of sequence reads obtained) for each library; the number of reads mapped to the *B. terrestris* genome; the percentage of total reads mapped to the *B. terrestris* genome; and a description of the library contents.**

| Library | Colony | Number of larvae | Read count | Mapped read count | Mapped % | Description |
|---|---|---|---|---|---|---|
| EW1 | QR-2 | 41 | 5151876 | 2898777 | 56.27 | Early-instar worker-destined library |
| EW2 | QR-7 | 44 | 6132675 | 3385243 | 55.20 | Early-instar worker-destined library |
| EW3 | QR-15 | 43 | 2659202 | 1478673 | 55.61 | Early-instar worker-destined library |
| EW4 | QR-17 | 36 | 3200163 | 1669350 | 52.16 | Early-instar worker-destined library |
| LW1 | QR-2 | 19 | 3387850 | 1743149 | 51.45 | Late-instar worker-destined library |
| LW2 | QR-7 | 21 | 3107558 | 1539630 | 49.54 | Late-instar worker-destined library |
| LW3 | QR-15 | 23 | 2650788 | 1116095 | 42.10 | Late-instar worker-destined library |
| LW4 | QR-17 | 22 | 3187550 | 2093390 | 65.67 | Late-instar worker-destined library |
| EQ1 | QL-10 | 21 | 4304218 | 2789484 | 64.81 | Early-instar queen-destined library |
| EQ2 | QL-14 | 10 | 5861751 | 3959008 | 67.54 | Early-instar queen-destined library |
| EQ3 | QL-15 | 25 | 4618104 | 2562376 | 55.49 | Early-instar queen-destined library |
| EQ4 | QL-19 | 18 | 4882244 | 3499260 | 71.67 | Early-instar queen-destined library |
| LQ1 | QL-10 | 22 | 3490982 | 2101928 | 60.21 | Late-instar queen-destined library |
| LQ2 | QL-14 | 5 | 6341100 | 3990237 | 62.93 | Late-instar queen-destined library |
| LQ3 | QL-15 | 27 | 3984868 | 2323650 | 58.31 | Late-instar queen-destined library |
| LQ4 | QL-19 | 11 | 2410775 | 1147637 | 47.60 | Late-instar queen-destined library |

**Table 3.4: Details of samples prepared from queen- and worker-destined *Bombus terrestris* larvae and used for Northern blots (see also Table 3.1).**

| Sample name | Colony | Number of larvae | Description |
|---|---|---|---|
| NB.EW1 | QR-21 | 59 | Early-instar worker-destined RNA sample |
| NB.EW2 | QR-22 | 43 | Early-instar worker-destined RNA sample |
| NB.EW3 | QR-23 | 40 | Early-instar worker-destined RNA sample |
| NB.EW4 | QR-24 | 47 | Early-instar worker-destined RNA sample |
| NB.EW5 | QR-25 | 34 | Early-instar worker-destined RNA sample |
| NB.LW1 | QR-21 | 20 | Late-instar worker-destined RNA sample |
| NB.LW2 | QR-22 | 6 | Late-instar worker-destined RNA sample |
| NB.LW3 | QR-23 | 22 | Late-instar worker-destined RNA sample |
| NB.LW4 | QR-24 | 25 | Late-instar worker-destined RNA sample |
| NB.LW5 | QR-25 | 19 | Late-instar worker-destined RNA sample |
| NB.EQ1 | QL-23 | 33 | Early-instar queen-destined RNA sample |
| NB.EQ2 | QL-27 | 31 | Early-instar queen-destined RNA sample |
| NB.EQ3 | QL-29 | 30 | Early-instar queen-destined RNA sample |
| NB.EQ4 | QL-32 | 11 | Early-instar queen-destined RNA sample |
| NB.EQ5 | QL-33 | 17 | Early-instar queen-destined RNA sample |
| NB.LQ1 | QL-23 | 17 | Late-instar queen-destined RNA sample |
| NB.LQ2 | QL-27 | 11 | Late-instar queen-destined RNA sample |
| NB.LQ3 | QL-29 | 10 | Late-instar queen-destined RNA sample |
| NB.LQ4 | QL-32 | 10 | Late-instar queen-destined RNA sample |
| NB.LQ5 | QL-33 | 8 | Late-instar queen-destined RNA sample |

**Table 3.5: MiRNAs from *Bombus terrestris* female larvae probed using Northern blots, their mature sequences and the corresponding probe sequences, along with corresponding sequences for the U6 control.**

| miRNA name | Mature miRNA sequence | Probe sequence |
|---|---|---|
| miR-6001-5p | GUAGGUAACGACUGAUGGGAACA | TGTTCCCATCAGTCGTTACCTAC |
| miR-6001-3p | UUCUCUUUGGUUGUUACCACUA | TAGTGGTAACAACCAAAGAGAA |
| miR-11 | CAUCACAGGCAGAGUUCUAGUU | AACTAGAACTCTGCCTGTGATG |
| miR-12 | UGAGUAUUACAUCAGGUACUGGU | ACCAGTACCTGATGTAATACTCA |
| miR-9a | UCUUUGGUUAUCUAGCUGUAUGA | TCATACAGCTAGATAACCAAAGA |
| miR-71 | UGAAAGACAUGGGUAGUGAGAUG | CATCTCACTACCCATGTCTTTCA |
| miR-184 | UGGACGGAGAACUGAUAAGG | CCTTATCAGTTCTCCGTCCA |
| miR-275 | UCAGGUACCUGAAGUAGCGCGCG | CGCGCGCTACTTCAGGTACCTGA |
| miR-283 | AAAUAUCAGCUGGUAAUUCUG | CAGAATTACCAGCTGATATTT |
| miR-315 | UUUUGAUUGUUGCUCAGAAAGCC | GGCTTTCTGAGCAACAATCAAAA |
| U6 control | GGAACGAUACAGAGAAGAUUAGC | GCTAATCTTCTCTGTATCGTTCC |

Chapter 4

# The role of miRNAs in social behaviour and reproductive differentiation in adult females of the bumble bee, *Bombus terrestris*

*Abstract*

A fundamental challenge in biology is explaining how different cell types and body forms are generated from a single genome. For example, cells in a developing embryo have different shapes and functions, yet all cells arise from the same genome. Likewise, eusocial insects such as bees, wasps and ants develop specialist queen and worker castes with different body forms and behaviours, and again these arise from the same genome. Therefore understanding how caste differences in eusocial insects are generated addresses a much more general biological problem. In previous experiments, we showed that the miRNAs Bte-miR-6001-5p and Bte-miR-6001-3p are differentially expressed between worker-destined and queen-destined *Bombus terrestris* larvae during caste determination. Adult *B. terrestris* exhibit two forms of reproductive division of labour, one that is between castes (queen versus worker) and one that is within the worker caste (inactive-ovary worker versus active-ovary worker). In the present experiments, we sought to: 1) isolate new and conserved miRNAs in the ovaries and brains of *B. terrestris* adult females; 2) test whether miRNAs regulate caste differences in adult females; 3) test whether miRNAs regulate the reproductive differences within the worker caste itself. We tested these aims using Illumina deep sequencing for miRNAs and Northern blot validation in adult queens and workers. We identified a set of miRNAs that differed in expression in the brains and ovaries of queens versus workers. We also identified miRNAs differing in expression in the ovaries of inactive-ovary workers versus active-ovary workers. Of particular interest were miR-184 and miR-133 (more highly expressed in inactive-ovary workers) and miR-279b and miR-279c (more highly expressed in active-ovary workers and queens), which have already been shown to regulate task specialisation in another eusocial insect, *Apis mellifera*, and to regulate reproductive maturation in *Drosophila*. This is the first study to show that miRNAs are associated with the reproductive division of labour in a eusocial insect.

## *4.1 Introduction*

### *Division of labour in eusocial insect societies*

Eusocial insect societies exhibit a reproductive division of labour (Wilson 1971), a phenomenon that has long fascinated biologists. To date, most of the attention has focused on the problem of how a sterile caste can continue to evolve when it does not directly pass on any traits, and, more fundamentally, how traits like sterility can be selected for in the first place (Ratnieks et al. 2011). The most convincing solution to these problems is Hamilton's inclusive fitness theory (Hamilton 1964a; 1964b). This showed how a gene that reduces the direct fitness of an individual (e.g. a gene for sterility) can still undergo selection if it increases the direct fitness of its co-bearers (Dawkins 1976; Bourke 2011a; 2011b). Empirical tests of Hamilton's theory have shown that, despite recent controversy (Abbot et al. 2011), it accurately predicts the conditions under which eusociality evolves (**Chapter 1**; Hughes et al. 2008; Abbot et al. 2011; Bourke 2011a).

Hamilton's inclusive fitness theory has successfully explained the evolution of a reproductive division of labour in eusocial insects at the ultimate level (Queller and Strassmann 1998). However, at the proximate level, the molecular mechanisms underlying the evolution of a sterile worker caste remain far less well understood and are still under active investigation (Robinson et al. 2005; Smith et al. 2008). Understanding these mechanisms has implications for the evolution of co-operation because it potentially helps to elucidate how complex forms such as multicellular organisms are able to evolve (Bourke 2011a). Furthermore, eusociality in insects is a prominent example of a polyphenism, where multiple phenotypes arise from the same genotype (Nijhout 2003). Therefore understanding social evolution in eusocial insects at the proximate level will also increase understanding of how polyphenisms evolve.

Several theories for the proximate mechanisms underlying the evolution of eusociality in insects have been proposed, based on the concept of the 'Ground Plan', which we here refer to as the collective predictions of two hypotheses. The first, the Ovarian Ground Plan Hypothesis (OGPH), states that sociality producing co-operative sibling care has evolved from maternal care in solitary insects. This proposal was based on the observation that solitary wasps undergo reproductive and non-reproductive phases resembling those

involved in social insect task partitioning (West-Eberhard 1987). In the non-reproductive phase the wasp females forage for food and in the reproductive phase they lay eggs and nurse brood. This was likely to have been the case for the ancestors of eusocial insects (West-Eberhard 1987; 1989). At the molecular level, a change to sociality is expected to have happened in eusocial lineages by the co-option of conserved genes present in solitary ancestors, i.e. the genes making up the ground plan (Linksvayer and Wade 2005; Woodard et al. 2013). Therefore a prediction of the OGPH is that, in eusocial insects, genes associated with sib-care behaviour such as nursing brood will also be associated with maternal behaviour such as activating ovaries. Meanwhile, genes associated with foraging behaviour will also be associated with maintaining inactive-ovaries. The second hypothesis, the Reproductive Ground Plan Hypothesis (RGPH), shows how reproductive and brood care traits become coupled even in species that no longer have maternal brood care, such as the honey bee, *Apis mellifera* (Amdam et al. 2004; 2006). The RGPH makes the specific prediction in *A. mellifera* that pollen foraging (a maternal brood care trait in solitary species) is associated with activating ovaries, while nectar foraging (which is necessary for survival even in species without brood care) is associated with maintaining inactive-ovaries. These associations have been shown to occur in workers of *A. mellifera* (Amdam et al. 2006) and there is evidence that the genes that cause ovaries to become active also affect social foraging behaviour in this species (Graham et al. 2011). However, other experiments have challenged some of these associations in *A. mellifera,* where strains of *A. mellifera* that were artificially selected for reproductive variation had none of the predicted differences in caste specialisation (e.g. Oldroyd and Beekman 2008). Therefore the RGPH remains controversial (Linksvayer and Wade 2005; Smith et al. 2008; Graham et al. 2011).

The ground plan hypotheses have been criticised because they do not explain the large expansion of task repertoire in eusocial insects, and many social tasks (e.g. Communication systems in eusocial insects) have no antecedents in solitary species, which means they must have evolved from novel pathways or genes (Johnson and Linksvayer 2010). Despite these issues the ground plan hypotheses still make specific predictions that some genes (that were important for the regulation and behavioural characteristics of solitary species) are regularly co-opted in the multiple evolutions of sociality, and this particular claim has broad support (Smith et al. 2008; Johnson and Linksvayer 2010; Woodard et al. 2011; Dolezal and

Toth 2014). Therefore, some authors have proposed the existence of a 'genomic tool kit' of genes that are essential, or at least regularly re-used, during each independent origin of eusociality (Toth and Robinson 2007; Smith et al. 2008; Toth et al. 2010; Woodard et al. 2013). One example of such genes are those in the insulin-signalling pathway, a pathway that has been co-opted repeatedly in the division of labour and other social processes in all of the major eusocial insect groups (ants, bees, wasps, and termites; Smith et al. 2008). This supports the ground plan and tool kit hypotheses.

*Transcriptomic studies of division of labour*

To test the predictions of the ground plan hypotheses it is important to investigate the transcriptomes of eusocial insects, i.e. the set of all genes that are transcribed into messenger RNA (mRNA). This approach also permits investigators to identify the differentially expressed genes (DEGs) occurring between queens and workers. Several studies of this kind have been performed in a range of eusocial taxa (e.g. Evans and Wheeler 1999; Evans and Wheeler 2001b; Evans and Wheeler 2001a; Linksvayer and Wade 2005; Pereboom et al. 2005; Grozinger et al. 2007; Thompson et al. 2007; Toth et al. 2007; Smith et al. 2008; Kocher et al. 2010; Woodard et al. 2011; Chen et al. 2012a; Ferreira et al. 2013; Woodard et al. 2013; Feldmeyer et al. 2014; Harrison et al. 2015). Many of these studies have sought to identify DEGs from the brains of adult females, and have shown that genes involved in behavioural differentiation are often expressed in brains (e.g. Grozinger et al. 2007; Toth et al. 2007; Kocher et al. 2010; Woodard et al. 2011; Zayed and Robinson 2012; Ferreira et al. 2013; Woodard et al. 2013) and that genes in the brain in social lineages evolve more rapidly than genes in the brain in non-social lineages (Woodard et al. 2011). While the brain is an important organ for investigating behavioural DEGs, this approach is limited because it ignores caste differentiation in other organs. For example, identifying DEGs in eusocial insect ovaries is also important because the most important difference between castes is reproductive capability (Thompson et al. 2007) and multiple tissues and organs should be investigated to understand the reproductive division of labour in eusocial insects (Thompson et al. 2006).

Previously, microarrays and PCR-based methods were used to isolate DEGs in eusocial taxa (Barchuk et al. 2007; Toth et al. 2007; Toth et al. 2010; Johnson and Tsutsui 2011; Woodard

et al. 2013). Some of these approaches have shown that caste-biased genes are conserved across eusocial lineages (Woodard et al. 2013), while others have shown instead that taxonomically-restricted genes (genes not conserved across independent eusocial lineages) are likely to be at least as important (Barchuk et al. 2007; Johnson and Tsutsui 2011). Microarray studies involve hybridising transcripts against pre-existing libraries of expressed genes, and will therefore tend to overlook genes that have not been previously investigated in 'model' species (Ferreira et al. 2013). This is an important shortcoming in the present context, because the alternative to the genomic tool kit and ground plan hypotheses is that the transition to eusociality involves the evolution of numerous new genes and pathways that have evolved *de novo* (Johnson and Tsutsui 2011; Ferreira et al. 2013; Simola et al. 2013; Feldmeyer et al. 2014; Sumner 2014). This is plausible given that taxonomically restricted genes are important in the evolution of novel functions, behaviours and structures in solitary species (Khalturin et al. 2009). Microarray studies therefore introduce a bias in favour of the genomic tool kit and ground plan hypotheses.

Recently, next-generation sequencing technologies have permitted the determination of entire transcriptomes and hence the comparison of gene expression profiles across several eusocial lineages at different stages of eusocial evolution (Toth et al. 2007; Woodard et al. 2011; Ferreira et al. 2013; Feldmeyer et al. 2014). In addition, the availability of a relatively large number of whole genome sequences in the eusocial insects (e.g. *A. mellifera* (Weinstock et al. 2006), *Camponotus floridanus* and *Harpegnathos saltator* (Bonasio et al. 2010), *Acromyrmex exhinatior* (Nygaard et al. 2011), *Atta cephalotes* (Suen et al. 2011), *Linepithema humile* (Smith et al. 2011a), *Pogonomyrmex barbatus* (Smith et al. 2011b), *Solenopsis invicta* (Wurm et al. 2011), *Bombus terrestris* and *B. impatiens* (Sadd et al. 2015)) now makes it easier to identify the genomic context of DEGS from transcriptomic studies, and therefore to discriminate whether conserved or novel genes are most important. Using such approaches, recent studies of eusocial insects have found that taxonomically-restricted genes make up a very large proportion of caste-associated DEGs (Ferreira et al. 2013; Feldmeyer et al. 2014). Most of these DEGs have no known function and so were inferred to have evolved recently in the transition to sociality. Many of these previously unidentified novel genes have been hypothesised to be non-coding RNA, implicating RNA-level regulatory processes in the differences among castes (Ferreira et al. 2013). Overall,

these studies do not appear to support the ground plan and tool kit hypotheses. For example, Ferreira et al. (2013) reported that only 6.5% of the genes found to be caste-biased in the *A. mellifera* were also caste-biased in the paper wasp *Polistes canadensis* (a primitively eusocial insect, i.e. a species with only a behavioural reproductive division of labour that is thought to represent an early state of eusociality). Even in closely related social insects there was little sign of caste-biased genes being shared. For example, *P. canadensis* shared only a very small proportion (roughly 5%) of caste-biased genes with *P. metricus*, despite the fact that these two species have recent common ancestors and similar social systems (Ferreira et al. 2013). However, these findings have not completely refuted the ground plan hypotheses because the functions of these DEGs are not known, and it could be that they have evolved as a consequence of eusocial adaptation, with the core genes for eusociality still having evolved from an ancestral ground plan. Therefore to distinguish between the hypotheses it is important to sequence DEGs in a range of social contexts and species. Most studies to date have focussed on protein coding genes, but this ignores the many non-coding novel DEGs that regulate the transcriptome in other ways (Ferreira et al. 2013). Hence understanding how these regulatory RNAs affect caste will be crucial for gaining a fuller understanding of how eusociality evolved in insects.

*MiRNAs are important regulators of division of labour in eusocial insects*

Few sociogenomic studies have addressed the role of one important class of non-coding regulatory RNAs, the microRNAs (miRNAs). MiRNAs are a class of small RNAs (roughly 21-23 bp in length) that regulate gene expression at the post-transcriptional level (Bartel 2004) and have been shown to play important roles in development (Kloosterman and Plasterk 2006), physiology (Chang and Mendell 2007) and evolution across plants and animals (Axtell et al. 2011). In addition, miRNAs regulate phenotypic differences in species that exhibit polyphenisms (Gutierrez et al. 2009; Legeai et al. 2010). Specifically, one study has shown that up to 17 miRNAs were differentially expressed between winged and wingless phenotypes of the pea aphid *Acyrthosiphon pisum*, wing presence or absence being a polyphenism that is dependent on population density (Legeai et al. 2010).

In previous work (**Chapter 3**), we showed that the expression of Bte-miR-6001-5p was associated with queen-worker caste differentiation in female larvae in the bumble bee

*Bombus terrestris*. We also showed that Bte-miR-6001-5p was tissue- and developmental stage-specific, which is also the case for other miRNAs in other insects such as *Drosophila melanogaster* and *A. mellifera* (Biemar et al. 2005; Weaver et al. 2007). However, we did not investigate whether the differences between the castes in miRNA expression were maintained into adulthood. Other authors have identified miRNAs in *A. mellifera* (Weinstock et al. 2006; Chen et al. 2010) and shown that some miRNAs are involved in caste differentiation in this species (Weaver et al. 2007; Guo et al. 2013). For example Weaver et al. (2007) dissected *A. mellifera* adults into head, thorax, and abdomen, and used qRT-PCR to show that Ame-miR-2 and Ame-miR-9a were more highly expressed in queen than worker abdomens, while Ame-miR-71 was more highly expressed in worker abdomens (Note that throughout this study we use the Ame- prefix to refer to *A. mellifera* miRNAs and the Bte- prefix to refer to *B. terrestris* miRNAs, when describing miRNAs in several species at the same time we will not add a prefix). Some studies have also shown that miRNAs are also associated with polyethisms in workers in *A. mellifera* (where workers initially specialise in nursing behaviour early in their adult life, then later specialise in foraging behaviour). In these studies, miRNAs such as Ame-miR-184 and Ame-miR-133 were found to be more highly expressed in the brains of workers that specialise in foraging (Behura and Whitfield 2010; Greenberg et al. 2012; Liu et al. 2012), while Ame-miR-279b and Ame-miR-279c were more highly expressed in the brains of workers that specialise in nursing (Liu et al. 2012).

*Caste differentiation and division of labour in* B. terrestris

Research on miRNAs and sociogenomics has focussed on *A. mellifera*, which is a good model species for the study of DEGs in eusocial insects. However, as an advanced eusocial insect (a eusocial insect with morphologically distinct castes and a complex reproductive division of labour, rather than simple, behaviourally distinct castes), it is less likely to be informative as regards explaining the evolution of a sterile insect caste from a solitary ancestor. The bumble bee, *B. terrestris*, represents another important model in the study of eusocial evolution in insects (e.g. Baer and Schmid-Hempel 1999; Duchateau et al. 2004; Lopez-Vaamonde et al. 2004; Sadd et al. 2015). *B. terrestris* is usually classified as primitively eusocial (e.g. Cardinal and Danforth 2011) because 1) workers express relatively high levels of reproductive behaviour; 2) queen-worker morphological and size

dimorphism, though present, is relatively low in degree; and 3) colony-founding queens perform both foraging and nursing behaviour (Goulson 2003). However, other primitively eusocial insects such as the *Polistes* wasps, have a worker caste whose members are capable of mating and assuming the role of queen (Liebig et al. 2005; Sumner et al. 2010). This is not the case in *Bombus* species where caste is fixed in adults and workers are not able to mate, a trait that is more characteristic of advanced eusocial insects (Goulson 2003). In addition, *B. terrestris* exhibits task specialisation in the worker caste (alloethism), with specialist individuals performing some tasks such as foraging behaviour, nursing, colony thermoregulation and guarding the colony against entry by intruders (Goulson et al. 2002; Goulson 2003; Gardner et al. 2007). Again, this is a trait often associated with advanced rather than primitive eusociality. Therefore, *B. terrestris* exhibits both primitive and advanced features of eusociality and so could be described as intermediately eusocial. This renders *B. terrestris* particularly suitable for the study of the molecular basis of caste evolution in insects. In addition, *Bombus* species share a common, primitively eusocial ancestor with *A. mellifera* (Cardinal and Danforth 2011). This means that the two groups have both common and independently evolved eusocial traits, making them good comparative species to study the molecular evolution of the reproductive division of labour in eusocial Hymenoptera (the taxonomic group that includes bees, ants, and wasps).

*Social regulation in* B. terrestris *colonies*

A second reason to study the reproductive division of labour in *B. terrestris* is because it has a relatively simple and well understood system of social regulation. Like most other bumble bee species, *B. terrestris* has an annual life cycle. The colony is founded by the queen, who does all of the initial rearing of larvae, before these duties are taken over by the first workers (Goulson 2003). Once enough workers have eclosed (emerged from the pupae), the queen starts to produce sexuals instead, first producing new queens, then undergoing a 'switch point' when she switches from laying diploid eggs (that develop as females) to haploid eggs (that develop as males; Duchateau and Velthuis 1988; Holland et al. 2013). A worker cannot produce diploid eggs, but soon after the queen switches to laying haploid eggs there is a 'competition point' where the workers become aggressive to the queen and to each other, and then start to activate their ovaries so they can produce haploid eggs (Duchateau and Velthuis 1988; Bloch and Hefetz 1999). Therefore, even if the

queen dies, there exists a second reproductive division of labour between the workers with activated ovaries (that produce unfertilised eggs) and inactive-ovary workers, which continue to engage in colony-related tasks (e.g. foraging and raising offspring).

An alternative approach to sequencing DEGs between queen and worker castes is to explore DEGs related to reproductive differentiation within the worker caste, as these are more likely to be related to the reproductive division of labour itself (since queens and workers differ in several traits aside from the ability to lay eggs). This approach has revealed genes in whose expression is related to workers' ovary activity in ants (Feldmeyer et al. 2014), *A. mellifera* (Thompson et al. 2006; Grozinger et al. 2007; Thompson et al. 2008; Cardoen et al. 2011), and *B. terrestris* (Pereboom et al. 2005; Harrison et al. accepted). *B. terrestris* is an informative study species for this purpose because worker reproduction is a regular part of its life cycle. In the worker caste, one might expect there to be fewer DEGs than are found between queens and workers, and therefore the remaining DEGs of inactive-ovary and active-ovary workers are more likely to affect reproductive differentiation directly. Three studies have addressed the molecular basis of worker reproduction in *B. terrestris*. Pereboom et al. (2005) used suppression subtractive hybridisation to test for differential expression between inactive-ovary workers, active-ovary workers and queens. The research showed that several genes were queen-worker differentiated. Although the research group showed that three genes (*chrymotrypsin*, *cytochrome oxidase I*, *peroxiredoxin*) were more highly expressed in inactive-ovary workers than active-ovary workers, there was no evidence of genes that were more highly expressed in reproductive workers and queens compared to non-reproductive workers. In a more recent study, Amarasinghe et al. (2014) showed that inactive-ovary *B. terrestris* workers had methylation profiles differing from those of active-ovary *B. terrestris* workers, and that, crucially, when the methylation profiles of workers were altered using 5-aza-2'-deoxycitidine (an inhibitor of methyltransferases), the workers were more likely to be aggressive and to activate their ovaries. This shows that the control of gene expression via methylation is also an important factor in the molecular regulation of ovarian activity in *B. terrestris*. Finally Harrison et al (accepted) used mRNAseq of RNA extracted from whole-body adult *B. terrestris* to test the hypothesis that active-ovary workers are more similar to

queens than to inactive-ovary workers in their expression profiles. Their results largely confirmed that this was the case for protein-coding genes.

So far the genes that are directly involved in ovary activation in *B. terrestris* have not been identified, and no study has shown whether worker reproduction is associated with changes in miRNA expression in *B. terrestris* or any other eusocial insect. This is important given that, as described above, miRNAs are associated with queen-worker differentiation (Weaver et al. 2007; Guo et al. 2013) and with age polyethism (Behura and Whitfield 2010; Greenberg et al. 2012; Liu et al. 2012). Therefore sequencing the miRNAs associated with worker reproduction in *B. terrestris* will help elucidate the proximate mechanisms underpinning eusociality in this species, and also provides the opportunity to test the possible roles of ground plan miRNAs in the evolution of reproductive division of labour.

*Study aims*

The first aim of this study was to use Illumina deep sequencing of miRNAs (miRNA-seq) and Northern blot validation to isolate new and conserved miRNAs in the ovaries and brains of adult inactive-ovary workers, active-ovary workers, and queens of *B. terrestris*. The second aim was to test whether miRNAs are differentially expressed between adult queens and workers and, if so, whether the miRNAs associated with queen-worker differences among adults are the same as those associated with queen-worker caste determination in larval females (**Chapter 3**). The third aim was to test whether miRNAs are differentially expressed between inactive-ovary workers and active-ovary workers and, if so, whether the miRNAs associated with worker-worker differences among adults are the same as those associated with queen-worker caste determination in adult and larval females. This last aim was addressed by Pereboom et al. (2005), who focussed on protein coding genes rather than miRNAs. In the absence of high-throughput sequencing techniques, the research group were not able to show whether there were any genes that were more highly expressed in reproductive workers and queens compared to non-reproductive workers. Since then Harrison et al. (accepted) have used high-throughput sequencing to test a variation of the same hypothesis, however once again they focussed on protein coding genes and sequenced RNA from whole body extracts rather than from specific tissues. By sequencing from both brains and ovaries we were able to identify tissue specific DEGs, though at the

cost of isolating DEGs associated with other parts of the insect. For both our second and third aims we used Northern blots to validate candidate miRNAs, focusing on miRNAs that have previously been associated with queen-worker differentiation and polyethism in *A. mellifera* (Chapter 3; Weaver et al. 2007; Chen et al. 2010; Behura and Whitfield 2010; Greenfield et al. 2012; Liu et al. 2012; Guo et al. 2013). This was to allow us to test some of the predictions of the ground plan hypotheses (West-Eberhard 1987; Amdam et al. 2004).

## *4.2 Methods*

### *Rearing of B. terrestris colonies and worker sampling*

We obtained 16 incipient *B. terrestris terrestris* colonies from a commercial supplier (Syngenta Bioline Bees B.V., Weert, The Netherlands) in April 2012 (**Table 4.1**). Each colony consisted of a foundress queen with brood (eggs, larvae, and worker-destined pupae) but no adult workers. Colonies were retained in the plastic nest-boxes provided by Syngenta (measuring 200 mm × 200 mm × 150 mm) and kept at constant environmental conditions (28°C, 60% relative humidity) under red light. Colonies were provided with sugar-syrup (Syngenta) and freeze-dried pollen (Koppert, Berkel en Rodenrijis, The Netherlands) *ad libitum*.

We monitored seven of the 16 colonies at 17:00 every day and individually marked newly-eclosed workers with a numbered disk stuck to the thorax (Buzzy Bees Shop Ltd, Leeds, UK), so permitting the age of these workers (days since eclosion as adult) to be known. We allowed the colonies to develop for two weeks, marking all worker bees as they eclosed. Any workers with missing disks were removed to prevent them from being confused with newly-eclosed workers. To induce worker ovary-activation and stimulate worker egg-laying, we removed the queen from each colony 14 days after the first worker eclosed.

Queens were also removed from the other nine colonies (i.e. those with unmarked workers) 14 days after the first worker eclosed. All removed queens were anaesthetised by freezing, weighed to the nearest 100 µg, and then killed by decapitation. We then dissected the queens' ovaries and brains and placed the dissected tissues into RNAlater (Life Technologies, Paisley, UK) for storage at -20°C.

Following removal of queens from the seven colonies with marked workers, we continued to mark all newly-eclosed workers, noting that each colony continued to produce eggs even in the absence of the queen. We let the colonies develop for a further four weeks, until the first males eclosed. We then removed all workers from all seven colonies and divided them into three age-groups, using the numbered disks to estimate the ages of each worker to the nearest day after eclosion: 1) group one consisted of workers more than four weeks old and that had eclosed in queenright conditions (i.e. while the queen was still present); 2) group two consisted of workers that were two to four weeks old and that had eclosed in queenless conditions; 3) group three consisted of workers that were less than two weeks old and that had eclosed in queenless conditions. We weighed the workers and then dissected the ovaries and brains of all of them following the methods used for the queens. We used ovary activation as our measure of worker reproductive behaviour. Although not all active-ovary workers are egg layers, ovary activation has often been used as a measure of a worker's reproductive ability in eusocial bees (Degroot and Voogd 1954; Duchateau and Velthuis 1989; Thompson et al. 2006; Backx et al. 2012). We classified each worker as an active-ovary worker or an inactive-ovary worker according to a previously-published scale (Duchateau and Velthuis 1989). In brief, active-ovary workers were those in which both ovaries contained mature oocytes (stages IV-VI in the scale), while inactive-ovary workers were those in which both ovaries lacked mature oocytes (stage I-II in the scale). Individuals at the intermediate grade (stage III in the scale), i.e. in which there was an intermediate level of ovary activation in one or both ovaries, were discounted. We also discounted bees in which only one of the two ovaries were at stage IV or above in the scale. Likewise, individuals that had activated ovaries and had then undergone ovary regression were also discounted. Ovary regression typically begins once the ovary reaches stage IV (Duchateau and Velthuis 1989). It is possible to distinguish fully regressed ovaries from active-ovaries and inactive-ovaries, as regressed ovaries are as large as active-ovaries but do not contain mature oocytes (Van Oystaeyen et al. 2014). Workers' ovaries and brains were placed in RNAlater and were stored at -20°C.

We compared the mean masses (mg) and ages (number of days between eclosion and dissection) of active-ovary workers and inactive-ovary workers within age-group three with independent-sample t-tests. In the comparison of masses, we excluded one active-ovary

worker from colony AC.2 and in the comparison of ages we excluded two inactive-ovary workers from colony AC.2, because the disks on these three workers had lost their numbers (through wear). Active-ovary workers had significantly higher mass (mean ± SD = 201.9 ± 43.6 mg) than inactive-ovary workers (mean ± SD = 166.1 ± 30.1 mg; $t_{130} = 5.5$, $p < 0.001$). Although we minimised age differences by only sequencing workers from one age group (further details below), active-ovary workers were significantly older (mean ± SD = 11.1 ± 2.7 days) than inactive-ovary workers (mean ± SD = 8.0 ± 3.5 days; $t_{129} = 5.1$, $p < 0.001$). We also compared the mean mass of the queens (mean ± SD = 804.8 ± 50.1 mg) to that of the workers and found that, as expected, the queens had significantly higher mass than active-ovary workers ($t_{52} = 40.9$, $p < 0.001$) and inactive-ovary workers ($t_{12.1} = 43.1$, $p < 0.001$).

*Tissue sample preparations for miRNA-seq and Northern blots*

For miRNA-Seq analysis we only used workers from age-group three because this age-group provided the highest numbers of inactive-ovary workers. Final sample sizes from this group were 12 queens (one each from 12 of the 16 colonies; Colonies AC.1,2,4-12 in **Table 4.1**), 90 inactive-ovary workers from three of the seven colonies with marked workers (Colonies AC.1-4 in **Table 4.1**) and 43 active-ovary workers from three of the seven colonies with marked workers (Colonies AC.2-4 in **Table 4.1**). The samples sizes were equal for the dissected brains and ovaries. Within each worker age-group, tissue, colony and ovary activation stage, we pooled tissue from different individuals before extracting RNA to construct cDNA libraries. MiRNA-seq requires 300-2000 ng of total RNA (depending on the tissue), while Northern blots require 10 µg of total RNA per blot. Meanwhile, the RNA extraction procedure that we used produced approximately 100-8000 ng of total RNA per worker from worker ovaries (depending on worker size and ovary-activation status), 30,000 ng per queen from queen ovaries, 50-500 ng per worker from worker brains, and 100-1000 ng per queen from queen brains. Therefore using pooled samples was necessary to ensure sufficient RNA was available to produce each library. Pooling also meant that expression differences between samples were averaged across individuals and so reflected differences between phenotypes rather than differences between individuals (Feldmeyer et al. 2014). In the worker samples only individuals from the same colonies were pooled, while in the queen samples individuals were pooled between colonies. Each pool of tissues was used to make a separate library. We used these samples to prepare 18 cDNA libraries for miRNA-

seq (**Table 4.2**). These included three replicates each for: 1) inactive-ovary worker ovaries (MS.IO.ov.1-3) and inactive-ovary worker brains (MS.IO.br.1-3); 2) active-ovary worker ovaries (MS.AO.ov.1-3) and active-ovary worker brains (MS.AO.ov.1-3); and 3) queen ovaries (MS.Q.ov.1-3) and queen brains (MS.Q.br.1-3).

We used tissues dissected from workers in age-group two for Northern blot analysis (because of insufficient samples in age-group three for both miRNA-Seq and Northern blot analysis). For age-group two workers, mean per-worker mass of active-ovary workers (mean ± SD = 239.1 ± 58.0 mg) was not significantly different from that of inactive-ovary workers (mean ± SD = 227.0 ± 41.3 mg; $t_{94}$ = -0.84, p = 0.403). Furthermore, mean age of active-ovary workers (mean ± SD = 23.9 ± 4.4 days) was not significantly different from that of inactive-ovary workers (mean ± SD = 24.8 ± 3.7 days; $t_{94}$ = 0.82, p = 0.416). However, because the RNA extracted for miRNA-seq and Northern blots came from different worker age groups, workers providing miRNA-seq samples were, on average, two weeks younger than workers providing Northern blot samples.

For the Northern blot analyses, the sample sizes were three queens from three colonies (AC.14, 15, 16 in **Table 4.1** and **Table 4.3**), 18 inactive-ovary workers from four colonies (AC. 3, 5, 6, 7 in **Table 4.1** and **Table 4.3**), and 78 active-ovary workers from five colonies (AC. 1, 2, 3, 4, 5 in **Table 4.1** and **Table 4.3**). These samples were again pooled prior to RNA extraction, with worker tissues being pooled within a single colony and queen tissues being pooled between colonies. As there were only three queens available for Northern blots, only one Northern blot sample was produced for the ovaries and brains of the three queens. We used our pooled samples to make 19 Northern blot samples (**Table 4.3**). These included: 1) four replicates each for the inactive-ovary worker ovaries (NB.IO.ov.1-4) and inactive-ovary worker brains (NB.IO.br.1-4); 2) five replicates for the active-ovary worker ovaries (NB.AO.ov.1-5) and four replicates for the active-ovary worker brains (NB.AO.br.1-4); and 3) one replicate for the queen ovaries (NB.Q.ov.1) and one replicate for the queen brains (NB.Q.br.1).

*RNA extractions*

To extract RNA, we removed the stored pooled tissue samples and ground them in liquid nitrogen with a mortar and pestle. We extracted RNA using tri-reagent (Sigma-Aldrich,

Gillingham, UK) according to the manufacturer's instructions. We checked RNA integrity by separating the samples on a 1.2% agarose gel and verifying the presence of a single band that contained the small ribosomal sub-unit and both fragments of the large ribosomal sub-unit (Winnebeck et al. 2010). None of the RNA extracts appeared to be degraded (data not shown).

We used the miRVana small RNA purification kit (Ambion, Foster City, California, USA) to isolate small RNAs (i.e. regulatory RNA sequences less than 100 bp in size, which would include all miRNAs) from the samples of total RNA according to the manufacturer's instructions. We then calculated the concentration of the small RNA samples using a Nanodrop 8000 spectrophotometer. All of the pooled samples produced more than 2000 ng of total RNA, which is sufficient for miRNA-seq library preparation (data not shown).

*cDNA library preparation*

We prepared the 18 libraries (**Table 4.2**) using the Scriptminer small RNA-seq library preparation kit (Epicentre, Maddison, Wisconsin, USA) according to a modified version of the manufacturer's protocol, using 1000 ng of input RNA from the pooled ovary samples and 300 ng of input RNA from the pooled brain samples for each library. We used a modified 3' adaptor (Sorefan et al. 2012) to reduce sequencing bias (**Chapter 2.2** *cDNA library preparation* for more details).

Having ligated both the 3' and 5' adaptors to the small RNA sequences and reverse transcribed the resulting products, we PCR-amplified the cDNA with unique Illumina index sequences (index sequences 1-12; **Table 4.2**) that annealed to the adaptor sequences. We tested the specificity of the PCRs by varying the cycle number for each library individually, choosing the cycle number that provided the most clear product when separated on an 8% polyacrylamide gel. We then used the selected cycle number to prepare four PCR reactions for each library, which we separated on a second 8% polyacrylamide gel. We used a razor to cut out the 21-24mer band of interest from the gel, as this size class corresponded to the size class of miRNAs.

We extracted the nucleic acids from the excised gel fragments by shaking them at 4°C overnight in New England Biolabs buffer 2 (NEB2), then precipitating them in ethanol. To reduce adaptor-adaptor contamination, we repeated the process of separating the

products on an 8% polyacrylamide gel, extracting them in NEB 2 and precipitating them with ethanol. We then resuspended each library in 12 µl water. The entire process was repeated for each of the 18 libraries.

*Sanger sequencing and Illumina deep sequencing*

To ensure the cDNA libraries contained the products of the intended size class (i.e. the RNA fraction that contained products of length 21-24 bp) we cloned the constructs into *Escherichia coli* and then sequenced them using Sanger chain-termination sequencing. We used a fifth of each of the libraries MS.AO.br.1 and MS.AO.ov.1 (**Table 4.2**) and ligated the products into pGEMTeasy (Clontech, Saint-Germain-en-Laye, France) vector. We then transformed the recombinant plasmids into DH5α super-competent *E. coli* using blue-white colony staining.

We removed ten white colonies from each library culture and extracted the plasmids using the Qiaprep Spin Mini-prep kit (Qiagen, Crawley, UK) according to the manufacturer's instructions. We sent the samples to the company Eurofins MWG Operon (Ebersberg, Germany) to perform the sequencing reactions.

We analysed the returned reads using the sequence analysis software *FinchTV*, and confirmed that each of the ten sequenced plasmids from each of the two libraries contained miRNAs or miRNA-length sequences (data not shown). The 18 prepared Illumina cDNA libraries were then sent to The Genome Analysis Centre (TGAC; Norwich, UK) for Illumina sequencing on two lanes of a Hiseq2500 using rapid-run mode. Once completed, the Illumina deep sequencing returned a total of 247,985,222 reads.

*Bioinformatics analysis*

We removed the adaptor sequences from the returned sequence reads by matching the 5' end nucleotides 5-12 of the 3' adaptor, and removing them together with the 1-4 degenerate nucleotide sequence.

All sequences were then mapped to the *B. terrestris* genome (Bter20110317) using *PatMaN* and every sequence that contained mismatches or gaps was excluded. Across all libraries, an average 64.9% of the total sequence reads were mapped to the genome (**Table 4.2**). These included redundant reads (sequences counted more than once within a single

library) and non-redundant reads (sequences that were only sequenced once within a single library). We quality-checked the mapped reads using log-sequence count scatter plots, MA plots, and a Jaccard index for the 500 most highly expressed sequences in each library (**Figure 4.1**). Log-sequence count scatter plots are the plots of the read counts (number of reads of a given sequence in a miRNA-seq library) of one replicate within a phenotype, against every other replicate within a phenotype, on a log scale. These plots can then be visually assessed for deviations of each read towards either of the two comparable replicates, with replicates that deviate appreciably then being excluded from further analysis. MA plots were used to assess the fold-change ratio ($M$) of each sequence between two replicates in relation to the average abundance of each sequence ($A$). For a sequence $i$ in samples $k$ and $k'$, the $M$ and $A$ values would be calculated as:

$$M_i = \log_2 \frac{k_i+O}{k_i'+O} \qquad \textbf{Equation 4.1}$$

$$A_i = \log_2 \frac{(k_i+O)(k_i'+O)}{2} \qquad \textbf{Equation 4.2}$$

$O$ is the offset, which was set at 20 in order to remove inherent biases towards fold-change differences in sequences with low read counts. MA plots naturally show a larger $M$ value spread at lower $A$ values, the distribution of which tightens towards zero from both directions at high $A$ values (Beckers et al. In prep). A deviation away from zero that is not corrected by normalisation indicates a scaling bias towards one of the two replicates, indicating that the data are not comparable between samples. The Jaccard index is a simple matrix that compares the proportion of reads for each sample with every other sample. A value of 1.00 indicates that 100% of the most highly expressed 500 reads in sample A are shared with the 500 most highly expressed reads in sample B. A value of 0.00 indicates that none of the 500 most highly expressed reads are shared between samples A and B. Using these quality checks, we determined that, from the scatter plots, none of the replicates deviated strongly from each other, that the MA plots did not show strong deviation from the expected pattern of zero between replicates, and that all libraries had 50% or more of the 500 most highly expressed sequences between replicates of the same phenotype (**Figure 4.1**).

On average we sequenced approximately 14 million reads per library. However, libraries MS.IO.br.1 and MS.AO.br.3 were large outliers, with 28 million and 347 thousand sequences in each library respectively (**Table 4.2**). Reasons for the discrepant numbers of reads in these libraries were unknown, since the pooling method should have returned libraries of similar sizes. To prevent any disproportionate influences of MS.IO.br.1 and MS.AO.br.3 on the quantile normalisation methods, we removed them from further analysis.

We also compared Jaccard indices between samples to identify differential patterns of gene expression between tissues (Lopez-Gomollon et al. 2012). The Jaccard indexes were very different between tissues, which provides evidence of tissue specificity in sequence expression (**Figure 4.1**). Because of this tissue-specific expression, we performed all subsequent analysis in the ovaries and brains separately.

As part of our first aim (to isolate new and conserved miRNAs), we normalised the data to make sequence expression comparable between libraries within each tissue. Prior to normalisation, we boot-strapped the libraries by reducing the number of reads to equal the library with the lowest read count (MS.IO.ov.2 for ovaries and MS.AO.br.1 for brains, **Table 4.2**), removing sequences from the other libraries by random sampling. This method reduces the effect of smaller libraries having a disproportionate influence on differential expression between libraries (Beckers et al. In prep). We normalised the read counts of the sixteen non-excluded libraries using quantile-normalisation (Bolstad et al. 2003), and then calculated the average number of reads of each sequence for each phenotype (across all replicates). As a further quality check, we used read-length abundance plots to compare the read lengths against the average read count between replicates. High abundances at 21-23 nucleotides indicated the presence of miRNAs in these plots, and therefore within the miRNA-seq libraries (**Figure 4.2**).

We used the automated miRNA annotation software *MapMi* (Guerra-Assuncao and Enright 2010) to predict all of the conserved miRNAs that were present in the sequencing data from each library. We then used the miRNA prediction software *miRCat* (Moxon et al. 2008) to predict new miRNAs in the adult data. We classified *miRCat* predictions as new miRNAs if they were sequenced more than 100 times on average in at least one of the phenotypes.

We used *Basic Local Alignment Search Tool* (BLAST) to identify whether the predicted precursors of new miRNAs were conserved in the *A. mellifera* genome, identifying homologues as sequences that had > 85% conservation between putative precursor structures and for which the putative precursor in *A. mellifera* formed a hairpin structure, using the software *RNAfold* (http://srna-tools.cmp.uea.ac.uk/animal/cgi-bin/srna-tools.cgi).

To meet our second aim (to identify miRNAs differentially expressed between worker and queen castes) and third aim (to identify miRNAs differentially expressed between inactive-ovary workers and active-ovary workers), we performed a three-way correlation analysis of the miRNA-seq sequences (including miRNAs) between inactive-ovary workers, active-ovary workers and queens for ovaries and brains separately. For each isolated miRNA, we calculated the fold change ratio ($M_i$, calculated using **Equation 4.1**) between two of the three phenotypes and plotted the values against $M_i$ between one of them and the third phenotype, setting the offset, $O$, at 20 as before. To determine correlative changes in expression related to caste, for each sequence, $i$, we calculated $M_i$ in inactive-ovary workers (relative to queens), where the read count in inactive-ovary workers was $K_i$ and the read count in queens was $K'_i$ from **Equation 4.1**. We then plotted this value against $M_i$ in active-ovary workers (relative to queens) where the read count in active-ovary workers was $K_i$ and the read count in queens was $K'_i$ from **Equation 4.1**. To determine correlative changes in expression relating to ovary-activity, we calculated $M_i$ in queens $K_i$ (relative to inactive-ovary workers $K'_i$) and plotted it against $M_i$ in active-ovary workers $K_i$ (relative to queens $K'_i$). We then calculated Pearson correlation for each obtained set of values.

In a second comparative analysis between phenotypes we compared the expression of individual miRNAs between the three phenotypes to isolate individual miRNAs that were differentially expressed. To meet our second aim, we compared the expression of worker miRNAs (including inactive-ovary workers and active-ovary workers) with that of queen miRNAs. To meet our third aim, we compared the expression of active-ovary workers and queens with that of inactive-ovary workers. For each miRNA, $i$, we calculated $M_i$ in every possible pairwise comparison between each replicate of each phenotype. We identified differentially expressed miRNAs as *miRCat* predictions for which more than half of the pairwise comparisons between replicates of each phenotype had a $M_i$ of > 1 or < -1 (i.e.

sequences with a two-fold or greater change in gene expression; Beckers et al. In prep). The most highly differentiated miRNAs were identified as those that produced the most positive and most negative values of $M_i$ (**Appendix 8** and **Appendix 9**).

*PAGE, Semi-dry transfer, and Northern Blotting*

We used Northern blots to verify the presence of miRNAs isolated by the Illumina deep sequencing and to confirm the expression patterns of selected miRNAs (**Table 4.4)**. To meet our second and third aims, we selected miRNAs for validation by cross-referencing the miRNAs that were differentially expressed between phenotypes with a set of miRNAs associated with social traits in *A. mellifera* derived from previous miRNA studies (**Table 4.4**). We also tested Bte-miR-6001-5p because it was shown to be caste differentiated between worker-destined and queen-destined larvae in **Chapter 3**. In ovaries, following the results of the literature comparison, we used Northern blots to test the expression of the following miRNAs: Bte-miR-184, Bte-miR-279b, Bte-miR-279c, Bte-miR-133, Bte-miR-9a, Bte-miR-71, Bte-miR-3718a, Bte-miR-92b, and Bte-miR-6001-5p (Weaver et al. 2007; Behura and Whitfield 2010; Chen et al. 2010; Greenberg et al. 2012; Liu et al. 2012; Guo et al. 2013). In brains, following the results of the literature comparison, we used Northern blots to test the expression of the following miRNAs: Bte-miR-14, Bte-miR-3718a, Bte-miR-317, and Bte-miR-6001-5p (Yamamoto et al. 2008; Chen et al. 2010; Varghese et al. 2010). For each miRNA we produced two Northern blots using independent colony replicates for the inactive-ovary and active-ovary workers and using technical replicates from the same pooled sample for the queens (**Table 4.3**). The expression pattern was considered to be validated if both Northern blots showed a pattern of differential gene expression in the same direction as that shown by the miRNA-seq results.

To conduct Northern blots, we separated 10 µg of RNA for each phenotype on a 15% denaturing polyacrylamide gel. We transferred the RNA from the gel to a nylon membrane using semi-dry transfer apparatus (Scie-plas), then chemically cross-linked the membrane using 1-ethyl-3-(dimethylaminopropyl) carbodiimide (EDC). We then hybridised the membranes in UltraHyb-Oligo hybridization buffer (Ambion) with probes that were reverse complementary to the miRNAs of interest, and were labelled with 32P using T4 polynucleotide kinase and $\gamma^{-32}$P ATP.

We then rinsed the membranes with a wash solution (0.2 x sodium chloride/sodium citrate, 0.1% (w/v) sodium dodecyl sulphate (SDS)), and exposed them on a blank phospho-imaging screen inside a radioactive cassette (Fujifilm). We scanned the screens on a Molecular Imager FX Pro-Plus to visualize the signal from the radiative probes bound to the membrane. In order to re-use the same membrane with probes for miRNAs of interest and for the loading control, we stripped the membranes of the radioactive signal by incubating them in a stripping solution (pH 8.5, 0.1% SDS, 5mM EDTA) at 95°C for 20-40 minutes. All membranes were re-probed with U6, which is a small nuclear RNA (snRNA) acting as a loading control (Lopez-Gomollon 2011). Further details of this method are described in **Appendix 1**.

## *4.3 Results*

*New miRNAs in B. terrestris*

Our first aim was to isolate the new and conserved miRNAs in *B. terrestris*. Using *miRCat* we predicted 24 new miRNA in the miRNA-seq dataset (**Table 4.5**). Six of these new miRNAs were predicted in our study of *B. terrestris* female larvae (**Chapter 2**; **Appendix 2**) and 18 are newly sequenced in this study. Of these new miRNAs, two (MA1199 and MA341) were conserved in the genome of *A. mellifera*.

The ten most highly expressed miRNAs were different between tissues, indicating a degree of tissue specificity in miRNA expression (**Table 4.6**). Some of the newly-predicted miRNAs were among the most highly expressed miRNAs across all phenotypes (e.g. MC1030, **Table 4.6** and **Appendix 8**).

*Correlation analysis of sequence reads*

Our second aim was to test for caste-biased miRNAs. In the ovaries, there was no correlation in sequence expression (across all miRNA-seq reads, including miRNAs) between inactive-ovary workers (relative to queens) and active-ovary workers (relative to queens), i.e. there was no evidence that a sequence that was highly expressed in the ovaries of inactive-ovary workers compared to queens was also more highly expressed in the ovaries of active-ovary workers compared to queens (**Figure 4.3a**). Hence there was

little evidence of caste-biasing in the ovaries. However, our third aim was to test for miRNAs associated with ovary activation, and this was supported by a positive correlation in sequence expression between active-ovary workers (relative to inactive-ovary workers) and queens (relative to inactive-ovary workers; **Figure 4.3b**). Therefore a sequence that was more highly expressed in the active-ovary workers (relative to inactive-ovary workers) was also likely to be more highly expressed in queens (relative to inactive-ovary workers). This association was found in the miRNAs and in all of the sequenced reads together (**Figure 4.3b**).

The brains exhibited an opposite trend compared to the ovaries. There was a relatively strong positive correlation in sequence expression between inactive-ovary workers (relative to queens) and active-ovary workers (relative to queens; **Figure 4.4a**), i.e. a sequence that was more highly expressed in inactive-ovary workers (relative to queens) was also more likely to be highly expressed in active-ovary workers (relative to queens). This finding addresses our second aim in the brains. Meanwhile, there was only a very weak correlation in sequence expression in active-ovary workers (relative to inactive-ovary workers) and queens (relative to inactive-ovary workers; **Figure 4.4b**). Therefore, as regards our third aim, we found little evidence that ovary activation in workers and queens was associated with differential expression of miRNAs in brains.

*Individual miRNA expression patterns*

In the ovaries there were 58 miRNAs that were differentially expressed (by two-fold or more) between inactive-ovary workers and active-ovary workers, 57 miRNAs that were differentially expressed between inactive-ovary workers and queens, and one miRNA that was differentially expressed between active-ovary workers and queens (**Appendix 8)**. Several of the miRNAs that were differentially expressed between reproductive phenotypes were miRNAs that have previously been found to exhibit differential expression between the castes and/or polyethism stages in *A. mellifera* (Weaver et al. 2007; Behura and Whitfield 2010; Greenberg et al. 2012; Liu et al. 2012; Guo et al. 2013). Of particular interest were Bte-miR-184, Bte-miR-133, Bte-miR-71, and Bte-miR-3718a (which were all more highly expressed in inactive-ovary workers compared to active-ovary workers and queens; **Figure 4.5**), and Bte-miR-279b, Bte-miR-279c, and Bte-miR-9a (which

were all more highly expressed in active-ovary workers and queens compared to inactive-ovary workers; **Figure 4.5**). These results are relevant to our third aim by showing differential expression between inactive-ovary workers and active-ovary workers in the ovaries. Bte-miR-92b was also more highly expressed in queens compared to inactive-ovary workers, but the change was less than a two-fold change in gene expression (**Figure 4.5h**).

The expression patterns of some of the miRNAs that were differentially expressed in ovaries were validated by Northern blot analysis (**Figure 4.6**, **Appendix 8**). Consistent with the miRNA-seq results, Bte-miR-184 was more highly expressed in inactive-ovary workers compared to active-ovary workers and queens (**Figure 4.6a**). However, in contrast with the miRNA-seq, the miRNA was also differentially expressed between queens and active-ovary workers, being most highly expressed in inactive-ovary workers, then expressed at an intermediate level in active-ovary workers, and expressed at the lowest level in queens (**Figure 4.6a**). The Northern blots were consistent with the miRNA-seq for Bte-miR-279b and Bte-miR-279c, which were more highly expressed in queens and active-ovary workers compared to inactive-ovary workers (**Figure 4.6b-c**). Like Bte-miR-184, and consistent with the miRNA-seq results, Bte-miR-133 was more highly expressed in inactive-ovary workers compared to queens and active-ovary workers (**Figure 4.6d**). Bte-miR-92b and Bte-miR-9a also had consistent patterns of miRNA expression between experiments. In both the miRNA-seq and the Northern blot experiments the two miRNAs were more highly expressed in queens and active-ovary workers compared to inactive-ovary workers (**Figure 4.6e,h**). By contrast, there was little evidence from the Northern blots that Bte-miR-71 or Bte-miR-3718a were differentially expressed between any of the adult phenotypes (**Figure 4.6f-g**). This was in contrast with the miRNA-seq results, which showed that both miRNAs were more highly expressed in inactive-ovary workers (**Figure 4.5f-g**).

In brains, we isolated one miRNA that was differentially expressed between inactive-ovary workers and active-ovary workers, 16 miRNAs that were differentially expressed between inactive-ovary workers and queens, and 5 miRNAs that were differentially expressed between active-ovary workers and queens (**Appendix 9**). Most of these have not previously been associated with social behaviour in *A. mellifera* and other eusocial insects (except for miR-279c). The majority of differentially expressed miRNAs (16 miRNAs in total) were differentially expressed between inactive-ovary workers and queens. For example both

Bte-miR-14 and Bte-miR-3718a were more highly expressed in queen brains than in worker brains (**Figure 4.7**). However, Bte-miR-317 was more highly expressed in inactive-ovary workers compared to active-ovary workers, and gene expression in the inactive-ovary workers was closer to the expression levels in queens (**Figure 4.7**). In general, miRNAs that were differentially expressed in ovaries showed little evidence of differential expression in brains.

The Northern blots did not validate the miRNA-seq results in the brains of the three phenotypes (**Figure 4.8a**). There was a low detection threshold for Bte-miR-14 and it had the same level of expression in workers compared to queens. Likewise there was no evidence that Bte-miR-3718a was differentially expressed between queens and workers (**Figure 4.8b**). Finally, Bte-miR-317 was more highly expressed in active-ovary worker brains and queens compared to inactive-ovary worker brains (**Figure 4.8c**). These results are relevant to our third aim for this one miRNA, however the patterns have only been shown in one colony replicate, and are not corroborated by the miRNA-seq results.

*Relationship between larval and adult differentiation*

In previous work (**Chapter 3**), we identified two miRNAs (Bte-miR-6001-5p and Bte-miR-6001-3p) that were more highly expressed in late-instar queen-destined larvae compared with worker-destined larvae. Our second and third aims involved comparisons of the expression patterns of caste-associated larval miRNAs with the miRNA expression patterns in adults. Bte-miR-6001-5p and Bte-miR-6001-3p were not expressed in the ovaries or brains of the adult *B. terrestris* (data not shown) and Northern blots analysis using probes for Bte-miR-600-5p failed to detect a signal (data not shown). Overall we found no evidence that these two miRNAs that were found to be associated with caste differentiation in female larvae were also differentially expressed in adults.

## *4.4 Discussion*

*Expression of miRNAs in B. terrestris*

The first aim of this study was to sequence miRNAs in ovaries and brains from inactive-ovary workers, active-ovary workers, and queens in the bumble bee *B. terrestris*. We then

validated the differentially expressed miRNAs using Northern blot analysis. Two miRNAs in the ovaries, Bte-miR-71 and Bte-miR-3718a, and two miRNAs in the brains, Bte-miR-14 and Bte-miR-3718a, showed different results between miRNA-seq and Northern blot analysis. Previous studies have also shown discrepancies between miRNA-seq and for Northern blot validation (see **Chapter 3**; Baker 2010; Greenberg et al. 2012). One possible reason for this is that the workers used for miRNA-seq were two weeks younger than the workers used for Northern blot analysis, and therefore differential expression of miRNAs in different phenotypes could be age-dependent. However, this was not the case for most ovary miRNAs, where the Northern blot analysis validated the miRNA-seq as predicted. Therefore we will focus the rest of this discussion on the miRNAs that showed consistent patterns of gene expression between the miRNA-seq and Northern blot experiments, and three miRNAs that showed divergent expression patterns between the two methods but were still differentially expressed in the Northern blot analysis (Bte-miR-184, Bte-miR-92b, and Bte-miR-317 respectively).

We used *miRCat* to predict the new miRNAs in *B. terrestris* (**Table 4.5**), isolating 24 new predicted miRNAs (of which 6 had previously been isolated by our study of larvae; **Chapter 2**). The present study is the first to sequence miRNAs in the ovaries of a eusocial Hymenopteran. While some of some of the new miRNAs were conserved in the genomes of both *B. terrestris* and *A. mellifera* (e.g. MA1199 and MA341, **Table 4.5**), most of them were found only in the genome of *B. terrestris*. Some of the new miRNAs were very highly expressed (e.g. MC1030 was the most highly expressed miRNA in the ovaries), implying that the new miRNAs have potentially important functions in *B. terrestris*. In **Chapter 2** we showed that 103 new miRNAs (which had previously only been sequenced in *A. mellifera*) were not conserved in *B. terrestris* and were therefore taxonomically restricted to *A. mellifera*. One explanation for this is that new miRNAs have evolved rapidly in *A. mellifera*, a phenomenon that has been recorded in other insect taxa (Lu et al. 2008; Marco et al. 2010). However, in **Chapter 2**, the number of taxonomically restricted miRNAs in *Bombus* species (*B. terrestris* and *B. impatiens*), at 15, was much lower than *A. mellifera*. In the present study, by sequencing miRNAs in the ovaries and brains of *B. terrestris*, the total number of miRNAs that are taxonomically restricted to *Bombus* was found to be 39 (15 isolated in **Chapter 2**, 24 isolated in the present study). This suggests that the evolution of

new miRNAs has potentially been rapid in both *Bombus* and *Apis* since the two lineages split. However, this conclusion comes with the caveat that all of the new miRNAs are predictions currently based on sequencing evidence alone, so further research, in both species (*B. terrestris* and *A. mellifera*), is needed to prove that they are functional miRNAs (**Appendix 10**).

Our results, in agreement with studies in other species, show that the miRNAs of *B. terrestris* have high tissue specificity (Lagos-Quintana et al. 2002; Biemar et al. 2005; Sood et al. 2006). For example, the Jaccard index showed that the majority of the 500 most highly expressed sequences in each library were not shared between tissues (**Figure 4.1**), and this was also the case for the ten most highly expressed miRNAs (**Table 4.6**). Many studies of miRNA expression in *A. mellifera* have focussed exclusively on the brains or heads (Behura and Whitfield 2010; Hori et al. 2011; Greenberg et al. 2012; Liu et al. 2012). Our results indicate that expression patterns vary between tissues, and therefore these studies have not addressed caste- and polyethism-specific changes in gene expression that occur outside the brain. In the present study, the ovaries showed caste- and reproduction-related differences between the three phenotypes under investigation, thus highlighting the importance of testing for caste differences in different tissues.

As miRNA expression was tissue specific, and there were different trends in different tissues, we will discuss each tissue separately for the second and third aim.

*MiRNAs are differentially expressed between queens and workers*

Our second aim was to identify whether miRNAs show differential expression between adult queens and workers and, if so, whether the miRNAs associated with reproductive differences among adult workers are the same genes as those associated with queen-worker caste determination in female larvae. In the ovaries, miRNA-seq revealed no miRNAs that were differentially expressed between queens and workers (combining inactive-ovary and active-ovary workers into one group, **Appendix 8**) and there was no correlation between caste and sequence expression in the ovaries (**Figure 4.3a**). Nevertheless, the Northern blots revealed that Bte-miR-184 was more highly expressed in both worker phenotypes compared to queens (**Figure 4.5a**). Bte-miR-184 exhibited a gradient of expression within the ovaries, being most highly expressed in inactive-ovary

workers, expressed at an intermediate level in active-ovary workers, and expressed at relatively low levels in queens (**Figure 4.5a**). In addition, Bte-miR-184 had the second highest overall expression in the ovaries of any miRNA (after MC1030; **Table 4.6**). These results strongly suggest that Bte-miR-184 has a role in maintaining different levels of ovary-activation and that it might also be partially responsible for regulating reproductive differences between the castes. Bte-miR-184 is already known to be an essential gene for reproductive development in *Drosophila* (Iovino et al. 2009) and it has also been associated with social regulation and caste differentiation in *A. mellifera* (Behura and Whitfield 2010; Greenberg et al. 2012; Guo et al. 2013). For example, one study showed that Ame-miR-184 is very highly expressed in the Worker Jelly of *A. mellifera* (compared to Royal Jelly), and experimentally demonstrated that it biases female larvae towards development as workers (Guo et al. 2013). Taken in combination with our results, these findings provide strong evidence that miR-184 affects queen-worker caste differentiation, and that it also potentially has a physiological effect on the reproductive differences between adults in both *B. terrestris* and *A. mellifera*.

The pattern of gene expression was different in the brains compared to the ovaries. In the brains there were several sequences (including sequences that were not classified as miRNAs) for which there was a strong association between caste and sequence expression (**Figure 4.4a**). Therefore the miRNA expression profiles in the brains of the two worker phenotypes were more similar to each other than either was to the queen. The sequencing showed caste differentiation in the brains for some miRNAs (e.g. Bte-miR-14 and Bte-miR-3718a being more highly expressed in queen brains compared to worker brains; **Figure 4.7**), but Northern blots on these miRNAs did not show consistent patterns with the miRNA-seq (**Figure 4.8**).

In previous work (**Chapter 3**), we identified two miRNAs (Bte-miR-6001-5p and Bte-miR-6001-3p) that were more highly expressed in queen-destined larvae. In that study we showed that the miRNA was not highly expressed in early-instar larvae but that its expression peaked in late-instar larvae before declining during pupa development. In the present work we isolated 57 miRNAs in the ovaries (**Appendix 8**) and 16 miRNAs in the brains (**Appendix 9**) that were differentially expressed between inactive-ovary workers and queens in adult *B. terrestris*. However, none of these miRNAs were differentially regulated

between caste pathways in larvae. We also found no evidence from miRNA-seq that Bte-miR-6001-5p and Bte-miR-6001-3p were expressed in the brains or ovaries of adult queens and workers at all. In addition, there was no detectable Northern blot signal from Bte-miR-6001-5p, implying that any potential influence of these miRNAs on caste differences occurs before eclosion. These results are not surprising given our evidence that levels of Bte-miR-6001-5p and Bte-miR-6001-3p decline throughout pupa development following peaks in expression in late-instar larvae (**Chapter 3**). Overall, there was no evidence that miRNAs associated with caste differentiation in female larvae were also associated with queen-worker differences at the adult stage.

*MiRNAs are differentially expressed between inactive-ovary and active-ovary bees*

Our third aim was to test whether miRNAs are differentially expressed between inactive-ovary workers and active-ovary workers, and, if so, whether the miRNAs associated with worker-worker differences among adults are also the same as those associated with queen-worker caste differentiation in larval females. As discussed above, we found no evidence that Bte-miR-6001-5p or Bte-miR-6001-3p were expressed in the adults, and therefore no evidence for the latter hypothesis. However, we found that miRNAs were differentially expressed in the ovaries between different ovary activity states. For example, the miRNAs that were more highly expressed in the ovaries of queens (relative to inactive-ovary workers) were also more highly expressed in the ovaries of active-ovary workers (relative to inactive-ovary workers; **Figure 4.3b**). This result complements a study by Harrison et al. (accepted) which explicitly tested the hypothesis that queens and active-ovary workers of *B. terrestris* would be more similar in their expression profiles than either would be to inactive-ovary workers. The authors showed that this was the case for mRNAs isolated by mRNAseq, and this study shows that the result is the same for miRNAs in the ovaries. However the pattern was less strong for miRNAs in the brains of the three castes (**Figure 4.4a**) which further emphasises the importance of extracting RNA from separate tissues to identify caste-specific gene expression differences.

Overall we identified and validated the expression of six miRNAs (Bte-miR-184, Bte-miR-279b, Bte-miR-279c, Bte-miR-133, Bte-miR-92b and Bte-miR-9a; Bte-miR-92b was differentially expressed in the Northern blots only) that were differentially expressed in

ovaries between active-ovary bees (queens and active-ovary workers) and inactive-ovary workers (**Figures 4.5**; **Figure 4.6**).

Of these miRNAs, the best studied in insects are miR-184 (Iovino et al. 2009) and miR-279 (Yoon et al. 2011; Luo and Sehgal 2012). As mentioned previously, Bte-miR-184 was caste differentiated, but the Northern blots revealed that it was also more highly expressed in inactive-ovary workers (**Figure 4.6a**). Both Bte-miR-279b and Bte-miR-279c were more highly expressed in the ovaries of active-ovary workers and queens, compared to the ovaries of inactive-ovary workers (**Figure 4.5b-c**, **Figure 4.6b-c**). In *Drosophila*, Dme-miR-279 plays an important role in regulation of the JAK-STAT signalling pathway, which is responsible for cell division during developmental processes such as follicle cell development (Yoon et al. 2011; Luo and Sehgal 2012). Specifically, knockout mutants are unable to produce mature oocytes (Yoon et al. 2011). This suggests a conserved role for this miRNA across insects in the regulation of ovarian development and activity. These miRNAs have also previously been associated with social behaviour. For example, miR-279c has been shown to be highly conserved in sequence similarity in eusocial Hymenoptera including *A. mellifera*, *B. terrestris*, *B. impatiens*, and several ant species (Greenberg et al. 2012, **Chapter 2**). By contrast, this miRNA is less conserved in solitary Hymenoptera such as *Nasonia vitripennis* (Greenberg et al. 2012). In addition, miR-279b and miR-279c are associated with polyethism in *A. mellifera*, being more highly expressed in nurse specialist workers compared to foraging specialist workers. The effect was the same even when age was controlled for by inducing younger bees to become forager specialist, and older bees to remain as nurse specialists (Greenberg et al. 2012).

Ame-miR-133 has also been associated with polyethism in workers of *A. mellifera*, where it is more highly expressed in forager specialists compared to nurse specialists. Its *B. terrestris* homologue, Bte-miR-133, showed a similar expression pattern to that of Bte-miR-184, i.e. it was most highly expressed in the ovaries of inactive-ovary workers compared to active-ovary workers and queens (**Figure 4.5d, Figure 4.6d**).

Interestingly, the expression patterns of these four miRNAs (Bte-miR-184, Bte-miR-279b, Bte- miR-279c, and Bte-miR-133) support predictions of the ground plan hypothesis. Bte-miR-279b and Bte-miR-279c were both more highly expressed in the active-ovary workers

and queens compared to the inactive-ovary workers, while the *Apis* homologues, Ame-miR-279b and Ame-miR-279c, have been shown to be more highly expressed in nursing *A. mellifera* workers (Liu et al. 2012). The association of miR-279b and miR-279c with ovarian activity in *B. terrestris*, and with nurse behaviour in *A. mellifera* workers, is consistent with the prediction of the ground plan hypotheses that genes associated with reproduction in the ancestors of eusocial insects have been co-opted to regulate division of labour within the worker caste (West-Eberhard 1987; 1989; Amdam et al. 2004; 2006). Another effect of miR-279c is that it controls circadian rhythmicity in *Drosophila*, again by regulation of the JAK-STAT signalling pathway (Luo and Sehgal 2012). This is relevant to the present study because rhythmicity is a trait that is affected by foraging and nursing behaviour in both *A. mellifera* and *B. terrestris* (Bloch et al. 2001; Eban-Rothschild et al. 2011; Bloch et al. 2013). However, it has not previously been linked to ovary activity in these species. The expression patterns of Bte-miR-184 and Bte-miR-133 are also consistent with the predictions of the reproductive ground plan hypothesis. Our results showed that Bte-miR-133 and Bte-miR-184 were more highly expressed in the ovaries of inactive-ovary workers in *B. terrestris*, whereas Greenberg et al. (2012), Liu et al. (2012), and Behura and Whitfield (2010) together showed that Ame-miR-133 and Ame-miR-184 were more highly expressed in old, foraging *A. mellifera* workers compared to nurse workers and younger workers. These findings complement the pattern of expression in Bte-miR-279b and Bte-miR-279c, since they suggest that genes downregulating reproduction are associated with foraging tasks as the hypothesis predicts (West-Eberhard 1987; 1989; Amdam et al. 2004; 2006).

These associations require several qualifications. Firstly, the differences were identified in different social lineages. While *B. terrestris* and *A. mellifera* share a common social ancestor (Cardinal and Danforth 2011), the social structures and caste determining mechanisms in these species are quite different. Therefore to truly establish whether miR-279b, miR-279c, miR-133, and miR-184 are ground plan genes, further studies will need to establish that these miRNAs are also associated with foraging and nursing behaviour in *B. terrestris*, and in ovary activity in *A. mellifera*. A second caveat is that relevant expression patterns in *A. mellifera* derive from brains but those for *B. terrestris* derive from ovaries, and the brains in *B. terrestris* did not show the corresponding expression differences. Finally these results only support the general version of the ground plan hypotheses that individual genes have

been co-opted into social evolution. There is no evidence that this process has co-opted entire pathways that link task specialisation and reproductive behaviour, as predicted by some versions of the ground plan hypotheses (Johnson and Linksvayer 2010). Further studies encompassing whole pathways would be needed to validate such predictions fully. However, these results support the prediction that, for miRNAs in the Apidae, genes associated with foraging behaviour are more highly expressed in inactive-ovary workers; and genes associated with nursing behaviour are more highly expressed in active-ovary workers. These results also provide evidence that caste and reproductive differentiation are controlled by miRNAs, which could target regulatory and developmental genes to produce some of the key changes underlying the evolution of caste systems in eusocial insects.

Besides Bte-miR-184, Bte-miR-279b, Bte-miR-279c, and Bte-miR-133, we also identified reproductive differentiation in Bte-miR-9a and Bte-miR-92b. Bte-miR-9a was more highly expressed in the ovaries of active-ovary workers and queens compared to inactive-ovary workers (**Figure 4.5e, Figure 4.6e**). Ame-miR-9a has been shown to be more highly expressed in the abdomens of workers compared to queens (Weaver et al. 2007). In that study the reproductive status of the workers was not identified, but since most *A. mellifera* workers are sterile, this result was the opposite of what we would expect given that the miRNA was more highly expressed in active-ovary workers in this study. Meanwhile Bte-miR-92b is a miRNA that has been associated with muscle development in insects (Chen et al. 2012b). The miRNA-seq did not show differential expression of the miRNA (**Figure 4.5h**), however the Northern blot analysis showed that it was more highly expressed in the ovaries of queens and active-ovary workers compared to inactive-ovary workers in *B. terrestris* (**Figure 4.6h**). There is not much known about miR-92b in terms of its effect on social behaviour in *A. mellifera* but, like miR-279c, it is more highly conserved in eusocial Hymenoptera (including *B. terrestris*), compared to solitary Hymenoptera such as *N. vitripennis* (Greenberg et al. 2012). Together with our results, this implies that it might play a specific role in social evolution.

In the brains there was also some evidence that one miRNA was differentially expressed between active-ovary bees and inactive-ovary bees. Bte-miR-317 showed differential gene expression between different ovary states. Dme-miR-317 has previously been linked with

sensory behaviour and brain morphogenesis in *Drosophila* (Yamamoto et al. 2008). The sequencing results in this experiment showed that the miRNA was apparently more highly expressed in inactive-ovary workers compared to active-ovary workers and queens (**Figure 4.7**). However, when we tested this with the Northern blots we found the opposite result: that the miRNA is instead at its highest level of expression in queens and active-ovary worker brains (**Figure 4.8**). A possible reason for this discrepancy is that the RNA that was used for Northern blots came from individuals that were from a different age group to the individuals that were used to extract RNA for deep sequencing. If Bte-miR-317 expression has age-specific patterns this could explain the difference in results. However, further experiments will be needed to test this.

Overall we identified seven miRNAs (six in the ovaries, and one in the brain) that were differentially expressed between active-ovary bees and inactive-ovary bees. One of the miRNAs, Bte-miR-184, was also linked to caste in this study. However, none of the isolated miRNAs was associated with caste differentiation in larvae in our previous study (**Chapter 3**). In addition, there was no evidence that Bte-miR-6001 was expressed in adult ovaries or brains at all, so it is unlikely to be responsible for maintaining reproductive differences within the worker caste of *B. terrestris*. Therefore there was no evidence that miRNAs associated with caste differentiation in larvae have an effect on reproductive differentiation once the larvae reach adulthood. Instead it is more likely that different sets of miRNAs are involved at different stages of caste differentiation in *B. terrestris*, with some miRNAs (such as Bte-miR-6001-5p) being associated with caste in larvae, while others (such as Bte-miR-184) being associated with caste in adult females.

## *4.5 References*

Abbot P Abe J Alcock J Alizon S Alpedrinha JAC Andersson M Andre JB van Baalen M Balloux F Balshine S et al. 2011. Inclusive fitness theory and eusociality. *Nature* **471**: E1-E4.

Amarasinghe HE, Clayton CI, Mallon EB. 2014. Methylation and worker reproduction in the bumble-bee (*Bombus terrestris*). *Proc R Soc B* **281**: 20132502.

Amdam GV, Csondes A, Fondrk MK, Page RE. 2006. Complex social behaviour derived from maternal reproductive traits. *Nature* **439**: 76-78.

Amdam GV, Norberg K, Fondrk MK, Page RE. 2004. Reproductive ground plan may mediate colony-level selection effects on individual foraging behavior in honey bees. *Proc Natl Acad Sci USA* **101**: 11350-11355.

Axtell MJ, Westholm JO, Lai EC. 2011. Vive la difference: biogenesis and evolution of microRNAs in plants and animals. *Genome Biol* **12**: 221.

Backx AG, Guzman-Novoa E, Thompson GJ. 2012. Factors affecting ovary activation in honey bee workers: a meta-analysis. *Insect Soc* **59**: 381-388.

Baer B, Schmid-Hempel P. 1999. Experimental variation in polyandry affects parasite loads and fitness in a bumble-bee. *Nature* **397**: 151-154.

Baker M. 2010. MicroRNA profiling: separating signal from noise. *Nat Methods* **7**: 687-692.

Barchuk AR, Cristino AS, Kucharski R, Costa LF, Simoes ZLP, Maleszka R. 2007. Molecular determinants of caste differentiation in the highly eusocial honeybee *Apis mellifera*. *BMC Dev Biol* **7**: 70.

Bartel DP. 2004. MicroRNAs: Genomics, biogenesis, mechanism, and function. *Cell* **116**: 281-297.

Behura SK, Whitfield CW. 2010. Correlated expression patterns of microRNA genes with age-dependent behavioural changes in honeybee. *Insect Mol Biol* **19**: 431-439.

Biemar F, Zinzen R, Ronshaugen M, Sementchenko V, Manak JR, Levine MS. 2005. Spatial regulation of microRNA gene expression in the *Drosophila* embryo. *Proc Natl Acad Sci USA* **102**: 15907-15911.

Bloch G, Barnes BM, Gerkema MP, Helm B. 2013. Animal activity around the clock with no overt circadian rhythms: patterns, mechanisms and adaptive value. *Proc R Soc B* **280**: 20130019.

Bloch G, Hefetz A. 1999. Regulation of reproduction by dominant workers in bumblebee (*Bombus terrestris*) queenright colonies. *Behav Ecol Sociobiol* **45**: 125-135.

Bloch G, Toma DP, Robinson GE. 2001. Behavioral rhythmicity, age, division of labor and period expression in the honey bee brain. *J Biol Rhythms* **16**: 444-456.

Bonasio R, Zhang G, Ye C, Mutti NS, Fang X, Qin N, Donahue G, Yang P, Li Q, Li C et al. 2010. Genomic comparison of the ants *Camponotus floridanus* and *Harpegnathos saltator*. *Science* **329**: 1068-1071.

Bolstad BM, Irizarry RA, Astrand M, Speed TP. 2003. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* **19**: 185-193.

Bourke AFG. 2011a. *Principles of Social Evolution*. Oxford University Press, Oxford, UK.

-. 2011b. The validity and value of inclusive fitness theory. *Proc R Soc B* **278**: 3313-3320.

-. 2014. Hamilton's rule and the causes of social evolution. *Phil Trans R Soc B* **369**: 20130362.

Cardinal S, Danforth BN. 2011. The antiquity and evolutionary history of social behavior in bees. *PLOS One* **6**: e21086.

Cardoen D, Wenseleers T, Ernst UR, Danneels EL, Laget D, De Graaf DC, Schoofs L, Verleyen P. 2011. Genome-wide analysis of alternative reproductive phenotypes in honeybee workers. *Mol Ecol* **20**: 4070-4084.

Chang TC, Mendell JT. 2007. MicroRNAs in vertebrate physiology and human disease. *Annu Rev Genomics Hum Genet* **8**: 215-239.

Chen X, Hu Y, Zheng HQ, Cao LF, Niu DF, Yu DL, Sun YQ, Hu SN, Hu FL. 2012a. Transcriptome comparison between honey bee queen- and worker-destined larvae. *Insect Biochem Mol Biol* **42**: 665-673.

Chen X, Yu X, Cai Y, Zheng H, Yu D, Liu G, Zhou Q, Hu S, Hu F. 2010. Next-generation small RNA sequencing for microRNAs profiling in the honey bee *Apis mellifera*. *Insect Mol Biol* **19**: 799-805.

Chen ZM, Liang SS, Zhao Y, Han Z. 2012b. miR-92b regulates *Mef2* levels through a negative-feedback circuit during *Drosophila* muscle development. *Development* **139**: 3543-3552.

Dawkins R. 1976. *The Selfish Gene*. Oxford University Press, Oxford, UK.

Degroot AP, Voogd S. 1954. On the ovary development in queenless worker bees (*Apis-mellifica* L). *Experientia* **10**: 384-385.

Dolezal AG, Toth AL. 2014. Honey bee sociogenomics: a genome-scale perspective on bee social behavior and health. *Apidologie* **45**: 375-395.

Duchateau MJ, Velthuis HHW. 1988. Development and reproductive strategies in *Bombus* colonies. *Behaviour* **107**:186-207.

-. 1989. Ovarian development and egg-laying in workers of *Bombus terrestris*. *Entomol Exp Appl* **51**: 199-213.

Duchateau MJ, Velthuis HHW, Boomsma JJ. 2004. Sex ratio variation in the bumblebee *Bombus terrestris*. *Behav Ecol* **15**: 71-82.

Eban-Rothschild A, Belluci S, Bloch G. 2011. Maternity-related plasticity in circadian rhythms of bumble-bee queens. *Proc R Soc B* **278**: 3510-3516.

Evans JD, Wheeler DE. 1999. Differential gene expression between developing queens and workers in the honey bee, *Apis mellifera*. *Proc Natl Acad Sci USA* **96**: 5575-5580.

-. 2001a. Expression profiles during honeybee caste determination. *Genome Biol* **2**:research0001.1–0001.6.

-. 2001b. Gene expression and the evolution of insect polyphenisms. *Bioessays* **23**: 62-68.

Feldmeyer B, Elsner D, Foitzik S. 2014. Gene expression patterns associated with caste and reproductive status in ants: worker-specific genes are more derived than queen-specific ones. *Mol Ecol* **23**: 151-161.

Ferreira PG, Patalano S, Chauhan R, Ffrench-Constant R, Gabaldon T, Guigo R, Sumner S. 2013. Transcriptome analyses of primitively eusocial wasps reveal novel insights into the evolution of sociality and the origin of alternative phenotypes. *Genom Biol* **14**:R20.

Gardner KE, Foster RL, O'Donnell S. 2007. Experimental analysis of worker division of labor in bumblebee nest thermoregulation (*Bombus huntii*, Hymenoptera : Apidae). *Behav Ecology and Sociobiol* **61**: 783-792.

Goulson D. 2003. *Bumblebees: Behaviour, ecology and conservation*. Oxford University Press, Oxford, UK.

Goulson D, Peat J, Stout JC, Tucker J, Darvill B, Derwent LC, Hughes WOH. 2002. Can alloethism in workers of the bumblebee, *Bombus terrestris*, be explained in terms of foraging efficiency? *Anim Behav* **64**: 123-130.

Graham AM, Munday MD, Kaftanoglu O, Page RE, Amdam GV, Rueppell O. 2011. Support for the reproductive ground plan hypothesis of social evolution and major QTL for ovary traits of Africanized worker honey bees (*Apis mellifera* L.). *BMC Evol Biol* **11**:95.

Greenberg JK, Xia J, Zhou X, Thatcher SR, Gu X, Ament SA, Newman TC, Green PJ, Zhang W, Robinson GE et al. 2012. Behavioral plasticity in honey bees is associated with differences in brain microRNA transcriptome. *Genes Brain Behav* **11**: 660-670.

Grozinger CM, Fan YL, Hoover SER, Winston ML. 2007. Genome-wide analysis reveals differences in brain gene expression patterns associated with caste and reproductive status in honey bees (*Apis mellifera*). *Mol Ecol* **16**: 4837-4848.

Guerra-Assuncao JA, Enright AJ. 2010. *MapMi*: automated mapping of microRNA loci. *BMC Bioinformatics* **11**: 133.

Guo XQ, Su SK, Skogerboe G, Dai SAJ, Li WF, Li ZG, Liu F, Ni RF, Guo Y, Chen SL et al. 2013. Recipe for a busy bee: microRNAs in honey bee caste determination. *PLOS ONE* **8**: e81661.

Gutierrez L, Bussell JD, Pacurar DI, Schwambach J, Pacurar M, Bellini C. 2009. Phenotypic plasticity of adventitious rooting in arabidopsis is controlled by complex regulation of *auxin* response factor transcripts and microRNA abundance. *Plant Cell* **21**: 3119-3132.

Hamilton WD. 1964a. The genetical evolution of social behaviour. I. *J Theor Biol* **7**: 1-16.

-. 1964b. The genetical evolution of social behaviour. II. *J Theor Biol* **7**: 17-52.

Harrison MC, Hammond RL, Mallon EB. (accepted). Reproductive workers show queen-like gene expression in an intermediately eusocial insect, the buff-tailed bumble bee Bombus terrestris. *Mol Ecol*.

Holland JG, Guidat FS, Bourke AFG. 2013. Queen control of a key life-history event in a eusocial insect. *Biol Letters* **9**: 20130056.

Hori S, Kaneko K, Saito TH, Takeuchi H, Kubo T. 2011. Expression of two microRNAs, ame-mir-276 and-1000, in the adult honeybee (*Apis mellifera*) brain. *Apidologie* **4**: 89-102.

Hughes WOH, Oldroyd BP, Beekman M, Ratnieks FLW. 2008. Ancestral monogamy shows kin selection is key to the evolution of eusociality. *Science* **320**: 1213-1216.

Iovino N, Pane A, Gaul U. 2009. MiR-184 has multiple roles in *Drosophila* female germline development. *Dev Cell* **17**: 123-133.

Johnson BR, Linksvayer TA. 2010. Deconstructing the superorganism: social physiology, groundplans, and sociogenomics. *Quar Rev Biol* **85**: 57-79.

Johnson BR, Tsutsui ND. 2011. Taxonomically restricted genes are associated with the evolution of sociality in the honey bee. *BMC Genomics* **12**: 164.

Khalturin K, Hemmrich G, Fraune S, Augustin R, Bosch TCG. 2009. More than just orphans: are taxonomically-restricted genes important in evolution? *Trends Genet* **25**: 404-413.

Kloosterman WP, Plasterk RHA. 2006. The diverse functions of microRNAs in animal development and disease. *Dev Cell* **11**: 441-450.

Kocher SD, Ayroles JF, Stone EA, Grozinger CM. 2010. Individual variation in pheromone response correlates with reproductive traits and brain gene expression in worker honey bees. *PLOS One* **5**: e9116.

Lagos-Quintana M, Rauhut R, Yalcin A, Meyer J, Lendeckel W, Tuschl T. 2002. Identification of tissue-specific microRNAs from mouse. *Curr Biol* **12**: 735-739.

Legeai F, Rizk G, Walsh T, Edwards O, Gordon K, Lavenier D, Leterme N, Mereau A, Nicolas J, Tagu D et al. 2010. Bioinformatic prediction, deep sequencing of microRNAs and expression analysis during phenotypic plasticity in the pea aphid, *Acyrthosiphon pisum*. *BMC Genomics* **11**: 281-289.

Liebig J, Monnin T, Turillazzi S. 2005. Direct assessment of queen quality and lack of worker suppression in a paper wasp. *Proc R Soc B* **272**: 1339-1344.

Linksvayer TA, Wade MJ. 2005. The evolutionary origin and elaboration of sociality in the aculeate hymenoptera: Maternal effects, sib-social effects, and heterochrony. *Q Rev Biol* **80**: 317-336.

Liu F, Peng W, Li Z, Li W, Li L, Pan J, Zhang S, Miao Y, Chen S, Su S. 2012. Next-generation small RNA sequencing for microRNAs profiling in *Apis mellifera*: comparison between nurses and foragers. *Insect Mol Biol* **21**: 297-303.

Lopez-Gomollon S. 2011. Detecting sRNAs by Northern Blotting. In *MicroRNAs in Development: Methods and Protocols*, (ed. T Dalmay). Humana Press, New York, USA.

Lopez-Gomollon S, Mohorianu I, Szittya G, Moulton V, Dalmay T. 2012. Diverse correlation patterns between microRNAs and their targets during tomato fruit development indicates different modes of microRNA actions. *Planta* **236**: 1875-1887.

Lopez-Vaamonde C, Koning JW, Brown RM, Jordan WC, Bourke AFG. 2004. Social parasitism by male-producing reproductive workers in a eusocial insect. *Nature* **430**: 557-560.

Lu J, Shen Y, Wu Q, Kumar S, He B, Shi S, Carthew RW, Wang SM, Wu CI. 2008. The birth and death of microRNA genes in *Drosophila*. *Nat Genet* **40**: 351-355.

Luo WY, Sehgal A. 2012. Regulation of circadian behavioral output via a microRNA-JAK/STAT circuit. *Cell* **148**: 765-779.

Marco A, Hui JHL, Ronshaugen M, Griffiths-Jones S. 2010. Functional shifts in insect microRNA evolution. *Genome Biol Evol* **2**: 686-696.

Moxon S, Schwach F, Dalmay T, MacLean D, Studholme DJ, Moulton V. 2008. A toolkit for analysing large-scale plant small RNA datasets. *Bioinformatics* **24**: 2252-2253.

Nijhout HF. 2003. Development and evolution of adaptive polyphenisms. *Evol Dev* **5**: 9-18.

Nygaard S, Zhang GJ, Schiott M, Li C, Wurm Y, Hu HF, Zhou JJ, Ji L, Qiu F, Rasmussen M et al. 2011. The genome of the leaf-cutting ant *Acromyrmex echinatior* suggests key adaptations to advanced social life and fungus farming. *Genome Res* **21**: 1339-1348.

Oldroyd BP, Beekman M. 2008. Effects of selection for honey bee worker reproduction on foraging traits. *PLOS Biol* **6**: 463-470.

Pereboom JJM, Jordan WC, Sumner S, Hammond RL, Bourke AFG. 2005. Differential gene expression in queen-worker caste determination in bumble-bees. *Proc R Soc B* **272**: 1145-1152.

Queller DC, Strassmann JE. 1998. Kin selection and social insects. *Bioscience* **48**: 165-175.

Ratnieks FLW, Foster KR, Wenseleers T. 2011. Darwin's special difficulty: the evolution of "neuter insects" and current theory. *Behav Ecol Sociobiol* **65**: 481-492.

Robinson GE, Grozinger CM, Whitfield CW. 2005. Sociogenomics: Social life in molecular terms. *Nature Rev Genet* **6**: 257-216.

Simola DF, Wissler L, Donahue G, Waterhouse RM, Helmkampf M, Roux J, Nygaard S, Glastad KM, Hagen DE, Viljakainen L et al. 2013. Social insect genomes exhibit dramatic evolution in gene composition and regulation while preserving regulatory features linked to sociality. *Genome Res* **23**: 1235-1247.

Smith CD, Zimin A, Holt C, Abouheif E, Benton R, Cash E, Croset V, Currie CR, Elhaik E, Elsik CG et al. 2011a. Draft genome of the globally widespread and invasive Argentine ant (*Linepithema humile*). *Proc Natl Acad Sci USA* **108**: 5673-5678.

Smith CR, Smith CD, Robertson HM, Helmkampf M, Zimin A, Yandell M, Holt C, Hu H, Abouheif E, Benton R et al. 2011b. Draft genome of the red harvester ant *Pogonomyrmex barbatus*. *Proc Natl Acad Sci USA* **108**: 5667-5672.

Smith CR, Toth AL, Suarez AV, Robinson GE. 2008. Genetic and genomic analyses of the division of labour in insect societies. *Nature Rev Genet* **9**: 735-748.

Sood P, Krek A, Zavolan M, Macino G, Rajewsky N. 2006. Cell-type-specific signatures of microRNAs on target mRNA expression. *Proc Natl Acad Sci USA* **103**: 2746-2751.

Sorefan K, Pais H, Hall AE, Kozomara A, Griffiths-Jones S, Moulton V, Dalmay T. 2012. Reducing ligation bias of small RNAs in libraries for next generation sequencing. *Silence* **3**: 4.

Suen G, Teiling C, Li L, Holt C, Abouheif E, Bornberg-Bauer E, Bouffard P, Caldera EJ, Cash E, Cavanaugh A et al. 2011. The genome sequence of the leaf-Cutter ant *Atta cephalotes* reveals insights into its obligate symbiotic lifestyle. *PLOS Genet* **7**. e1002007.

Sumner S. 2014. The importance of genomic novelty in social evolution. *Mol Ecol* **23**: 26-28.

Sumner S, Kelstrup H, Fanelli D. 2010. Reproductive constraints, direct fitness and indirect fitness benefits explain helping behaviour in the primitively eusocial wasp*, Polistes canadensis*. *Proc R Soc B* **277**: 1721-1728.

Thompson GJ, Kucharski R, Maleszka R, Oldroyd BP. 2006. Towards a molecular definition of worker sterility: differential gene expression and reproductive plasticity in honey bees. *Insect Mol Biol* **15**: 637-644.

-. 2008. Genome-wide analysis of genes related to ovary activation in worker honey bees. *Insect Mol Biol* **17**: 657-665.

Thompson GJ, Yockey H, Lim J, Oldroyd BP. 2007. Experimental manipulation of ovary activation and gene expression in honey bee (*Apis mellifera*) queens and workers: Testing hypotheses of reproductive regulation. *J Exp Zool Part A* **307A**: 600-610.

Toth AL, Robinson GE. 2007. Evo-devo and the evolution of social behavior. *Trends in Genetics* **23**: 334-341.

Toth AL, Varala K, Henshaw MT, Rodriguez-Zas SL, Hudson ME, Robinson GE. 2010. Brain transcriptomic analysis in paper wasps identifies genes associated with behaviour across social insect lineages. *Proc Roy Soc B* **277**: 2139-2148.

Toth AL, Varala K, Newman TC, Miguez FE, Hutchison SK, Willoughby DA, Simons JF, Egholm M, Hunt JH, Hudson ME et al. 2007. Wasp gene expression supports an evolutionary link between maternal behavior and eusociality. *Science* **318**: 441-444.

Van Oystaeyen A, Oliveira RC, Holman L, van Zweden JS, Romero C, Oi CA, d'Ettorre P, Khalesi M, Billen J, Wackers F et al. 2014. Conserved class of queen pheromones stops social insect workers from reproducing. *Science* **343**: 287-290.

Varghese J, Lim SF, Cohen SM. 2010. *Drosophila* miR-14 regulates insulin production and metabolism through its target, sugarbabe. *Genes Dev* **24**: 2748-2753.

Weaver DB, Anzola JM, Evans JD, Reid JG, Reese JT, Childs KL, Zdobnov EM, Samanta MP, Miller J, Elsik CG. 2007. Computational and transcriptional evidence for microRNAs in the honey bee genome. *Genom Biol* **8**: R97.

Weinstock GM Robinson GE Gibbs RA Worley KC Evans JD Maleszka R Robertson HM Weaver DB Beye M Bork P et al. 2006. Insights into social insects from the genome of the honeybee *Apis mellifera*. *Nature* **443**: 931-949.

West-Eberhard MJ. 1987. Flexible strategy and social evolution. In *Animal societies: theories and facts*, (ed. Y Ito, J Brown, J Kikkawa), pp. 35-51. Scientific Society Press, Tokyo, Japan.

-. 1989. Phenotypic plasticicty and the origins of diversity. *Annu Rev Ecol Syst* **20**: 249-278.

Wilson EO. 1971. *The Insect Societies*. Belknap Press of Harvard University Press, Cambridge, Massachusetts, USA.

Winnebeck EC, Millar CD, Warman GR. 2010. Why does insect RNA look degraded? *J Insect Sci* **10**: 159.

Woodard SH, Bloch G, Band MR, Robinson GE. 2013. Social regulation of maternal traits in nest-founding bumble bee (*Bombus terrestris*) queens. *J Exp Biol* **216**: 3474-3482.

Woodard SH, Fischman BJ, Venkat A, Hudson ME, Varala K, Cameron SA, Clark AG, Robinson GE. 2011. Genes involved in convergent evolution of eusociality in bees. *Proc Natl Acad Sci USA* **108**: 7472-7477.

Wurm Y, Wang J, Riba-Grognuz O, Corona M, Nygaard S, Hunt BG, Ingram KK, Falquet L, Nipitwattanaphon M, Gotzek D et al. 2011. The genome of the fire ant *Solenopsis invicta*. *P Natl Acad Sci USA* **108**: 5679-5684.

Yamamoto A, Zwarts L, Callaerts P, Norga K, Mackay TFC, Anholt RRH. 2008. Neurogenetic networks for startle-induced locomotion in *Drosophila melanogaster*. *Proc Natl Acad Sci USA* **105**: 12393-12398.

Yoon WH, Meinhardt H, Montell DJ. 2011. MiRNA-mediated feedback inhibition of JAK/STAT morphogen signalling establishes a cell fate threshold. *Nature Cell Biology* **13**: 1062-U1082.

Zayed A, Robinson GE. 2012. Understanding the relationship between brain gene expression and social behavior: Lessons from the honey bee. *Annu Rev Genet* **46**: 591-615.

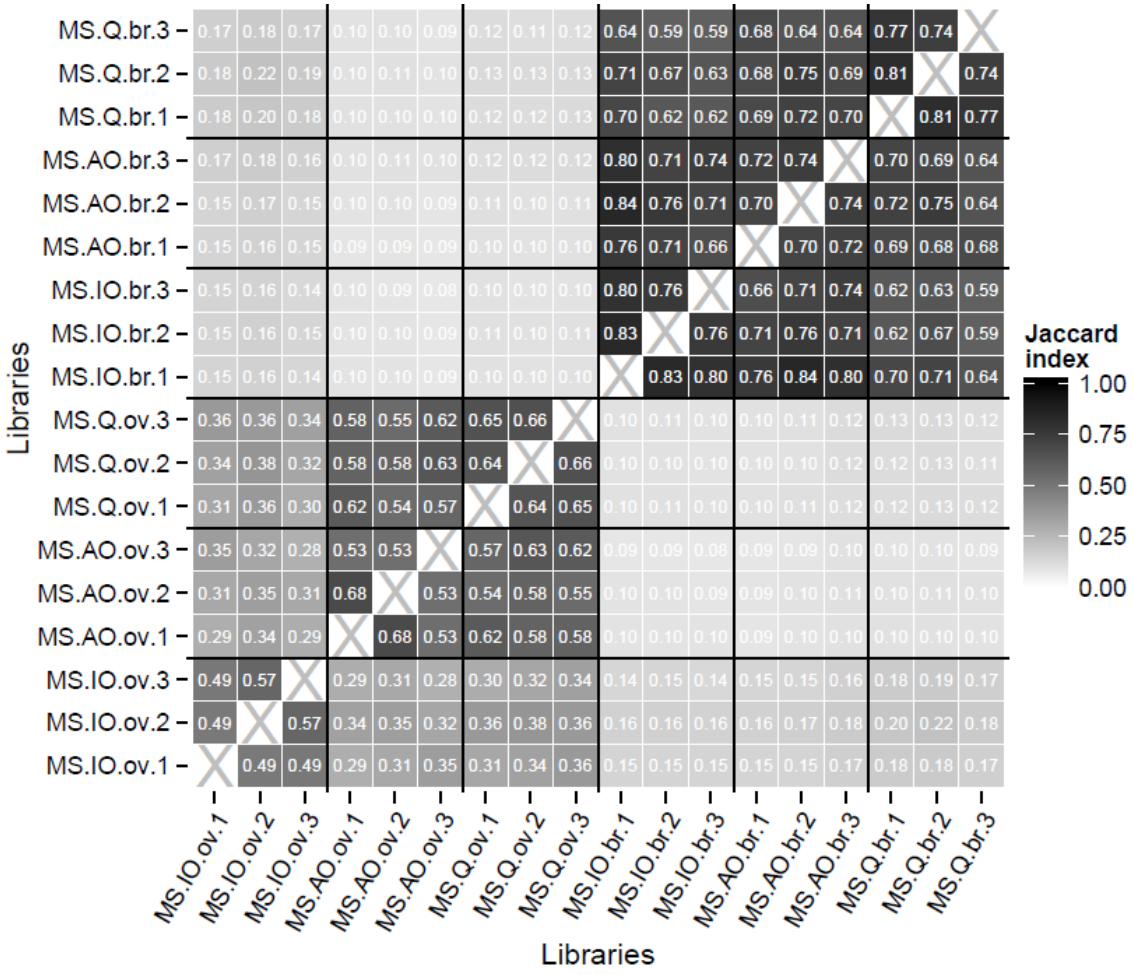


**Figure 4.1: Matrix of Jaccard indices comparing 18 miRNA-enriched libraries prepared from adult queen and worker brains and ovaries in the bumble bee *Bombus terrestris* (three replicates from two tissues from three phenotypes). For the 500 most abundant sequences, the Jaccard index denotes the proportion of shared sequences between pairs of libraries. The index ranges from 0.00 to 1.00 with 0.00 representing libraries containing no shared sequences and 1.00 representing libraries containing exactly the same sequences. Library names correspond to those in Table 4.2.**

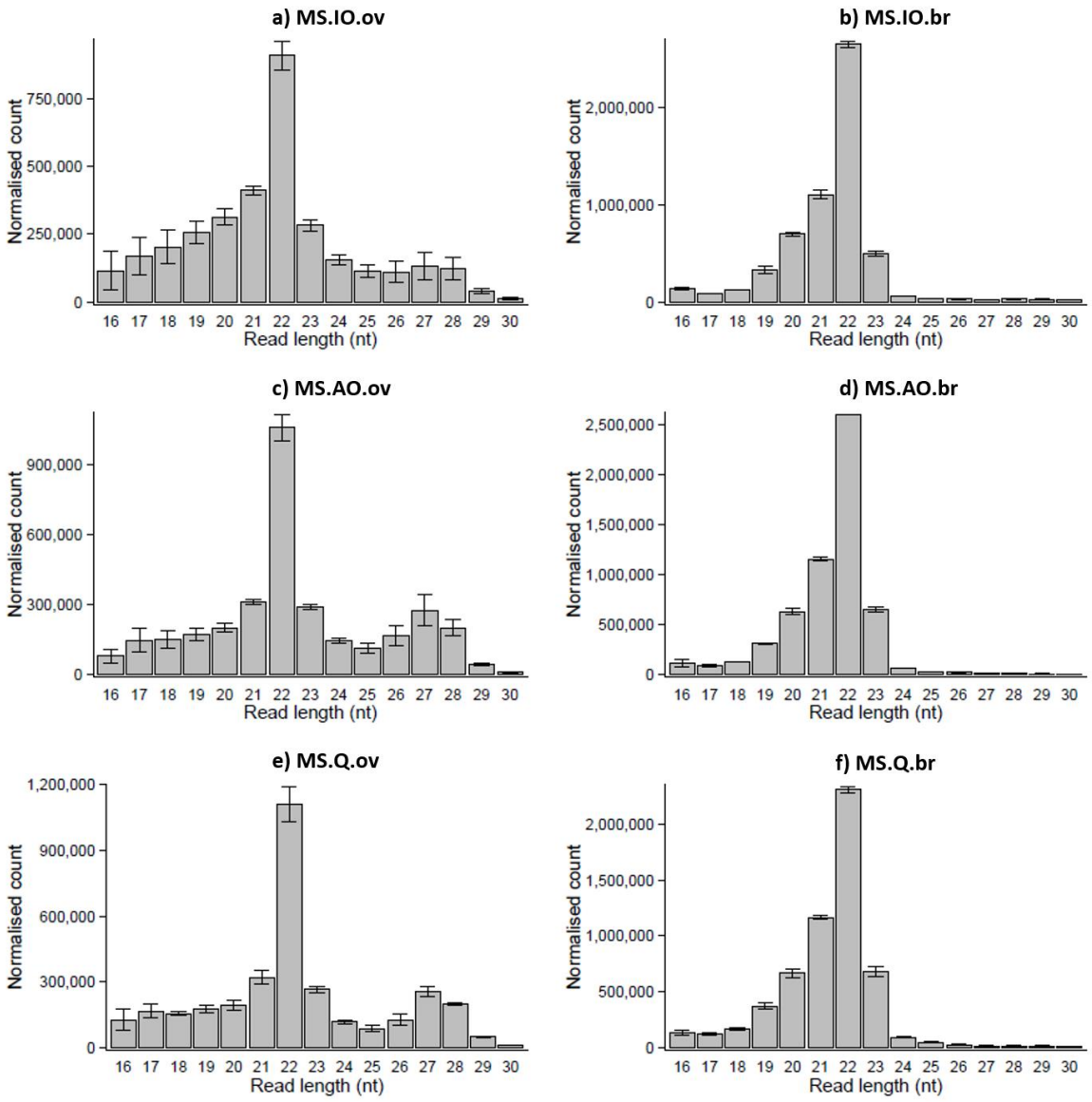


**Figure 4.2: Read length-abundance distribution (following quantile normalisation) of RNA sequences from the 18 miRNA-enriched libraries prepared from *Bombus terrestris* adult females (three replicates from two tissues from the three phenotypes, i.e. ovaries and brains from inactive-ovary workers, active-ovary workers, and queens, with replicates averaged between phenotypes). Error bars represent the range of each read count across the three replicate libraries for each phenotype. The largest read count peaks in all libraries were at 21-22 bp. Peaks were also observed at 27 bp in the libraries prepared from ovary RNA samples. Library names correspond to those in Table 4.2.**

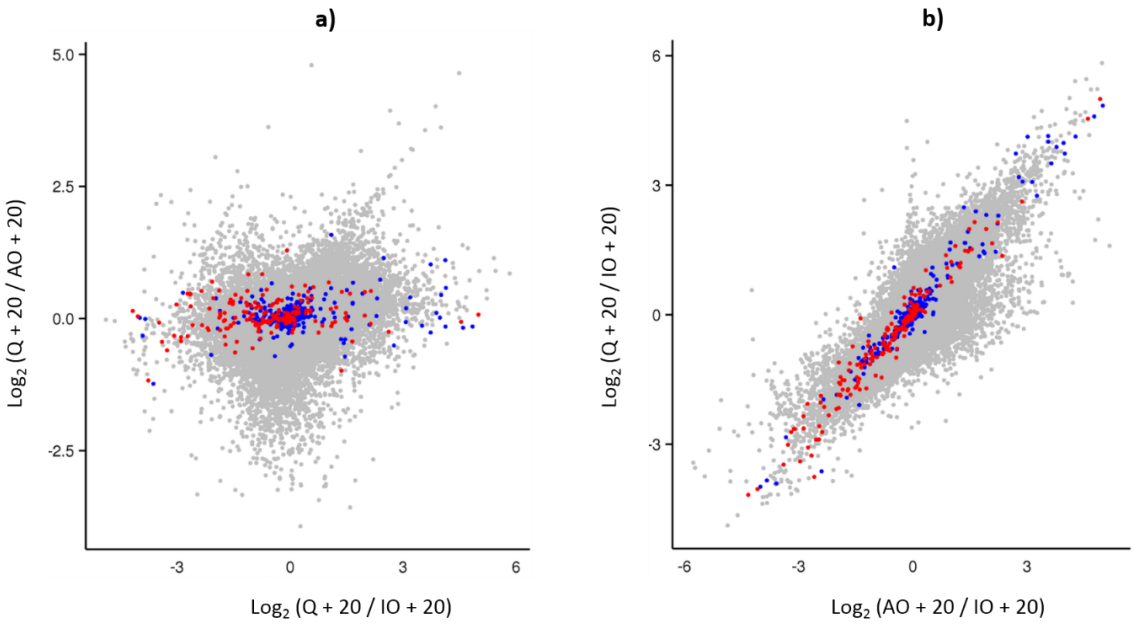


**Figure 4.3: MiRNA-enriched libraries from ovaries in *Bombus terrestris* adult females, a) the average read count for sequences found in queens (relative to average read counts from active-ovary workers) as a function of average read count for sequences found in queens (relative to average read counts from inactive-ovary workers). Across all points Pearson correlation = 0.30, N = 3090328, P < 0.001; b) the average read count for sequences found in active-ovary workers (relative to average read counts from inactive-ovary workers) as a function of average read count for sequences found in queens (relative to average read counts from inactive-ovary workers), across all points Pearson correlation = 0.84, N = 3090328, P < 0.001. Averages are across replicated libraries. Red points, known miRNA sequences that have previously been identified in *A. mellifera* and submitted to *miRBase* (www.miRBase.com); blue points, new miRNA predictions using the miRNA prediction software *miRCat*; grey points, all other sequences. IO = Inactive-ovary worker; AO = active-ovary worker; Q = queen.**

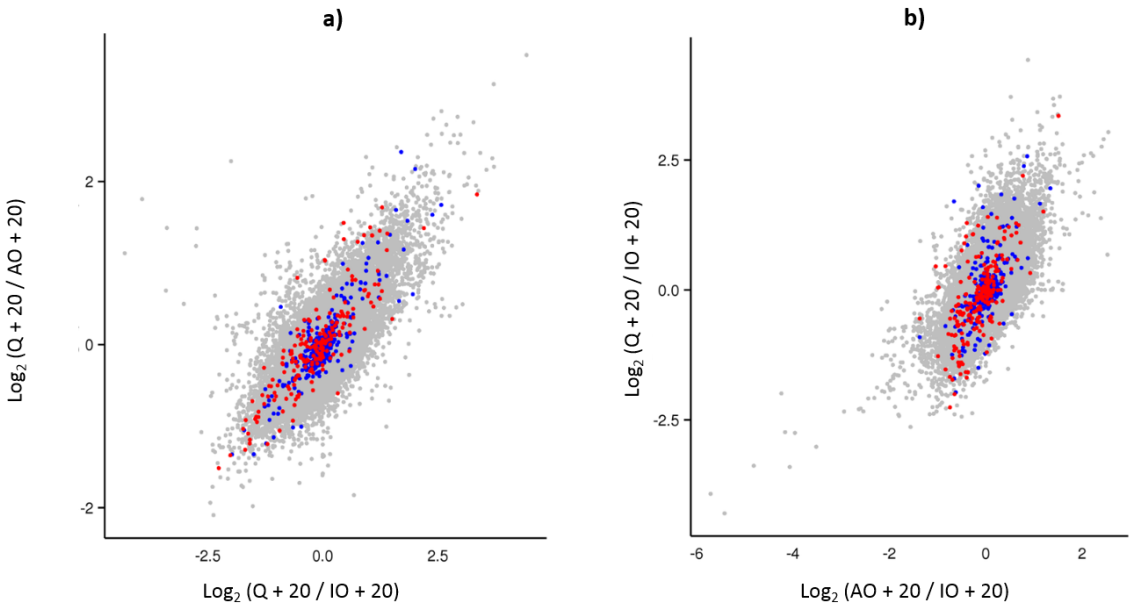


**Figure 4.4: MiRNA-enriched libraries from brain in *Bombus terrestris* adult females, a) the average read count for sequences found in queens (relative to average read counts from active-ovary workers) as a function of average read count for sequences found in queens (relative to average read counts from inactive-ovary workers), across all points Pearson correlation = 0.71, N = 886344, P < 0.001; b) the average read count for sequences found in active-ovary workers (relative to average read counts from inactive-ovary workers) as a function of average read count for sequences found in queens (relative to average read counts from inactive-ovary workers), across all points Pearson correlation = 0.61, N = 886344, P < 0.001. Averages are across replicated libraries. Red points, known miRNA sequences that have previously been identified in *A. mellifera* and submitted to *miRBase* (www.miRBase.com); blue points, new miRNA predictions using the miRNA prediction software *miRCat*; grey points, all other sequences. IO = Inactive-ovary worker; AO = active-ovary worker; Q = queen.**

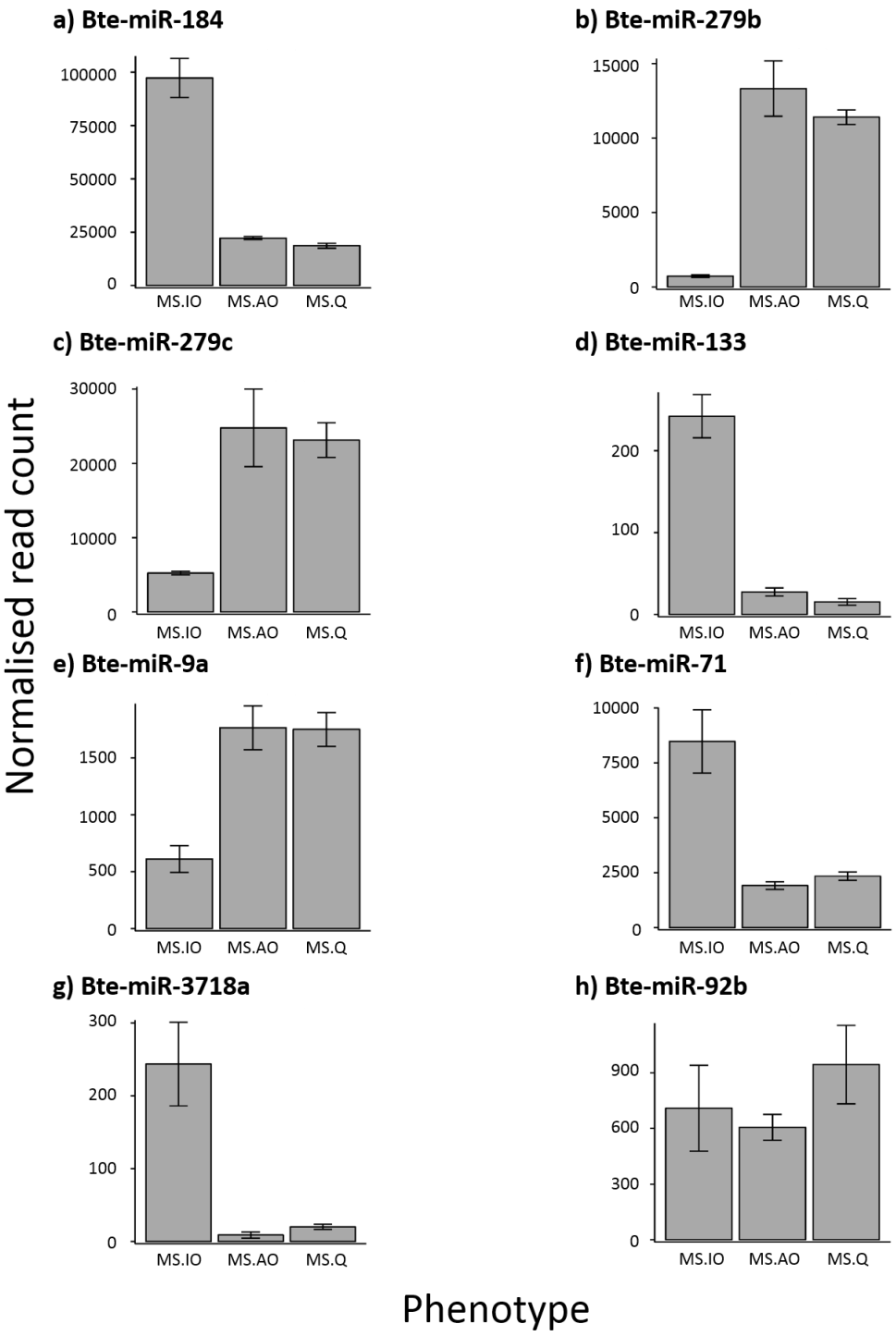


**Figure 4.5: Normalised read counts (reads per million) in the ovaries of adult females of *Bombus terrestris* for (a-g) seven differentially expressed miRNAs and (h) one miRNA that was associated with sociality in a literature search (Greenberg et al. 2012). Error bars represent the range of each read count across all three replicate libraries for each phenotype (excluding libraries MS.IO.br.1 and MS.AO.br.3).MS.IA= inactive-ovary worker, MS.AO = active-ovary worker, MS.Q= queen.**

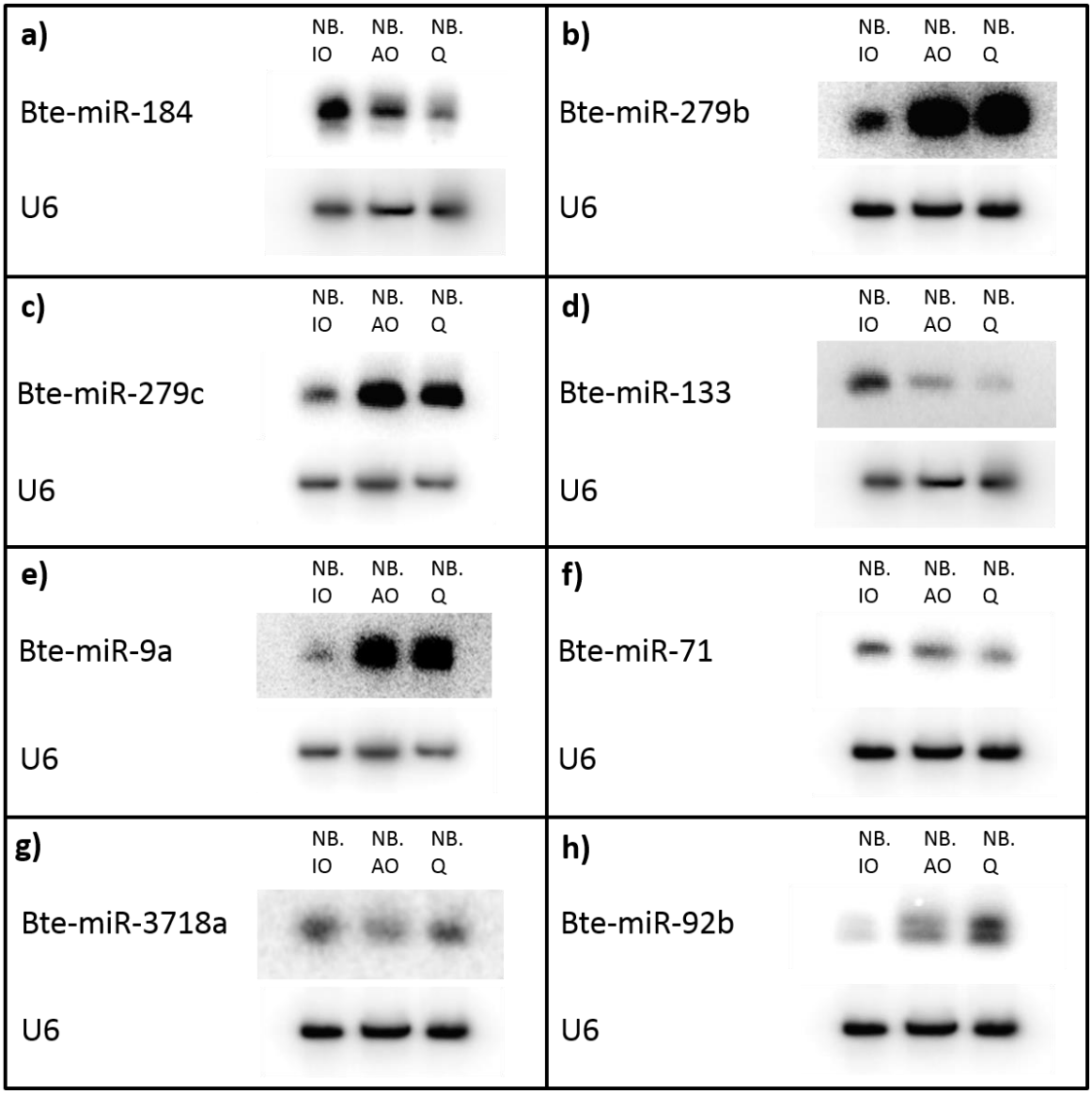


**Figure 4.6: Northern blot expression profiles for seven miRNAs (a-g) that were classified as differentially expressed in ovaries from inactive-ovary workers, active-ovary workers, and queens in *Bombus terrestris* adult females according to miRNA-seq, and one miRNA (h) that was associated with sociality in a literature search (Greenberg et al. 2012). The U6 panel (control) demonstrates equal loading for each sample. MS.IO, inactive-ovary worker; MS.AO, active-ovary worker; MS.Q, Queen.**

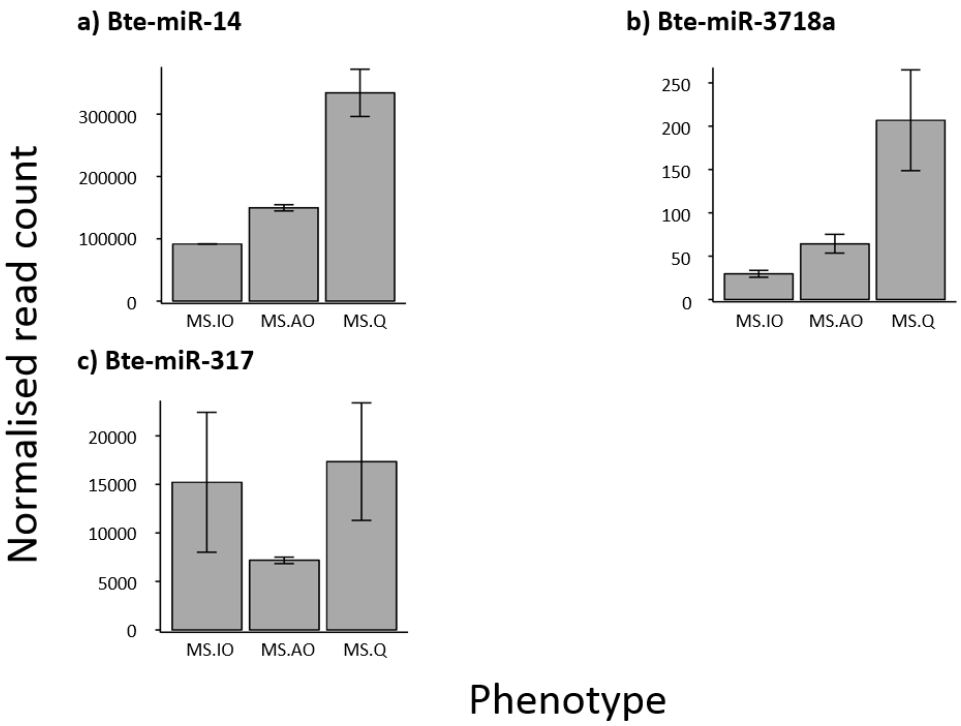


**Figure 4.7: Normalised read counts (reads per million) for three differentially expressed miRNAs in brains from inactive-ovary workers, active-ovary workers and queens in *Bombus terrestris* adult females. Error bars represent the range of each read count across all three replicate libraries for each phenotype (Excluding MS.IO.br1 and MS.AO.br3)**. **MS.IA= inactive-ovary worker, MS.AO = active-ovary worker, MS.Q= queen.**

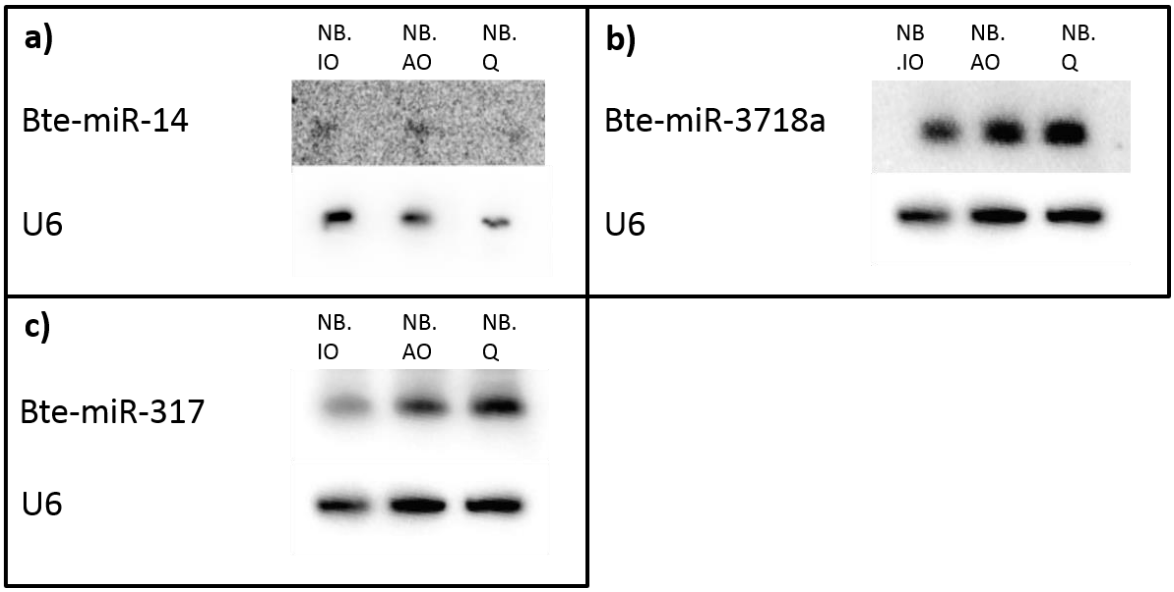


**Figure 4.8: Northern blot expression profiles for three miRNAs that were classified as differentially expressed in brains from inactive-ovary workers, active-ovary workers, and queens in *Bombus terrestris* adult females. The U6 panel (control) demonstrates equal loading for each sample. NB.IO= inactive-ovary worker, NB.AO= active-ovary worker, NB.Q= Queens.**

**Table 4.1: Details of samples of *Bombus terrestris* adult workers used for miRNA-seq of libraries and Northern blot analysis. Samples were divided into age-groups two and three, age-group 2 was used to create Northern blot samples, age-group 3 was used to create miRNA-seq libraries. Columns of numerical data represent: total number of inactive-ovary workers (IO) from age-group two and age-group three, total number of active-ovary workers from age-group two and age-group three (AO). The two far-right columns show the miRNA-seq libraries and Northern blot samples to which each colony contributed. In addition to workers, one queen was collected from each colony, note that no workers were collected from AC.8-16 as these were used as a source of queens only (see text).**

| Colony | Number of IO (group 2) | Number of IO (group 3) | Number of AO (group 2) | Number of AO (group 3) | Sequencing samples | Northern blot samples |
|---|---|---|---|---|---|---|
| AC.1 | 0 | 1 | 17 | 14 | MS.AO.ov.1, MS.AO.br.1, MS.Q.ov.1, MS.Q.br.1 | NB.AO.ov.1, NB.AO.br.1 |
| AC.2 | 0 | 35 | 21 | 13 | MS.IO.ov.2, MS.IO.br.2, MS.AO.ov.2, MS.AO.br.2, MS.Q.ov.1, MS.Q.br.1 | NB.AO.ov.2, NB.AO.br.2 |
| AC.3 | 6 | 27 | 17 | 2 | MS.IO.ov.1, MS.IO.br.1 | NB.IO.ov.1, NB.IO.br.1, NB.AO.ov.3, NB.AO.br.3 |
| AC.4 | 1 | 28 | 13 | 16 | MS.AO.ov.3, MS.IO.br.3, MS.AO.ov.3, MS.AO.br.3, MS.Q.ov.1, MS.Q.br.1 | NB.AO.ov.4 |
| AC.5 | 4 | 0 | 10 | 0 | MS.Q.ov.2, MS.Q.br.2 | NB.IO.ov.2, NB.IO.br.2, NB.AO.ov.5, NB.AO.br.4 |
| AC.6 | 5 | 0 | 3 | 0 | MS.Q.ov.2, MS.Q.br.2 | NB.IO.ov.3, NB.IO.br.3 |
| AC.7 | 3 | 0 | 2 | 0 | MS.Q.ov.2, MS.Q.br.2 | NB.IO.ov.4, NB.IO.br.4 |
| AC.8 | N/A | N/A | N/A | N/A | MS.Q.ov.2, MS.Q.br.2 | None |
| AC.9 | N/A | N/A | N/A | N/A | MS.Q.ov.3, MS.Q.br.3 | None |
| AC.10 | N/A | N/A | N/A | N/A | MS.Q.ov.3, MS.Q.br.3 | None |
| AC.11 | N/A | N/A | N/A | N/A | MS.Q.ov.3, MS.Q.br.3 | None |
| AC.12 | N/A | N/A | N/A | N/A | MS.Q.ov.3, MS.Q.br.3 | None |
| AC.13 | N/A | N/A | N/A | N/A | MS.Q.ov.3, MS.Q.br.3 | None |
| AC.14 | N/A | N/A | N/A | N/A | None | NB.Q.ov.1, NB.Q.br.1 |
| AC.15 | N/A | N/A | N/A | N/A | None | NB.Q.ov.1, NB.Q.br.1 |
| AC.16 | N/A | N/A | N/A | N/A | None | NB.Q.ov.1, NB.Q.br.1 |

**Table 4.2: Details of libraries prepared from *Bombus terrestris* inactive-ovary workers, active-ovary workers and queen adult females and used for miRNA-seq (see also Table 4.1). Columns show: the phenotype and tissue used to produce each library; the colony that each library was used to construct each library; the number of individuals pooled to make each library; the index identifier sequence of each library; the read count (total number of reads) for each library; and the percentage of total reads mapped to the *B. terrestris* genome. Indexes were re-used between ovary samples and brain samples.**

| Library name | Phenotype (tissue) | Source Colony (s) | Number of individuals | Illumina Index sequence | Total number of reads | Percent mapped |
|---|---|---|---|---|---|---|
| MS.IO.ov.1 | Inactive-ovary worker (ovary) | AC.3 | 27 | 2 | 13015233 | 71.7 |
| MS.IO.ov.2 | Inactive-ovary worker (ovary) | AC.2 | 35 | 4 | 8612909 | 45.7 |
| MS.IO.ov.3 | Inactive-ovary worker (ovary) | AC.4 | 28 | 5 | 12273898 | 70.6 |
| MS.AO.ov.1 | Active-ovary worker (ovary) | AC.1 | 14 | 6 | 14683990 | 74.8 |
| MS.AO.ov.2 | Active-ovary worker (ovary) | AC.2 | 13 | 8 | 17655625 | 71.5 |
| MS.AO.ov.3 | Active-ovary worker (ovary) | AC.4 | 16 | 9 | 13040341 | 74.4 |
| MS.Q.ov.1 | Queen (ovary) | AC.1, 2, 4 | 3 | 10 | 17310301 | 68.4 |
| MS.Q.ov.2 | Queen (ovary) | AC.5, 6, 7, 8 | 4 | 12 | 19886641 | 71.9 |
| MS.Q.ov.3 | Queen (ovary) | AC.9, 10, 11, 12, 13 | 5 | 11 | 11259681 | 75.0 |
| MS.IO.br.1 | Inactive-ovary worker (brain) | AC.3 | 27 | 1 | 27850234 | 62.7 |
| MS.IO.br.2 | Inactive-ovary worker (brain) | AC.2 | 35 | 3 | 13929475 | 59.5 |
| MS.IO.br.3 | Inactive-ovary worker (brain) | AC.4 | 28 | 4 | 14197055 | 65.2 |
| MS.AO.br.1 | Active-ovary worker (brain) | AC.1 | 14 | 5 | 9682659 | 62.9 |
| MS.AO.br.2 | Active-ovary worker (brain) | AC.2 | 13 | 7 | 12486030 | 59.8 |
| MS.AO.br.3 | Active-ovary worker (brain) | AC.4 | 16 | 8 | 347722 | 60.6 |
| MS.Q.br.1 | Queen (brain) | AC.1, 2, 4 | 3 | 9 | 15614722 | 63.8 |
| MS.Q.br.2 | Queen (brain) | AC.5, 6, 7, 8 | 4 | 11 | 14809035 | 51.6 |
| MS.Q.br.3 | Queen (brain) | AC.9, 10, 11, 12, 13 | 5 | 12 | 11329671 | 59.1 |

**Table 4.3: Details of samples prepared from the ovaries and brains in inactive-ovary workers, active-ovary workers and queens of *Bombus terrestris* adults, and used for Northern blots (see also Table 4.1).**

| Northern blot sample name | Phenotype (tissue) | Source Colony | Number of individuals |
|---|---|---|---|
| NB.IO.ov.1 | Inactive-ovary worker (ovary) | AC.3 | 6 |
| NB.IO.ov.2 | Inactive-ovary worker (ovary) | AC.5 | 3 |
| NB.IO.ov.3 | Inactive-ovary worker (ovary) | AC.6 | 5 |
| NB.IO.ov.4 | Inactive-ovary worker (ovary) | AC.7 | 3 |
| NB.AO.ov.1 | Active-ovary worker (ovary) | AC.1 | 17 |
| NB.AO.ov.2 | Active-ovary worker (ovary) | AC.2 | 21 |
| NB.AO.ov.3 | Active-ovary worker (ovary) | AC.3 | 17 |
| NB.AO.ov.4 | Active-ovary worker (ovary) | AC.4 | 13 |
| NB.AO.ov.5 | Active-ovary worker (ovary) | AC.5 | 10 |
| NB.Q.ov.1 | Queen (ovary) | AC.14,15,16 | 3 |
| NB.IO.br.1 | Inactive-ovary worker (brain) | AC.3 | 6 |
| NB.IO.br.2 | Inactive-ovary worker (brain) | AC.5 | 3 |
| NB.IO.br.3 | Inactive-ovary worker (brain) | AC.6 | 5 |
| NB.IO.br.4 | Inactive-ovary worker (brain) | AC.7 | 3 |
| NB.AO.br.1 | Active-ovary worker (brain) | AC.1 | 17 |
| NB.AO.br.2 | Active-ovary worker (brain) | AC.2 | 21 |
| NB.AO.br.3 | Active-ovary worker (brain) | AC.3 | 17 |
| NB.AO.br.4 | Active-ovary worker (brain) | AC.5 | 10 |
| NB.Q.br.1 | Queen (brain) | AC.14,15,16 | 3 |

**Table 4.4: MiRNAs from the ovaries and brains in inactive-ovary workers, active-ovary workers and queens in *Bombus terrestris* adult females probed using Northern blots, their mature sequences, the corresponding probe sequence, a literature summary, along with corresponding sequences for the U6 control.**

| miRNA name | Mature miRNA sequence | Probe sequence | Literature description |
|---|---|---|---|
| miR-184 | UGGACGGAGAACUGAUAAGG | CCTTATCAGTTCTCCGTCCA | More highly expressed in foragers versus nursers during *A. mellifera* polyethism (Behura and Whitfield 2010; Greenberg et al. 2012); causes worker traits in *A. mellifera larvae*; reproductive development in Drosophila (Iovino et al. 2009). |
| miR-279b | UGACUAGAUCGAAAUACUCGUCCU | AGGACGAGTATTTCGATCTAGTCA | More highly expressed in nursers versus foragers during *A. mellifera* polyethism (Liu et al. 2012); ovary maturation and circadian rhythmicity in *Drosophila* (Yoon et al. 2011; Luo and Sehgal 2012). |
| miR-279c | UGACUAGAGUCACACUCGUCCA | TGGACGAGTGTGACTCTAGTCA | More highly expressed in nursers versus foragers during *A. mellifera* polyethism (Liu et al. 2012); more highly conserved in social Hymenoptera compared to solitary Hymenoptera (Greenberg et al. 2012); ovary maturation and circadian rhythmicity in *Drosophila*(Yoon et al. 2011; Luo and Sehgal 2012). |
| miR-133 | UUGGUCCCCUUCAACCAGCUGU | ACAGCTGGTTGAAGGGGACCAA | More highly expressed in foragers versus nursers during *A. mellifera* polyethism (Liu et al. 2012). |
| miR-9a | UCUUUGGUUAUCUAGCUGUAUGA | TCATACAGCTAGATAACCAAAGA | More highly expressed in adult workers compared to queens in *A. mellifera* (Weaver et al. 2007). |
| miR-71 | UGAAAGACAUGGGUAGUGAGAUG | CATCTCACTACCCATGTCTTTCA | More highly expressed in adult workers compared to queens in *A. mellifera* (Weaver et al. 2007). |
| miR-3718a | UCCCCUGUCCUGUCCCGAUAGU | ACTATCGGGACAGGACAGGGGA | Taxonomically restricted to *A. mellifera* (Chen et al. 2010) and *Bombus* species (*Chapter 2*). |
| miR-92b | AAUUGCACCCGUCCCGGCCUGA | TCAGGCCGGGACGGGTGCAATT | More highly conserved in social Hymenoptera compared to solitary Hymenoptera (Greenberg et al. 2012). |
| miR-14 | GGGGGUGAGAAACUGGCUUGGCU | AGCCAAGCCAGTTTCTCACCCCC | Involved in neurophysiology in *Drosophila* through regulation of insulin secretion (Varghese et al. 2010). |
| miR-317 | UGAACACAGCUGGUGGUAUCUCAGU | ACTGAGATACCACCAGCTGTGTTCA | Involved in behaviour in *Drosophila* where mutations of the miRNA have epistatic effects on startle-induced locomotion (Yamamoto et al. 2008). |
| miR-6001-5p | GUAGGUAACGACUGAUGGGAACA | TGTTCCCATCAGTCGTTACCTAC | More highly expressed in queen-destined larvae compared to worker-destined larvae in *B. terrestris* (**Chapter 3**). |
| U6 | UGAGGUAGUAGGUUGUAUAGU | ACTATACAACCTACTACCTCA | Widely used as a small RNA loading control (Lopez-Gomollon 2011). |

**Table 4.5: New miRNAs in *Bombus terrestris* adult females Columns include the predicted mature sequence, genomic co-ordinates, and tissue that each miRNA was sequenced in. MA# includes *miRCat* predictions in adult *B. terrestris*. MC# includes miRNAs that were previously predicted by *miRCat* from miRNA-seq data in *B. terrestris* larvae (Chapter 2) that are also expressed in *B. terrestris* adults. MiRNAs highlighted in bold are conserved between *Bombus* and *Apis*.**

| Name | Mature sequence | Scaffold (strand) | Start co-ordinates | End co-ordinates | Tissue |
|---|---|---|---|---|---|
| MA1051 | AUCUGUAUGAUUCCACGUUUUG | 14.8 (+) | 1434315 | 1434412 | Brains |
| **MA1199** | **UCACCGGGUUGAAUUCAUCCAG** | **15.6 (-)** | **5457818** | **5457742** | **Ovaries** |
| MA1204 | AAGCGAUUGCGGGUUUUGUCAU | 15.6 (+) | 5628617 | 5628685 | Ovaries |
| MA1388 | AAGCGAUUGCGGGUUUUGUCA | 17.2 (+) | 2243078 | 2243152 | Ovaries |
| MA1408 | AUCUACAACAGCUCUGUCACCA | 17.2 (-) | 2706398 | 2706309 | Ovaries |
| MA1495 | ACAAGAUCCGUAAUUUAGUUGC | 18.1 (+) | 2547307 | 2547384 | Ovaries |
| MA1532 | UUGAUCCCUAGUUUAUCCGAAU | 2.1 (+) | 1244746 | 1244814 | Ovaries |
| MA1875 | AGCCCUUCAUCCUUGUAGCACG | 3.3 (+) | 727635 | 727722 | Brains |
| MA2776 | UAUCAGUGUAGCCAGAUAUCCAU | 9.3 (+) | 1380891 | 1380952 | Ovaries |
| MA2830 | CAAAGAAUCACUAAUCAUCCUA | 9.4 (+) | 221398 | 221476 | Ovaries |
| MA2944 | UGACCCAUCGCCGCCCCUCGGG | Un1082 (-) | 1005 | 923 | Brains |
| MA3051 | CUGAUUGUCGCCUCGAGCCCGA | Un1950 (-) | 735 | 642 | Brains |
| MA3157 | CUAACCUUUCAACCUUGGCGCU | Un3897 (-) | 1488 | 1394 | Ovaries |
| MA3224 | ACGAGAAGGCGAACACGAGGG | Un513 (+) | 20077 | 20169 | Brains |
| MA3271 | UCAGGUUGUAGUCCUCCUUGG | Un59 (-) | 43369 | 43296 | Ovaries |
| MA3290-3p | UGUCGGUAGCAAAGAGGUGGAAG | Un633 (+) | 1439407 | 1439485 | Brains |
| MA3290-5p | CCCACUCUUUGACUACCCGACA | Un633 (+) | 1439409 | 1439484 | Brains |
| **MA341** | **AAUUAAGUGGUUGCGGAUUUUG** | **10.1** (+) | **8133504** | **8133580** | **Ovaries** |
| MC1030 | AACUCCGUAGUCUCUUAGUUGA | 18.1 (+) | 2545808 | 2545882 | Ovaries |
| MC24 | UGUGGGGCGGCGUCCGGGUCACU | 9.1 (+) | 889091 | 889165 | Ovaries /Brains |
| MC485 | UGUAGGUACUAGUCGUCUCUAUA | 11.5 (+) | 1346061 | 1346125 | Ovaries |
| MC712 | UAGCGCUUUAGGUUUCGAACGU | 1.4 (-) | 1146 | 1087 | Ovaries /Brains |
| MC753-3p | UAUGUUUUGUAGGGCCUUGCGU | Un1079 (-) | 2121 | 2044 | Ovaries /Brains |
| MC753-5p | ACAAGGCCCUACAAAAUAUAGUA | Un1079 (-) | 2120 | 2045 | Ovaries /Brains |

**Table 4.6: Ten most highly expressed miRNAs from miRNA-seq in ovaries and brains in *Bombus terrestris* adult females, averaged between all libraries.**

| Tissue | Name | Sequence | Average read count |
|---|---|---|---|
| Ovaries | MC1030 | AACUCCGUAGUCUCUUAGUUGA | 54122.35 |
| | Bte-miR-184 | UGGACGGAGAACUGAUAAGGG | 46061.93 |
| | Bte-miR-275 | UCAGGUACCUGAAGUAGCGCG | 38184 |
| | MA1199 | UCACCGGGUUGAAUUCAUCCAG | 38177.04 |
| | Bte-miR-279a | UGACUAGAUCCACACUCAUUA | 33824.35 |
| | Bte-miR-137 | UAUUGCUUGAGAAUACACGUAG | 28301.84 |
| | MC762 | UGACUAGAUCCACACUCAUCCA | 26202.13 |
| | Bte-miR-2944 | UAUCACAGCAGUAGUUACCUAGG | 20637.69 |
| | Bte-miR-279c | UGACUAGAGUCACACUCGUCCA | 17716.7 |
| | Bte-miR-1 | UGGAAUGUAAAGAAGUAUGGAG | 16282.43 |
| Brain | Bte-miR-184 | UGGACGGAGAACUGAUAAGGG | 369837.55 |
| | Bte-miR-2796 | GUAGGCCGGCGGAAACUACUUGC | 313120.18 |
| | Bte-miR-14 | GGGGGUGAGAAACUGGCUUGG | 192019.8 |
| | Bte-miR-252 | AUAAGUACUAGUGCCGCAGGA | 148140.24 |
| | Bte-miR-927 | UUAGAAUUCCUACGCUUUACCG | 55870.34 |
| | Bte-miR-275 | UCAGGUACCUGAAGUAGCGCG | 50170.015 |
| | Bte-miR-252a | AUAAGUACUAGUGCCGCAGGA | 39554.23 |
| | Bte-miR-14 | UCAGUCUUUUUCUCUCUCCUAU | 20062.54 |
| | Bte-miR-276a | UAGGAACUUCAUACCGUGCUCU | 15567.65 |
| | Bte-miR-10a | UACCCUGUAGAUCCGAAUUUGU | 14228.32 |

Chapter 5

# A test for experimental knockdown of gene expression in *Bombus terrestris* and the relationship between *foraging* gene expression and queen activity

*Abstract*

One way to induce gene expression changes in a species of interest is to 'knock down' (i.e. prevent the expression of) the target gene using double-stranded RNA (dsRNA). In the bumble bee *Bombus terrestris*, there is no established protocol for knocking down gene expression. Therefore, the aim of this study was to: 1) induce knockdown of a target gene, *foraging kinase* (*for*); and 2) elucidate the effects of the gene on foraging behaviour and circadian rhythmicity in *B. terrestris* queens. We did this by feeding dsRNA targeting *for* to *B. terrestris* queens and then using tracking software to measure their overall activity levels and so determine whether they exhibited circadian rhythmicity. We also verified the effects of the knockdown treatment on *for* expression levels using qRT-PCR. Overall we found no effect of the dsRNA treatment on *for* expression levels and we found no evidence that *for* is related to foraging activity. However, the reproductive status of queens was correlated with their levels of circadian activity and egg-laying queens had significantly higher levels of *for* than non-laying queens. These results show that there is a complex relationship between the stage of a queen in her life history, expression levels of *for*, queens' activity levels, and whether or not queens exhibit circadian rhythmicity.

## *5.1 Introduction*

### *The importance and applications of RNA interference*

RNA interference (RNAi) is the process where small RNAs (sRNA) bind to and induce silencing of the messenger RNA (mRNA) transcript of a target gene, ultimately causing a reduction in gene expression where protein-coding genes produce less protein because they can no longer be synthesised from the silenced transcripts (Fire et al. 1998; Wilson and Doudna 2013). This pathway is used to control and fine-tune gene expression during an organism's development and physiology. There are numerous types of regulatory RNA that target gene expression in this way: they include small interfering RNAs (siRNA), microRNAs (miRNA), piwi-interacting RNAs (piRNA), and several others (Carthew and Sontheimer 2009). The long forms of these genes are recognised by an RNase III-type enzyme called *Dicer* that binds to and induces cleavage in double-stranded RNA (dsRNA) and in hairpin-shaped loops of single-stranded RNA (ssRNA) (Bartel 2004). This leaves a 21-25 bp dsRNA molecule with a 2 bp overhang at the 3' end. These short sequences are the mature forms of siRNAs and miRNAs and can control gene expression by associating with a large complex of proteins known as the *RNA-induced silencing com*plex (RISC). This complex of proteins contains *argonaute* enzymes, which have catalytic activity (Carthew and Sontheimer 2009). Once bound to the RNAi molecule, the strands dissociate from each other. One of the two strands is less thermodynamically stable at the 5' end and is often termed the 'pilot strand' because it remains in the complex and guides *RISC* to a complementary target mRNA transcript, while the other strand (termed the 'passenger strand') is degraded (Carthew and Sontheimer 2009).

After the pilot strand has guided *RISC* to its target, the transcript becomes less accessible to the ribosomes and its translation into protein is greatly reduced. Effectively the transcript has been silenced. The mechanisms of RNAi were first established in plants (Ratcliff et al. 1997), and research into the individual components of RNAi has been ongoing. It has included identifying: 1) the most important proteins in the pathway (Sen and Blau 2006) and 2) the sequences of the crucially important regulatory RNAs (e.g. new miRNA sequences are recorded in www.miRBase.com). Since the discovery of RNAi, researchers have found that this process can be manipulated to experimentally downregulate expression of a target gene *in vivo*, in a process termed 'knockdown'. This

can be done at any time during a target organism's life cycle. RNAi is a potentially powerful tool for understanding the direct function of a gene, and the consequences of an immediate reduction in the gene's expression levels (i.e. by preventing the translation of mRNA into protein). Hence the technique has been routinely applied to understand the function of particular genes and gene pathways in model organisms such as *Caenorhabditis elegans* (Kamath and Ahringer 2003), and *Drosophila melanogaster* (Misquitta and Paterson 1999; Boutros et al. 2004). These methods have also established tools for researchers to carry out whole-genome functional screens in a variety of research applications ranging from drug discovery (Hong-Geller and Micheva-Viteva 2010), through identifying cancerous genes and tumour suppressors (Westbrook et al. 2005), to discovering new genetic pathways (Ashrafi et al. 2003; Berns et al. 2004; Zhang et al. 2009).

*RNA knockdown in eusocial insects*

Eusocial insects are species with a reproductive division of labour (Wilson 1971). In most cases they have a queen caste that specialises in reproduction, and a worker caste that specialises in foraging for food and nursing developing brood (Wilson 1971). In recent years RNAi mechanisms in eusocial insects have been studied and found to directly influence social development (Guo et al. 2013). For example, miRNAs have been studied in the social processes in the honey bee *Apis mellifera*, and several miRNAs have been correlated with caste differentiation (Weaver et al. 2007; Guo et al. 2013), and worker task specialisation and aging (Behura and Whitfield 2010; Greenberg et al. 2012; Liu et al. 2012a). Of particular interest, one study has shown that *A. mellifera* workers possibly feed miRNAs to developing worker-destined larvae in order to prevent them from developing into queens. One miRNA in particular (miR-184) was very highly expressed in worker gland secretions, and when artificially fed to queen-destined larvae it caused them to develop worker-like traits (Guo et al. 2013).

In previous chapters we showed that, in the bumble bee *Bombus terrestris*, miRNAs are associated with caste differentiation in larvae (**chapter 3**) and with the reproductive division of labour in adults (**chapter 4**). To further investigate the causative effects of miRNAs and other social genes (e.g. *foraging kinase*; Kodaira et al. 2009; Tobback et al. 2011)) on caste differentiation processes and other social processes in *B. terrestris*, RNAi

would be a very powerful tool. To date, there is no well-established protocol for inducing knockdown in the miRNAs of any eusocial insect; however, there has been some success using RNAi to target protein-coding genes in *A. mellifera* (Jarosch and Moritz 2011; Kamakura 2011), the termite *Reticulitermes flavipes* (Zhou et al. 2008), in the larvae of the social wasp *Polistes metricus* (Hunt et al. 2011), and recently in adult *B. terrestris* (Deshwal and Mallon 2014). In these protocols a 400-600bp length of dsRNA that was complementary to the gene of interest was delivered *in vivo* to the target species. This was then recognised by the *Dicer* enzyme which cleaves the gene into several lengths of 21-23bp dsRNA (Terenius et al. 2011; Scott et al. 2013). These were then incorporated into the RNAi pathway, where they prevented translation of the complementary target transcripts (Zamore et al. 2000).

Delivery of the dsRNA into the organism of interest can be done in several ways, with the two most common being by direct injection (e.g. Misquitta et al. 2008; Belles 2010) or by ingestion where the dsRNA is delivered orally or supplied in the organism's food (Scott et al. 2013; Singh et al. 2013). Oral delivery of dsRNA is the easiest of the two methods and the least traumatic to the target organism (Yu et al. 2013). Previous studies have achieved some level of success using dsRNA oral delivery methods in insects, including the Triatomine bug *Rhodnius prolixus* (Araujo et al. 2006), *R. flavipes* (Zhou et al. 2008), *A. mellifera* (Jarosch and Moritz 2011; Kamakura 2011), brown plant hopper *Nilaparvata lugens* (Li et al. 2011), eusocial wasp *P. metricus* (Hunt et al. 2011), Western corn rootworm *Diabrotica virgifera* (Bolognesi et al. 2012) , yellow fever mosquito *Aedes aegypti* (Singh et al. 2013), and numerous others (Reviewed by Terenius et al. 2011; Scott et al. 2013; Yu et al. 2013). Oral delivery of dsRNA is most likely to downregulate gene expression if the genes being targeted are mostly expressed in the gut epithelium, because these are the tissues that the dsRNA molecules first come into contact with. However, this delivery method has been shown to have systemic effects on gene expression (i.e. dsRNA is transmitted widely and targets tissues throughout the whole organism), even in tissues such as the brain (Singh et al. 2013).

## Foraging kinase *and queen foraging behaviour*

Bumble bees are annual eusocial insects that thrive in temperate climates with seasonal variation (Williams 1998). In these species, the mated queen comes out of hibernation at the beginning of the season and searches for a suitable nest site (Goulson 2003; 2013). Once she has found one, she lays an initial clutch of eggs and, upon the eggs hatching, she provisions the larvae by periodically leaving the nest to forage externally at flowers for pollen and nectar. Eventually the larvae pupate and eclose (emerge from pupation) as the first brood of workers, thus forming a colony. The queen is an extremely valuable resource to a young bumble bee colony because only she can produce new workers and reproductive females (Goulson 2003). Therefore, even though the queen rears the first batch of brood by herself, once workers eclose they begin to take over foraging duties, allowing the queen to concentrate on reproduction (Oster and Wilson 1978). Upon the eclosion of the first workers, the queen undergoes a behavioural change and no longer leaves the colony to forage. The proximate factors that cause the shift from the queen foraging to the queen remaining in the nest could be time dependent, since other changes during the colony cycle have been shown to be a function of time. An example is the 'switch point', which occurs when queens change from laying purely diploid eggs to laying purely haploid eggs (e.g. Duchateau and Velthuis 1988; Holland et al. 2013). Another possibility is that the timing of queen cessation of foraging is influenced by social factors. Evidence for this comes from the finding that the queen's foraging behaviour (but not the workers') can be experimentally manipulated, whereby adding brood and/or workers causes queens to have attenuated circadian rhythms, while removing brood causes them to have strong circadian rhythmicity (Eban-Rothschild et al. 2011).

Foraging behaviour in bees has been linked to circadian rhythmicity, the process whereby animals exhibit an endogenous oscillation in behavioural state on a day-length basis (Giebultowicz 2000). In *A. mellifera* workers, foraging specialists have circadian activity, where they still display locomotor behaviour in regularly cycles, even if they are kept in complete darkness (Bloch et al. 2001). Meanwhile, nursing specialists are completely arrhythmic, even when they are kept under day-night cycles (Bloch et al. 2013). Similar patterns to these are found in the workers of harvester ants (Ingram et al. 2005). In *B. terrestris*, the queen does not exhibit circadian rhythmicity while in hibernation (because

she is almost completely inactive), but she exhibits strong circadian rhythmicity upon emergence from hibernation during her foraging phase early in the colony cycle (Eban-Rothschild et al. 2011). She then becomes arrhythmic again when she has developing brood, soon before the workers take over her foraging duties (Eban-Rothschild et al. 2011).

The behavioural features of the queen's cessation of foraging behaviour are reasonably well established but far less is known about the molecular mechanisms involved. We hypothesized that the gene *foraging kinase* (*for*) plays a role in this behavioural change. *For* codes for a cGMP-dependent protein kinase (PKG), which is a serine/threonine-specific protein kinase activated by cGMP (Kalderon and Rubin 1989; Osborne et al. 1997). *For* has been shown to influence numerous forms of foraging behaviour. For example, in *Drosophila*, *for* exists as a naturally occurring polymorphism with two distinct phenotypes (rover and sitter) that manifest as different behavioural strategies during the larval and adult stages (Debelle and Sokolowski 1987; Pereira and Sokolowski 1993; Osborne et al. 1997). In these species, during feeding, rover phenotypes are more active and venture further than sitter phenotypes, which remain closer to their starting location. In both *A. mellifera* and *B. terrestris* workers, *for* has been shown to be positively associated with foraging activity, being more highly expressed in foraging specialists compared to nursing specialists (Ben-Shahar et al. 2002; Tobback et al. 2011; Rodriguez-Zas et al. 2012). There has been some suggestion of a causal link between foraging activity in *A. mellifera* workers and *for* gene expression because artificially increasing PKG levels by feeding workers an analogue of the gene caused workers to shift to a foraging phenotype (Ben-Shahar et al. 2002). However, the relationship between foraging behaviour and *for* expression is the opposite in two ant species and the wasp *Vespula vulgaris*, where *for* is more highly expressed in brood care specialists (Ingram et al. 2005; Tobback et al. 2008; Lucas and Sokolowski 2009). This was also the case in workers of *Bombus ignites*, but this result may have been confounded by the ages of the bees studied (Kodaira et al. 2009). It is known that *for* is age-related in *B. terrestris* because overall levels of the gene tend to decline with age in *B. terrestris* workers (Tobback et al. 2011). Finally, *for* has been shown to vary between *B. terrestris* queens as a function of social context, in that queens kept without social contact with adult workers had higher expression levels of *for* than control queens (Edd Almond, personal communication). Therefore, in *Bombus*, *for* has been implicated in

queen behaviour in general, and in foraging behaviour in particular. However, so far no study has established whether there is a link between the queen cessation of foraging and the gene *for*.

*Aims*

In the present study, our first aim was to test whether we could induce gene knockdown in *B. terrestris* by feeding queens with dsRNA targeting a behaviourally-linked gene. We chose to test this by knocking down the gene *for,* a gene influencing foraging behaviour in *Drosophila* (Debelle and Sokolowski 1987; Kaun et al. 2007) and eusocial Hymenoptera such as the honey bee (Ben-Shahar et al. 2002; 2003), bumble bees (Kodaira et al. 2009; Tobback et al. 2011), wasps (Tobback et al. 2008), and ants (Ingram et al. 2005; Lucas and Sokolowski 2009). Our second aim was to test the specific hypothesis that *for* has a causal role in regulating foraging behaviour and circadian rhythmicity in *B. terrestris* queens. Specifically whether the gene is downregulated in queens during the early part of the colony cycle when queens make the behavioural transition from external foraging to remaining in the nest while workers take over all foraging. This would establish whether the gene is involved in the queen cessation of foraging during the early phases of colony growth, and establish whether *for* is directly involved in regulating foraging behaviour in eusocial insects, in contrast to merely being correlated with it.

## *5.2 Methods*

*dsRNA production*

We performed two dsRNA treatments, one (experimental treatment) in which the dsRNA had a sequence complementary to a section of the target gene (*for*), and the other (control) in which the dsRNA was complementary to green fluorescent protein (*gfp*). Since bumble bees do not possess *gfp*, the second treatment acted as a nonsense sRNA control to measure the effect on queens of the dsRNA treatment itself (rather than targeted inhibition by *for* dsRNA). For both sequences we developed dsRNA by first identifying the sequences and structures of the two genes on the *National Centre for Biotechnology Information* (NCBI) website (http://www.ncbi.nlm.nih.gov/). We then extracted DNA from male *B.*

*terrestris* larvae using phenol-chloroform extraction, and PCR-amplified the target regions. For each gene we designed four primers, including a pair of primers that were complementary to the region of interest and a pair with nearly identical sequences to which a T7 promoter sequence was attached on each 5' end (**Table 5.1**). Both sets of primers produced a product length of approximately 400 bp, which was within the optimum range for RNAi (Scott et al. 2013). In addition, the resulting products had little complementarity (< 15 bp) to other sections of the genome, thereby minimising the chance of off-target effects on the treatment bee's gene expression (Scott et al. 2013). For the *for* gene, the primers encompassed a 400 bp segment of *for* exon 7, which was the largest exon in the gene in *B. terrestris*. The *gfp* primers amplified a 400 bp segment of the gene from the plasmid pEGFP-N1 (Clontech, Saint-Germain-en-Laye, France). For each gene we performed two PCR reactions, pairing the T7 primers with the primers that did not contain T7; therefore the T7 forward primer was paired with the non-T7 reverse primer and vice versa (**Table 5.1**). This produced two strands of DNA for each gene, one with the T7 promoter at the 5' end and one with the T7 promoter at the 3' end. We recovered the products from these four PCR reactions by separating them on a 1.2% agarose gel. We then cut the PCR products of these reactions from the gel and purified them using a Zymoclean Gel DNA Recovery kit (Zymo Research, Irvine, USA) according to the manufacturer's instructions.

We replicated the extracted strands by molecular cloning, ligating the T7-containing products into a pGEMTeasy vector (Promega, Southampton, UK). We transformed the vector containing the products into DH5-α transformation competent *Escherichia coli* cultures. We then used blue-white colony staining to pick out the transformed *E. coli* colonies, and extracted the plasmids using a Qiagen miniprep kit according to the manufacturer's instructions (Qiagen, Manchester, UK). We sent the extracted plasmids to Eurofins genomics (Ebersberg, Germany) for Sanger chain termination sequencing. The sequencing confirmed that each plasmid contained the correct sequence, with no nucleotide substitutions and the T7 end still intact on either the 5' end or 3' end of each sequence. We then converted the DNA to ssRNA with T7-dependent RNA polymerase using the MEGAscript RNAi kit (Invitrogen, Carlsbad, USA) according to the manufacturer's instructions for generating dsRNA from DNA plasmids. We annealed the complementary

ssRNA strands for each product by heating them at 75°C and then cooled them to room temperature. This step produced the two dsRNA treatments, one targeting *gfp* and the other targeting *for*. We calculated the concentration of the dsRNA on a nanodrop 8000 spectrophotometer (Thermo Scientific, Wilmington, Delaware, USA). Overall we produced approximately 1000 µg of *for* dsRNA from 10 kit reactions, and 800 µg of *gfp* dsRNA from 10 kit reactions. We visualised the band integrity by separating a sample of the products on a 1.2% agarose gel to ensure that the RNA was still intact. We found no evidence that the RNA had degraded in any of the reactions (data not shown). We stored the remaining dsRNA at -20°C prior to use.

*Bumble bee rearing and husbandry*

We obtained 199 mated, post-diapause *B. terrestris audax* queens from a commercial supplier (Biobest, Westerlo, Belgium) in January 2013. The queens were housed individually in plastic boxes measuring 140 × 79 × 60 mm and kept at 28°C and 60% humidity. We supplied them with sugar solution (1:1 water: syrup; Koppert, Berkel en Rodenrijis, The Netherlands) and freeze dried pollen (Koppert) ad libitum. The queens were checked daily for the presence of queen-laid eggs. The sugar solution was replaced every day and the pollen every two days. In order to measure age related expression levels of *for* we removed and immediately dissected eight queens from the original 199 and classified them as 'early untreated' queens.

*Tracking of locomotor activity*

Because queens were kept enclosed, their locomotor activity was hypothesized to represent a proxy for propensity to forage because previous studies have shown that caged foraging specialist honey bee workers still have circadian rhythms in their locomotor behaviour (Bloch et al. 2001). We selected *Object Tracker* freeware (available at http://iEthology.com/) for the automated recording of queen locomotor behaviour and in a preliminary trial we tested its usefulness for estimating locomotor activity and circadian rhythmicity in locomotor activity.

*Object Tracker* uses positional information on each individual, taken every second, therefore potentially allowing an estimate of each queen's locomotor activity levels to be made. A previous study used *Object Tracker* to measure activity levels of *Drosophila* at a

small spatial scale (Donelson et al. 2012). We connected the *Object Tracker* software to a webcam ('Xbox Live Vision Camera', Microsoft, Reading, UK) whose field of view was able to capture and simultaneously monitor up to 30 randomly placed nest boxes containing queens. We placed opaque cardboard dividers between each nest box to minimise the chance of the behaviour of one bee affecting that of the others. Beginning on 7 February 2013, we randomly selected 17 non-laying queens and monitored their locomotor activity levels in this set-up for 7-10 days. Queens were fed *ad libitum* sugar solution and pollen throughout the trial period. We found that *Object Tracker* could monitor locomotor activity levels of these queens simultaneously and that ten days of monitoring was better for ascertaining the level of circadian rhythmicity than seven days.

We performed two main behavioural experiments. In the first, we monitored each of the *B. terrestris* non-experimental queens every day, checking for signs of egg-laying activity. In the initial experimental design we believed that it would be important to use egg-laying queens because in nature egg-laying queens are the ones that actively forage in order to provision their offspring. We hypothesised that, because they could still be in a hibernating state, non-egg laying queens would show lower overall levels of locomotor activity and would not display circadian rhythmicity. Therefore all of the queens that laid eggs were used as experimental queens in the first of the two experiments. Each queen was randomly assigned either a *for* dsRNA treatment, a *gfp* dsRNA treatment, or a water treatment (see '*dsRNA treatment*' below). These queens' treatment period and locomotor activity monitoring took place for seven days, in each case from the day her first egg was observed. Seven days was originally chosen because the preliminary experiments showed that this was enough to establish rhythmicity; however we also found that ten days was more effective so subsequent experiments were carried out for ten days, the first queen laid eggs on 6 February 2013, and, by 6 March 2013, a total of ten queens were laying eggs. Queens were randomly assigned to treatment with *for* dsRNA, *gfp* dsRNA, or water, but only six survived until the end of the treatment period. Two of these queens (one with gfp treatment, one with water treatment) were also given a worker pupa to test for the effect of brood on queens' locomotor activity, but this treatment was later discontinued because relatively few of the study queens laid eggs. On the 16 March 2013 an 11$^{th}$ queen started

to lay eggs; she was monitored for ten days in the same way as the queens that were observed in the second experiment (see below).

In order to increase sample sizes, in the second experiment it became necessary to use queens that had not yet started laying eggs. Although this set up was not ideal, because of the aforementioned problems with non-egg laying queens, we subsequently found that non-egg laying queens were still highly active (personal observation) prior to experimental treatment. Therefore we aimed to test whether their activity levels could be manipulated by experimental knockdown of *for*. Beginning on 13 March 2013, a random sample of 46 non-laying queens was used to test the effect of RNAi treatment on locomotor activity and circadian rhythmicity. For each experimental day, we randomly selected eight non-laying queens and fed them treatment solutions; three were treated with *for* dsRNA, two were treated with *gfp* dsRNA, and three were treated with water, we then continued to monitor them on subsequent days using the *Tracker Software*. We continued this on successive days until up to 24 treatment queens and eight control queens were being monitored at once.

We monitored the experimental queens for 20 hours each day, starting at 17:00-17:45 and finishing at 13:00-13:45 the next day (which we defined as the tracking period). To minimise disturbance, no-one was present in the room containing the experimental bees except between 13:00 and 17:45 each day, which was timed to coincide with the hours when the *Object Tracker* software was not running. In this period, we randomised the positions of the queen boxes and provided treatment solutions to the queens (we therefore defined this period as the 'treatment period'). For both experiments we monitored each treatment queen for ten days (i.e. ten tracking sessions) except in six cases in which treatment queens died before the experiment had been completed (data from these queens were not used in further analysis). By calculating the Euclidean distance between the positions of each queen over successive one second intervals, we estimated the distance travelled per minute for each queen. This estimate formed our metric of overall locomotor activity level for each queen. We also used *Object Tracker* to record four dead queens placed in boxes identical to the ones housing the live queens. This was to control for the effect on the software's measurement of queen movement of the constant minute light fluctuations in the room housing the queens. The mean movement rate recorded for these queens was taken to represent the minimum threshold for detecting locomotion. Therefore, following

a procedure recommended by other studies (Eban-Rothschild et al. 2011), we subtracted this value from the estimated movement rates of tracked queens.

*dsRNA treatment*

We introduced dsRNA into the target queens by feeding it directly to them as liquid sugar solution. Ingestion of dsRNA has been shown to cause knockdown of target genes in a number of organisms (Timmons and Fire 1998; Araujo et al. 2006; Garbian et al. 2012). The solutions contained either 15 µl (or on one occasion 21 µl due to an error) pure sugar syrup and an equal amount of each treatment solution. We provided each queen with a dose of 5 µg of dsRNA per treatment, summing to a total dose of 50 µg over ten days (assuming the queen consumed every dose). Relative to body size, this was a similar dose of dsRNA to the dose sizes used in other studies of RNAi in insects (Araujo et al. 2006; Borgio 2010). The treatment solutions were: 1) water containing 5 µg of *for* dsRNA, 2) water containing 5 µg of *gfp* dsRNA, 3) pure water. We provided the treatments in 250 µl Eppendorf tubes which had been cut to an approximately 50 µl volume using a scalpel. We placed the Eppendorf tubes containing the treatments upright in the dishes that contained the pollen, and during the treatment period we removed any other sources of syrup to encourage the queens to drink the treatments. Each day, before tracking began, we recorded whether they had consumed the treatment, with the consumption rate being approximately 90% before each tracking session. At the start of each tracking period we provided the queens with 500 µl of syrup, which we had previously discovered was enough to feed them overnight without them starving, but not so much that they were completely satiated when we provided them with treatment solution.

*Dissection and qRT-PCR*

We dissected the treatment queens ten days after treatment started, and removed their brains and digestive tracts within seven hours after their last tracking period. To dissect queens we first anaesthetised them by placing them on ice. We then removed their heads and gently detached the brain tissue and placed it in 500 µl of RNAlater (Life Technologies, Paisley, UK). We also dissected the digestive tract, from the crop ('honey stomach') to the rectum (Prys-Jones and Corbet 1987), and then placed it in 1000 µl of RNAlater. We dissected the brain because we were interested in testing the direct effect of the *for* gene

on behaviour. Other studies of the effects of *for* have also focussed on expression of the gene in brain (Ben-Shahar et al. 2002; Tobback et al. 2011; Rodriguez-Zas et al. 2012). We dissected the digestive tract because the treatments were administered orally and the dsRNA is in direct contact with the digestive tract. Therefore even if the treatment did not produce systemic effects including effects on brain, we might still have detected an effect on *for* gene expression in the digestive tract (Scott et al. 2013).

In addition to the treatment queens, we dissected brains and digestive tracts from ten non-laying queens at the start of the experiment and ten at the end (**Table 5.2** 'early' and 'late' queens respectively) for qRT-PCR. By dissecting these queens we were able to test whether age alone had an effect on *for* expression. These differed in food intake compared to the treatment queens because they were fed *ad libitum* throughout the experiment, while the treatment queens were given a restricted supply of sugar solution during the tracking periods.

We extracted total RNA from the stored tissue samples by briefly drying them, and homogenising them in Trizol reagent (Invitrogen, Carlsbad, California, USA) with a mortar and pestle. To separate the organic phase we added an appropriate amount of chloroform, and then precipitated the RNA in isopropanol. We resuspended the extracted RNA in RNase-free water. We treated the samples with the TURBO DNAfree kit (Ambion, Foster City, California, USA) according to the manufacturer's instructions to remove any residual DNA that had been carried over from the RNA extractions. We then used the Omega GoScript reverse transcription system (Promega, Fitchburg, Wisconsin, USA) and a universal primer (T(20)VN, **Table 5.1**) to reverse transcribe the RNA to cDNA. We used the cDNA for qRT-PCR with the SYBR Green assay system, which was performed by the company qStandard (Division of Surgery and Interventional Science, University College London, UK). QRT-PCR was performed to ascertain the expression levels (copy number) of three genes in each treatment queen, namely *for*, *ArgK* and *PLA2* with either pre-designed (*PLA2*) or custom-made (*for* and *ArgK*, **Table 5.1**) primers and probes. The *ArgK* and *PLA2* genes were used as reference genes because they have previously been shown to be relatively stable in expression level in *B. terrestris* (Hornakova et al. 2010). To estimate the expression levels of *for*, we obtained the copy number from 2 µl of cDNA, using standard

curves that were then divided by the geometric mean of the expression levels of the two reference genes (which was used as a normalisation factor).

*Statistical analysis*

We calculated the proportion of queens that exhibited circadian rhythmicity and the length of the circadian rhythms using Lomb-Scargle periodograms. This method was chosen because it does not require the input time-series data to be evenly spaced (Ruf 1999), which was important because each day there was a four-five hour gap in the tracking activity during the experimental treatment downtime. Since the periodograms showed extremely high periodicity at almost every measured time interval, we created models that only considered data from one random minute, per tracked hour per queen (e.g. around 200 random minutes per queen). This method has been used in previous studies to yield a single clear peak in periodicity or otherwise no peak for each individual (Bloch et al. 2001). We treated queens that showed a peak of periodicity between every 20 to 28 hours, which was significant at $p < 0.01$ (Ruf 1999), as having circadian rhythms.

Statistical analysis was carried out using the *R* statistical programming platform. Our first aim was to induce knockdown of *for* using *for* dsRNA. For this aim, we tested the effects of dsRNA treatment on *for* gene expression levels and on overall activity levels in the 'treated' queens in **Table 5.2** using ANOVA linear models. We used treatment as a predictor variable, whilst activity levels and *for* gene expression in either of the collected tissues were the response variables in each separate model. We excluded two outlier queens (that had been treated with water) because their *for* gene expression values differed from the mean *for* gene expression values by a factor of greater than three times the standard deviation.

Our second aim was to test for an association between *for* expression and queen behaviour. We therefore tested whether there were any significant effects of *for* gene expression on overall locomotor activity levels and on circadian rhythmicity. In addition, we tested whether any other variables in the experiment including age, egg-laying status, and feeding regime had an effect on these behavioural traits. We used a generalised linear model with a binomial error distribution and logit link function, with gene expression as the predictor variable and the probability of circadian activity (using the p-value from the Lomb-Scargle periodogram) as the response variable. We also used a chi-squared test to

compare the number of rhythmic queens among the *ad libitum* fed non-laying queens; the food-restricted non-laying queens; and the food-restricted laying queens ('preliminary trial' queens, 'treatment' queens, and 'laying' queens in **Table 5.2**, respectively). We then used a post-hoc chi-squared test to make each comparison individually. Although the laying queens differed in treatments and whether a pupae was present, we treated them as a single group as there was no effect of treatment (see **Section 5.3**). We also compared rhythmicity across all queens for the final seven days to ensure the shorter time intervals for some of the treatment queens were not artificially reducing the number of individuals that showed circadian activity (the laying queens were only treated for seven rather than ten days). We used a linear model with *for* expression, age, and laying status as predictor variables, and locomotor activity as the response variable. We also used a linear mixed model to test the effects of age and laying status on locomotor activity, using queen as a random effect (which accounted for queens that were tracked once as non-treated queens, then again as treated queens). We also tested whether there was a relationship between overall locomotor activity and rhythmicity using a Spearman's rank correlation between the mean locomotor activity levels calculated using *Object Tracker* freeware and the probability of being arrhythmic in the final seven days of tracking (p-value from Lomb-Scargle periodogram).

Finally, we determined whether there was an effect of age, egg-laying status and feeding regime on *for* expression in both the brains and the digestive tracts of all of the queens for which expression data were available (n = 72; **Table 5.2**). We constructed two linear models with age, laying status, and feeding regime as predictor variables, and with the *for* expression levels in each tissue as the response variable for each model. We used a Pearson's product moment correlation to test for a relationship between *for* gene expression in the brain and digestive tract to determine whether gene expression in either organ was independent of the other.

## *5.3 Results*

*dsRNA that targeted* for *did not induce for knockdown*

Our first aim was to induce *for* knockdown using dsRNA. Following treatment, a mean of 88% of the queens consumed the treated sugar solution (*for*) or control solutions (*gfp* or water) each day. Four queens consumed less than half of their total treatment doses over the entire treatment period; these were excluded from further analysis. We found no relationship between *for* expression in the brains and digestive tract in any of the analyses (Pearson's product moment correlation, $r$ = -0.03, $n$ = 70, $p$ = 0.814), so we considered the tissues independently for each analysis. Overall we found no effect of dsRNA treatment on *for* expression in the brains of non-laying queens (means ± SE: *for* queens = 2244 ± 56 copy number, *gfp* queens = 2054 ± 46 copy number; water queens = 2114 ± 72 copy number; ANOVA, $F_{2,36}$ = 1.80, $n$ = 39, $p$ = 0.180; **Figure 5.1a**); and no effect of dsRNA on *for* expression in the digestive tract of non-laying queens (means ± SE: *for* queens = 12933 ± 766 copy number; *gfp* queens = 10133 ± 874 copy number; water queens = 11219 ± 914 copy number; ANOVA, $F_{2,39}$ = 2.34, $n$ = 42, $p$ = 0.110; **Figure 5.1b**). These results do not support the hypothesis that *for* expression was knocked down by dsRNA treatments.

Given the above results, it was not surprising that we also found no effect of treatment on the recorded behaviour of non-laying queens, i.e. there was no measurable effect on rhythmicity because all of the treated queens had circadian rhythms (see '*Evidence of circadian rhythmicity in* B. terrestris *queens'* below), and there was no significant effect on locomotor activity levels, defined as movement in (arbitrary units) per second (means ± SE: *for* queens = 2.7 ± 0.4; *gfp* queens = 3.0 ± 0.2; water queens = 2.7 ± 0.1; ANOVA, $F_{2,39}$ = 0.59, $n$ = 42, $p$ = 0.558).

*For expression had no effect on locomotor activity levels*

Our second aim was to test whether the gene *for* affected the locomotor activity levels and circadian rhythmicity in *B. terrestris* queens. We found no effect on locomotor activity levels of *for* expression in brain (ANOVA, $F_{1, 40}$ = 0.12, $n$ = 50, $p$ = 0.731; **Figure 5.2a**) or digestive tract ($F_{1, 45}$ = 0.76, $n$ = 52 $p$ = 0.388; **Figure 5.2b**). We did find that age had a significant linear negative relationship with the locomotor activity levels across both sets of feeding regimes (ad libitum or food-restricted), i.e. the youngest queens had the highest activity levels (linear mixed model, $\chi^2$ = 101.5, $n$ = 70, $p$ <0.001; **Figure 5.2c**). Although there

was a trend for laying queens to have greater activity levels, this difference was not significant ($\chi^2$ = 2.18, $p$ = 0.140; **Figure 5.2d**).

*Evidence of circadian rhythmicity in* B. terrestris *queens*

Actograms showing an example of a rhythmic queen and an arrhythmic queen are depicted in **Figure 5.3**. Rhythmic queens always had a burst of activity shortly following the tracking downtimes from about 1700-1745 (e.g. **Figure 5.3a**), indicating that some factor at these times was causing these individuals to maintain circadian rhythmicity. Examples could be the burst of white light caused by entering the room, or a small change in temperature or disturbance when the locations of the queen viewing boxes were randomised and they had their food and syrup replaced. However, this was unlikely to be a simple response to disturbance because a) the peak activity levels were usually several hours after the tracking cameras had been set up during the downtime period; and b) some of the queens were arrhythmic and showed no response to disturbance (e.g. **Figure 5.3b**). We recorded the mean activity levels as being significantly positively correlated with the probability of queens being arrhythmic (Spearman rank correlation, $\rho$ = 0.70, $n$ = 70, $S$ = 17058, $p$ < 0.001). By establishing which queens were rhythmic we were able to investigate our second aim.

For *expression had no effect on circadian rhythmicity, but circadian rhythmicity is significantly affected by laying status and feeding regime in* B. terrestris *queens*

Contrary to the hypothesis advanced in our second aim, we found no significant effect of *for* gene expression on whether queens were rhythmic or arrhythmic in either brain (generalised linear model; $\chi^2$ = 0.35, $n$ = 70, $p$ = 0.551) or digestive tract ($\chi^2$ = 0.65, $n$ = 70, $p$ = 0.422); we also found no evidence of an interaction between these factors ($\chi^2$ = 0.39, $n$ = 70, $p$ = 0.528). Interestingly we did find significant interactions between the treatment regimes and laying status of queens in their likelihood to exhibit circadian rhythmicity (chi-squared test, $\chi^2$ = 40.3, df = 2, p < 0.001; **Figure 5.4**). For example, all of the food-restricted, non-laying queens (i.e. the 'treatment' queens in **Table 5.2**) were rhythmic, but only 43% of the food-restricted laying queens (n = 7) and 29% of the ad libitum-fed non-laying queens ('laying' queens and 'preliminary trial' queens in **Table 5.2**, respectively) were rhythmic. Pairwise comparisons revealed that the food-restricted, non-laying queens had significantly higher numbers of individuals with circadian rhythmicity than either the food-

restricted, laying queens (chi-squared test, $\chi^2$ = 20.8, df = 1, $p$ < 0.001) or the *ad libitum*-fed non-laying queens (chi-squared test, $\chi^2$ = 35.7, df = 1, $p$ < 0.001). The last two groups did not differ significantly from each other (chi-squared test, $\chi^2$ = 0.03, df = 1, $p$ = 0.874). Therefore, we found no evidence to support the hypothesis that *for* has an effect on overall activity levels, or on circadian rhythms in *B. terrestris* queens.

For *gene expression was effected by reproductive status and feeding regime, but not age*

We found no significant effect of age on *for* gene expression in brain (ANOVA, $F_{1, 63}$ = 2.39, $n$ = 68, $p$ = 0.123) or digestive tract (ANOVA, $F_{1,66}$ = 0.02, $n$ = 72, $p$ = 0.878). Laying queens had significantly higher *for* gene expression than non-laying queens in both brain (means ± SE: laying queens = 2453 ± 113 copy number, $n$ = 6; non-laying queens = 2240 ± 36 copy number, n = 62; $F_{1,63}$ = 6.44, $p$ = 0.014; **Figure 5.5a**) and digestive tract (means ± SE: laying queens = 15492 ± 1189 copy number, $n$ = 6; non-laying queens = 10816 ± 401 copy number, $n$ = 66; $F_{1,69}$ = 13.03, $p$ < 0.001; **Figure 5.5b**). However, it should be noted that the magnitude of the effect was small (a less than two-fold change in gene expression between treatment regimes).

Feeding regime was found to have an effect on *for* levels in the treatment queens, as queens that were fed ad libitum ('early' and 'late' queens in **Table 5.2**) had higher expression levels of *for* than queens with a limited food supply (the 'laying' queens and 'treatment' queens in **Table 5.2**) in brain (means ± SE: ad libitum queens = 2435 ± 69 copy number, $n$ = 20; limited queens = 2180 ± 37 copy number, n = 50; $F_{1,63}$ = 14.21, $p$ < 0.001; **Figure 5.6a**; Note that 'preliminary trial' queens were not included in this analysis because we had no data on their *for* expression). However, the opposite effect occurred in the digestive tract , in which queens fed ad libitum had lower *for* expression levels (means ± SE: ad libitum queens = 9000 ± 376 copies, n = 20; limited queens = 12055 ± 500 copies, n = 52; $F_{1,69}$ = 10.26, $p$ = 0.002; **Figure 5.6b**). We also found that feeding regime had a significant interaction with age, so that *for* expression increased with age in the queens that were fed ad libitum, but the opposite effect occurred in queens with a limited food supply ($F_{2,63}$ = 3.25, $p$ = 0.045**; Figure 5.7a**). The effect was only found in brain, with no significant effect being recorded in digestive tract ($F_{2,67}$ = 0.04, $p$ = 0.958; **Figure 5.7b**). We

found no significant interaction between age and laying status on *for* gene expression in either brain ($F_{1,62}$ = 0.15, $p$ = 0.703) or in digestive tract ($F_{2,66}$ = 0.72, $p$ = 0.490).

## *5.4 Discussion*

### *RNAi by dsRNA ingestion had no effect on* Bombus *queens*

Our first aim was to test whether dsRNA could induce knockdown in *B. terrestris* queens. We tested this using a feeding experiment where we fed non-laying *B. terrestris* queens dsRNA that could target the gene *for*, a gene known to be involved in regulating foraging behaviour, and possibly circadian rhythmicity. Feeding *for* dsRNA to queens had no measurable effect on *for* expression levels (**Figure 5.1**) and consequently there was no measurable effect of treatment on either type of behaviour (foraging or rhythmicity). This means that it is likely that the knockdown experiments were unsuccessful.

There are a number of possible reasons for this. One reason is that the method of dsRNA delivery was through ingestion rather than injection. Ingestion of dsRNA is one of the simplest methods of RNA knockdown, and is routinely used in species with simple digestive systems such as *C. elegans*. The method has been shown to work in higher eukaryotes, including insects (Araujo et al. 2006; Li et al. 2011; Bolognesi et al. 2012; Singh et al. 2013). However, results are often very variable and dosage dependent (Terenius et al. 2011; Scott et al. 2013; Yu et al. 2013), with very high doses usually working more effectively. We provided queens with 5 µg of dsRNA per dose, a level of treatment based on those in other studies where orally administered dsRNA induced gene knockdown. Because dsRNA application by oral ingestion has variable effects between species, it is possible that a larger dose would have been more effective (Scott et al. 2013). In addition, oral administration of dsRNA might be less effective if the target insects are able to produce dsRNA-degrading enzymes (Scott et al. 2013). This has been shown to be the case in at least one species of moth, *Bombyx mori* (Arimatsu et al. 2007; Liu et al. 2012b).

In addition to the problems highlighted above, another potential issue with the experimental mechanism of delivery was that the dsRNA might not have been able to target the correct organ. Ideally the treatment would have induced knockdown of *for* in the brain

of the target queens by systemically knocking down expression across the whole organism. This method has been effective in experiments in other insects (Terenius et al. 2011; Scott et al. 2013; Singh et al. 2013). Even if effects of RNAi had not been delivered to other parts of the body, we should still expect to see transcriptional repression in the digestive tract, which is the first organ in direct contact with dsRNA following ingestion. In our experiment we measured *for* expression in both brain and digestive tract, and found no evidence of transcriptional decrease in *for* expression in either tissue. This suggests that the dsRNA did not induce knockdown at all, in which case delivery would not have been the main problem with the treatment. But it is also possible that barriers in the digestive tract of the queens prevented us from identifying transcriptional repression in that organ too.

One potential solution to these issues would be to induce knockdown by direct injection of dsRNA into the target organs. Direct injection of dsRNA has been shown to work more effectively, more specifically, and more consistently than knockdown by dsRNA ingestion (Scott et al. 2013; Yu et al. 2013). Furthermore this method of dsRNA delivery has previously been shown to induce knockdown in *B. terrestris*, where it was used to reduce the gene expression of antimicrobial peptides to test their effect on *Crithidia bombi* infection in the bumble bee (Deshwal and Mallon 2014). We did not employ this method in this experiment because injection was likely to be traumatic to queens and therefore likely to have affected their behaviour. It could also have affected their mortality; for example Holehouse et al. (2003) showed that injecting *B. terrestris* workers (to extract haemolymph for DNA sampling from live animals) caused a significant increase in their mortality. However, other experiments that target different genes and that do not directly measure a behavioural response might be an effective means of establishing a protocol for RNAi in *Bombus* species.

An alternative reason why the RNAi did not affect *for* treatment could be due to the construction of the dsRNA itself. We designed the dsRNA to target a 400bp exon on the *for* gene. It is possible that interference at this region did not prevent the production of a *for* transcript, albeit one with the 3' end missing. In such cases the *for* mRNA transcript might still be functional, or alternatively it might be non-functional but still be detectable using the qRT-PCR assays. Either of these remains a possibility because the qRT-PCR probes targeted a different section of the gene to the dsRNA. This arrangement was deliberate and

was designed to prevent us from inadvertently detecting *for* gene expression from the dsRNA strands themselves; this possibility could be verified by performing qRT-PCRs targeting different parts of the gene and observing whether these produced similar results. It also needs noting that the *for* gene is not well characterised in *B. terrestris*, and its annotation could be inaccurate. Therefore, it is possible that we could have targeted a section of the gene that was spliced out following transcription, either because it contained unidentified intronic regions or because the transcript had undergone alternative splicing. With the present data we cannot distinguish between these possibilities, but changing the section of the gene that dsRNA targets, or knocking down another gene entirely, could help establish a protocol for RNAi in *Bombus* species in future studies.

For *gene expression had no effect on activity levels or circadian rhythms*

Our second aim was to establish whether *for* affected behaviour in *B. terrestris* queens. Previous studies have shown that the gene *for* has been associated with differences in foraging behaviour in both *A. mellifera* (Ben-Shahar et al. 2002; Rodriguez-Zas et al. 2012) and *B. terrestris* (Tobback et al. 2011). These studies provided our rationale for knocking down *for* and then monitoring changes in the levels of overall locomotor activity and in circadian rhythmicity. We found no evidence that *for* expression was related to overall levels of locomotor activity (**Figure 5.2a-b**) or circadian rhythmicity in *B. terrestris* queens. It therefore seems unlikely that the gene plays a central role in regulating foraging behaviour in *B. terrestris*, which stands in contrast to the role of *for* in *Drosophila* (Debelle and Sokolowski 1987; Pereira and Sokolowski 1993). However, our results do not rule out the possibility that *for* subtly affects foraging state in other ways, because we did not allow our queens to forage outside of the nest boxes. We also used queens that had recently ended diapause, and so it remains unknown whether foraging is related to queen activity in more established nests in which the queen provisions her first brood of workers.

*The relationship between laying status and queen behaviour*

We did not successfully knock down *for*, and we found no relationship between *for* expression and overall activity levels; however, we did find that there was a significant effect of both age and queen laying status on some of these traits. Specifically, we found that older queens were less active (**Figure 5.2c**), implying that queens are less likely to

forage as they get older. This is known to happen in most bumble bee species, since the queen stops foraging after the first few workers emerge (see Introduction). Of particular interest was a general trend for laying queens to be more active than non-laying queens (although the effect was not significant; **Figure 5.2d**). Our results also showed that queen status is related to circadian rhythmicity, with laying queens being more likely to be arrhythmic (**Figure 5.4**). This result is in agreement with a previous study where queens became arrhythmic just prior to laying eggs (Eban-Rothschild et al. 2011). However, the result only held when the queens were kept on a restricted diet. One interpretation for these results is that as queens become reproductively active they forage less often, focusing more on laying eggs and rearing brood. The dependency of the effect on diet might be expected if the queens that were fed *ad libitum* used the presence of a large amount of stored food as a stimulus to focus more on rearing brood rather than foraging. These results might also explain why the foraging treatment had no effect on queen activity and rhythmicity as we could only test the effect of dsRNA treatment on queens that were non-egg layers and therefore had reduced activity levels and levels of rhythmicity. However because the reduced level of activity in non-egg layers was not-significant, and because both groups still exhibited measurable levels of activity and circadian rhythmicity we should still have identified some effect on these characteristics if the treatment had induced a change in locomotor behaviour.

In this study we show that laying status also had a significant effect on *for* expression in both brain and digestive tract (**Figure 5.5**), with laying queens having significantly higher levels of *for* than non-laying queens. However, these results should be interpreted with some caution because: 1), although statistically significant, the expression differences between queen laying states were quite small; 2) the number of 'laying' queens was low, and the analysis was therefore lacking in statistical power; and 3) diet was also shown to have an effect on *for* expression, with opposite trends being shown in different tissues (**Figure 5.6**). There was no effect of age on *for* expression (**Figure 5.7**), so even though the laying queens tended to be younger these effects seemed to occur independently of age. One explanation for these results is that *for* is one of many genes that affect queen reproductive status, and the gene becomes highly expressed just prior to when the queens start to lay eggs. While we found no effect of *for* expression on locomotor activity, we

studied only a small number of laying queens, and most of the queens were not reproductively active. In this study we have shown that egg laying queens have higher levels of rhythmicity (and trend towards higher overall levels of activity), and that they expressed *for* at a higher level than non-laying queens. Therefore further studies should focus on the effect of *for* when it is expressed at its highest level (i.e. when queens are laying eggs) which would allow us to determine which of these factors are most important for queen cessation of foraging.

## *5.5 References*

Araujo RN, Santos A, Pinto FS, Gontijo NF, Lehane MJ, Pereira MH. 2006. RNA interference of the salivary gland nitrophorin 2 in the triatomine bug *Rhodnius prolixus* (Hemiptera : Reduviidae) by dsRNA ingestion or injection. *Insect Biochem Molec* **36**: 683-693.

Arimatsu Y, Kotani E, Sugimura Y, Furusawa T. 2007. Molecular characterization of a cDNA encoding extracellular dsRNase and its expression in the silkworm, *Bombyx mori*. *Insect Biochem Molec* **37**: 176-183.

Ashrafi K, Chang FY, Watts JL, Fraser AG, Kamath RS, Ahringer J, Ruvkun G. 2003. Genome-wide RNAi analysis of *Caenorhabditis elegans* fat regulatory genes. *Nature* **421**: 268-272.

Bartel DP. 2004. MicroRNAs: Genomics, biogenesis, mechanism, and function. *Cell* **116**: 281-297.

Behura SK, Whitfield CW. 2010. Correlated expression patterns of microRNA genes with age-dependent behavioural changes in honeybee. *Insect Mol Biol* **19**: 431-439.

Belles X. 2010. Beyond *Drosophila*: RNAi in vivo and functional genomics in insects. *Annu Rev Entomol* **55**: 111-128.

Ben-Shahar Y, Leung HT, Pak WL, Sokolowski MB, Robinson GE. 2003. cGMP-dependent changes in phototaxis: a possible role for the foraging gene in honey bee division of labor. *J Exp Biol* **206**: 2507-2515.

Ben-Shahar Y, Robichon A, Sokolowski MB, Robinson GE. 2002. Influence of gene action across different time scales on behavior. *Science* **296**: 741-744.

Berns K, Hijmans EM, Mullenders J, Brummelkamp TR, Velds A, Heimerikx M, Kerkhoven RM, Madiredjo M, Nijkamp W, Weigelt B et al. 2004. A large-scale RNAi screen in human cells identifies new components of the p53 pathway. *Nature* **428**: 431-437.

Bloch G, Barnes BM, Gerkema MP, Helm B. 2013. Animal activity around the clock with no overt circadian rhythms: patterns, mechanisms and adaptive value. *Proc R Soc B* **280**: 20130019.

Bloch G, Toma DP, Robinson GE. 2001. Behavioral rhythmicity, age, division of labor and period expression in the honey bee brain. *J Biol Rhythm* **16**: 444-456.

Bolognesi R, Ramaseshadri P, Anderson J, Bachman P, Clinton W, Flannagan R, Ilagan O, Lawrence C, Levine S, Moar W et al. 2012. Characterizing the mechanism of action of double-stranded RNA activity against western corn rootworm (*Diabrotica virgifera virgifera* LeConte). *PLOS One* **7**: e47534.

Borgio JF. 2010. RNAi mediated gene knockdown in sucking and chewing insect pests. *J Biopest* **3**: 386 - 393.

Boutros M, Kiger AA, Armknecht S, Kerr K, Hild M, Koch B, Haas SA, Paro R, Perrimon N, Consortium HFA. 2004. Genome-wide RNAi analysis of growth and viability in *Drosophila* cells. *Science* **303**: 832-835.

Carthew RW, Sontheimer EJ. 2009. Origins and mechanisms of miRNAs and siRNAs. *Cell* **136**: 642-655.

Debelle JS, Sokolowski MB. 1987. Heredity of rover Sitter - alternative foraging strategies of *Drosophila melanogaster* larvae. *Heredity* **59**: 73-83.

Deshwal S, Mallon EB. 2014. Antimicrobial peptides play a functional role in bumblebee anti-trypanosome defense. *Dev Comp Immunol* **42**: 240-243.

Donelson NC, Kim EZ, Slawson JB, Vecsey CG, Huber R, Griffith LC. 2012. High-resolution positional tracking for long-term analysis of *Drosophila* sleep and locomotion using the "*Tracker*" program. *PLOS One* **7**: e37250.

Duchateau MJ, Velthuis HHW. 1988. Development and reproductive strategies in *Bombus* colonies. *Behaviour* **107**: 186-207.

Eban-Rothschild A, Belluci S, Bloch G. 2011. Maternity-related plasticity in circadian rhythms of bumble-bee queens. *Proc R Soc B* **278**: 3510-3516.

Garbian Y, Maori E, Kalev H, Shafir S, Sela I. 2012. Bidirectional transfer of RNAi between honey bee and *Varroa destructor*: *Varroa* gene silencing reduces *Varroa* population. *PLOS Pathogens* **8**: e1003035.

Giebultowicz JM. 2000. Molecular mechanism and cellular distribution of insect circadian clocks. *Annu Rev Entomol* **45**: 769-793.

Goulson D. 2003. *Bumblebees: Behaviour, ecology and conservation*. Oxford University Press, Oxford, UK.

Goulson D. 2013. *A Sting in the Tale*. Jonathan Cape, London, UK.

Greenberg JK, Xia J, Zhou X, Thatcher SR, Gu X, Ament SA, Newman TC, Green PJ, Zhang W, Robinson GE et al. 2012. Behavioral plasticity in honey bees is associated with differences in brain microRNA transcriptome. *Genes Brain Behav* **11**: 660-670.

Guo XQ, Su SK, Skogerboe G, Dai SAJ, Li WF, Li ZG, Liu F, Ni RF, Guo Y, Chen SL et al. 2013. Recipe for a busy bee: microRNAs in honey bee caste determination. *PLOS One* **8**: e81661.

Holehouse KA, Hammond RL, Bourke AFG. 2003. Non-lethal sampling of DNA from bumble bees for conservation genetics. *Insect Soc* **50**: 277-285.

Holland JG, Guidat FS, Bourke AFG. 2013. Queen control of a key life-history event in a eusocial insect. *Biology Letters* **9**: 20130056.

Hong-Geller E, Micheva-Viteva S. 2010. Functional gene discovery using RNA interference-based genomic screens to combat pathogen infection. *Curr Drug Disc Technol* **7**: 86-94.

Hornakova D, Matouskova P, Kindl J, Valterova I, Pichova I. 2010. Selection of reference genes for real-time polymerase chain reaction analysis in tissues from *Bombus terrestris* and *Bombus lucorum* of different ages. *Anal Biochem* **397**: 118-120.

Hunt JH, Mutti NS, Havukainen H, Henshaw MT, Amdam GV. 2011. Development of an RNA interference tool, characterization of its target, and an ecological test of caste differentiation in the eusocial wasp *Polistes*. *PLOS One* **6**: e26641.

Ingram KK, Oefner P, Gordon DM. 2005. Task-specific expression of the foraging gene in harvester ants. *Mol Ecol* **14**: 813-818.

Jarosch A, Moritz RFA. 2011. Systemic RNA-interference in the honeybee *Apis mellifera*: Tissue dependent uptake of fluorescent siRNA after intra-abdominal application observed by laser-scanning microscopy. *J Insect Physiol* **57**: 851-857.

Kalderon D, Rubin GM. 1989. cGMP-dependent protein kinase genes in *Drosophila*. *J Biol Chem* **264**: 10738-10748.

Kamakura M. 2011. Royalactin induces queen differentiation in honeybees. *Nature* **473**: 478-483.

Kamath RS, Ahringer J. 2003. Genorne-wide RNAi screening in *Caenorhabditis elegans*. *Methods* **30**: 313-321.

Kaun KR, Riedl CAL, Chakaborty-Chatterjee M, Belay AT, Douglas SJ, Gibbs AG, Sokolowski MB. 2007. Natural variation in food acquisition mediated via a *Drosophila* cGMP-dependent protein kinase. *J Exp Biol* **210**: 3547-3558.

Kodaira Y, Ohtsuki H, Yokoyama J, Kawata M. 2009. Size-dependent foraging gene expression and behavioral caste differentiation in *Bombus ignitus*. *BMC Res Notes* **2**: 184-188.

Li J, Chen QH, Lin YJ, Jiang TR, Wu G, Hua HX. 2011. RNA interference in *Nilaparvata lugens* (Homoptera: Delphacidae) based on dsRNA ingestion. *Pest Manag Sci* **67**: 852-859.

Liu F, Peng W, Li Z, Li W, Li L, Pan J, Zhang S, Miao Y, Chen S, Su S. 2012a. Next-generation small RNA sequencing for microRNAs profiling in *Apis mellifera*: comparison between nurses and foragers. *Insect Mol Biol* **21**: 297-303.

Liu JH, Swevers L, Iatrou K, Huvenne H, Smagghe G. 2012b. *Bombyx mori* DNA/RNA non-specific nuclease: Expression of isoforms in insect culture cells, subcellular localization and functional assays. *J Insect Physiol* **58**: 1166-1176.

Lucas C, Sokolowski MB. 2009. Molecular basis for changes in behavioral state in ant social behaviors. *Proc Natl Acad Sci USA* **106**: 6351-6356.

Misquitta L, Paterson BM. 1999. Targeted disruption of gene function in *Drosophila* by RNA interference (RNA-i): A role for nautilus in embryonic somatic muscle formation. *Proc Natl Acad Sci USA* **96**: 1451-1456.

Misquitta L, Wei Q, Paterson BM. 2008. Targeted disruption of gene function in *Drosophila* by RNA interference. In *Drosophila Protocols*, (ed. W Sullivan, M Ashburner, R Scott Hawley). Cold Spring Harbor Laboratory Press, New York, USA.

Osborne KA, Robichon A, Burgess E, Butland S, Shaw RA, Coulthard A, Pereira HS, Greenspan RJ, Sokolowski MB. 1997. Natural behavior polymorphism due to a cGMP-dependent protein kinase of *Drosophila*. *Science* **277**: 834-836.

Oster GF, Wilson EO. 1978. *Caste and ecology in the social insects*. Princeton University Press, Princeton, USA.

Pereira HS, Sokolowski MB. 1993. Mutations in the larval *foraging* gene affect adult locomotory behavior after feeding in *Drosophila melanogaster*. *Proc Natl Acad Sci USA* **90**: 5044-5046.

Prys-Jones OE, Corbet SA. 1987. *Bumblebees*. Cambridge University Press, Cambridge, UK.

Ratcliff F, Harrison BD, Baulcombe DC. 1997. A similarity between viral defense and gene silencing in plants. *Science* **276**: 1558-1560.

Rodriguez-Zas SL, Southey BR, Shemesh Y, Rubin EB, Cohen M, Robinson GE, Bloch G. 2012. Microarray analysis of natural socially regulated plasticity in circadian rhythms of honey bees. *J Biol Rhythm* **27**: 12-24.

Ruf T. 1999. The Lomb-Scargle periodogram in biological rhythm research: analysis of incomplete and unequally spaced time-series. *Biol Rhythm Res* **30**: 178-201.

Scott JG, Michel K, Bartholomay LC, Siegfried BD, Hunter WB, Smagghe G, Zhu KY, Douglas AE. 2013. Towards the elements of successful insect RNAi. *J Insect Physiol* **59**: 1212-1221.

Sen GL, Blau HM. 2006. A brief history of RNAi: the silence of the genes. *FASEB J* **20**: 1293-1299.

Singh AD, Wong S, Ryan CP, Whyard S. 2013. Oral delivery of double-stranded RNA in larvae of the yellow fever mosquito, *Aedes aegypti*: Implications for pest mosquito control. *J Insect Sci* **13**: 69.

Terenius O, Papanicolaou A, Garbutt JS, Eleftherianos I, Huvenne H, Kanginakudru S, Albrechtsen M, An CJ, Aymeric JL, Barthel A et al. 2011. RNA interference in Lepidoptera: An overview of successful and unsuccessful studies and implications for experimental design. *J Insect Physiol* **57**: 231-245.

Timmons L, Fire A. 1998. Specific interference by ingested dsRNA. *Nature* **395**: 854-854.

Tobback J, Heylen K, Gobin B, Wenseleers T, Billen J, Arckens L, Huybrechts R. 2008. Cloning and expression of *PKG*, a candidate foraging regalating gene in *Vespula vulgaris*. *Anim Biol* **58**: 341-351.

Tobback J, Mommaerts V, Vandersmissen HP, Smagghe G, Huybrechts R. 2011. Age- and task-dependent foraging gene expression in the bumble bee *Bombus terrestris*. *Arch Insect Biochem Physiol* **76**: 30-42.

Weaver DB, Anzola JM, Evans JD, Reid JG, Reese JT, Childs KL, Zdobnov EM, Samanta MP, Miller J, Elsik CG. 2007. Computational and transcriptional evidence for microRNAs in the honey bee genome. *Genom Biol* **8**: R97.

Westbrook TF, Martin ES, Schlabach MR, Leng YM, Liang AC, Feng B, Zhao JJ, Roberts TM, Mandel G, Hannon GJ et al. 2005. A genetic screen for candidate tumor suppressors identifies *REST*. *Cell* **121**: 837-848.

Williams PH. 1998. An annotated checklist of bumble bees with an analysis of patterns of description. *Bulletin of the Natural History Museum (Entomology)* **67**: 79-152.

Wilson EO. 1971. *The Insect Societies*. Belknap Press of Harvard University Press, Cambridge, Massachusetts, USA.

Wilson RC, Doudna JA. 2013. Molecular mechanisms of RNA interference. *Annu Rev Biophys* **42**: 217-239.

Yu N, Christiaens O, Liu JS, Niu JZ, Cappelle K, Caccia S, Huvenne H, Smagghe G. 2013. Delivery of dsRNA for RNAi in insects: an overview and future directions. *Insect Sci* **20**: 4-14.

Zamore PD, Tuschl T, Sharp PA, Bartel DP. 2000. RNAi: Double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. *Cell* **101**: 25-33.

Zhang EE, Liu AC, Hirota T, Miraglia LJ, Welch G, Pongsawakul PY, Liu XZ, Atwood A, Huss JW, Janes J et al. 2009. A genome-wide RNAi screen for modifiers of the circadian clock in human cells. *Cell* **139**: 199-210.

Zhou XG, Wheeler MM, Oi FM, Scharf ME. 2008. RNA interference in the termite *Reticulitermes flavipes* through ingestion of double-stranded RNA. *Insect Biochem Molec* **38**: 805-815.

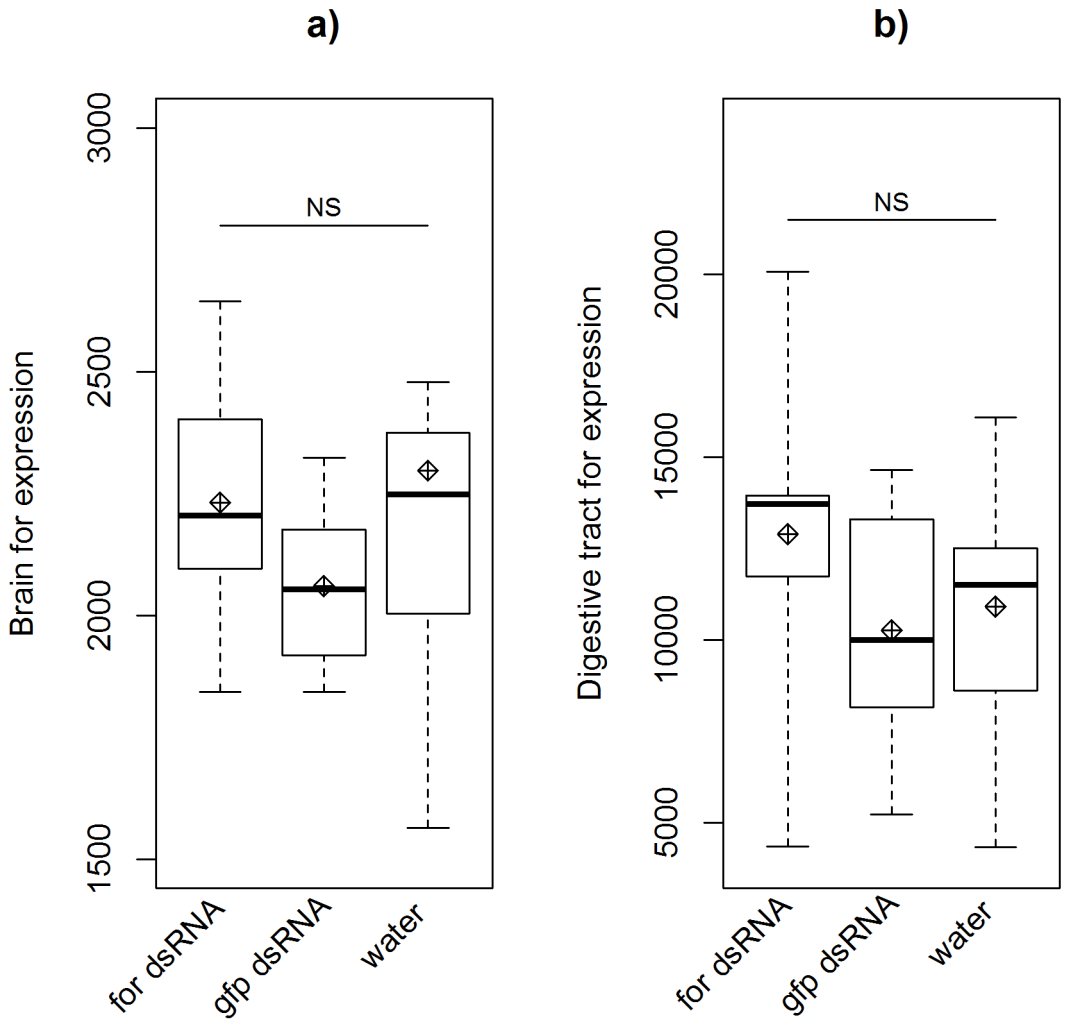


**Figure 5.1: Expression levels (copy number) of *for* in RNAi treatments in *Bombus terrestris* queens. a) In brain. b) In digestive tract. Gene expression level (copy number) was estimated using qRT-PCR. Expression level for *for* was calculated relative to mean expression level of two reference genes, *ArgK* and *PLA2*. Sample sizes of treatments: *for* dsRNA (n = 18), *gfp* dsRNA control (n = 14) and water control (n = 14). Diamonds indicate mean expression values, thick horizontal bars median values, boxes interquartile ranges, and dotted lines ranges excluding outliers. NS, not significant (ANOVA). *Image reproduced and adapted from Holland. 2013. PhD thesis, with permission from Dr Jacob Holland.***

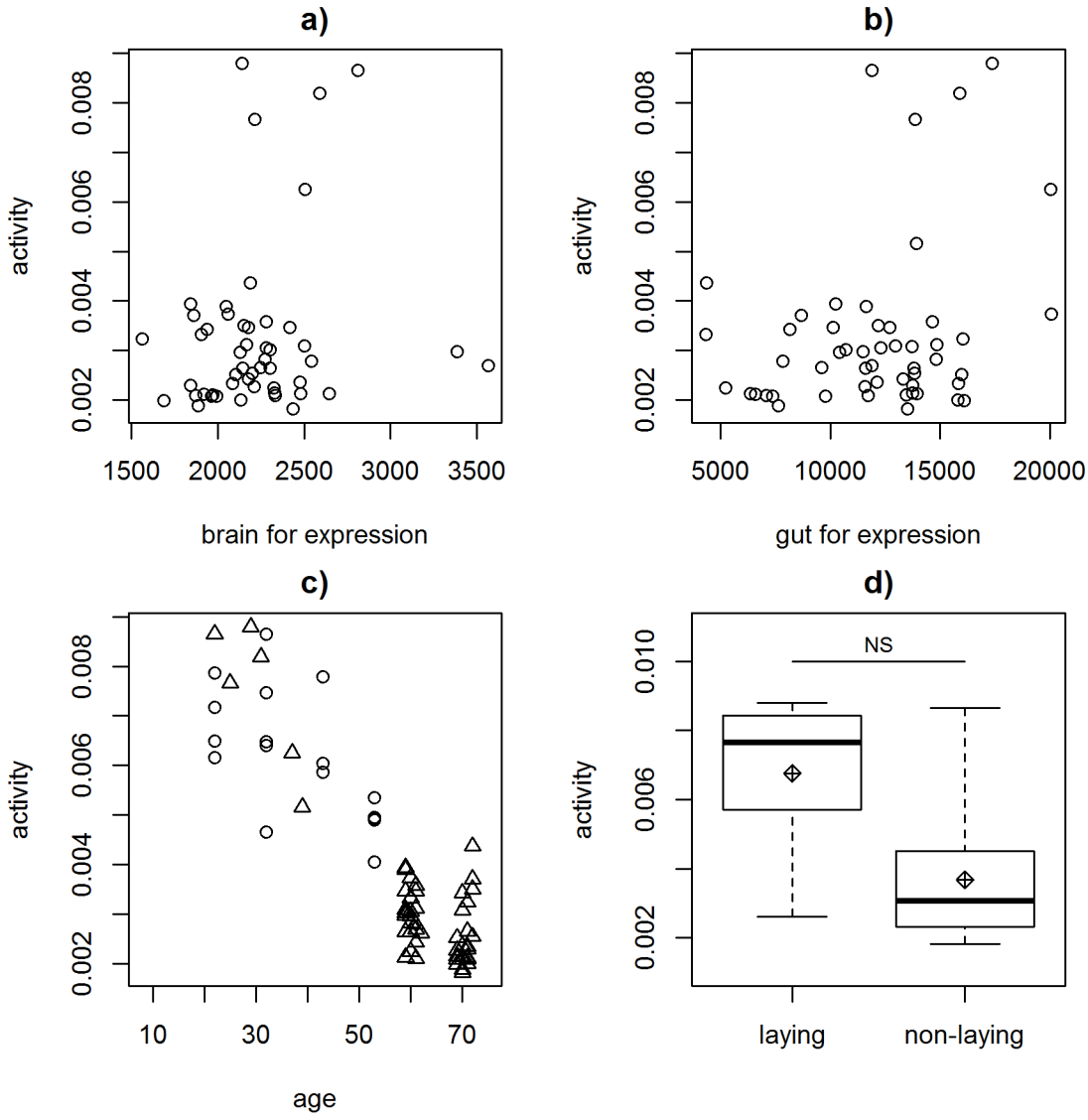


**Figure 5.2: Effect of various factors on overall locomotor activity in movement (arbitrary units) per second in *Bombus terrestris* queens. a) Effect of *for* gene expression level in brain (n = 50). b) Effect of *for* gene expression level in digestive tract (n = 52). In a) and b), gene expression was estimated using qRT-PCR and was calculated relative to mean expression level of two reference genes, *ArgK* and *PLA2*. c) Effect of queen age in days since the start of the experiment (when the ordered queens arrived) until the day of dissection (n = 70), with triangles representing queens on a restricted diet and circles queens fed ad libitum. d) Effect of queen egg-laying status (n = 70). In d), diamonds indicate mean expression values, thick horizontal bars median values, boxes interquartile ranges, and dotted lines ranges excluding outliers. NS, not significant (chi-squared test). *Image reproduced and adapted from Holland. 2013. PhD thesis, with permission from Dr Jacob Holland.***

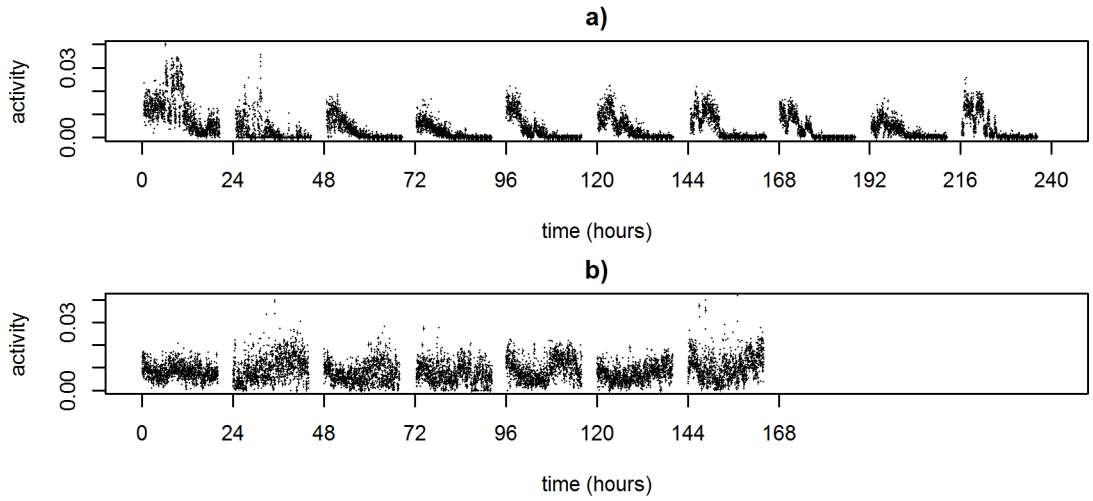


**Figure 5.3: Actograms of two *Bombus terrestris* queens whose activity levels were monitored using Object Tracker Software. a) A non-laying queen with significant circadian rhythmicity that was monitored for ten days. b) A laying queen with no circadian rhythmicity that was monitored for seven days. The regularly spaced zero activity values represent the downtime between each activity recording session (four-hour periods between 1300-1345 and 1700-1745 each day). *Image reproduced and adapted from Holland. 2013. PhD thesis, with permission from Dr Jacob Holland.***

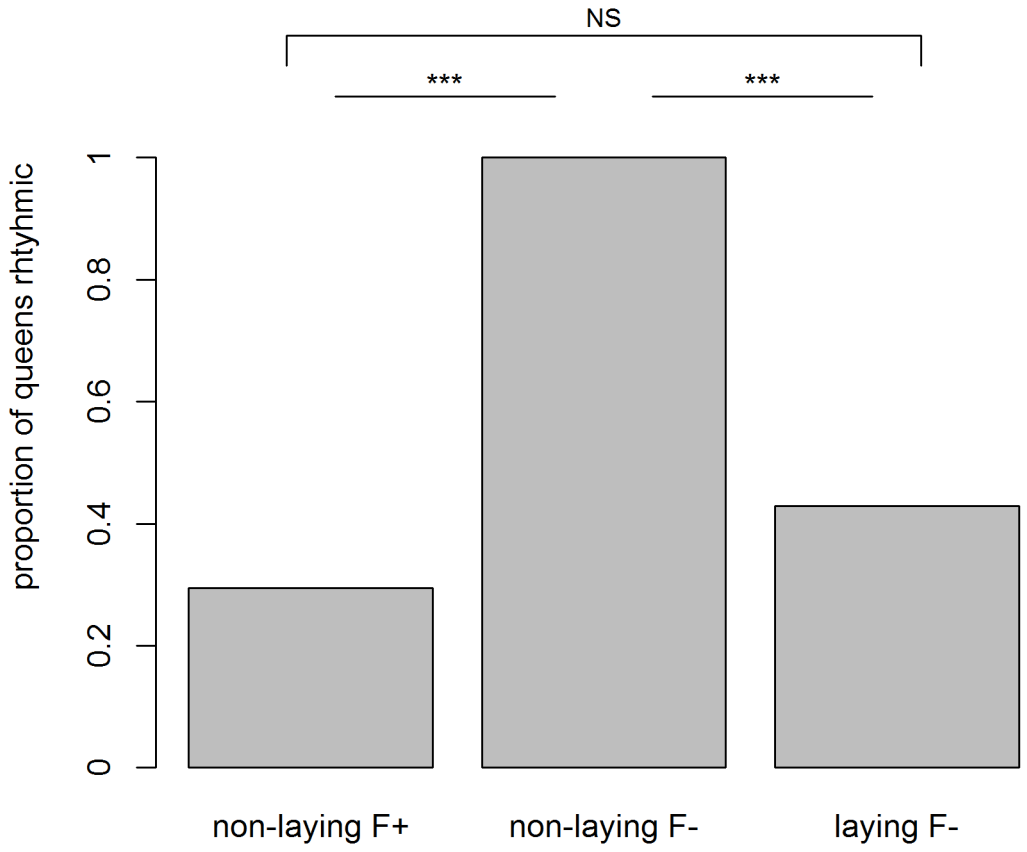


**Figure 5.4: Circadian rhythmicity in treatment groups of *Bombus terrestris* queens with differences in laying status and feeding regime. Non-laying F+ are non-laying queens with ad libitum food (n = 17, 'Preliminary trial' queens in Table 5.1); non-laying F- are non-laying queens with restricted food (n = 46, 'Treatment' queens in Table 5.1); laying F- are laying queens with restricted food (n = 6, 'laying' queens in Table 5.1). NS represents differences that are not significant (chi-squared test); *** $p < 0.001$ (chi-squared test). *Image reproduced and adapted from Holland. 2013. PhD thesis with permission, from Dr Jacob Holland.***

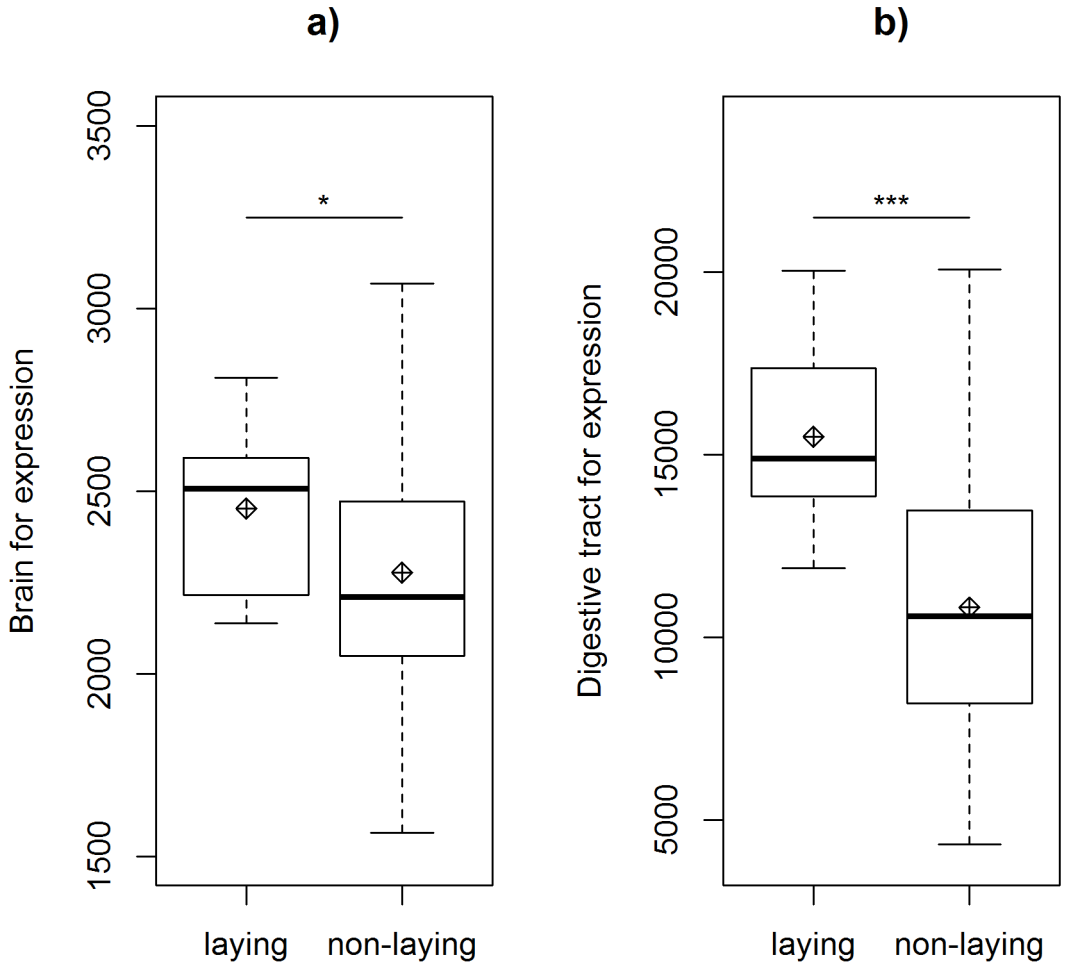


**Figure 5.5: Expression levels (copy number) of *for* in *Bombus terrestris* queens with different laying status's. a) In brain. b) In digestive tract. Gene expression level (copy number) was estimated using qRT-PCR. Expression level for *for* were calculated relative to mean expression level of two reference genes, *ArgK* and *PLA2*. Sample sizes of treatments: laying queens (n = 6), non-laying queens (n=66). The brain expression of one laying queen was not measured, and so is not included in a). Diamonds indicate mean expression values, thick horizontal bars median values, boxes interquartile ranges, and dotted lines ranges excluding outliers. * $p < 0.05$ (ANOVA); *** $p < 0.001$ (ANOVA). *Image reproduced and adapted from Holland. 2013. PhD thesis, with permission from Dr Jacob Holland.***

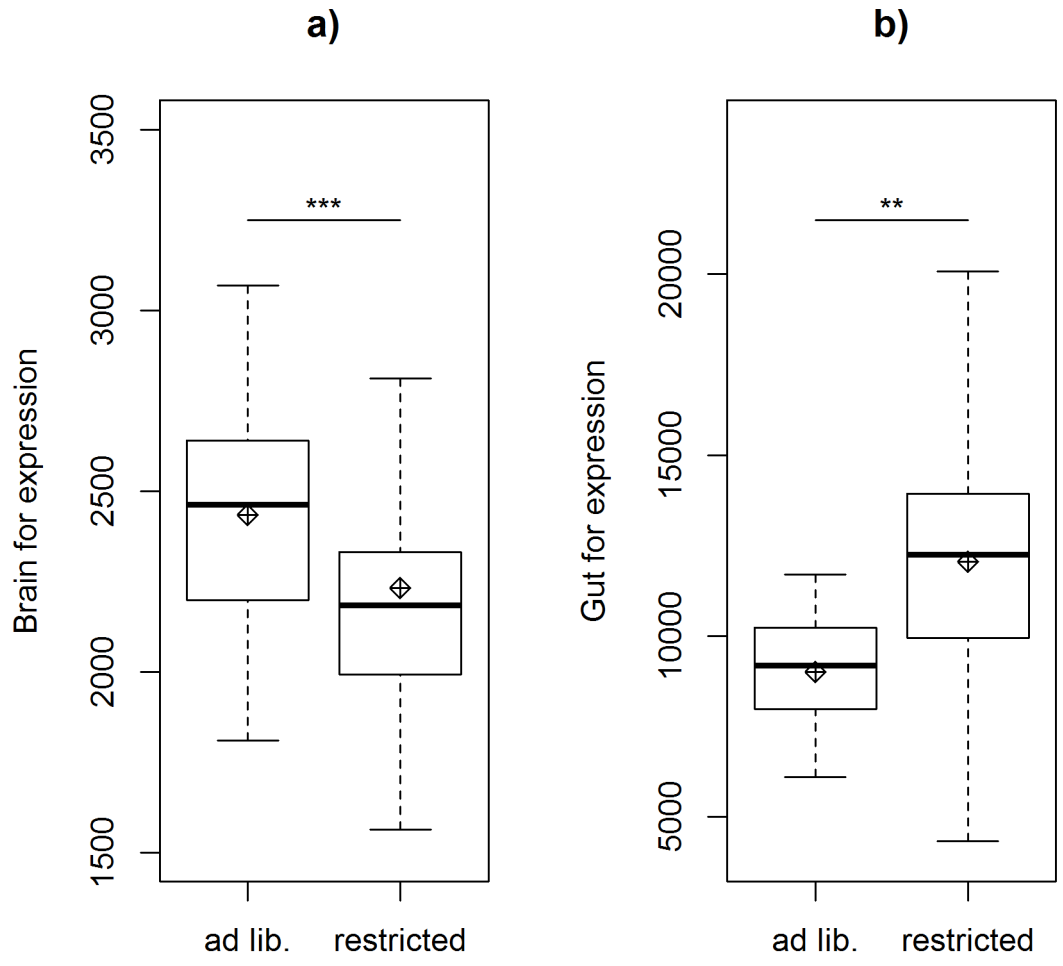


**Figure 5.6: Expression levels (copy number) of *for* in *Bombus terrestris* queens with different feeding regimes. a) In brain. b) In digestive tract. Gene expression level (copy number) was estimated using qRT-PCR. Expression level for *for* were calculated relative to mean expression level of two reference genes, *ArgK* and *PLA2*. Sample sizes of treatments: ad libitum (n = 20), restricted intake (n=52). We measured for expression in the a) brains and b) digestive tract of queens fed ad libitum ('Early' and 'Late' queens on Table 5.1, n = 20) and queens fed with restricted nectar sources ('Laying' and 'Treated,' queens on Table 5.1, n = 52). Diamonds indicate mean expression values, thick horizontal bars median values, boxes interquartile ranges, and dotted lines ranges excluding outliers. * $p < 0.05$ (ANOVA); *** $p < 0.001$ (ANOVA). ** $p < 0.01$; *** $p < 0.001$. *Image reproduced and adapted from Holland. 2013. PhD thesis, with permission from Dr Jacob Holland.***

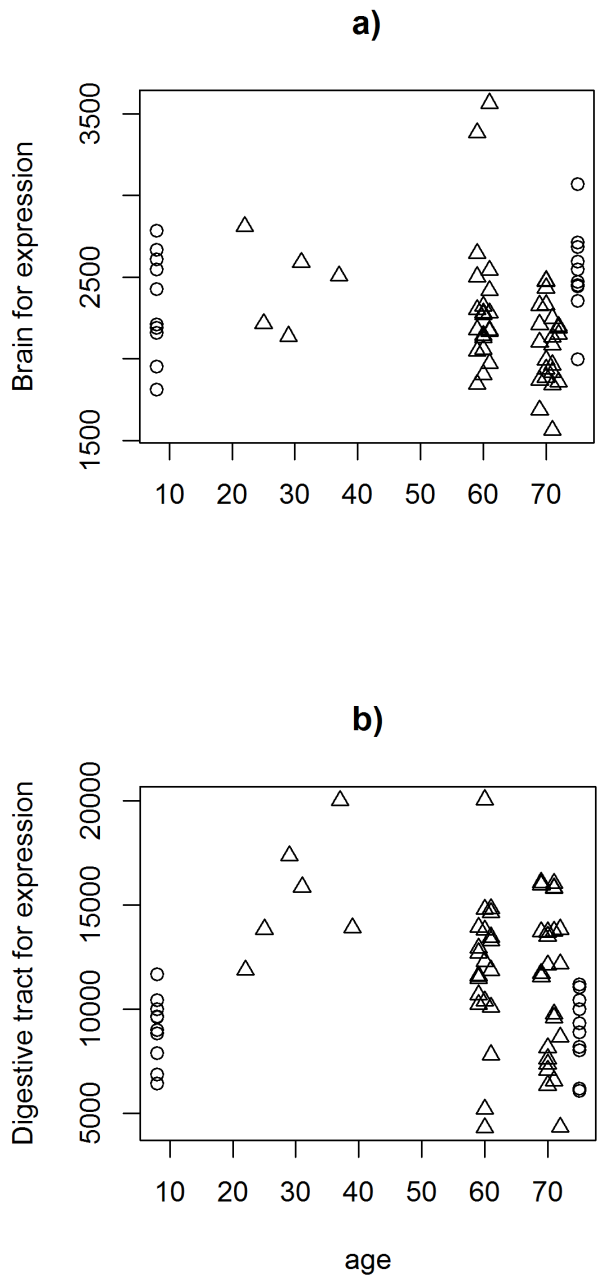


**Figure 5.7: The relationship between age in days since the start of the experiment (when the ordered queens arrived) until the day of dissection, and *for* gene expression (copy number) in *Bombus terrestris* queens. a) In brain. b) In digestive tract. Gene expression level (copy number) was estimated using qRT-PCR. Expression level for *for* were calculated relative to mean expression level of two reference genes, *ArgK* and *PLA2*. Triangles represent queens on a restricted diet (n = 52) and circles queens fed ad libitum (n = 20). *Image reproduced and adapted from Holland. 2013. PhD thesis, with permission from Dr Jacob Holland.***

**Table 5.1: Primers used for RNAi and qRT-PCR in the experimental *Bombus terrestris* queens**

| Primer | Sequence |
|---|---|
| *for* forward | GGATCTTCGACCACTGGCTA |
| *for* reverse | GCAAAAGATTCTCCGGTTTG |
| T7 *for* forward | TAATACGACTCACTATAGGGAGAGGATCTTCGACCACTGGCTA |
| T7 *for* reverse | TAATACGACTCACTATAGGGAGAGCAAAAGATTCTCCGGTTTG |
| *gfp* forward | GCCACAAGTTCAGCGTGTCC |
| *gfp* reverse | TTCTGCTTGTCGGCCATGAT |
| T7 *gfp* forward | TAATACGACTCACTATAGGGAGAGCCACAAGTTCAGCGTGTCC |
| T7 *gfp* reverse | TAATACGACTCACTATAGGGAGATTCTGCTTGTCGGCCATGAT |
| Reverse transcription universal primer | T(20)VN |
| *ArgK* qPCR forward | TGACCCAGCTGGTGAGTTC |
| *ArgK* qPCR reverse | TACTGAGCTTCAGTCAGGCAG |
| *ArgK* qPCR probe | TGTAAGATGCGGCCGTTCCCTGGAA |
| *for* qPCR forward | ATCGATCACACGAAATGCAACT |
| *for* qPCR reverse | TGTATTTCGCTAATGCCACCTT |
| *for* qPCR probe | CAGGATTATCCCTGCACAGCTT |

**Table 5.2: Sample sizes, feeding regimes, age range (days since the start of the experiment, when the ordered queens arrived, until the day of dissection) and data collected for different experimental groups of *Bombus terrestris* queens. Preliminary trial queens were. Preliminary trial queens were used in the initial trials to verify that the tracker software could detect queen locomotor activity. Laying queens include all of the queens in experiment one that laid eggs. Treatment queens include all of the queens that were treated with dsRNA (*for* or *gfp*) or water and that were then tracked following treatment. Early queens include the ten queens that were dissected, and that had their *for* gene expression measured approximately one week after the start of the experiment, and late queens were dissected and had their *for* gene expression measured at the end of the experiment.**

| | **Preliminary trial queens** | **Laying queens** | **Treatment queens** | **Early queens** | **Late queens** |
|---|---|---|---|---|---|
| Sample size | 17 | 7 | 46 | 10 | 10 |
| Feeding regime | ad libitum | treatment with restricted syrup | treatment with restricted syrup | ad libitum | ad libitum |
| Age range | 15-43 | 22-62 | 59-72 | 8 | 75 |
| Data collected | activity levels | activity levels and *for* expression | activity levels and *for* expression | *for* expression | *for* expression |

# Chapter 6

# General conclusions

In this thesis I aimed to investigate RNA interference (RNAi) and its relationship with caste and sociality in eusocial insects, using the bumble bee, *Bombus terrestris* as a model system. The results of this research have contributed to existing knowledge of sociogenomics and the role of RNAi mechanisms in eusocial insects, in particular the role of microRNAs (miRNA) in caste determination and differentiation in *Bombus* larvae and adults. In this chapter I provide an overview of the main findings of this thesis. For each chapter, I will also give a general discussion of the importance of the findings to the field of social genomics, and relate them to the questions considered about miRNA research in **Chapter 1**. Finally I will provide some suggestions for future work given the results of the studies outlined in this chapter.

*Conclusions from Chapter 2: miRNA annotation*

MiRNAs are small RNAs that regulate gene expression at the post-transcriptional level (Bartel 2004). Previous studies have implicated them in the social behaviour of eusocial insects (species with reproducer and worker phenotypes) such as *A. mellifera* (Weaver et al. 2007; Behura and Whitfield 2010; Greenberg et al. 2012; Guo et al. 2013). *B. terrestris* is another eusocial insect that has become an important model for eusociality (Goulson 2003). Recently the genomes of *B. terrestris* and another bumble bee, *B. impatiens*, have been sequenced to permit the study of social processes at the genomic level (Stolle et al. 2011; Sadd et al. 2015). In **Chapter 2** we used deep sequencing of the miRNAs (from now on we refer to this as miRNA-seq) of *B. terrestris* larvae and bioinformatics prediction programs to annotate the miRNAs of the two newly-sequenced genomes, establishing a comprehensive list of 115 conserved and 16 new miRNAs in *B. terrestris*, and 115 conserved and 11 new miRNAs in *B. impatiens*. We compared these miRNAs against the miRNAs in the previously sequenced honey bee (*Apis mellifera*) genome and found that *A. mellifera* had a large number (103) of miRNAs that are not conserved in either of the bumble bee genomes. We attributed this result to two non-mutually exclusive possibilities: 1) the miRNAs of *A. mellifera* have rapidly evolved as part of the transition towards a more

complex form of eusociality; or 2) the miRNAs of *A. mellifera* have been poorly annotated, which has inflated the apparent number of miRNAs in this species.

The results presented in this chapter contributed to the ongoing *Bombus* genome project (Sadd et al. 2015), which aimed to sequence the genomes of two *Bombus* species and then compare them with the genomes of another social bee, *A. mellifera* (Weinstock et al. 2006; Elsik et al. 2014). This was important for the field of socio-genomics because most of the insect genomes that have been sequenced so far are from advanced eusocial species (species with a complex division of labour, large colonies, and morphologically differentiated castes) or from solitary species. Bumble bees are interesting because they represent an intermediate stage between primitive eusociality (species with a simple division of labour, and castes that are not morphologically differentiated) and advanced eusociality. They have morphologically differentiated castes, and task specialisation between nurse specialist workers and forager specialist workers, however they also have small colonies and the reproductive division of labour is less pronounced than in other eusocial insects (Goulson 2003). They also share a common eusocial ancestor with *A. mellifera* (Cardinal and Danforth 2011). Therefore genomic differences between the two groups are of great interest because they might be of importance in generating their very different social structures. Our finding that 103 of the miRNAs of *A. mellifera* were not found in the *Bombus* genomes could therefore suggest that miRNAs are important in the evolution of advanced eusociality in insects, and be informative as regards the very different social systems of the two groups. Weaver et al. (2007) have shown that some miRNAs (e.g. Ame-miR-2, Ame-miR-9a, and Ame-miR-71) are differentially expressed between queen-destined larvae, pupae, and adults, while Guo et al. (2013) have shown that a miRNA (Ame-miR-184) can induce queen-destined larvae to acquire worker-like traits. MiRNAs have also been implicated in the evolution of complexity more generally (Lee et al. 2007; Berezikov 2011). One way miRNAs could be involved in the transition towards advanced eusociality is through the evolution of a tighter system of regulation during development in the evolution of morphologically differentiated castes. Therefore it is possible that the advanced eusocial *A. mellifera* has evolved a larger number of miRNAs to make development more highly regulated in this species.

There has been little work on the functions of new miRNAs, so a further test of the hypothesis that miRNAs are involved in the evolution of advanced eusociality should focus on understanding the genes that the new miRNAs of both groups (*Apis* and *Bombus*) target. Furthermore, while we sequenced new miRNAs in the larvae of *B. terrestris*, we have not investigated miRNA expression in the range of tissues and phenotypes that have been investigated in *A. mellifera*. Further research could therefore profitably investigate miRNA expression in adult *B. terrestris* and *B. impatiens*. This process has been started in *B. terrestris* (see **Chapter 4**), and these results could be used to discover new miRNAs that assist in adult functions.

*Conclusions from Chapter 3: miRNAs in caste determination*

The existence of differentiated queen and worker castes in eusocial insects represents an interesting phenomenon in biology because in most cases all the individuals start off totipotent (capable of developing along either caste pathway), but their caste fate becomes fixed in later development (Schwander et al. 2010). In **Chapter 3** we used miRNA-seq and Northern blot validation to show that this loss in totipotency is associated with differences in miRNA gene expression, with miR-6001-5p being much more highly expressed in late-instar queen-destined larvae than in worker-destined larvae. These differences in miRNA expression only occurred after the point at which larvae change from being totipotent to having their caste set along a single caste pathway (Cnaani et al. 2000), implying that miRNA expression at that point is closely associated with queen-worker determination. This represents the first time that a miRNA (Bte-miR-6001) has been unambiguously associated with the loss of totipotency in a eusocial insect, the first well validated miRNA that has been associated with caste differences in a eusocial insect, and the first time that miRNAs have been associated with caste in a bumble bee. This provides evidence that miRNAs have a role in the caste determining and differentiating mechanisms in eusocial insects.

This work follows on from previous experiments in *A. mellifera* and *B. terrestris*, which showed that several genes are differentially expressed between queens and workers, with many of these genes being likely to form part of the caste determination process itself (Evans and Wheeler 1999; Evans and Wheeler 2001; Pereboom et al. 2005; Colgan et al. 2011; Johnson and Tsutsui 2011; Chen et al. 2012; Harrison et al. accepted). Similar studies

have identified the miRNAs that are differentially expressed in adult *A. mellifera* nurse specialist workers and forager specialist workers (Behura and Whitfield 2010; Greenberg et al. 2012; Liu et al. 2012), and together with this study they show that post-transcriptional mechanisms are likely to be important for caste development. This is an important insight for sociogenomics because to fully understand the molecular mechanisms that underpin the development and evolution of castes, the post-transcriptional mechanisms, as well as the transcriptional mechanisms, should be considered.

This study showed that miRNAs are associated with caste determination, but did not test for the specific functions of caste-associated miRNAs. Further work should focus on the association between the miRNAs that were differentially expressed between queen-destined and worker-destined larvae in this study, with the genes that were differentially expressed between the adult castes in other studies (Pereboom et al. 2005; Colgan et al. 2011; Woodard et al. 2013; Harrison et al. accepted). Once an association has been found between the two groups of genes it will be possible to identify whether miRNAs have a direct causal effect on caste determining mechanisms in eusocial insects by targeting the causative caste genes for regulation. One way to do this is through RNA knockdown, using dsRNA to target the miRNA target genes (Terenius et al. 2011; Scott et al. 2013; Yu et al. 2013), or antagomirs which target the miRNAs themselves (Krutzfeldt et al. 2005), to test whether either of these has a direct effect on caste determination.

*Conclusions from chapter 4: miRNAs in reproductive division of labour*

Eusocial insect castes have a reproductive division of labour whereby most of the reproduction is performed by queens, while colony maintenance and food provisioning are performed by the workers. These are the most important biological differences between the castes, and, in bumble bees, they are only manifested in adulthood. While our previous study (**Chapter 3**) showed that the differences in larval development are associated with miRNA expression, there had been no research on whether miRNAs are associated with reproductive differentiation in adulthood. Instead, previous studies on the miRNAs in adult *A. mellifera* have used miRNA-seq to test whether miRNAs are differentially expressed between nurse specialist workers and forager specialist workers. They have shown that several miRNAs (e.g. Ame-miR-184, Ame-miR-133) are more highly expressed in forager

specialists (Behura and Whitfield 2010; Greenberg et al. 2012; Liu et al. 2012), and other miRNAs (e.g. Ame-miR-279b, Ame-miR-279c) being more highly expressed in nurse specialists (Liu et al. 2012), although the absolute number tends to vary between studies. Therefore, the first main aim of **Chapter 4** was to use miRNA-seq to sequence the miRNAs expressed in the ovaries and brains in *B. terrestris* adult females. Our second aim was to use the miRNA-seq results, combined with Northern blot validation, to test whether miRNAs are associated with the reproductive division of labour in adulthood, again using *B. terrestris* as a model system. We showed that the queen-worker distinction was also associated with miRNA expression differences in adult bumble bees; of particular interest was Bte-miR-184, which was upregulated in ovaries of workers compared to queens.

In addition to this there exists a second reproductive division of labour between the workers themselves in many eusocial Hymenopteran species. In these species the workers can activate their ovaries and lay unfertilised eggs that develop into males (Duchateau and Velthuis 1988; Duchateau and Velthuis 1989; Goulson 2003). Bumble bee workers are capable of activating their ovaries, and towards the end of the colony cycle a number of workers start to lay eggs. Therefore the third main aim of **Chapter 4** was to test whether miRNAs were involved in regulating the reproductive division of labour that occurs between reproductive and non-reproductive workers, and whether this is related to the miRNAs that are associated with queen-worker caste differentiation. We found evidence that some miRNAs, including Bte-miR-184, Bte-miR-279b, Bte-miR-279c, Bte-miR-133, and Bte-miR-9a are associated with maintaining different ovary activation states in workers. In particular, Bte-miR-184 was interesting because, as well as being associated with queen-worker caste differentiation, it proved to be associated with the reproductive division of labour within the worker caste, and was therefore most highly expressed in inactive-ovary workers compared to active-ovary workers, and more highly expressed in active-ovary workers compared to queens.

This work is of particular interest to the field of socio-genomics because the expression patterns of many of the miRNAs of interest (e.g. Bte-miR-184, Bte-miR-279b, Bte-miR-279c, and Bte-miR-133) fulfil some of the collective predictions of the ovarian and reproductive ground plan hypotheses (West-Eberhard 1987; Amdam et al. 2006). These predict that genes associated with reproduction are linked to maternal care behaviour, while genes

associated with maintaining non-active ovaries will be linked to behaviours that are unique to social species, e.g. foraging behaviour. We found that Bte-miR-184 and Bte-miR-133, whose homologues have previously been shown to be upregulated in forager specialist *A. mellifera* workers compared to nurse specialist workers, were upregulated in inactive-ovary *B. terrestris* workers compared to active-ovary workers and queens. We also found that the *A. mellifera* homologues of Bte-miR-279b and Bte-miR-279c, have previously been shown to be upregulated in nurse specialist *A. mellifera* workers compared to forager specialist workers, and were upregulated in *B. terrestris* queens and active-ovary workers compared to inactive-ovary workers. More generally, if this hypothesis was shown to be correct, it would show how alternative body types and behaviours evolve through the co-option of ancestral genes with different functions. An alternative but non-mutually exclusive hypothesis is that eusociality evolved through the rapid evolution of new genes which perform most of the functions associated with eusociality (Ferreira et al. 2013). Our prediction methods did not specifically investigate the importance of new miRNAs in the caste differentiation processes of *Bombus*, but further work would help to elucidate 1) whether new miRNAs are involved in the caste differentiation processes of *B. terrestris* and 2) the relative importance of new miRNAs and conserved miRNAs such as miR-184 and miR-279 in the transition to eusociality from a solitary ancestor. One way to test the importance of each set of genes would be to identify their targets (using target prediction software such as *TargetScan*) and test whether these targets have important causative roles in the evolution of eusociality, using *Luciferase* assays which would validate which mRNAs are the most likely targets of caste and reproduction associated miRNAs such as miR-184 and miR-279b (Akgül and Göktaş 2014), and miRNA knockdown experiments which would show their effects on *B. terrestris* reproductive differentiation (Krutzfeldt et al. 2005).

*Conclusions from Chapter 5: RNA interference*

In addition to being a part of miRNA regulation, RNAi can be used as a laboratory tool to control gene expression of a target gene experimentally. This has been achieved in many insects, including eusocial bees such as *A. mellifera* (Jarosch and Moritz 2011), but until recently it has not yet been achieved in bumble bees (Deshwal and Mallon 2014). In **Chapter 5** our aim was to establish a protocol for inducing gene knockdown (artificial down-regulation of a target gene using RNAi) in *B. terrestris* by oral administration of double-

stranded RNA to queens. Specifically we targeted the gene *foraging* (*for*), which has previously been associated with feeding behaviour in solitary species such as *Drosophila* (Debelle and Sokolowski 1987; Pereira and Sokolowski 1993) and with the division of labour in eusocial insects (Tobback et al. 2011; Rodriguez-Zas et al. 2012). We hypothesised that the gene was downregulated in queens during the early part of the colony cycle when queens make the behavioural transition from external foraging to remaining in the nest while workers take over all foraging (Goulson 2003). We tested the effect of dsRNA treatment on queen behaviour (measuring locomotor activity and circadian rhythmicity). The dsRNA treatment did not induce knockdown, and there was no evidence of an association between foraging and queen behaviour. However, there was evidence of an association between *for* gene expression and the laying status of queens, showing that laying queens had a higher degree of *for* expression in brains and digestive tracts compared to non-laying queens. The factors affecting for gene expression were complex because we also showed that the feeding regime of queens had significant effects on *for* expression, with opposite effects being found in different tissues.

It is still not clear why the RNAi strategy did not induce knockdown, and further experiments will need to consider alternative methods and possibly different gene targets in order to utilise this technology in *B. terrestris*. This would greatly assist the development of this group as a model system for the study of the molecular evolution of sociality. In our study we had low sample sizes and only a small effect size. Therefore the associations found should be explored in further detail to identify whether *for* expression has a causal role in queen laying behaviour.

*Concluding remarks*

In **Chapter 1** I outlined some of most important areas that have not been addressed with regards to miRNAs and social insect genomics, and throughout this thesis I have attempted to address some of these problems. To recap, these problems, and the relevant Chapters that address them, are as follows: 1) previously, there had been no comparative work on miRNAs in social species other than *A. mellifera* (and some brief work in *Camponotus* species), therefore, whether miRNA regulation of caste polyphenism is confined to this single eusocial lineage was unclear. In **Chapters 2**, **3** and **4**, we used deep sequencing, and

bioinformatics to identify 131 miRNAs in *B. terrestris* and have shown that the miRNAs of *A. mellifera* differ considerably from *B. terrestris* and that they are therefore potentially important in eusocial evolution (**Chapter 2**). In Chapter 3 and Chapter 4 we showed that miRNAs (e.g. Bte-miR-6001-5p and Bte-miR-184) are differentiated between workers and queens, showing that caste differentiation of miRNAs is not confined to *A. mellifera*; 2) No studies of miRNAs in social insects had characterised whether miRNAs were differentially expressed at the point when their castes first start to differentiate. In **Chapter 3**, we showed using miRNA-seq and validation with Northern blots that miR-6001-5p and miR-6001-3p were more highly expressed in the late-instar queen-destined larvae relative to worker-destined larvae in *B. terrestris*, following the loss of totipotency during larval development; 3) No studies of miRNAs had addressed the reproductive division of labour that occurs within the worker caste. In **Chapter 4**, we used miRNA-seq and Northern blot validation to show that Bte-miR-184 was more highly expressed in worker adults than in queens, and that Bte-miR-184 and Bte-miR-133 were more highly expressed in inactive-ovary workers compared to active-ovary workers and queens, while Bte-miR-279b, Bte-miR-279c, Bte-miR-92b, and Bte-miR-9a were more highly expressed in active-ovary workers and queens compared to inactive-ovary workers; 4) Several previous studies of miRNAs in social insects have shown differential expression through miRNA-seq data, but they used a single replicate per phenotype and their results were not validated. Throughout this thesis, in **Chapters 2**, **3**, and **4**, we have used a combination of miRNA-seq to identify differentially expressed sequences, and Northern blots to validate their expression patterns in independent colonies. We have expanded the number of replicates (four colony replicates per phenotype in **Chapter 3**, three colony replicates per phenotype in **Chapter 4**) for both sets of methods to account for colony to colony variation. Our methods have confirmed that at least Bte-miR-6001 and Bte-miR-184 are associated with caste differentiation in *B. terrestris* and further work should likewise be able to confirm the validity of miRNA-associated caste differentiation in *A. mellifera* and other eusocial insects; 5) Finally, very few studies have characterised the role of miRNAs and identified the tissue- and stage-specific patterns of gene expression in eusocial insects. In **Chapter 3** and **4**, we addressed this issue by identifying expression of miRNAs in specific tissues in *B. terrestris* (in **Chapter 3** we used Northern blots to qualify Bte-miR-6001 in head, digestive tract, and cuticle; in **Chapter 4** we used miRNA-seq and Northern blots to qualify miRNA expression

in brains and ovaries). In **Chapter 3**, we also used Northern blots to identify the expression of Bte-miR-6001 throughout larval and pupal development.

In this thesis I have shown that miRNAs are associated (in some cases strongly) with many of the behavioural and developmental mechanisms of reproductive division of labour in eusocial insects, using *B. terrestris* as a model system. I have identified miRNAs that are newly evolved in diverging eusocial lineages (**Chapter 2**), associated with the caste determination mechanisms of *B. terrestris* larvae (**Chapter 3**), and associated with caste and reproduction in *B. terrestris* adults (**Chapter 4**). These results indicate that miRNAs have a major role in the regulation of eusocial traits and that they might therefore be important factors in the proximate mechanisms involved in the evolution of eusociality.

*References*

Akgül B, Göktaş C. 2014. Gene reporter assay to validate microRNA targets in *Drosophila* S2 cells. In *miRNomics: MicroRNA Biology and Computational Analysis Methods in Molecular Biology*, Vol 1107 (ed. Y M, A J), pp. 233-242. Humana Press, New York, USA.

Amdam GV, Csondes A, Fondrk MK, Page RE. 2006. Complex social behaviour derived from maternal reproductive traits. *Nature* **439**: 76-78.

Bartel DP. 2004. MicroRNAs: Genomics, biogenesis, mechanism, and function. *Cell* **116**: 281-297.

Behura SK, Whitfield CW. 2010. Correlated expression patterns of microRNA genes with age-dependent behavioural changes in honeybee. *Insect Mol Biol* **19**: 431-439.

Berezikov E. 2011. Evolution of microRNA diversity and regulation in animals. *Nat Rev Genet* **12**: 846-860.

Cardinal S, Danforth BN. 2011. The antiquity and evolutionary history of social behavior in bees. *PLOS One* **6**: e21086

Chen X, Hu Y, Zheng HQ, Cao LF, Niu DF, Yu DL, Sun YQ, Hu SN, Hu FL. 2012. Transcriptome comparison between honey bee queen- and worker-destined larvae. *Insect Biochem and Mol Biol* **42**: 665-673.

Cnaani J, Robinson GE, Hefetz A. 2000. The critical period for caste determination in Bombus terrestris and its juvenile hormone correlates. *J Comp Physiol A* **186**: 1089-1094.

Colgan TJ, Carolan JC, Bridgett SJ, Sumner S, Blaxter ML, Brown MJF. 2011. Polyphenism in social insects: insights from a transcriptome-wide analysis of gene expression in the life stages of the key pollinator, *Bombus terrestris*. *BMC Genomics* **12**: 623.

Debelle JS, Sokolowski MB. 1987. Heredity of rover sitter - alternative foraging strategies of *Drosophila melanogaster* larvae. *Heredity* **59**: 73-83.

Deshwal S, Mallon EB. 2014. Antimicrobial peptides play a functional role in bumblebee anti-trypanosome defense. *Dev Comp Immunol* **42**: 240-243.

Duchateau MJ, Velthuis HHW. 1988. Development and reproductive strategies in *Bombus* colonies. *Behaviour* **107**: 186-207.

-. 1989. Ovarian development and egg-laying in workers of *Bombus terrestris*. *Entomol Exp Appl* **51**: 199-213.

Elsik CG, Worley KC, Bennett AK, Beye M, Camara F, Childers CP, de Graaf DC, Debyser G, Deng JX, Devreese B et al. 2014. Finding the missing honey bee genes: lessons learned from a genome upgrade. *BMC Genomics* **15**: 86.

Evans JD, Wheeler DE. 1999. Differential gene expression between developing queens and workers in the honey bee, *Apis mellifera*. *Proc Natl Acad Sci USA* **96**(10): 5575-5580.

-. 2001. Expression profiles during honeybee caste determination. *Genome Biol* **2**: 1.

Ferreira PG, Patalano S, Chauhan R, Ffrench-Constant R, Gabaldon T, Guigo R, Sumner S. 2013. Transcriptome analyses of primitively eusocial wasps reveal novel insights into the evolution of sociality and the origin of alternative phenotypes. *Genom Biol* **14**: R20.

Goulson D. 2003. *Bumblebees: Behaviour, ecology and conservation*. Oxford University Press, Oxford, UK.

Greenberg JK, Xia J, Zhou X, Thatcher SR, Gu X, Ament SA, Newman TC, Green PJ, Zhang W, Robinson GE et al. 2012. Behavioral plasticity in honey bees is associated with differences in brain microRNA transcriptome. *Genes Brain Behav* **11**: 660-670.

Guo XQ, Su SK, Skogerboe G, Dai SAJ, Li WF, Li ZG, Liu F, Ni RF, Guo Y, Chen SL et al. 2013. Recipe for a busy bee: microRNAs in honey bee caste determination. *PLOS One* **8**: e81661.

Harrison MC, Hammond RL, Mallon EB. (accepted). Reproductive workers show queen-like gene expression in an intermediately eusocial insect, the buff-tailed bumble bee Bombus terrestris. *Mol Ecol*.

Jarosch A, Moritz RFA. 2011. Systemic RNA-interference in the honeybee *Apis mellifera*: Tissue dependent uptake of fluorescent siRNA after intra-abdominal application observed by laser-scanning microscopy. *J Insect Physiol* **57**: 851-857.

Johnson BR, Tsutsui ND. 2011. Taxonomically restricted genes are associated with the evolution of sociality in the honey bee. *BMC Genomics* **12**: 164.

Krutzfeldt J, Rajewsky N, Braich R, Rajeev KG, Tuschl T, Manoharan M, Stoffel M. 2005. Silencing of microRNAs in vivo with 'antagomirs'. *Nature* **438**: 685-689.

Lee CT, Risom T, Strauss WM. 2007. Evolutionary conservation of microRNA regulatory circuits: An examination of microRNA gene complexity and conserved microRNA-target interactions through metazoan phylogeny. *DNA Cell Biol* **26**: 209-218.

Liu F, Peng W, Li Z, Li W, Li L, Pan J, Zhang S, Miao Y, Chen S, Su S. 2012. Next-generation small RNA sequencing for microRNAs profiling in *Apis mellifera*: comparison between nurses and foragers. *Insect Mol Biol* **21**: 297-303.

Pereboom JJM, Jordan WC, Sumner S, Hammond RL, Bourke AFG. 2005. Differential gene expression in queen-worker caste determination in bumble-bees. *Proc R Soc B* **272**: 1145-1152.

Pereira HS, Sokolowski MB. 1993. Mutations in the larval *foraging* gene affect adult locomotory behavior after feeding in *Drosophila melanogaster*. *Proc Natl Acad Sci USA* **90**: 5044-5046.

Rodriguez-Zas SL, Southey BR, Shemesh Y, Rubin EB, Cohen M, Robinson GE, Bloch G. 2012. Microarray analysis of natural socially regulated plasticity in circadian rhythms of honey bees. *J Biol Rhythm* **27**: 12-24.

Sadd BM, Barribeau SM, Bloch G, de Graff DC, Dearden P, Elsik CG, Gadau J, Grimmelikhuijzen CJP, Hasselmann M, Lozier JD et al. 2015. The genomes of two key bumblebee species with primitive eusocial organization. *Genome Biol* **16**:76.

Schwander T, Lo N, Beekman M, Oldroyd BP, Keller L. 2010. Nature versus nurture in social insect caste differentiation. *Trends Ecol Evol* **25**: 275-282.

Scott JG, Michel K, Bartholomay LC, Siegfried BD, Hunter WB, Smagghe G, Zhu KY, Douglas AE. 2013. Towards the elements of successful insect RNAi. *J Insect Physiol* **59**: 1212-1221.

Stolle E, Wilfert L, Schmid-Hempel R, Schmid-Hempel P, Kube M, Reinhardt R, Moritz RFA. 2011. A second generation genetic map of the bumblebee *Bombus terrestris* (Linnaeus, 1758) reveals slow genome and chromosome evolution in the Apidae. *BMC Genomics* **12**: 48.

Terenius O, Papanicolaou A, Garbutt JS, Eleftherianos I, Huvenne H, Kanginakudru S, Albrechtsen M, An CJ, Aymeric JL, Barthel A et al. 2011. RNA interference in Lepidoptera: An overview of successful and unsuccessful studies and implications for experimental design. *J Insect Physiol* **57**: 231-245.

Tobback J, Mommaerts V, Vandersmissen HP, Smagghe G, Huybrechts R. 2011. Age- and task-dependent *foraging* gene expression in the bumble bee *Bombus terrestris*. *Arch Insect Biochem Physiol* **76**: 30-42.

Weaver DB, Anzola JM, Evans JD, Reid JG, Reese JT, Childs KL, Zdobnov EM, Samanta MP, Miller J, Elsik CG. 2007. Computational and transcriptional evidence for microRNAs in the honey bee genome. *Genome Biol* **8**: R97.

Weinstock GM Robinson GE Gibbs RA Worley KC Evans JD Maleszka R Robertson HM Weaver DB Beye M Bork P et al. 2006. Insights into social insects from the genome of the honeybee *Apis mellifera*. *Nature* **443**: 931-949.

West-Eberhard MJ. 1987. Flexible strategy and social evolution. In *Animal societies: theories and facts*, (ed. Y Ito, J Brown, J Kikkawa), pp. 35-51. Scientific Society Press, Tokyo, Japan.

Woodard SH, Bloch G, Band MR, Robinson GE. 2013. Social regulation of maternal traits in nest-founding bumble bee (*Bombus terrestris*) queens. *J Exp Biol* **216**: 3474-3482.

Yu N, Christiaens O, Liu JS, Niu JZ, Cappelle K, Caccia S, Huvenne H, Smagghe G. 2013. Delivery of dsRNA for RNAi in insects: an overview and future directions. *Insect Sci* **20**: 4-14.

## Glossary of key terms

Active-ovary worker: A worker bee that has activated its ovaries, and is therefore able to lay haploid eggs. This is usually to a much lesser extent than the queen.

Caste: The reproducer and non-reproducer phenotypes in eusocial organisms. These are often referred to as queen and worker castes respectively. Caste determination is the process where development towards a particular caste becomes irreversible. Caste differentiation is the process where development causes an individual to acquire characteristics that are unique to one caste.

cDNA: Complementary DNA, DNA which has been reverse transcribed from RNA.

Circadian rhythmicity: The state of having cycles in an activity (such as a behaviour) which has periods between cycles of approximately one day (normally 18 to 30 hours).

Competition point: In annual bumble bees, the process when workers begin to activate their ovaries and produce haploid eggs. They may also attack each other and the queen in order to establish dominance.

Gene: A conserved functional length of DNA on the genome of an organism. Genes can exist as regulatory elements (e.g. 3'UTRs), or can produce functional products such as interfering RNA (e.g. miRNAs) or proteins.

Eusociality: The state of having colonial living between con-specifics, non-overlapping generations, and a reproductive division of labour between members of the group. Includes primitive eusociality where the differences between the castes are behavioural rather than morphological, and advanced eusociality which have morphologically, physiologically, and behaviourally distinct castes.

Gene expression: The process where a functional product is synthesised from a gene. In the case of protein-coding genes this product is a protein. Throughout this thesis we have referred to reductions in 'gene expression' due to RNA interference. This equates to a reduction in the amount of protein produced from a target gene.

Haplodiploidy: A sex determination system where unfertilised eggs are haploid and hatch into males, while fertilised eggs are diploid and hatch into females. All members of the Hymenoptera order have a haplodiploid sex determination system.

Hymenoptera: The insect order that includes bees, ants, and wasps.

Inactive-ovary worker: A worker that maintains its ovaries in an inactive state, and is therefore unable to lay eggs.

MicroRNA (miRNA): 21-23bp length of RNA that induces gene silencing by binding to a complementary site on a target mRNA transcript and repressing it, preventing it from being translated into protein. MiRNAs are a sub-class of small RNA (sRNA) which are small non-coding lengths of RNA that have functions related to regulation of gene expression.

MiRNA-seq: High throughput sequencing using the Illumina method (formerly the Solexa method) to sequence reads between 16 nucleotides to 30 nucleotides in length.

Phenotypic plasticity: The ability of an organism to change phenotype in response to its external environment.

Polyethism: The functional specialisation of individuals in a social insect colony. One of the most well-known examples is age-polyethism in the honey bee *Apis mellifera*, the process by which worker task repertoire shifts with age, with workers acting act as brood nurses when young and switching as they grow older to external foraging.

Polyphenism: The ability of an organism to develop into multiple discreet phenotypes in response to changes in its external environment. These processes are often irreversible.

Queen: The female reproducer caste of a eusocial insect.

RNA interference (RNAi): The process where RNA molecules cause inhibition of gene expression by either repressing or degrading mRNA.

Switch point: The process, in annual bumble bees, where the queen switches from producing diploid eggs (which produce females) to haploid eggs (which produce males).

Worker: The female non-reproducer caste of a eusocial insect.

# Appendices

## Appendix 1

Supplementary methods

*RNA extraction (**Chapter 2** and **Chapter 3**)*

Prior to beginning the extractions we ensured that all of the work surfaces and pipettes were clean by wiping them with 10% Tris buffer (Fisher Scientific, Loughborough, UK), then again with 70% ethanol (Sigma, Gillingham, UK). All the work, reactions and centrifugation steps were performed at room temperature unless otherwise stated.

The RNA from the larvae-homogenate mixtures was extracted separately for each colony using Trizol reagent (Invitrogen, Carlsbad, California, USA). We extracted total RNA from the homogenate mixtures according to the Trizol manufacturer's instructions with some modifications. In brief, the tubes containing the homogenates were defrosted on ice then centrifuged at 10,000 × g. The supernatant was removed to a second tube. We added chloroform (200 µl for every 1000 µl of Trizol) to the tubes and shook the mixture vigorously for 15 seconds to ensure complete mixing of the phases. The mixture was left at room temperature for five minutes and then centrifuged 10,000 × g, at 4°C for 15 minutes. The upper aqueous phase containing the RNA was removed to a new tube, with care taken not to touch the white interphase. The upper phase was shaken again for 15 seconds with the same amount of chloroform (200 µl for every 1000 µl of Trizol). The centrifugation step was repeated once and the upper phase again removed to a new tube. Isopropanol was added to the solution (500 µl for every 1000 µl of Trizol used in the initial extraction step) and the tube was inverted several times to ensure complete mixing. The solution was kept at -20°C overnight.

The next morning we centrifuged the samples at 10,000 × g at 4°C for 20 minutes. The supernatant was removed and discarded. The pellet was washed twice with 80% ethanol by adding 1 ml, briefly flicking the tube to dislodge the pellet, then spinning in a centrifuge at 10,000 × g at 4°C for five minutes and removing the ethanol. After the second ethanol wash, the pellet was air-dried in a 25°C heating block and suspended in analytical reagent grade (ARG) water (ThermoFisher Scientific).

We quantified the total RNA using a Nanodrop 8000 spectrophotometer (ThermoFisher Scientific). We measured the concentration (ng/µl) and optical density of the RNA, in each case the second extraction produced RNA with a 260/280 ratio greater than 1.6 ensuring that the RNA was sufficiently pure for library preparations.

We measured the integrity of the total RNA by separating it on an agarose gel. To do this we diluted a sample of the RNA (>500 ng) 1:1 with stop mix (95% (v/v) formamide, 5 mM EDTA, 1.9 mM xylene cyanol, 1.5 mM bromophenol blue), heated it to 70°C for two minutes and kept the sample stored on ice. We prepared a 1.2% agarose gel using 0.5 × (v/v) Tris/Borate/EDTA (TBE) buffer (45 mM Tris, 45 mM Boric acid, 1 mM EDTA; pH 8.3) and stained with 10 mg/ml ethidium bromide. We samples separated the samples on the gel for one hour at 120 V. We scanned the gels using a transilluminator (UPV, Upland, California, USA) and analyzed the ribosomal bands using *Labworks Image Acquisition and Analysis software* (UPV). We used the presence of a single band corresponding to the small ribosomal sub-unit as an indicator of the amount of degradation in the RNA samples (Winnebeck et al. 2010). This confirmed that the RNA used for library preparations and Northern blots was not degraded.

*Illumina library sample preparation (**Chapter 2** and **Chapter 3**)*

To make the cDNA libraries we used a modified version of the TruSeq RNA sample preparation protocol (Illumina, Chesterford, UK). We used a modified 3' adaptor to reduce sequencing bias (Sorefan et al. 2012). To produce the modified adaptor, we obtained the non-adenylated 3' adaptor sequence (NNNNTGGAATTCTCGGGTGCCAAGG) from Sigma. We phosphorylated the sequence incubating 600 pmol of the adaptor for 30 minutes at 37°C with 20 units of T4 polynucleotide kinase (New England Biolabs, Ipswich, UK), 1 × kinase buffer, and 1 mM Adenosine-triphosphate (ATP). We precipitated the solution by adding 10 µg of glycogen (Ambion, Foster City, California, USA), 250 µl of ethanol, and Sodium Acetate (NaOAc) to a final concentration of 120 mM. We stored the solution at -20°C overnight.

The next day we centrifuged the solution at 10,000 × g and removed the supernatant. We washed the pellet with 250 µl of 80% ethanol, then centrifuged the solution for five minutes and re-suspended the pellet in 12 µl of ARG water.

We took a 100 pmol (4.5 µl) aliquot of the phosphorylated oligonucleotide, and then adenylated it combining it with 200 pmol of Mth RNA ligase (New England Biolabs), 1 × DNA Adenylation Reaction Buffer, and 100 µM ATP. We incubated the solution at 65°C for one hour, and then inactivated the enzyme by heating it to 85°C for five minutes. This process was repeated with a second aliquot of 100 pmol of phosphorylated oligonucleotide and we mixed the two solutions together. We mixed the resulting 80 µl solution with 200 µl acid equilibrated 1:1 phenol: chloroform, and 120 µl of ARG water, shaken for 15 seconds and centrifuged at 10,000 × g, at 4°C, for ten minutes. We removed 170 µl of the aqueous phase, taking care not to touch the white interphase, and made it up to 200 µl with ARG water. We then added 20 µg of glycogen, 120 mM NaOAc and 500 µl of ethanol to the solution. We mixed the solution by inverting the tube several times, and then left it at -20°C overnight.

The next day we centrifuged the solution at 10,000 × g for 15 minutes. We removed the supernatant, and washed the pellet with 500 µl of 80% ethanol, we centrifuged it again at 10,000 × g for five minutes. We removed the ethanol, and left the pellets in a 25°C heating block to air dry for five minutes. We suspended the pellets in 12 µl of ARG water. We then used a nanodrop to calculate the concentration of the adaptor and adjusted the final concentration to 10 µM with ARG water.

To test whether the adenylation was successful we loaded 0.5µl of the adenylated product on a 15% denaturing polyacrylamide gel alongside 50 pmols of the non-adenylated oligonucleotide, and a mixture of the two. The samples were prepared by making them up to 5 µl with ARG water, and then diluting them 1:1 with stop mix. They were heated to 70°C for two minutes, and then kept on ice. We prepared the gel using the mini PROTEAN III system (Bio-Rad, Hercules, California, USA). To make the gel we dissolved 6.3 g of urea in 8 ml of ARG water with 1.5 ml of 5 × TBE. We heated the solution for up to 30 seconds, until the urea had completely dissolved. We then added 5.5 ml of 40% (v/v) 19:1 acrylamide: bis solution (Bio-Rad). To induce polymerization we added 150 µl of 10% Ammonium persulphate (APS) and 7.5 µl of tetra-methyl-ethylenediamine (TEMED) to the solution. We then poured the entire solution in between the two glass plates on the gel dock system and added a 1 mm ten-well gel-comb. The gels were left to set for at least 20 minutes. We removed the combs and ran the gels by themselves at

100 volts for 30 minutes in 0.5 × TBE. We washed the urea out of the wells of each gel with 0.5 × TBE and loaded the samples alongside each other onto the gels. We ran the gels for another 60 minutes, making sure to re-wash the wells with 0.5 × TBE 20 minutes after the reaction started. The gels were then run for another 40 minutes. Once the reaction had finished we removed the gels and stained them with ethidium bromide to a final concentration of 1 mg/ml. We then photographed them on a Molecular imager FX pro plus (BIORAD) using the software *Quantity One*. A size shift in the adenylated product showed that the customized adaptor oligonucleotide was successfully adenylated.

We prepared the libraries themselves using the Truseq library preparation kit version 1.5 according to the manufacturer's protocol with some modifications. In all cases we prepared the libraries using half the volume recommended in the TruSeq library preparation kit's protocol. Unless otherwise stated all the reagents were provided with the kit. For each sample we used 1000 ng of the sRNA-enriched RNA. We added 1 µl of the modified 3' adaptor to the RNA and made up the solution to 3 µl using ARG water. We then heated the solution to 70°C for two minutes and kept it stored on ice. We incubated the RNA and the 3' adaptor sequence with 5 units of truncated RNA ligase 2 (New England Biolabs), 1 X T4 RNA ligation buffer and 10 units of RNase inhibitor (Ambion) at 28°C for one hour. The enzymes were heat denatured by incubating the reaction at 70°C for ten minutes. We then heated 1 µl of the 5' adaptor sequence (GUUCAGAGUUCUACAGUCCGACGAUC) to 70°C for two minutes, then immediately added 0.5 µl of 10 mM ATP and 5 units of RNA ligase 1 to the adaptor. We then combined the two solutions containing the 5' adaptor and the 3' adaptor-ligated RNA, incubating them at 28°C for one hour.

We added the reverse transcription primer (RTP; Illumina) to the adaptor ligated RNA and heated it at 70°C for two minutes. For each sample we prepared a reverse transcription master mix which was made from 1 × first strand RT buffer, 0.5 µl of 480 µM dNTPs (New England Biolabs), 1 × DTT, 20 units of RNase inhibitor, and 20 units of reverse transcriptase. We added the master mix to the adaptor ligated RNA and incubated the solution at 50°C for one hour; we then either kept the synthesised cDNA on ice or froze it at -80°C for long term storage.

We used PCR to amplify the cDNA libraries with unique Illumina index primers (1-16); making sure to use a different index for each library. For each library we made 20 µl PCR reactions using 2 µl of cDNA for each reaction and a PCR master mix containing 0.4 units of Phusion polymerase (New England Biolabs), 1 × Phusion HF buffer (New England Biolabs), 625 µM dNTPs, 1 µM of the Illumina RP1 forward primer and 1 µM of the Illumina index reverse primer in 20 µl reactions. To test the specificity of the PCRs, we made three reactions for each sample and varied the number of cycles at 13, 15 and 17 cycles. We loaded these onto an 8% non-denaturing polyacrylamide gel to identify the number of cycles required to produce the clearest product in each library. To make the gel we added 3 ml of 40% (v/v) 19:1 acrylamide: bis solution, 1.5 ml of 5 × TBE, and 10.5 ml of water to a 15 ml Falcon tube. We then induced polymerization by adding 150 µl of APS and 7.5 µl of TEMED and shaking the tubes. We poured the solution in between two glass plates of the mini PROTEAN III system with a 10 well 1 mm gel comb placed into them. We left the gel for at least 20 minutes to set, and then loaded the samples. We separated the samples on the gel for one hour 30 minutes at 120 V and then stained them with SyberGold. We then scanned the gels with a Molecular imager FX pro plus using the software *Quantity One*.

We selected the appropriate number of cycles for each sample independently, based on the clarity of the signal for each of the PCR products, making sure that the 21-23mer miRNA band was clearly visible and that the amount of adaptor-adaptor ligated product was kept to a minimum. We used the rest of the cDNA to make four more 20 µl PCR reactions with the appropriate index primers and PCR amplified them for the optimal number of cycles given from the results of the previous reaction. We separated the PCR products on an 8% Polyacrylamide gel, made in the same way as before, and then stained the gel with SyberGold. We scanned the gels using a Molecular imager FX pro plus and Quantity One software. We printed an exact sized copy of the scanned gel on standard printer paper, and then placed the gel on a transparent glass plate on top of the printout. As the gel and the plate were both transparent, we could infer where on the gel 21-23mer products were located using the printed copy of the scanned gel underneath them both. This meant that we could cut the correct bands from the gel using a razor. We rescanned the gel after cutting the bands to ensure that the entire product had been removed in the gel fragments.

The nucleic acid products were extracted from the gel in 1 X NEB buffer 2 (New England Biolabs). To do this we made a makeshift gel breaker column by piercing a 500 µl Eppendorf tube (Starlab, Milton Keynes, UK) twice at the bottom of the tube with a 25G 5/8 inch syringe needle (Fisher) and placing the tube into a 2 ml Eppendorf tube. We placed the cut gel fragments into the column and centrifuged it at 12,000 X G for 5 minutes until all of the gel fragments were collected in the 2ml Eppendorf tube. We discarded the makeshift column and added 300 µl of 1 X NEB buffer 2 to the tube; we then left it shaking overnight at 4°C.

The next morning we removed the gel debris from the elute by pipetting the entire solution containing the gel fragments into a Spin-x column (Fisher) which had been inserted into a clean Eppendorf tube. We centrifuged the solution for 2-5 minutes at 2800 X G, until all of the eluted nucleic acid was collected in the bottom tube.

We used ethanol-precipitation to concentrate the solutions to 12 µl. To do this we added 10 µg of Glycogen, 30 µl of 3M NaOAc, and 975 µl of 100% ethanol to the eluted solution. We incubated the solution at -80°C for 30 minutes, and then centrifuged it at 20,000 × G for 20 minutes at 4°C. Following this we removed the supernatant and washed the pellet with 500 µl of 80% ethanol. We re-centrifuged the solution again at 20,000 × G for five minutes, removed the supernatant, and then dried the pellet in a 25°C heating block for five minutes. We re-suspended the pellet in 12 µl of ARG water.

To reduce the chances of adaptor-adaptor contamination we repeated the entire process by separating the sample on a second 8% polyacrylamide gel, cutting it from the gel, and extracting it from the gel debris using all of the same procedures. We then separated 1/6$^{th}$ of the sample on a third 8% polyacrylamide gel to ensure the correct band was still present. We stored the rest of the sample at -80°C. We repeated the entire procedure for all 16 libraries.

*Sanger sequencing and Illumina deep sequencing (**Chapter 2** and **Chapter 3**)*

To ensure that the prepared cDNA libraries contained the correct RNA fragments we cloned some of the products into DH5α super-competent *Escherichia coli*, and employed Sanger chain-termination sequencing. To do this, we removed a fifth of library EW1 (this was the most highly concentrated library) and adjusted to it to 4 µl with ARG water. We then added

adenine overhangs to each end of the construct. To do this we incubated the sample with 5 units of GoTaq polymerase (Promega, Southampton, UK), 1 X GoTaq PCR buffer (Promega), 200 µM of dATP (New England Biolabs), and 1.5 mM of Magnesium Chloride (Promega) in a 10 µl reaction which was placed in a heat block at 70°C for 30 minutes. After adding the adenine overhangs the product was kept stored on ice.

We then ligated the cDNA library (with adenine overhangs) into the plasmid pGEMTeasy (Clontech, Saint-Germain-en-Laye, France). To do this we took 2 µl of the previous reaction and incubated it with 1 µl of the pGEMTeasy plasmid, 1 × pGEMT ligation buffer (Clontech) and 3 units of T4 ligase (Promega). We then transformed the plasmid into DH5α super-competent *E.coli* cells. For each transformation reaction, we defrosted a 25 µl tube of DH5α cells on ice. We added a volume of 5 µl of each ligation reaction to each tube of cells and gently pipetted them until the mixture was homogenous. We incubated the tubes on ice for 30 minutes. We then 'heat-shocked' the samples at 42°C for 50 seconds. We then incubated the samples on ice again for another two minutes. We diluted the samples in 950 µl of luria broth (LB) medium (1.0% Tryptone, 0.5% yeast extract, 1.0% Sodium chloride; pH 7.0).

We placed the solutions containing the LB and the cells on a shaker for 1 hour 30 minutes at 37°C. During this time we prepared agar plates for blue-white colony staining. These were made by melting 250 ml of agar in a microwave, cooling the solution to 60°C, and adding Ampicillin to a final concentration of 100 µg/ml, isopropyl-D-1-thiogalactopyranoside ( IPTG) to a final concentration of 80 µg/ml, and 5-bromo-4-chloro-3-indolyl-D-galactoside (X-GAL) to a final concentration of 0.5 mM. The solution was mixed by swirling gently. It was then poured into petri-dishes (roughly 25 ml per plate) and the plates left to cool for one hour. To prevent the plates becoming contaminated from external sources of microbes, they were kept in aseptic conditions next to a roaring Bunsen flame. We removed a 200 µl isolate of the cell suspensions and pipetted it onto the agar plates, which we then left in a 37°C incubator for 18 hours.

We selected ten white colonies from the agar plates using pipette tips (Starlab), which we then dropped into 15 ml falcon tubes containing LB and 100 µg/ml ampicillin. We placed these on a shaker for 12 hours at 37°C. We then centrifuged the tubes at 4000 rpm for ten

minutes. To extract the plasmid from the bacteria, we used the Qiaprep Spin Mini-prep kit (Qiagen, Manchester, UK) according to the manufacturer's instructions. At the end of the procedure we eluted the plasmid from a Qiagen Mini-prep spin column in 50 µl of ARG water.

For each extracted bacteria colony we prepared the samples using 'ready reactions' and sent the reactions to The Genome Analysis Centre (TGAC) for Sanger chain-termination sequencing. To prepare the ready reactions, we added 1 µl of DNA (approximately 100ng) to a PCR reaction with a final volume of 10% (v/v) Half BigDye Terminator reagent, 10% (v/v) BigDye Terminator reagent, 1 x BigDye Sequencing Buffer and 1 µM of M13 sequencing primer (CGCCAGGGTTTTCCCAGTCACGAC).

Once the sequences were returned from TGAC, we analysed them using the sequence analysis software *FinchTV*. We verified that of the ten colonies that were sent for sequencing, six of them contained miRNA length sequences which mapped to the *B. terrestris* genome (Bter20110317) and one contained the sequence for Bte-miR-1 (data not shown). This confirmed the presence of miRNAs in the Illumina libraries.

We sent prepared Illumina cDNA libraries to BaseClear B.V on dry ice for Illumina sequencing on a HiSeq2000.

*PAGE electrophoresis, Semi-dry transfer and Northern Blotting (**Chapter 2, Chapter 3 and Chapter 4**)*

We extracted RNA using the methods described above except that this time we did not sRNA enrich them with the miRVana kit, instead preparing the blots with total RNA. Following extraction we took 10 µg of total RNA from each treatment and diluted each with stop mix. We heated the solution at 70°C for 2 minutes and then stored the samples on ice prior to loading. We separated the total RNA on a 15% denaturing polyacrylamide gel. We left the gels to set for at least 20 minutes. We removed the combs and ran a current through the gels at 100 volts for 30 minutes in 0.5 X TBE. We then washed the urea out of the wells of each gel with 0.5 X TBE and loaded the samples onto the gel. We separated the RNA on the gels for 120 minutes at 100V, making sure to wash the wells with 0.5 X TBE a second time 20 minutes into the reaction. To confirm equal

loading, we stained the gels with 1 mg/ml ethidium bromide and photographed on a Molecular imager FX pro plus using the software *Quantity One*.

After the scan we transferred the RNA to a Hybond-NX nylon membrane (GE Healthcare) using Semi-dry membrane transfer apparatus (Scie-plas, Cambridge, UK). We soaked three BLT2 filter papers (Munktel, Falun, Sweden) in 0.5 X TBE and placed them on top of each other on the transfer apparatus. We labelled the nylon membranes with pencil and then soaked them in ARG water, followed by 0.5 X TBE, and then placed them on top of the three filter papers. We placed the gels containing the separated total RNA on top of the membranes, and then placed another three BLT2 filter papers were placed on top of them. Throughout the procedure we took care to avoid or remove air bubbles. We transferred the RNA to the membrane by running the semi-dry apparatus at 350 mA, 4°C for 35 minutes per gel. After the procedure we scanned the gel on a Molecular imager FX pro plus using Quantity One for a second time to ensure the transfer of RNA was complete.

We bound the RNA to the membrane by chemical crosslinking. To make the crosslinking solution we added 122.5 µL of 12.5 M 1- methylimidazole and 6 µL of 6M hydrochloric acid to 10ml of ARG water. We shook the solution vigorously and added it to a 15ml falcon tube containing 373mg of 1-ethyl-3-(dimethylaminopropyl) carbodiimide (EDC), we then adjusted the volume to 12ml with ARG water. We soaked a piece of BLT3 filter paper that was slightly larger than the membrane with the crosslinking solution, and we wiped off the excess solution with a Kleenex tissue. We placed the membrane RNA-side up on the soaked filter paper and wrapped them both in Saran wrap (Dow Chemical). We incubated the membrane in a 60°C oven for 1-2 hours. Following the crosslinking, we washed the membrane in ARG water for ten minutes and either kept it stored at -20°C or started the Northern blot immediately afterwards.

For the Northern blots, we pre-hybridised the membranes by placing them into a hybridization tube with the RNA side facing inwards and with 10ml of UltraHyb-Oligo hybridization buffer (Ambion). We placed the tubes into a HB-1000 hybridization oven (UVP), and rotated them at 37°C for one to two hours. Meanwhile we radiolabelled a probe that was reverse complementary to the miRNA of interest with $\gamma^{-32}$P from adenosine 5'-$\gamma^{-32}$P triphosphate (ATP) triethylammonium salt (Perkin Elmer) made up to a final

reaction concentration of 1 X Kinase buffer, 0.5mM oligonucleotide, 10 units of T4 polynucleotide kinase and 1.1 MBq of $\gamma^{-32}$ ATP. We placed the reaction at 37°C in a heating block for one hour. We then diluted the probe solution with 30 µl of ARG water and added it directly to the hybridisation tubes containing the membranes. We left the membranes rotating in the oven at 37°C overnight.

The next morning we removed the membranes, rinsed them with a wash solution (0.2 x sodium chloride/sodium citrate, 0.1% (w/v) sodium dodecyl sulphate (SDS)), and then left them rotating at 37°C for 20-40 minutes in the solution. We repeated this process a second time. We rinsed the membranes one final time, and then removed them and stored them in saran wrap. To detect the radioactive signal on the membranes we placed them in a radioactive cassette (Fujifilm) and placed a blank phospho-imaging screen (Fujifilm) on top of them. We exposed the screens to the membrane for between five hours and seven days depending on the strength of the radioactive signal. We then scanned the screens on a Molecular imager FX pro plus, and used the software *Quantity One* to visualize the signal.

After attaining the required images we stripped the radioactive probes from the membranes (and therefore removed their radioactive signals) by boiling them in 200ml of stripping solution (0.1% SDS, 5mM EDTA) using a microwave. We waited until the solution had cooled back to room temperature then wrapped the membranes in saran wrap. We placed them into a radioactive cassette, placed a blank phospho-imaging screen on top and exposed the screens for seven days. We scanned the screens and if there was any detectable signal on them, we repeated the process until the signal was removed. If there was no detectable signal on the screen we stored the membranes in saran wrap at -20°C or re-hybridized them with another probe. Once the membranes were stripped, we then re-probed each one with a probe that was reverse-complementary to U6, U6 is a snRNA that is often used as both a positive control and as a loading control (Lopez-Gomollon 2011). Following subsequent membrane strips, we re-probed the membranes with different miRNAs.

*Appendix 1 References*

Lopez-Gomollon S. 2011. Detecting sRNAs by Northern Blotting. In *MicroRNAs in Development: Methods and Protocols*, (ed. T Dalmay). Humana Press, New York, USA.

Sorefan K, Pais H, Hall AE, Kozomara A, Griffiths-Jones S, Moulton V, Dalmay T. 2012. Reducing ligation bias of small RNAs in libraries for next generation sequencing. *Silence* **3**: 4.

Winnebeck EC, Millar CD, Warman GR. 2010. Why does insect RNA look degraded? *J Insect Sci* **10**: 159.

Appendix 2

**MiRNAs (n = 101) that were sequenced by miRNA-seq in *Bombus terrestris* and that had a copy number of more than 100 read counts. Shown are the suggested miRNA name; the mature sequence; the chromosome group, the strand identity, the genomic co-ordinates of each precursor sequence (on the current assembly of the *Bombus* genome, Sadd et al. 2015); the prediction method (see main text); and additional information including whether the miRNA is located in an intron, whether it is clustered closely with other miRNAs and whether it is found in *B. terrestris* or both species of *Bombus* with sequenced genomes. Unless otherwise stated, all miRNAs were found in *B. terrestris*, *B. impatiens*, and *A. mellifera* (Chen et al. 2010, Liu et al. 2012, Sadd et al. 2015). The mature sequence was deduced from the high throughput sequencing libraries and may be on the opposite arm to the mature sequence of the honey bee homologues. MiRNAs with names starting MC are new miRNAs that have not previously been sequenced in any other species.**



| MiRNA name | Mature sequence | Chromosome Group | Strand (+/n) | Start co-ordinates | End co-ordinates | Prediction method | Additional information |
|---|---|---|---|---|---|---|---|
| Bte-bantam | UGAGAUCAUUGUGAAAGCUGAUU | 14.1 | + | 690929 | 691002 | *miRCat* | |
| Bte-let-7 | UGAGGUAGUAGGUUGUAUAGUA | 16.3 | - | 579914 | 579826 | *miRCat* | In a cluster containing let-7, miR-100, miR-125. Overlaps exon of Ets97D-like |
| Bte-miR-1 | UGGAAUGUAAAGAAGUAUGGAG | un4 | - | 169896 | 169818 | *miRCat* | The homologs of bte-miR-1-1 and bte-miR-1-2 are the same gene in *Bombus*. |
| Bte-miR-10 | ACCCUGUAGAUCCGAAUUUGU | 18.1 | + | 1482702 | 1482783 | *miR-abela* | Two copies of this miRNA on *Apis* but not *Bombus* genome |
| Bte-miR-100 | AACCCGUAGAUCCGAACUUGUG | 16.3 | - | 580511 | 580432 | *miRCat* | In a cluster containing let-7, miR-100, miR-125. |
| Bte-miR-1000 | AUAUUGUCUUGUCACAGCAGUA | un1114 | + | 686664 | 686747 | *miR-abela* | |
| Bte-miR-11 | CAUCACAGGCAGAGUUCUAGUU | 9.6 | + | 700510 | 700592 | *miRCat* | In the 7th intron of transcription factor E2F3 (LOC100741567). |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Bte-miR-1175 | AAGUGGAGAAGUGGUCUCU | un639 | + | 85154 | 85237 | *miR-abela* | In the 10th intron of SZT2; the precursor is much shorter than in Apis, evidence of mature sequence arm switching compared to *Apis* |
| Bte-miR-12 | UGAGUAUUACAUCAGGUACUGGU | 2.1 | - | 721884 | 721810 | *miRCat* | miR-12, miR-283, miR-3477 are in a cluster. |
| Bte-miR-124 | UAAGGCACGCGGUGAAUGCCA | 11.4 | - | 2824256 | 2824180 | *A. mellifera* homology | |
| Bte-miR-125 | CCCCUGAGACCCUAACUUGUGA | 16.3 | - | 579408 | 579334 | *miRCat* | In a cluster containing let-7, miR-100, miR-125. |
| Bte-miR-133 | UUGGUCCCCUUCAACCAGCUGU | 3.4 | - | 2250584 | 2250502 | *miRCat* | |
| Bte-miR-137 | UAUUGCUUGAGAAUACACGUAG | 1.7 | + | 304270 | 304353 | *miR-abela* | |
| Bte-miR-13a | ACAUCAAAUUGGUUGUGGAAUG | 7.3 | - | 1116178 | 1116119 | *A. mellifera* homology | In a miRNA cluster which contains miR-13a, miR-13b, miR-2-1,miR-2-2, miR-2-3, miR-71. In the intron of serine/threonine-protein phosphatase 4 regulatory subunit 1-like Evidence of mature sequence arm switching compared to Apis |
| Bte-miR-13b | UAUCACAGCCAUUUUUGACGAUU | 7.3 | - | 1115769 | 1115690 | *miRCat* | In a cluster which contains miR13a, miR-13b, miR-2-1,miR-2-2, miR-2-3, miR-71. In the intron of serine/threonine-protein phosphatase 4 regulatory subunit 1-like |
| Bte-miR-14 | GGGGGUGAGAAACUGGCUUGGCU | 7.1 | + | 745311 | 745391 | *miR-abela* | Evidence of mature sequence arm switching compared to *Apis,* both arms are highly expressed in larvae |
| Bte-miR-184 | UGGACGGAGAACUGAUAAGG | 13.5 | - | 3607498 | 3607419 | *miRCat* | |
| Bte-miR-190 | AGAUAUGUUUGAUAUUCUUGGUUG | 13.5 | - | 620922 | 620842 | *miR-abela* | In the intron of talin 2 like |
| Bte-miR-193 | UACUGGCCUGCUAAGUCCCAAG | 7.3 | - | 7628149 | 7628076 | *miRCat* | miR-193, miR-2788 are in a cluster. |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Bte-miR-2-1 | UAUCACAGCCAGCUUUGAUGAG | 7.3 | - | 1115504 | 1115443 | *miRCat* | In a cluster which contains miR13a, miR-13b, miR-2-1,miR-2-2, miR-2-3, miR-71. In the intron of serine/threonine-protein phosphatase 4 regulatory subunit 1-like |
| Bte-miR-2-2 | UAUCACAGCCAGCUUUGAUGAG | 7.3 | - | 1114233 | 1114159 | *miRCat* | In a cluster which contains miR13a, miR-13b, miR-2-1,miR-2-2, miR-2-3, miR-71. In the intron of serine/threonine-protein phosphatase 4 |
| Bte-miR-2-3 | UAUCACAGCCAGCUUUGAUGAGCG | 7.3 | - | 1116550 | 1116466 | *A. mellifera* homology | In a cluster which contains miR13a, miR-13b, miR-2-1,miR-2-2, miR-2-3, miR-71. In the intron of serine/threonine-protein phosphatase 4 regulatory subunit 1-like |
| Bte-miR-252a | AUAAGUACUAGUGCCGCAGGA | 15.5 | + | 2545890 | 2545988 | *miR-abela* | Homologue referred to as bte-miR-252 on NCBI and bte-miR-252a on *miRBase*. |
| Bte-miR-252b | UUAAGUAGUAGUGCCGUAGAUG | 15.5 | + | 2540154 | 2540226 | *miRCat* | |
| Bte-miR-263a | AAUGGCACUGGAAGAAUUCACG | 10.1 | - | 2721692 | 2721611 | *miR-abela* | |
| Bte-miR-263b | CUUGGCACUGGAAGAAUUCACAG | 4.5 | - | 480196 | 480123 | *miRCat* | |
| Bte-miR-275 | UCAGGUACCUGAAGUAGCGCGCG | un154 | - | 584382 | 584306 | *miRCat* | In a microRNA cluster containing miR-305, miR-275. |
| Bte-miR-276 | UAGGAACUUCAUACCGUGCUCU | 7.3 | + | 9509079 | 9509153 | *miRCat* | |
| Bte-miR-2765 | UGGUAACUCCACCACCGUUGGC | 9.3 | + | 455308 | 455376 | *miRCat* | |
| Bte-miR-277 | UAAAUGCACUAUCUGGUACGACA | 5.1 | + | 3556319 | 3556391 | *miRCat* | In a cluster containing miR-277, miR-34. |
| Bte-miR-278 | CCGGAUGAGGUCUCCAUCGACC | 9.1 | + | 1925309 | 1925397 | *miR-abela* | Evidence of mature sequence arm switching compared to Apis |
| Bte-miR-2788 | CAAUGCCCUUCGAAAUCCCAAA | 7.3 | - | 7618563 | 7618486 | *miRCat* | miR-193, miR-2788 are in a cluster. |
| Bte-miR-2796 | GUAGGCCGGCGGAAACUACUUGC | 4.2 | - | 1333735 | 1333659 | *miRCat* | In the intron of 1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase epsilon-1. |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Bte-miR-279a | UGACUAGAUCCACACUCAUUAA | 5.1 | - | 7184550 | 7184475 | *miRCat* | In a cluster containing miR-279a, miR996. In the intron of Dnase alpha sub-unit B. |
| Bte-miR-279b | UGACUAGAUCGAAAUACUCGUCCU | 15.6 | - | 5457693 | 5457615 | *miRCat* | In a cluster containing miR-279b, miR-2944, miR-318, miR 3791, miR-9c. In the intron of cytoplasmic aconitate hydratase. |
| Bte-miR-279c | UGACUAGAGUCACACUCGUCCA | 2.4 | + | 365840 | 365914 | *miRCat* | Located in the intron of enhancer of mRNA-decapping protein 3-like (LOC 100642439). |
| Bte-miR-281 | AAGAGAGCUAUCCAUCGACAGU | 1.1 | + | 646807 | 646882 | *miRCat* | Evidence of mature sequence arm switching compared to Apis |
| Bte-miR-282 | UAGCCUCUCCUAGGCUUUGUCU | 14.5 | + | 1727732 | 1727814 | *miR-abela* | Long mature miRNA may not be real |
| Bte-miR-283 | AAAUAUCAGCUGGUAAUUCUG | 2.1 | - | 723110 | 723039 | *miRCat* | In cluster containing miR-12, miR-283, miR-3477. |
| Bte-miR-2944 | UAUCACAGCAGUAGUUACCUAGG | 15.6 | - | 5457341 | 5457261 | *miRCat* | In a cluster containing miR-279b, miR-2944, miR-318, miR 3791, miR-9c. In the intron of cytoplasmic aconitate hydratase. |
| Bte-miR-29b | UAGCACCAUUUGAAAUCAGUAC | 16.3 | + | 68298 | 68392 | *A. mellifera* homology | |
| Bte-miR-3049 | UCGGGAAGGUAGUUGCGGCGGAU | 8.4 | - | 1144371 | 1144236 | *miR-abela* | |
| Bte-miR-305 | AUUGUACUUCAUCAGGUGCUCUG G | Un154 | - | 584091 | 584178 | *miR-abela* | In a microRNA cluster containing miR-305, miR-275 |
| Bte-miR-306 | UCAGGUACUGAGUGACUCUGAG | 15.5 | - | 1390107 | 1390034 | *miRCat* | In a cluster containing miR-79, miR-306, miR-9b. |
| Bte-miR-307 | CACAACCUUUUUGAGUGAGCGA | 9.6 | - | 1128303 | 1128211 | *miRCat* | |
| Bte-miR-315 | UUUUGAUUGUUGCUCAGAAAGCC | 2.3 | - | 5087664 | 5087592 | *miRCat* | |
| Bte-miR-316 | CCAGCAAAGGGGAACAGGCCGA | 10.1 | - | 6104103 | 6104032 | *miRCat* | Evidence of mature sequence arm switching compared to *Apis* |
| Bte-miR-317 | UGAACACAGCUGGUGGUAUCUCA GU | 5.1 | + | 3530331 | 3530420 | *A. mellifera* homology | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Bte-miR-31a | GGCAAGAUGUCGGCAUAGCUGA | 8.1 | - | 2737984 | 2737903 | *A. mellifera* homology | |
| Bte-miR-33 | CAAUACUUCUACAGUGCAACUC | 1.1 | + | 493988 | 494069 | *miR-abela* | In the intron of sterol regulatory element-binding protein 1. Evidence of mature sequence arm switching compared to *Apis* |
| Bte-miR-34 | UGGCAGUGUUGUUAGCUGGUUGUG | 5.1 | + | 3562441 | 3562532 | *miR-abela* | In a cluster containing miR-277, miR-34. |
| Bte-miR-3477 | UAAUCUCAUGCGGUAACUGUGA | 2.1 | - | 722341 | 722266 | *miRCat* | MiR-12, miR-283, miR-3477 are in a cluster. |
| Bte-miR-3715 | UAUUAUGCUCCGUUUAUCGUUG | 2.2 | + | 2860166 | 2860252 | *A. mellifera* homology | Evidence of mature sequence arm switching compared to *Apis* |
| Bte-miR-3718a | UCCCCUGUCCUGUCCCGAUAGU | 3.5 | + | 3729008 | 3729128 | *A. mellifera* homology | In the 2nd intron of cytochrome b5 reductase 4. |
| Bte-miR-3718b | UCGGGACAGGACAGGGGACAGA | 3.5 | - | 3729114 | 3729008 | *A. mellifera* homology | |
| Bte-miR-3730 | ACAACGAGGGUGAAGAUCGCG | 5.1 | + | 3557089 | 3557167 | *A. mellifera* homology | |
| Bte-miR-375 | UUUGUUCGUUCGGCUCGAGUUA | 3.1 | - | 3558480 | 3558403 | miRCat | |
| Bte-miR-3759 | CGGGACUCACGUUGACUGGGCG | 5.1 | - | 11230153 | 11230062 | *A. mellifera* homology | |
| Bte-miR-3770 | AAUCCUGCAUCAAGUGCGUUGC | 7.1 | + | 2328078 | 2328161 | *miR-abela* | |
| Bte-miR-3785 | UACCCUGUAACGUCCUGAGACU | 10.1 | + | 4344632 | 4344712 | *miRCat* | |
| Bte-miR-3786 | UCUGUAUGGCUCAGGACGAUAC | 14.7 | - | 1851595 | 1851521 | *miRCat* | Evidence of mature sequence arm switching compared to *Apis* |
| Bte-miR-3791 | UCACCGGGUUGAAUUCAUCCA | 15.6 | - | 5457821 | 5457739 | *miR-abela* | In a cluster containing miR-279b, miR-2944, miR-318, miR 3791, miR-9c. In the intron of cytoplasmic aconitate hydratase. Final mature sequence length and sequence not verified. |
| Bte-miR-6001 | GUAGGUAACGACUGAUGGGAACA | 13.5 | - | 4555864 | 4555800 | *miRCat* | In the 4th intron of of vitellogenin 6 (100643258) |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Bte-miR-6037 | UAAGCUCUGUGUACUUUUUACA | 2.4 | + | 578247 | 578329 | *A. mellifera* homology | In the intron of Inositol 1,4,5-trisphosphate receptor. |
| Bte-miR-6039 | AAUCGAACGCGUGAGUUUACGU | Un377 | + | 848 | 929 | *miRCat* | |
| Bte-miR-6040 | UAGUACGGGUAGUACUGGGGA | 11.5 | + | 1345670 | 1345744 | *A. mellifera* homology | In a cluster containing miR-600a, miR-6000b and miR-6040, In the fourth intron of Sialin Precursor contains two mature miRNAs (the 5' sequence is shown here) |
| Bte-miR-7 | UGGAAGACUAGUGAUUUUGUUGU | 9.4 | + | 221395 | 221476 | *A. mellifera* homology | In the 4th intron of La-related protein. |
| Bte-miR-71 | UGAAAGACAUGGGUAGUGAGAU | 7.3 | - | 1117033 | 1116964 | *miRCat* | In a cluster which contains miR13a, miR-13b, miR-2-1,miR-2-2, miR-2-3, miR-71. In the intron of serine/threonine-protein phosphatase 4 regulatory subunit 1-like |
| Bte-miR-750 | CCAGAUCUAACUCUUCCAGCUCA | Un639 | + | 84933 | 85059 | *A. mellifera* homology | In the 10th intron of SZT2 |
| Bte-miR-79 | CUUUGGUAAUACAGCUCUAUGA | 15.5 | - | 1389911 | 1389830 | *miR-abela* | In a cluster containing miR-79, miR-306, miR-9b. In the 5th intron of serine/threonine kinase. Evidence of mature arm switching compared to *Apis* |
| Bte-miR-8 | UAAUACUGUCAGGUAAAGAUGUC | 11.5 | - | 2450627 | 2450542 | *A. mellifera* homology | |
| Bte-miR-87 | GUGAGCAAAGUUUCAGGUGUGU | 13.5 | - | 424550 | 424451 | *A. mellifera* homology | Only one copy of miR-87 in *Bombus*, two copies in *Apis* |
| Bte-miR-927a | UUUUAGAAUUCCUACGCUUUACC | 16.1 | - | 1188029 | 1187958 | *miR-abela* | The honeybee homolog of this microRNA is referred to as bte-miR-927 on NCBI and bte-miR-927a on *miRBase*. |
| Bte-miR-927b | UUUUAGAAUUUGUACGUUCUGU | 15.6 | - | 92067 | 91988 | *miR-abela* | |
| Bte-miR-929 | AUUGACUCUAGUAGGGAGUCC | 15.5 | + | 2490373 | 2490459 | *miR-abela* | |
| Bte-miR-92a | AUUGCACUUGUCCCGGCCUAU | 15.5 | - | 1205883 | 1205817 | *miR-abela* | MiR-92a and miR-92c are in a cluster. |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Bte-miR-92b | AAUUGCACCCGUCCCGGCCUGA | 9.4 | + | 128769 | 128845 | *miRCat* | bte-miR-92b-1 and bte-miR-92b-2 have the same co-ordinates when *BLAST*ed against the bumble bee genome, located in the intron of a hypothetical protein. |
| Bte-miR-92c | AGGUUGGGAUGUGGGCAUUAUUUG | 15.5 | - | 1205709 | 1205620 | *miR-abela* | In a cluster containing MiR-92a, miR-92c. |
| Bte-miR-932 | UCAAUUCCGUAGUGCAUUGCAG | 1.1 | - | 311001 | 310905 | *A. mellifera* homology | In the 1st intron of neuroligin-1. |
| Bte-miR-965 | UAAGCGUAUAGCUUUUUCCCUU | 12.2 | - | 3439907 | 3439842 | *miRCat* | |
| Bte-miR-980 | AAGCUGCCUUUUGAAGGGCAACA | un981 | - | 1586952 | 1586876 | *miRCat* | |
| Bte-miR-993 | GAAGCUCGUCUCUACAGGUAUCU | 18.1 | - | 1550572 | 1550491 | *miRCat* | |
| Bte-miR-996 | UGACUAGAUACAUACUCGUCUA | 5.1 | - | 7184350 | 7184262 | *A. mellifera* homology | In a cluster containing miR-279a, miR996. In the intron of Dnase alpha sub-unit B. |
| Bte-miR-9a | UCUUUGGUUAUCUAGCUGUAUGA | 1.3 | + | 3684166 | 3684240 | *miRCat* | |
| Bte-miR-iab-4 | ACGUAUACUGAAUGUAUCCUGA | 18.1 | - | 669622 | 669551 | *miRCat* | |
| MC1030 | AACUCCGUAGUCUCUUAGUUGA | 18.1 | + | 2545808 | 2545882 | *miRCat* | Found on the two *Bombus* species only; MC65 and MC1030 are paralogs |
| MC1154 | GGUCCGUUGUAAUUGUCACUGG | 1.4 | + | 868570 | 868663 | *miRCat* | Found in *Bombus terrestris* only |
| MC1607 | GUGGAUUAUGGAAAGAAAUCG | 12.2 | + | 2992588 | 2992656 | *miRCat* | Found on the two *Bombus* species only |
| MC218 | UAAGUGAUUGCGGGUUUUGCCA | 16.3 | + | 1898069 | 1898139 | *miRCat* | Found on the two *Bombus* species only; MC64 and MC218 are paralogs |
| MC24 | UGUGGGGCGGCGUCCGGGUCACU | 1.4 | + | 889091 | 889165 | *miRCat* | Found in *Bombus terrestris* only |
| MC295 | UGUUGCGAGGUCGAAGGGCAGU | 3.3 | + | 727632 | 727723 | *miRCat* | Found on the two *Bombus* species only |
| MC314 | AUCAGGUCGCUCGAUAGAGGC | 3.6 | - | 320088 | 320021 | *miRCat* | Found on the two *Bombus* species only |
| MC37 | GAUGGAAGCUUUCUUUGAUACUU | 1.9 | + | 356062 | 356137 | *miRCat* | Found on the two *Bombus* species only |
| MC476 | UUAUUCCGGUGAACCACUCUUC | 9.6 | - | 796215 | 796137 | *miRCat* | Found on the two *Bombus* species only |
| MC485 | UGUAGGUACUAGUCGUCUCUAUA | 11.5 | + | 1346061 | 1346125 | *miRCat* | Found in *Bombus terrestris* only |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| MC64 | UAAGUGAUUGCGGGUUUUGCCA | 10.1 | + | 8011624 | 8011694 | *miRCat* | Found on the two *Bombus* species only; MC64 and MC218 are paralogs |
| MC65 | AAUUAAGUGGUUGCGGAUUUUG | 10.1 | + | 8133504 | 8133580 | *miRCat* | Found on the two *Bombus* species only; MC65 and MC1030 are paralogs |
| MC712 | UAGCGCUUUAGGUUUCGAACGU | 1.4 | - | 1146 | 1087 | *miRCat* | Found in *Bombus terrestris* only |
| MC732 | UAUUUGUACAUGCGUGGUUAGGA | 7.3 | + | 662445 | 662505 | *miRCat* | |
| MC753 | UAUGUUUUGUAGGGCCUUGCGU | un1079 | - | 2121 | 2044 | *miRCat* | Found in *Bombus terrestris* only |
| MC762 | UGACUAGAUCCACACUCAUCCA | 9.5 | - | 360148 | 360079 | *miRCat* | MC762 is a paralog of miR-279a, both paralogs exist in all three bee species |

Appendix 3

**MiRNAs (n = 29) that were predicted by computational methods in *Bombus terrestris* and that had a copy number of less than 100 read counts. Shown are the suggested miRNA name; the mature sequence; the chromosome group, the strand identity, the genomic co-ordinates of each precursor sequence (on the current assembly of the *Bombus* genome, Sadd et al. 2015); the prediction method (see main text); and additional information including whether the miRNA is located in an intron, whether it is clustered closely with other miRNAs. All miRNAs were found in *B. terrestris*, *B. impatiens*, and *Apis mellifera*. The mature sequence was predicted from the mature sequence the homologous miRNA in *A. mellifera*.**



| MiRNA name | Mature sequence | Chromosome Group | Strand (+/n) | Start co-ordinates | End co-ordinates | Prediction method | Additional information |
|---|---|---|---|---|---|---|---|
| Bte-miR-210 | UUGUGCGUGUGACAGCGGCUA | 2.3 | - | 393356 | 393282 | *miR-abela* | |
| Bte-miR-219 | UGAUUGUCCAAACGCAAUUCUUG | 7.1 | + | 230412 | 230483 | miRCat | |
| Bte-miR-318 | UCACUGGGUUUUGUUUGUCCC | 15.6 | - | 5457159 | 5457079 | *miR-abela* | In a cluster containing miR-279b, miR-2944, miR-318, miR 3791, miR-71, miR-9c. In the intron of cytoplasmic aconitate hydratase. |
| Bte-miR-3719 | UACGGAUUGCGUGACUUUUCGA | 3.5 | + | 2702602 | 2702675 | miRCat | |
| Bte-miR-3726 | AACGAGUGCUGGAUGCCAGCGU | 4.5 | - | 1372078 | 1371956 | *A. mellifera* homology | |
| Bte-miR-3727 | UUCGCGAACGCGCUCCUGCUAA | un154 | + | 783607 | 783546 | *A. mellifera* homology | Located in the fifth intron of radial spoke head protein 9 |
| Bte-miR-3734 | UUUCGCACGGGAGGACGUUAG | 6.3 | - | 5335050 | 5334978 | *A. mellifera* homology | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Bte-miR-3736 | CUAUUACCCUGAGUAUGCUGC | un892 | - | 4267 | 4162 | *A. mellifera* homology | |
| Bte-miR-3743 | UCGUGACUUUCUUUUCAGCCCCU | 8.4 | - | 491658 | 491596 | *A. mellifera* homology | |
| Bte-miR-3772 | GAUCUAGAUUGAGAGGAACGAA | 1.9 | - | 139137 | 1391313 | *A. mellifera* homology | Mature miRNA is one bp shorter than in *Apis* |
| Bte-miR-3777 | UGGAAACAUUUCUUCCUCGGGA | 12.1 | + | 3977790 | 3977914 | *A. mellifera* homology | |
| Bte-miR-3783 | UACUUUCAAUUGUUUGAUGAGG | 13.3 | - | 1821043 | 1820966 | miRCat | In the intron of Glutamate decarboxylase. |
| Bte-miR-3788 | GUUCUGUUACCCUGUCCCUC | 14.5 | - | 2402570 | 2402489 | *A. mellifera* homology | |
| Bte-miR-6000a | UAGGUACUAGUCGUCUCUAUA | 11.5 | + | 1346053 | 1346130 | *miR-abela* | In a cluster containing miR-600a, miR-6000b and miR-6040, In the fourth intron of Sialin, Precursor contains two mature miRNAs (the 5' sequence is shown here), MiR-6000a and miR-6000b are complementary to each other |
| Bte-miR-6000b | UAGAGACGACUAGUACCUACAAG | 11.5 | - | 1346130 | 1346053 | *miR-abela* | In a cluster containing miR-6000a, miR-6000b and miR-6040, This miRNA is on the opposite strand of the fourth intron of Sialin (100645295), MiR-6000a and miR-6000b are complementary to each other |
| Bte-miR-6003 | CACUCGGGGAAAAGAGGCC | 2.1 | + | 1926597 | 1926660 | *A. mellifera* homology | |
| Bte-miR-6012 | UUCGGCGAUGAGAUCAGCCUGU | 4.5 | - | 2212704 | 2212626 | miRCat | |
| Bte-miR-6038 | UAUGUUUCUGUCUUAUUUCAUU | 2.3 | - | 397530 | 397463 | miRCat | In a cluster containing miR-210, miR-6038. |
| Bte-miR-6046 | UCGACGACAGUUCCGUGAUCGGU | | + | 4383421 | 4383490 | *A. mellifera* homology | Quite dissimilar to it's equivalent sequence in *Apis* |
| Bte-miR-6048 | AGAGGCGAGGUGGACCGACUGGU | 8.1 | + | 2889672 | 2889755 | miRCat | |
| Bte-mir-6052 | AUGCAAAAGCUAGAACUCAUAGA | 6.3 | + | 187133 | 187201 | *A. mellifera* homology | Unusual precursor structure, In the intron of uncharacterised protein. possibly La-related protein. |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Bte-miR-6053 | AUACGAAAGACCGCGCGGAUGU | 1.9 | - | 1550889 | 1550804 | *A. mellifera* homology | |
| Bte-miR-6065 | CUGGAAUGCGAUCCCCCCGGU | 11.1 | - | 1440365 | 1440290 | *A. mellifera* homology | |
| Bte-miR-6067 | AAACGGAUCAAGCUUUUUGUGA | | - | 5910889 | 5910966 | *A. mellifera* homology | |
| Bte-miR-928 | CUGGCUGUGGAAGCUGGCGAA | 11.4 | + | 6986178 | 6986257 | *A. mellifera* homology | |
| Bte-miR-971 | UUGGUGUUCUACCUUACAGUGAG | 10.1 | + | 5257445 | 5257519 | miRCat | |
| Bte-miR-981 | UUCGUUGUCAACGAAACCUGCA | 2.3 | + | 1937553 | 1937639 | *A. mellifera* homology | |
| Bte-miR-9b | GCUUUGGUAAUCUAGCUUUAUGA | 15.5 | + | 1389829 | 1389912 | *miR-abela* | In a cluster containing miR-79, miR-306, miR-9b. In the 5th intron of serine/threonine kinase. |
| Bte-miR-9c | UAAAGCUAGACCAGCAAGGUG | 15.6 | - | 5457557 | 5457474 | *miR-abela* | In a cluster containing miR-279b, miR-2944, miR-318, miR 3791, miR-71, miR-9c. In the intron of cytoplasmic aconitate hydratase. Final mature sequence length and sequence not verified. |

Appendix 4

**MiRNAs (n = 115) that were predicted by computational methods in *Bombus impatiens*. Shown are the suggested miRNA name; the mature sequence; the chromosome group, the strand identity, the genomic co-ordinates of each precursor sequence (on the current assembly of the *Bombus* genome, Sadd et al. 2015); the prediction method (see main text); and additional information including whether the miRNA is located in an intron and whether it is clustered closely with other miRNAs. All miRNAs were found in *B. terrestris*, *B. impatiens*, and *Apis mellifera* unless otherwise stated (Chen et al. 2010, Liu et al. 2012, Sadd et al. 2015). The mature sequence was predicted from the mature sequence of the homologous miRNA in *A. mellifera*.**



| MiRNA name | Mature sequence | Chromosome Group | Strand (+/n) | Start co-ordinates | End co-ordinates | Prediction method | Additional information |
|---|---|---|---|---|---|---|---|
| Bim-bantam | UGAGAUCAUUGUGAAAGCUGAUU | scf_4016 | - | 8003 | | *miR-abela* | |
| Bim-let-7 | UGAGGUAGUAGGUUGUAUAGUA | scf_1334 | + | 242548 | 242544 | *A. mellifera* homology | In a cluster containing let-7, miR-100, miR-125. Overlaps exon of Ets97D-like |
| Bim-miR-1 | UGGAAUGUAAAGAAGUAUGGAG | scf_0155 | - | 158512 | 158432 | *A. mellifera* homology | The homologs of ame-miR-1-1 and ame-miR-1-2 are the same gene in *Bombus*. |
| Bim-miR-10 | ACCCUGUAGAUCCGAAUUUGU | scf_0152 | + | 747519 | 757600 | *miR-abela* | |
| Bim-miR-100 | AACCCGUAGAUCCGAACUUGUG | scf_1334 | + | 241951 | 242030 | *A. mellifera* homology | In a cluster containing let-7, miR-100, miR-125. |
| Bim-miR-1000-1 | AUAUUGUCUUGUCACAGCAGUA | scf_0872 | + | 229215 | 229296 | *A. mellifera* homology | miR-1000-1, miR-1000-2 are in a cluster. |
| Bim-miR-1000-2 | AUAUUGUCUUGUCACAGCAGUA | scf_0872 | + | 230042 | 230123 | *A. mellifera* homology | miR-1000-1, miR-1000-2 are in a cluster. miR-1000-2 not previously described in other species. |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Bim-miR-11 | CAUCACAGGCAGAGUUCUAGUU | scf_0051 | - | 582943 | 582857 | *miR-abela* | In the 7th intron of transcription factor E2F3. |
| Bim-miR-1175 | AAGUGGAGAAGUGGUCUCU | scf_1268 | - | 21923 | 21999 | *A. mellifera* homology | In the 10th intron of SZT2 |
| Bim-miR-12 | UGAGUAUUACAUCAGGUACUGGU | scf_0837 | + | 278904 | 278986 | *miR-abela* | miR-12, miR-283, miR-3477 are in a cluster. |
| Bim-miR-124 | UAAGGCACGCGGUGAAUGCCA | scf_0142 | - | 2804235 | 2804161 | *miR-abela* | |
| Bim-miR-125 | CCCCUGAGACCCUAACUUGUGA | scf_1334 | + | 243402 | 243508 | *A. mellifera* homology | In a cluster containing let-7, miR-100, miR-125. |
| Bim-miR-133 | UUGGUCCCCUUCAACCAGCUGU | scf_0114 | + | 79778 | 79866 | *A. mellifera* homology | |
| Bim-miR-137 | UAUUGCUUGAGAAUACACGUAG | scf_0095 | - | 619878 | 619795 | *miR-abela* | |
| Bim-miR-13a | ACAUCAAAUUGGUUGUGGAAUG | scf_0197 | - | 413788 | 413689 | *A. mellifera* homology | In a cluster which contains miR13a, miR-13b, miR-2-1,miR-2-2, miR-2-3, miR-71. In the intron of serine/threonine-protein phosphatase 4 regulatory subunit 1-like phosphotase 4 regulatory sub-unit like. |
| Bim-miR-13b | UAUCACAGCCAUUUUUGACGAUU | scf_0197 | - | 413380 | 413302 | *A. mellifera* homology | In a cluster which contains miR13a, miR-13b, miR-2-1,miR-2-2, miR-2-3, miR-71. In the intron of serine/threonine-protein phosphatase 4 regulatory subunit 1-like phosphotase 4 regulatory sub-unit like. |
| Bim-miR-14 | UCAGUCUUUUUCUCUCUCCUA | scf_0019 | + | 1443005 | 1443091 | *miR-abela* | |
| Bim-miR-184 | UGGACGGAGAACUGAUAAGG | scf_2222 | - | 2453954 | 2458313 | *miR-abela* | |
| Bim-miR-190 | AGAUAUGUUUGAUAUUCUUGGUUG | scf_1055 | + | 605780 | 605866 | *miR-abela* | In the intron of talin 2 like |
| Bim-miR-193 | UACUGGCCUGCUAAGUCCCAAG | scf_3007 | - | 515031 | 514950 | *miR-abela* | miR-193, miR-2788 are in a cluster. |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Bim-miR-2-1 | UAUCACAGCCAGCUUUGAUGAGCG | scf_0197 | - | 413116 | 413042 | *A. mellifera* homology | In a cluster which contains miR13a, miR-13b, miR-2-1,miR-2-2, miR-2-3, miR-71. In the intron of serine/threonine-protein phosphatase 4 regulatory subunit 1-like phosphotase 4 regulatory sub-unit like. |
| Bim-miR-210 | UUGUGCGUGUGACAGCGGCUA | scf_0463 | - | 1322403 | 1322329 | *miR-abela* | In a cluster containing miR-210, miR-6038. |
| Bim-miR-219 | UGAUUGUCCAAACGCAAUUCUUG | scf_0197 | + | 225485 | 225564 | *miR-abela* | |
| Bim-miR-2-2 | UAUCACAGCCAGCUUUGAUGAGC | scf_0197 | - | 411858 | 411788 | *miR-abela* | In a cluster which contains miR13a, miR-13b, miR-2-1,miR-2-2, miR-2-3, miR-71. In the intron of serine/threonine-protein phosphatase 4 regulatory subunit 1-like phosphotase 4 regulatory sub-unit like. |
| Bim-miR-2-3 | UAUCACAGCCAGCUUUGAUGAGC | scf_0197 | - | 414165 | 414087 | *miR-abela* | In a cluster which contains miR13a, miR-13b, miR-2-1,miR-2-2, miR-2-3, miR-71. In the intron of serine/threonine-protein phosphatase 4 regulatory subunit 1-like phosphotase 4 regulatory sub-unit like. |
| Bim-miR-252a | AUAAGUACUAGUGCCGCAGGA | scf_0452 | - | 769650 | 769552 | *miR-abela* | |
| Bim-miR-252b | UUAAGUAGUAGUGCCGUAGAUGA | scf_0452 | - | 775400 | 775328 | *miR-abela* | |
| Bim-miR-263a | AAUGGCACUGGAAGAAUUCACG | scf_0317 | - | 2739621 | 2739515 | *miR-abela* | |
| Bim-miR-263b | CUUGGCACUGGAAGAAUUCAC | scf_0248 | + | 1661960 | 1662050 | *miR-abela* | |
| Bim-miR-275 | UCAGGUACCUGAAGUAGCGCGCG | scf_0129 | - | 633506 | 633430 | *miR-abela* | In a cluster containing miR-275, miR-305. |
| Bim-miR-276 | UAGGAACUUCAUACCGUGCUCU | scf_0899 | + | 1117661 | 1117733 | *miR-abela* | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Bim-miR-2765 | UGGUAACUCCACCACCGUUGGC | scf_0264 | + | 1015172 | 1015255 | *A. mellifera* homology | |
| Bim-miR-277 | UAAAUGCACUAUCUGGUACGACA | scf_0263 | + | 1072969 | 1073052 | *A. mellifera* homology | In a cluster containing miR-277, miR-34. |
| Bim-miR-278 | CCGGAUGAGGUCUCCAUCGACC | scf_0551 | - | 124145 | 124061 | miR-abela | |
| Bim-miR-2788 | CAAUGCCCUUCGAAAUCCCAAA | scf_3007 | - | 505927 | 506023 | *A. mellifera* homology | MiR-193, miR-2788 are in a cluster. |
| Bim-miR-2796 | GUAGGCCGGCGGAAACUACUUGC | scf_0064 | - | 1770316 | 1770236 | *miR-abela* | In the intron of 1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase epsilon-1. |
| Bim-miR-279a | UGACUAGAUCCACACUCAUUAA | scf_0548 | - | 1946377 | 1946296 | *miR-abela* | In a cluster containing miR-279a, miR996. In the intron of Dnase alpha sub-unit B. |
| Bim-miR-279b | UGACUAGAUCGAAAUACUCGUC | scf_0151 | + | 697633 | 697717 | *miR-abela* | In a cluster containing miR-279b, miR-2944, miR-318, miR 3791, miR-71, miR-9c. In the intron of cytoplasmic aconitate hydratase. |
| Bim-miR-279c | UGACUAGAGUCACACUCGUCCA | scf_0089 | + | 367708 | 367783 | *A. mellifera* homology | In the intron of hypothetical protein LOC100746806 . |
| Bim-miR-281 | UGUCAUGGAGUUGCUCUCUUUGU | scf_0377 | + | 650303 | 65405 | *A. mellifera* homology | |
| Bim-miR-282 | UAGCCUCUCCUAGGCUUUGUCU | scf_0027 | + | 1411105 | 1411187 | *miR-abela* | |
| Bim-miR-283 | AAAUAUCAGCUGGUAAUUCUGGG | scf_0837 | + | 277676 | 277762 | *miR-abela* | miR-12, miR-283, miR-3477 are in a cluster. |
| Bim-miR-2944 | UAUCACAGCAGUAGUUACCUAGG | scf_0151 | + | 697992 | 698057 | *miR-abela* | In a cluster containing miR-279b, miR-2944, miR-318, miR 3791, miR-71, miR-9c. In the intron of cytoplasmic aconitate hydratase. |
| Bim-miR-29b | UAGCACCAUUUGAAAUCAGUAC | scf_0107 | - | 353561 | 353467 | *A. mellifera* homology | |
| Bim-miR-3049 | UCGGGAAGGUAGUUGCGGCGGAUU | scf_872 | - | 119371 | 119288 | *miR-abela* | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Bim-miR-305 | AUUGUACUUCAUCAGGUGCUCUG G | scf_0129 | - | 633299 | 633217 | *miR-abela* | In a cluster containing miR-275, miR-305. |
| Bim-miR-306 | UCAGGUACUGAGUGACUCUGAG | scf_0135 | - | 2797616 | 2797547 | *miR-abela* | In a cluster containing miR-79, miR-306, miR-9b. In the 5th intron of serine/threonine kinase. |
| Bim-miR-307 | CACAACCUUUUUGAGUGAGCGA | scf_0051 | + | 139177 | 139290 | *A. mellifera* homology | |
| Bim-miR-315 | UUUUGAUUGUUGCUCAGAAAGCC | scf_0181 | + | 67677 | 67757 | *A. mellifera* homology | |
| Bim-miR-316 | CCAGCAAAGGGGAACAGGCCGA | scf_0311 | + | 333278 | 333363 | *A. mellifera* homology | |
| Bim-miR-317 | UGAACACAGCUGGUGGUAUCUCAG U | scf_0263 | + | 1047266 | 1047349 | *miR-abela* | |
| Bim-miR-318 | UCACUGGGUUUUGUUUGUCCC | scf_0151 | + | 698165 | 698266 | *A. mellifera* homology | In a cluster containing miR-279b, miR-2944, miR-318, miR 3791, miR-71, miR-9c. In the intron of cytoplasmic aconitate hydratase. |
| Bim-miR-31a | GGCAAGAUGUCGGCAUAGCUGA | scf_0217 | - | 1038242 | 1038161 | *A. mellifera* homology | |
| Bim-miR-33 | CAAUACUUCUACAGUGCAACUU | scf_0377 | + | 499612 | 499703 | *miR-abela* | In the intron of sterol regulatory element-binding protein 1. |
| Bim-miR-34 | UGGCAGUGUUGUUAGCUGGUUGU G | scf_0263 | + | 1079167 | 1079258 | *miR-abela* | In a cluster containing miR-277, miR-34. |
| Bim-miR-3477 | UAAUCUCAUGCGGUAACUGUGA | scf_0837 | + | 278457 | 278528 | *miR-abela* | miR-12, miR-283, miR-3477 are in a cluster. |
| Bim-miR-3715 | UAUUAUGCUCCGUUUAUCGUUG | scf_1112 | - | 350773 | 350683 | *miR-abela* | |
| Bim-miR-3718a | UCCCCUGUCCUGUCCCGAUAGU | scf_0049 | - | 134278 | 134158 | *A. mellifera* homology | Unusual precursor structure, In a miRNA cluster containing miR-3718a and miR-3718b, In the intron of cytochrome b5 reductase 4. |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Bim-miR-3718b | UCGGGACAGGACAGGGGACAGA | scf_0049 | + | 134172 | 134278 | *A. mellifera* homology | In a miRNA cluster containing miR-3718a and miR-3718b, In the intron of cytochrome b5 reductase 4. |
| Bim-miR-3719 | UACGGAUUGCGUGACUUUUCGA | scf_0525 | - | 144022 | 144349 | *miR-abela* | |
| Bim-miR-3726 | AACGAGUGCUGGAUGCCAGCGU | scf_0248 | + | 753929 | 754051 | *A. mellifera* homology | |
| Bim-miR-3727 | UUCGCGAACGCGCUCCUGCUAA | scf_0129 | + | 834070 | 834131 | *A. mellifera* homology | |
| Bim-miR-3730 | ACAACGAGGGUGAAGAUCGCG | scf_0263 | + | 1073744 | 1073822 | *A. mellifera* homology | |
| Bim-miR-3734 | UUUCGCACGGGAGGACGUUAG | scf_0418 | - | 1668629 | 1668557 | *A. mellifera* homology | |
| Bim-miR-3736 | CUAUUACCCUGAGUAUGCUGC | scf_0543 | + | 617976 | 618081 | *A. mellifera* homology | |
| Bim-miR-3743 | UCGUGACUUUCUUUUCAGCCCCU | scf_0633 | - | 642590 | 642528 | *A. mellifera* homology | |
| Bim-miR-375 | UUUGUUCGUUCGGCUCGAGUUA | scf_0321 | + | 108171 | 108250 | *miR-abela* | |
| Bim-miR-3759 | CGGGACUCACGUUGACUGGGCG | scf_0259 | + | 199736 | 199826 | *A. mellifera* homology | |
| Bim-miR-3770 | AAUCCUGCAUCAAGUGCGUUGC | scf_0582 | - | 980524 | 980441 | *miR-abela* | |
| Bim-miR-3772 | GAUCUAGAUUGAGAGGAACGAA | scf_0319 | - | 495623 | 495539 | *A. mellifera* homology | |
| Bim-miR-3777 | UGGAAACAUUUCUUCCUCGGGA | scf_0491 | - | 737511 | 737387 | *A. mellifera* homology | |
| Bim-miR-3783 | UACUUUCAAUUGUUUGAUGAGG | scf_2645 | + | 51743 | 51828 | *miR-abela* | In the intron of Glutamate decarboxylase. |
| Bim-miR-3785 | UACCCUGUAACGUCCUGAGACU | scf_0091 | - | 400093 | 400023 | *miR-abela* | |
| Bim-miR-3786 | UCUGUAUGGCUCAGGACGAUAC | scf_0914 | - | 262331 | 262238 | *A. mellifera* homology | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Bim-miR-3788 | GUUCUGUUACCCUGUCCCUC | scf_0268 | + | 238759 | 238840 | *A. mellifera* homology | |
| Bim-miR-3791 | UCACCGGGUAGGAUUCAUCCA | scf_0151 | + | 697507 | 697591 | *miR-abela* | In a cluster containing miR-279b, miR-2944, miR-318, miR 3791, miR-71, miR-9c. In the intron of cytoplasmic aconitate hydratase. |
| Bim-miR-6000a | UGGGUACUAGUCGUCUCUAU | scf_0218 | + | 62512 | 62578 | *miR-abela* | In a cluster containing miR-600a, miR-6000b and miR-6040, In the fourth intron of Sialin, Precursor contains two mature miRNAs (the 5' sequence is shown here), MiR-6000a and miR-6000b are complementary to each other |
| Bim-miR-6000b | UAGAGACGACUAGUACCCACGAG | scf_0219 | - | 62579 | 62502 | *miR-abela* | In a cluster containing miR-600a, miR-6000b and miR-6040, In the fourth intron of Sialin, Precursor contains two mature miRNAs (the 5' sequence is shown here), MiR-6000a and miR-6000b are complementary to each other |
| Bim-miR-6001 | GUAGGUAACGACUGAUGGGAACAU | scf_2222 | - | 3420261 | 3420194 | *A. mellifera* homology | |
| Bim-miR-6003 | CACUCGGGAAAAGAGGCC | scf_0044 | - | 1042260 | 1042198 | *A. mellifera* homology | |
| Bim-miR-6012 | UUCGGCGAUGAGAUCAGCCUGU | scf_1060 | - | 91421 | 91339 | miR-abela | |
| Bim-miR-6037 | UAAGCUCUGUGUACUUUUUACA | scf_0089 | + | 586144 | 586229 | *A. mellifera* homology | In the intron of Inositol 1,4,5-trisphosphate receptor. |
| Bim-miR-6038 | UAUGUUUCUGUCUUAUUUCAUU | scf_0463 | - | 1326581 | 1326497 | *miR-abela* | In a cluster containing miR-210, miR-6038. |
| Bim-miR-6039 | AAUCGAACGCGUGAGUUUACGU | scf_0749 | - | 71885 | 71791 | *miR-abela* | |
| Bim-miR-6040 | UagUacgggcagUacUgagga | scf_0219 | + | 62119 | 62194 | *A. mellifera* homology | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Bim-miR-6046 | UcgacgacagUUccgUgacUggU | scf_0091 | - | 360968 | 360899 | *A. mellifera* homology | |
| Bim-miR-6048 | AGAGGCGAGGUGGACCGACUGGU | scf_0217 | + | 1191883 | 1191970 | *A. mellifera* homology | |
| Bim-miR-6052 | AUGCAAAAGCUAGAACUCAUAGA | scf_0023 | + | 220261 | 220329 | *A. mellifera* homology | Unusual precursor structure, In the intron of uncharacterised protein possibly La-related protein. |
| Bim-miR-6053 | AUACGAAAGACCGCGCGGAUGU | scf_0208 | - | 169083 | 168998 | *A. mellifera* homology | |
| Bim-miR-6065 | CUGGAAUGCGAUCCCCCCGGU | scf_0096 | - | 520314 | 520239 | *A. mellifera* homology | |
| Bim-miR-6067 | AAACGGAUCAAGCUUUUUGUGA | scf_0060 | + | 164782 | 164859 | *A. mellifera* homology | |
| Bim-miR-7 | UGGAAGACUAGUGAUUUUGUUGU | scf_0264 | - | 323528 | 323442 | *A. mellifera* homology | |
| Bim-miR-71 | UGAAAGACAUGGGUAGUGAGAUG | scf_0197 | - | 414649 | 4114649 | *A. mellifera* homology | In a cluster which contains miR13a, miR-13b, miR-2-1,miR-2-2, miR-2-3, miR-71. In the intron of serine/threonine-protein phosphatase 4 regulatory subunit 1-like phosphotase 4 regulatory sub-unit like. |
| Bim-miR-750 | CCAGAUCUAACUCUUCCAACUCA | scf_1268 | - | 22227 | 22101 | *A. mellifera* homology | In the intron of an uncharacterised protein. Possibly SZT2 |
| Bim-miR-79 | CUUUGGUAAUACAGCUCUAUGA | scf_0135 | - | 2796850 | 2796767 | *miR-abela* | In a cluster containing miR-79, miR-306, miR-9b. In the 5th intron of serine/threonine kinase. |
| Bim-miR-8 | UAAUACUGUCAGGUAAAGAUGUC | scf_0434 | - | 373187 | 373102 | *miR-abela* | |
| Bim-miR-87 | GUGAGCAAAGUUUCAGGUGUGU | scf_1055 | + | 798331 | 798430 | *A. mellifera* homology | |
| Bim-miR-927a | UUUUAGAAUUCCUACGCUUUACC | scf_0637 | - | 60847 | 60776 | *miR-abela* | |
| Bim-miR-927b | UUUUAGAAUUUGUACGCUCUGU | scf_0065 | + | 18696 | 18767 | *miR-abela* | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Bim-miR-928 | CUGGCUGUGGAAGCUGGCGAA | scf_0203 | - | 1098750 | 1098671 | *A. mellifera* homology | |
| Bim-miR-929 | AUUGACUCUAGUAGGGAGUCC | scf_0452 | - | 825061 | 824975 | *miR-abela* | |
| Bim-miR-92a | AUUGCACUUGUCCCGGCCUAU | scf_0135 | - | 2613750 | 2613684 | *miR-abela* | In a cluster containing MiR-92a, miR-92c. |
| Bim-miR-92b | AAUUGCACCCGUCCCGGCCUGA | scf_0264 | - | 419570 | 419490 | *miR-abela* | |
| Bim-miR-92c | AGGUUGGGAUGUGGGCAUUAUUUG | scf_0135 | - | 2613571 | 2613494 | *miR-abela* | In a cluster containing MiR-92a, miR-92c. |
| Bim-miR-932 | UCAAUUCCGUAGUGCAUUGCAG | scf_0377 | - | 312063 | 311967 | *A. mellifera* homology | In the intron of neurolignin-1. |
| Bim-miR-965 | UAAGCGUAUAGCUUUUUCCCUU | scf_1275 | - | 519079 | 518998 | *A. mellifera* homology | |
| Bim-miR-971 | UUGGUGUUCUACCUUACAGUGAG | scf_0815 | - | 348384 | 348302 | *miR-abela* | |
| Bim-miR-980 | AAGCUGCCUUUUGAAGGGCAACA | scf_0320 | - | 558365 | 558282 | *miR-abela* | |
| Bim-miR-981 | UUCGUUGUCAACGAAACCUGCA | scf_3474 | - | 1271907 | 1271840 | *A. mellifera* homology | Annotation extends into unsequenced region, so the predicted precursor is less likely to be true |
| Bim-miR-993 | GAAGCUCGUCUCUACAGGUAUCU | scf_0152 | - | 820922 | 820841 | *miR-abela* | |
| Bim-miR-996 | UGACUAGAUACAUACUCGUCUA | scf_0548 | - | 1946174 | 1946086 | *A. mellifera* homology | In a cluster containing miR-279a, miR-996. In the intron of Dnase alpha sub-unit B. |
| Bim-miR-9a | UCUUUGGUUAUCUAGCUGUAUGA | scf_0647 | - | 428601 | 428515 | *miR-abela* | |
| Bim-miR-9b | GCUUUGGUAAUCUAGCUUUAUGA | scf_0135 | + | 2796772 | 2796845 | *miR-abela* | In a cluster containing miR-79, miR-306, miR-9b. In the 5th intron of serine/threonine kinase. |
| Bim-miR-9c | UAAAGCUAGACCAGCAAGGUG | scf_0151 | + | 697752 | 697864 | *A. mellifera* homology | In a cluster containing miR-279b, miR-2944, miR-318, miR 3791, miR-71, miR-9c. In the intron of cytoplasmic aconitate hydratase. |
| Bim-miR-iab-4 | ACGUAUACUGAAUGUAUCCUGA | scf_1506 | - | 121620 | 121684 | *miR-abela* | |

Appendix 5

**List of new *Apis mellifera* miRNAs (n = 18) that were first published by Greenberg et al. (2012) but not uploaded to *miRBase*. Columns represent each mature sequence, whether the sequence is conserved in *Bombus terrestris*, the *B. terrestris* genomic co-ordinates of each sequence if they are conserved, and whether they were predicted by *miR-Abela*. MiRNAs #22, #32, and #6 were all predicted by *miR-Abela*, the other six putative miRNAs were not predicted by *miR-Abela* but were still found to have conserved sequences in the *B. terrestris* genome.**

| miRNA name | Mature sequence | In *Bombus*? | Chromosome Group | Strand (+/-) | Start co-ordinates | End co-ordinates | Additional information |
|---|---|---|---|---|---|---|---|
| ame-miR#10 | UUCGGGCGGGCUCGGGCGAGA | n | N/A | N/A | N/A | N/A | N/A |
| ame-miR#15 | UAGUACGGGCAGUACUGGGA | n | N/A | N/A | N/A | N/A | N/A |
| ame-miR#20 | UUUGUGACUGUAACAACAAAU | y | 12.1 | - | 481023 | 481008 | |
| ame-miR#22 | AAUCGAACGCGUGAGUUUACGU | y | un377 | + | 856 | 857 | Predicted by *miR-Abela*. In the 2nd intron of an uncharacterised gene. |
| ame-miR#26 | AGUGGACUGACUGGCCUGUGCU | n | N/A | N/A | N/A | N/A | N/A |
| ame-miR#27 | UAAUUACUCUGACAAUAGUAGUGG | n | N/A | N/A | N/A | N/A | N/A |
| ame-miR#30 | AUUGCACUCGUCCCGGCCU | y | 15.5 | - | 1205654 | 1205636 | |
| ame-miR#32 | UUAAGUAGUAGUGUCGUAGAUGA | y | 15.5 | + | 2540162 | 2540184 | Predicted by *miR-Abela* |
| ame-miR#35 | CGAAAGUCGUGGGAUAGUCGU | n | N/A | N/A | N/A | N/A | N/A |
| ame-miR#36 | UAAGCUCUGUGUACUUUUUAC | y | 2.4 | + | 578302 | 578322 | |
| ame-miR#37 | AUUCACUGGACGGCAAUGGGCU | n | N/A | N/A | N/A | N/A | N/A |
| ame-miR#38 | UAUUGGGUCAGGAUAGGGCAG | y | 3.5 | + | 3728833 | 3728853 | In the 2nd intron of cytochrome b5 reductase 4 |
| ame-miR#39 | AAAUAUCAUCGAGAUCGAAGGA | n | N/A | N/A | N/A | N/A | N/A |
| ame-miR#40 | AGAGGCGAGGUGGACCGACUGGU | y | 8.1 | + | 2889727 | 2889749 | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| ame-miR#41 | UUCGGAAAGUAAUUUCGUUUUU | y | un1188 | + | 492999 | 493020 | |
| ame-miR#6 | UUCGGCGAUGAGAUCAGCCUGU | y | 4.5 | - | 2212655 | 2212634 | Predicted by *miR-Abela*. |
| ame-miR#7 | AAGAGGCUUCUUAGGCAAGUGAU | n | N/A | N/A | N/A | N/A | N/A |
| ame-miR#9 | AUAUGAACUCUUAUGUACGUGA | n | N/A | N/A | N/A | N/A | N/A |

Appendix 6

**Details of samples of *Bombus terrestris* larvae used for miRNA-seq and Northern blot analysis, expanding on information provided in Table 3.1. A total of 60 colonies were sampled, divided into two cohorts, 40 colonies (cohort 1) and 20 colonies (cohort 2). Colonies were sampled for larvae either after queen removal (queenless) or with the queen still present (queenright). Columns represent: the date the first worker eclosed; the date queen was removed in the queenless colonies; the number of days after the first worker eclosed that the queen was removed (queen removal day); the date the first male eclosed; the experimental duration (the period when brood development was being monitored by photographing colonies); the number of earl-instar larvae sampled (removed); the number of late-instar larvae sampled (removed); the caste fate of the unsampled larvae; and additional notes about each colony. This final column included whether the queen died, causing the colony to be excluded from further analysis, and whether the colony was used to produce libraries (Table 3.3) or Northern blot samples (Table 3.4).**



| Cohort | Colony | Queen status | 1st Worker Eclosion | Queen removal date | Queen removal day | First male eclosion | Experiment duration | Early larvae removed | Late larvae removed | Adult caste | Notes |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | QR-1 | Queenright | 19/05/2011 | N/A | N/A | 22/07/2011 | 64 | 46 | 20 | Workers | |
| **1** | **QR-2** | **Queenright** | **23/05/2011** | **N/A** | **N/A** | **24/08/2011** | **93** | **41** | **26** | **Workers** | **RNA colony (libraries EW1, LW1)** |
| 1 | QR-3 | Queenright | N/A | N/A | N/A | N/A | N/A | 0 | 0 | N/A | Queen death |
| 1 | QR-4 | Queenright | 29/05/2011 | N/A | N/A | 08/08/2011 | 71 | 19 | 7 | Workers | |
| 1 | QR-5 | Queenright | 27/05/2011 | N/A | N/A | 20/07/2011 | 54 | 109 | 20 | Queens/ Workers | |
| 1 | QR-6 | Queenright | N/A | N/A | N/A | N/A | N/A | 0 | 0 | N/A | Queen death |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **1** | **QR-7** | **Queenright** | **22/05/2011** | **N/A** | **N/A** | **19/07/2011** | **58** | **44** | **21** | **Workers** | **RNA colony (libraries: EW2, LW2)** |
| 1 | QR-8 | Queenright | 25/05/2011 | N/A | N/A | 22/06/2011 | 28 | 15 | 0 | Workers/ Males | |
| 1 | QR-9 | Queenright | N/A | N/A | N/A | N/A | N/A | 0 | 0 | N/A | Queen death |
| 1 | QR-10 | Queenright | 22/05/2011 | N/A | N/A | 19/07/2011 | 58 | 72 | 31 | Workers/ Males | |
| 1 | QR-11 | Queenright | N/A | N/A | N/A | N/A | N/A | 0 | 0 | N/A | Queen death |
| 1 | QR-12 | Queenright | 22/05/2011 | N/A | N/A | Not recorded | N/A | 29 | 13 | Workers/ Males | |
| 1 | QR-13 | Queenright | 30/05/2011 | N/A | N/A | 15/08/2011 | 77 | 9 | 15 | Workers | |
| 1 | QR-14 | Queenright | N/A | N/A | N/A | N/A | N/A | 0 | 0 | N/A | Queen death |
| **1** | **QR-15** | **Queenright** | **02/06/2011** | **N/A** | **N/A** | **19/07/2011** | **47** | **43** | **23** | **Workers** | **RNA colony (libraries: EW3, LW3)** |
| 1 | QR-16 | Queenright | 15/06/2011 | N/A | N/A | 22/07/2011 | 37 | 39 | 21 | Workers/ Males | |
| **1** | **QR-17** | **Queenright** | **16/06/2011** | **N/A** | **N/A** | **19/08/2011** | **64** | **36** | **22** | **Workers** | **RNA colony (libraries: EW4, LW4)** |
| 1 | QR-18 | Queenright | 11/06/2011 | N/A | N/A | 25/08/2011 | 75 | 33 | 22 | Workers | |
| 1 | QR-19 | Queenright | 14/06/2011 | N/A | N/A | 15/06/2011 | 1 | 0 | 0 | N/A | Males eclosed early |
| 1 | QR-20 | Queenright | 18/07/2011 | N/A | N/A | Not recorded | N/A | 0 | 0 | N/A | Workers eclosed before arrival |
| **2** | **QR-21** | **Queenright** | **02/06/2012** | **N/A** | **N/A** | **12/07/2012** | **40** | **59** | **20** | **Workers** | **RNA colony (Northern blots)** |
| **2** | **QR-22** | **Queenright** | **04/06/2012** | **N/A** | **N/A** | **12/07/2012 (2nd male eclosion)** | **38** | **43** | **6** | **Workers/ Males** | **RNA colony (Northern blots)** |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **2** | **QR-23** | **Queenright** | **05/06/2012** | **N/A** | **N/A** | **16/07/2012** | **41** | **40** | **22** | **Workers** | **RNA colony (Northern blots)** |
| **2** | **QR-24** | **Queenright** | **03/06/2012** | **N/A** | **N/A** | **28/07/2012** | **N/A** | **47** | **25** | **Workers** | **RNA colony (Northern blots)** |
| **2** | **QR-25** | **Queenright** | **25/06/2012** | **N/A** | **N/A** | **12/08/2012** | **48** | **34** | **19** | **Workers** | **RNA colony (Northern blots)** |
| 1 | QL-1 | Queenless | 19/05/2011 | 29/05/2011 | 10 | 30/06/2011 | 42 | 71 | 0 | Workers | |
| 1 | QL-2 | Queenless | 19/05/2011 | 29/05/2011 | 10 | 07/07/2011 | 49 | 58 | 0 | Workers | |
| 1 | QL-3 | Queenless | 21/05/2011 | 01/06/2011 | 11 | 07/07/2011 | 47 | 46 | 9 | Queens/ Worker | |
| 1 | QL-4 | Queenless | 21/05/2011 | 01/06/2011 | 11 | 07/07/2011 | 47 | 45 | 0 | Workers | |
| 1 | QL-5 | Queenless | 22/05/2011 | 30/05/2011 | 8 | 13/07/2011 | 52 | 55 | 0 | Workers | |
| 1 | QL-6 | Queenless | 25/05/2011 | 01/06/2011 | 7 | 08/07/2011 | 44 | 46 | 0 | Workers | |
| 1 | QL-7 | Queenless | 11/06/2011 | 29/06/2011 | 18 | 02/08/2011 | 52 | 0 | 0 | N/A | Didn't produce enough larvae |
| 1 | QL-8 | Queenless | 11/06/2011 | 24/06/2011 | 13 | 27/07/2011 | 46 | 19 | 0 | Workers | |
| 1 | QL-9 | Queenless | 19/06/2011 | 03/07/2011 | 14 | 04/08/2011 | 46 | 17 | 13 | Queens/ Workers | |
| **1** | **QL-10** | **Queenless** | **11/07/2011** | **25/07/2011** | **14** | **23/08/2011** | **43** | **23** | **22** | **Queens** | **RNA colony (libraries: EQ1, LQ1)** |
| 1 | QL-11 | Queenless | 11/07/2011 | 25/07/2011 | 14 | 24/08/2011 | 44 | 14 | 14 | Queens/ Workers | |
| 1 | QL-12 | Queenless | 11/07/2011 | 29/07/2011 | 18 | 25/08/2011 | 45 | 28 | 0 | Workers | |
| **1** | **QL-13** | **Queenless** | **11/07/2011** | **25/07/2011** | **14** | **24/08/2011** | **44** | **10** | **5** | **Queens** | **RNA colony (libraries: EQ2, LQ2)** |
| **1** | **QL-14** | **Queenless** | **11/07/2011** | **25/07/2011** | **14** | **27/08/2011** | **47** | **25** | **27** | **Queens/ Workers** | **RNA colony (libraries: EQ3, LQ3)** |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | QL-15 | Queenless | 11/07/2011 | 23/07/2011 | 12 | 24/08/2011 | 44 | 38 | 0 | Workers | |
| 1 | QL-16 | Queenless | 11/07/2011 | 23/07/2011 | 12 | 22/08/2011 | 42 | 14 | 0 | Workers | |
| 1 | QL-17 | Queenless | 11/07/2011 | 23/07/2011 | 12 | 20/08/2011 | 40 | 25 | 7 | Queens/ Workers | |
| 1 | QL-18 | Queenless | 18/07/2011 | 29/07/2011 | 11 | 29/08/2011 | 42 | 33 | 0 | Workers | |
| **1** | **QL-19** | **Queenless** | **18/07/2011** | **29/07/2011** | **11** | **29/08/2011** | **42** | **18** | **11** | **Queens** | **RNA colony (libraries: EQ4, LQ4)** |
| 1 | QL-20 | Queenless | 18/07/2011 | 30/07/2011 | 12 | Not recorded | N/A | 33 | 0 | Workers | |
| 2 | QL-21 | Queenless | N/A | N/A | N/A | N/A | N/A | 0 | 0 | N/A | Queen death |
| 2 | QL-22 | Queenless | 28/04/2012 | 08/05/2012 | 10 | 10/06/2012 | N/A | 28 | 0 | Workers | |
| 2 | QL-23 | Queenless | 26/04/2012 | 10/05/2012 | 14 | 16/06/2012 | 51 | 33 | 17 | Queens/ Workers | **RNA colony (Northern blots)** |
| 2 | QL-24 | Queenless | 28/04/2012 | 11/05/2012 | 13 | 30/05/2012 | 32 | 5 | 0 | Workers | |
| 2 | QL-25 | Queenless | 30/04/2012 | 17/05/2012 | 17 | 17/06/2012 | N/A | 28 | 0 | Workers | |
| 2 | QL-26 | Queenless | N/A | N/A | N/A | N/A | N/A | 44 | 0 | Workers | Queen death |
| **2** | **QL-27** | **Queenless** | **02/05/2012** | **21/05/2012** | **19** | **24/06/2012** | **N/A** | **31** | **11** | **Queens** | **RNA colony (Northern blots)** |
| 2 | QL-28 | Queenless | 28/04/2012 | 11/05/2012 | 13 | 31/05/2012 | 33 | 29 | 0 | Workers | |
| **2** | **QL-29** | **Queenless** | **01/05/2012** | **11/05/2012** | **10** | **14/06/2012** | **44** | **30** | **10** | **Queens** | **RNA colony (Northern blots)** |
| 2 | QL-30 | Queenless | 01/05/2012 | 21/05/2012 | 20 | 11/05/2012 | 10 | 0 | 0 | Workers | |
| 2 | QL-31 | Queenless | 02/05/2012 | 16/05/2012 | 14 | 16/06/2012 | N/A | 0 | 0 | Workers | |
| **2** | **QL-32** | **Queenless** | **01/05/2012** | **11/05/2012** | **10** | **16/06/2012** | **46** | **11** | **10** | **Queens/ Workers** | **RNA colony (Northern blots)** |
| **2** | **QL-33** | **Queenless** | **03/05/2012** | **17/05/2012** | **14** | **31/05/2012** | **28** | **17** | **8** | **Queens/ Workers** | **RNA colony (Northern blots)** |
| 2 | QL-34 | Queenless | 04/05/2012 | 16/05/2012 | 12 | 17/06/2012 | N/A | 54 | 0 | Workers | |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 2 | QL-35 | Queenless | 12/05/2012 | 22/05/2012 | 10 | 27/06/2012 | N/A | 34 | 3 | Queens/ Workers |

# Appendix 7

**Details of the differentially expressed miRNAs in the larvae of *Bombus terrestris*. Columns show the average expression in each phenotype across all library replicates (reads per million; EW = early-instar worker-destined, LW = late-instar worker-destined, EQ = early-instar queen-destined, LQ = late-instar queen-destined), the pair of phenotypes that each miRNA is differentially expressed between, and the calculated $M$ value for each miRNA using Equation 3.1 (miRNAs are differentially expressed when $M > 1$ and $M < 1$, see Chapter 3.2 'bioinformatics analysis'). MiRNAs are separated by differentially expressed phenotypes. New miRNAs described in Chapter 2 are marked with the prefix 'MC'. MicroRNAs shown in Figure 3.4 are highlighted in bold. MiRNAs that are differentially expressed in more than one phenotype are shown for each pair of comparable phenotypes.**

| miRNA name | EW expression | LW expression | EQ expression | LQ expression | Differentially expressed phenotypes | Fold change, $M$ |
|---|---|---|---|---|---|---|
| MC1607-3p | 29.28 | 124.31 | 140.89 | 81.97 | EQ:EW | 1.71 |
| **miR-11-3p** | **257.72** | **960.27** | **260.61** | **191.53** | **LQ:LW** | **-2.21** |
| **miR-12-3p** | **1,032.51** | **1,552.47** | **852.09** | **392.21** | **LQ:LW** | **-1.08** |
| **miR-6001-3p** | **0.00** | **18.89** | **0.00** | **208.70** | **LQ:LW** | **2.54** |
| **miR-6001-5p** | **0.75** | **169.74** | **0.88** | **1,822.12** | **LQ:LW** | **3.28** |
| miR-263a-5p | 1,161.47 | 216.12 | 1,891.36 | 244.79 | LW:EW | -2.32 |
| miR-31-5p | 8,901.70 | 1,971.74 | 5,203.68 | 2,693.71 | LW:EW | -2.16 |
| miR-252-5p | 2,140.22 | 608.80 | 1,527.11 | 609.24 | LW:EW | -1.78 |
| miR-993-3p | 1,017.06 | 288.71 | 910.25 | 242.47 | LW:EW | -1.75 |
| miR-13b-3p | 407.29 | 112.49 | 229.94 | 53.53 | LW:EW | -1.69 |
| **miR-9a-5p** | **22,815.18** | **7,873.90** | **10,467.50** | **10,574.50** | **LW:EW** | **-1.53** |
| **miR-184-3p** | **68,886.57** | **24,997.10** | **41,729.95** | **19,148.52** | **LW:EW** | **-1.46** |
| MC1607-3p | 29.28 | 124.31 | 140.89 | 81.97 | LW:EW | 1.55 |
| miR-2765-5p | 106.95 | 365.81 | 62.34 | 365.13 | LW:EW | 1.60 |
| **miR-6001-5p** | **0.75** | **169.74** | **0.88** | **1,822.12** | **LW:EW** | **3.19** |
| miR-263a-5p | 1,161.47 | 216.12 | 1,891.36 | 244.79 | LQ:EQ | -2.85 |
| **miR-71-5p** | **1,927.69** | **875.38** | **3,919.40** | **732.24** | **LQ:EQ** | **-2.39** |
| miR-993-3p | 1,017.06 | 288.71 | 910.25 | 242.47 | LQ:EQ | -1.83 |
| miR-13b-3p | 407.29 | 112.49 | 229.94 | 53.53 | LQ:EQ | -1.77 |
| miR-993-5p | 238.68 | 75.26 | 261.10 | 66.41 | LQ:EQ | -1.70 |
| miR-263b-5p | 105.93 | 80.62 | 305.27 | 97.68 | LQ:EQ | -1.47 |
| miR-92b-3p | 329.32 | 125.16 | 382.75 | 135.20 | LQ:EQ | -1.38 |
| **miR-275-3p** | **1,570.37** | **2,569.15** | **1,518.94** | **4,321.03** | **LQ:EQ** | **1.50** |
| miR-100-5p | 49.42 | 104.07 | 56.49 | 227.98 | LQ:EQ | 1.70 |
| miR-2765-5p | 106.95 | 365.81 | 62.34 | 365.13 | LQ:EQ | 2.23 |
| **miR-6001-3p** | **0.00** | **18.89** | **0.00** | **208.70** | **LQ:EQ** | **3.54** |
| **miR-6001-5p** | **0.75** | **169.74** | **0.88** | **1,822.12** | **LQ:EQ** | **6.46** |

# Appendix 8

**Details of the differentially expressed miRNAs in the ovaries of *Bombus terrestris*. Columns show for the average expression in each phenotype across all miRNA-seq library replicates (reads per million; IO = inactive-ovary worker, AO = active-ovary worker Q = queen), the pair of phenotypes that each miRNA is differentially expressed between, and the calculated $M$ value for each miRNA using Equation 4.1 (miRNAs are differentially expressed when $M$ > 1 and $M$ < 1, see Chapter 4.2 'bioinformatics analysis'). Thick black lines show miRNAs that were differentially expressed between different phenotypes. New miRNAs that were previously discovered in larvae are designated 'MC' (Chapter 2), new predicted miRNAs in this study are designated 'MA'. MiRNAs that were tested for validation using Northern blots are highlighted in bold. MiRNAs that are differentially expressed in more than one phenotype are shown for each pair of comparable phenotypes.**

| miRNA Name | IO | AO | Q | Differentially expressed phenotypes | Fold change, $M$ |
|---|---|---|---|---|---|
| Bte-miR-87a | 2584.89 | 110.51 | 124.33 | Q:IO | -4.17 |
| Bte-miR-252a | 1317.09 | 59.21 | 61.42 | Q:IO | -4.04 |
| Bte-miR-316 | 6983.19 | 417.46 | 421.88 | Q:IO | -3.99 |
| Bte-miR-252 | 4587.83 | 351.97 | 278.72 | Q:IO | -3.95 |
| Bte-miR-87 | 14461.73 | 994.37 | 998.32 | Q:IO | -3.83 |
| Bte-miR-316 | 2511.30 | 349.55 | 168.32 | Q:IO | -3.75 |
| Bte-miR-281 | 1207.52 | 213.67 | 79.31 | Q:IO | -3.63 |
| Bte-miR-2796 | 697.97 | 93.80 | 54.99 | Q:IO | -3.26 |
| Bte-miR-2765 | 1455.04 | 199.34 | 155.06 | Q:IO | -3.07 |
| Bte-miR-34 | 1071.43 | 78.72 | 124.85 | Q:IO | -2.91 |
| **Bte-miR-133** | **242.10** | **27.53** | **15.34** | **Q:IO** | **-2.89** |
| Bte-miR-31 | 841.15 | 75.19 | 110.65 | Q:IO | -2.72 |
| **Bte-miR-3718a** | **243.37** | **8.86** | **19.99** | **Q:IO** | **-2.72** |
| Bte-miR-3715 | 163.99 | 15.33 | 7.99 | Q:IO | -2.72 |
| Bte-miR-2b | 8222.59 | 1481.96 | 1355.59 | Q:IO | -2.58 |
| **Bte-miR-184** | **97279.21** | **22246.56** | **18660.04** | **Q:IO** | **-2.38** |
| Bte-miR-276 | 4202.89 | 556.91 | 807.15 | Q:IO | -2.35 |
| Bte-miR-iab | 115.85 | 7.39 | 10.83 | Q:IO | -2.14 |
| Bte-miR-277 | 2750.48 | 639.81 | 631.33 | Q:IO | -2.09 |
| Bte-miR-276a | 15821.78 | 2942.06 | 3847.93 | Q:IO | -2.03 |
| Bte-miR-971 | 87.76 | 13.29 | 7.87 | Q:IO | -1.95 |

| | | | | | |
|---|---|---|---|---|---|
| Bte-miR-317 | 3632.84 | 686.72 | 944.01 | Q:IO | -1.92 |
| Bte-miR-3715 | 112.84 | 19.97 | 15.06 | Q:IO | -1.92 |
| Bte-miR-2c | 7250.15 | 2325.41 | 1993.65 | Q:IO | -1.85 |
| **Bte-miR-71** | **8469.72** | **1913.06** | **2341.64** | **Q:IO** | **-1.85** |
| Bte-miR-375 | 1216.34 | 256.83 | 342.51 | Q:IO | -1.77 |
| bantam | 3369.37 | 927.96 | 998.87 | Q:IO | -1.73 |
| Bte-miR-8 | 2728.33 | 829.54 | 811.20 | Q:IO | -1.73 |
| Bte-miR-927 | 168.51 | 47.25 | 41.71 | Q:IO | -1.61 |
| Bte-miR-2a | 15248.96 | 6341.89 | 5231.78 | Q:IO | -1.54 |
| Bte-miR-190 | 348.20 | 111.09 | 109.41 | Q:IO | -1.51 |
| Bte-miR-2779 | 122.28 | 18.24 | 32.22 | Q:IO | -1.45 |
| MA2776 | 18.57 | 122.83 | 83.57 | Q:IO | 1.43 |
| Bte-miR-263b | 2195.15 | 10006.07 | 6050.22 | Q:IO | 1.45 |
| **Bte-miR-9a** | **611.55** | **1765.80** | **1751.70** | **Q:IO** | **1.49** |
| MA2943 | 1589.81 | 4562.96 | 4667.19 | Q:IO | 1.54 |
| MA2940 | 36.33 | 180.89 | 154.48 | Q:IO | 1.63 |
| MA1866 | 80.79 | 240.11 | 299.33 | Q:IO | 1.66 |
| Bte-miR-92a | 1356.68 | 3753.96 | 5307.75 | Q:IO | 1.95 |
| Bte-miR-92 | 1822.68 | 5321.68 | 7764.74 | Q:IO | 2.08 |
| **Bte-miR-279c** | **5240.70** | **24781.81** | **23127.59** | **Q:IO** | **2.14** |
| Bte-miR-2788 | 15.39 | 163.78 | 141.06 | Q:IO | 2.19 |
| MA1204 | 13.72 | 140.23 | 145.43 | Q:IO | 2.29 |
| MA2735 | 93.73 | 406.79 | 541.02 | Q:IO | 2.30 |
| MA1421 | 2.92 | 38.18 | 108.64 | Q:IO | 2.49 |
| Bte-miR-193 | 6.48 | 190.31 | 160.08 | Q:IO | 2.77 |
| MC485 | 233.44 | 2384.71 | 1706.42 | Q:IO | 2.77 |
| Bte-miR-2944 | 3380.56 | 29946.07 | 28586.44 | Q:IO | 3.07 |
| MA3205 | 1011.74 | 7608.70 | 8710.81 | Q:IO | 3.08 |
| MA1199 | 4596.56 | 57552.96 | 52381.63 | Q:IO | 3.50 |
| **Bte-miR-279b** | **742.61** | **13328.19** | **11411.63** | **Q:IO** | **3.91** |
| MC1030 | 5229.11 | 76465.37 | 80672.59 | Q:IO | 3.94 |
| MA1388 | 38.74 | 896.12 | 906.33 | Q:IO | 3.98 |
| MA1495 | 372.04 | 4614.11 | 6290.56 | Q:IO | 4.01 |
| MA1532 | 81.88 | 1179.70 | 1774.67 | Q:IO | 4.14 |
| MA1408 | 67.71 | 721.49 | 1541.87 | Q:IO | 4.15 |
| Bte-miR-137 | 1313.87 | 42349.76 | 41241.91 | Q:IO | 4.95 |
| Bte-miR-87a | 2584.89 | 110.51 | 124.33 | AO:IO | -4.32 |
| Bte-miR-252a | 1317.09 | 59.21 | 61.42 | AO:IO | -4.08 |
| Bte-miR-316 | 6983.19 | 417.46 | 421.88 | AO:IO | -4.00 |
| Bte-miR-87 | 14461.73 | 994.37 | 998.32 | AO:IO | -3.84 |
| Bte-miR-252 | 4587.83 | 351.97 | 278.72 | AO:IO | -3.63 |
| Bte-miR-34 | 1071.43 | 78.72 | 124.85 | AO:IO | -3.47 |
| **Bte-miR-3718a** | **243.37** | **8.86** | **19.99** | **AO:IO** | **-3.19** |

| | | | | | |
|---|---|---|---|---|---|
| Bte-miR-31 | 841.15 | 75.19 | 110.65 | AO:IO | -3.18 |
| Bte-miR-276 | 4202.89 | 556.91 | 807.15 | AO:IO | -2.87 |
| Bte-miR-316 | 2511.30 | 349.55 | 168.32 | AO:IO | -2.78 |
| Bte-miR-2765 | 1455.04 | 199.34 | 155.06 | AO:IO | -2.75 |
| Bte-miR-2796 | 697.97 | 93.80 | 54.99 | AO:IO | -2.66 |
| **Bte-miR-133** | **242.10** | **27.53** | **15.34** | **AO:IO** | **-2.46** |
| Bte-miR-2b | 8222.59 | 1481.96 | 1355.59 | AO:IO | -2.46 |
| Bte-miR-276a | 15821.78 | 2942.06 | 3847.93 | AO:IO | -2.42 |
| Bte-miR-281 | 1207.52 | 213.67 | 79.31 | AO:IO | -2.39 |
| Bte-miR-3715 | 163.99 | 15.33 | 7.99 | AO:IO | -2.38 |
| Bte-miR-317 | 3632.84 | 686.72 | 944.01 | AO:IO | -2.37 |
| Bte-miR-iab | 115.85 | 7.39 | 10.83 | AO:IO | -2.31 |
| Bte-miR-375 | 1216.34 | 256.83 | 342.51 | AO:IO | -2.16 |
| **Bte-miR-71** | **8469.72** | **1913.06** | **2341.64** | **AO:IO** | **-2.13** |
| **Bte-miR-184** | **97279.21** | **22246.56** | **18660.04** | **AO:IO** | **-2.13** |
| Bte-miR-277 | 2750.48 | 639.81 | 631.33 | AO:IO | -2.07 |
| Bte-miR-2779 | 122.28 | 18.24 | 32.22 | AO:IO | -1.90 |
| bantam | 3369.37 | 927.96 | 998.87 | AO:IO | -1.84 |
| MC712 | 181.00 | 39.36 | 53.33 | AO:IO | -1.76 |
| Bte-miR-3715 | 112.84 | 19.97 | 15.06 | AO:IO | -1.73 |
| Bte-miR-971 | 87.76 | 13.29 | 7.87 | AO:IO | -1.69 |
| Bte-miR-8 | 2728.33 | 829.54 | 811.20 | AO:IO | -1.69 |
| Bte-miR-307 | 625.39 | 179.92 | 237.29 | AO:IO | -1.69 |
| Bte-miR-10 | 3206.37 | 981.09 | 1547.88 | AO:IO | -1.69 |
| Bte-miR-11 | 3125.85 | 963.15 | 1265.67 | AO:IO | -1.68 |
| Bte-miR-2c | 7250.15 | 2325.41 | 1993.65 | AO:IO | -1.63 |
| Bte-miR-190 | 348.20 | 111.09 | 109.41 | AO:IO | -1.49 |
| Bte-miR-927 | 168.51 | 47.25 | 41.71 | AO:IO | -1.49 |
| MA1866 | 80.79 | 240.11 | 299.33 | AO:IO | 1.37 |
| Bte-miR-92a | 1356.68 | 3753.96 | 5307.75 | AO:IO | 1.45 |
| **Bte-miR-9a** | **611.55** | **1765.80** | **1751.70** | **AO:IO** | **1.50** |
| MC753 | 1589.81 | 4562.96 | 4667.19 | AO:IO | 1.51 |
| Bte-miR-92 | 1822.68 | 5321.68 | 7764.74 | AO:IO | 1.54 |
| MA2940 | 36.33 | 180.89 | 154.48 | AO:IO | 1.83 |
| MA2776 | 18.57 | 122.83 | 83.57 | AO:IO | 1.89 |
| MA2735 | 93.73 | 406.79 | 541.02 | AO:IO | 1.91 |
| Bte-miR-263b | 2195.15 | 10006.07 | 6050.22 | AO:IO | 2.18 |
| **Bte-miR-279c** | **5240.70** | **24781.81** | **23127.59** | **AO:IO** | **2.24** |
| MA1204 | 13.72 | 140.23 | 145.43 | AO:IO | 2.25 |
| Bte-miR-2788 | 15.39 | 163.78 | 141.06 | AO:IO | 2.38 |
| MA3205 | 1011.74 | 7608.70 | 8710.81 | AO:IO | 2.89 |
| Bte-miR-193 | 6.48 | 190.31 | 160.08 | AO:IO | 2.99 |
| MA1408 | 67.71 | 721.49 | 1541.87 | AO:IO | 3.08 |
| Bte-miR-2944 | 3380.56 | 29946.07 | 28586.44 | AO:IO | 3.14 |

| | | | | | |
|---|---|---|---|---|---|
| MC485 | 233.44 | 2384.71 | 1706.42 | AO:IO | 3.25 |
| MA1532 | 81.88 | 1179.70 | 1774.67 | AO:IO | 3.56 |
| MA1495 | 372.04 | 4614.11 | 6290.56 | AO:IO | 3.56 |
| MA1199 | 4596.56 | 57552.96 | 52381.63 | AO:IO | 3.64 |
| MC1030 | 5229.11 | 76465.37 | 80672.59 | AO:IO | 3.87 |
| MA1388 | 38.74 | 896.12 | 906.33 | AO:IO | 3.96 |
| **Bte-miR-279b** | **742.61** | **13328.19** | **11411.63** | **AO:IO** | **4.13** |
| Bte-miR-137 | 1313.87 | 42349.76 | 41241.91 | AO:IO | 4.99 |
| MA341 | 240.76 | 166.08 | 537.97 | Q:AO | 1.58 |

# Appendix 9

**Details of the differentially expressed miRNAs in the brains of *Bombus terrestris*. Columns show for each miRNA the average expression in the miRNA-seq library for each phenotype (IO = inactive-ovary worker, AO = active-ovary worker Q = queen), the phenotypes that the miRNA is differentially expressed between, and the calculated $M$ value for each miRNA using Equation 4.1 (miRNAs are differentially expressed when $M$ > 1 and $M$ < 1, see Chapter 4.2 'bioinformatics analysis'). Thick black lines show miRNAs that were differentially expressed between different phenotypes. New miRNAs that were previously discovered in larvae are designated 'MC' (Chapter 2), new predicted miRNAs in this study are designated 'MA'. MiRNAs that were tested for validation using Northern blots are highlighted in bold. IO = inactive-ovary worker, AO = active-ovary worker, Q = Queen.**

| MiRNA Name | IO | AO | Q | Differentially expressed phenotypes | Fold change $M$ |
|---|---|---|---|---|---|
| Bte-miR-29b | 3477.00 | 2068.68 | 711.14 | Q:IO | -2.26 |
| Bte-miR-3783 | 737.21 | 474.02 | 173.59 | Q:IO | -1.97 |
| Bte-miR-137 | 8464.90 | 5281.40 | 2598.05 | Q:IO | -1.70 |
| Bte-miR-6038 | 98.95 | 70.44 | 17.00 | Q:IO | -1.68 |
| Bte-miR-13b | 1293.50 | 891.93 | 408.01 | Q:IO | -1.62 |
| Bte-miR-263a | 411.09 | 274.75 | 131.90 | Q:IO | -1.50 |
| MA362 | 10.26 | 13.15 | 69.95 | Q:IO | 1.57 |
| Bte-miR-316 | 1334.32 | 2928.93 | 4257.33 | Q:IO | 1.66 |
| MC24 | 176.27 | 281.84 | 628.21 | Q:IO | 1.72 |
| **Bte-miR-14** | **91708.93** | **149989.57** | **334360.90** | **Q:IO** | **1.87** |
| Bte-miR-124 | 818.18 | 2099.29 | 3231.81 | Q:IO | 1.96 |
| MC753-3p | 32.26 | 50.06 | 189.80 | Q:IO | 2.01 |
| MC753-5p | 32.11 | 27.05 | 189.25 | Q:IO | 2.01 |
| **Bte-miR-3718a** | **29.51** | **64.35** | **206.98** | **Q:IO** | **2.20** |
| MA3290 | 161.24 | 308.86 | 1058.55 | Q:IO | 2.57 |
| Bte-miR-3049 | 385.69 | 1133.14 | 4114.00 | Q:IO | 3.35 |
| Bte-miR-29b | 3477.00 | 2068.68 | 711.14 | Q:AO | -1.51 |
| MA3051 | 36.88 | 34.50 | 151.21 | Q:AO | 1.65 |
| MA3290 | 161.24 | 308.86 | 1058.55 | Q:AO | 1.71 |
| Bte-miR-3049 | 385.69 | 1133.14 | 4114.00 | Q:AO | 1.84 |
| MC753-5p | 32.11 | 27.05 | 189.25 | Q:AO | 2.15 |

# Appendix 10

The following appendix was not examined in the viva for this thesis. We identified the six new miRNAs with the greatest changes in gene expression between inactive-ovary workers, active-ovary workers, and queens (**Appendix 8**). We then used the same RNA that was generated to produce the miRNAseq and Northern blots in Chapter 4 to produce more Northern blot membranes (**Chapter 4** methods). These membranes were hybridised with probes that were reverse complimentary to the six selected new miRNAs. The Northern blots validated the miRNAseq results for five miRNAs. Bte-MC485, Bte-MC1030, Bte-MA1199, Bte-MA1408, and Bte-MA1532 were all more highly expressed in active-ovary workers and queens compared to inactive-ovary workers (**Figure A10.1**). Meanwhile BteMC753 was relatively similarly expressed in all three phenotypes. This is the first experimental evidence for the existence of these six miRNAs, and the first experimental evidence that new miRNAs are differentially expressed between reproductive phenotypes in *Bombus terrestris*. Previous authors have shown that taxonomically-restricted genes are differentially expressed between social insect castes, and have hypothesised that the most important genes that cause caste differentiation evolve de novo during the evolution of eusociality (Barchuk et al. 2007; Johnson and Tsutsui 2011; Ferreira et al. 2013; Feldmeyer et al. 2014). Throughout this thesis we have previously focussed on differentially expressed miRNAs that are conserved across multiple eusocial insect lineages, the results presented in this appendix show that new miRNAs in *B. terrestris* are also associated with caste differentiation.

*Appendix 10 References*

Barchuk AR, Cristino AS, Kucharski R, Costa LF, Simoes ZLP, Maleszka R. 2007. Molecular determinants of caste differentiation in the highly eusocial honeybee *Apis mellifera*. *BMC Dev Biol* **7**: 70.

Johnson BR, Tsutsui ND. 2011. Taxonomically restricted genes are associated with the evolution of sociality in the honey bee. *BMC Genomics* **12**: 164.

Feldmeyer B, Elsner D, Foitzik S. 2014. Gene expression patterns associated with caste and reproductive status in ants: worker-specific genes are more derived than queen-specific ones. *Mol Ecol* **23**: 151-161.

Ferreira PG, Patalano S, Chauhan R, Ffrench-Constant R, Gabaldon T, Guigo R, Sumner S. 2013. Transcriptome analyses of primitively eusocial wasps reveal novel insights into the evolution of sociality and the origin of alternative phenotypes. *Genom Biol* **14**: R20.

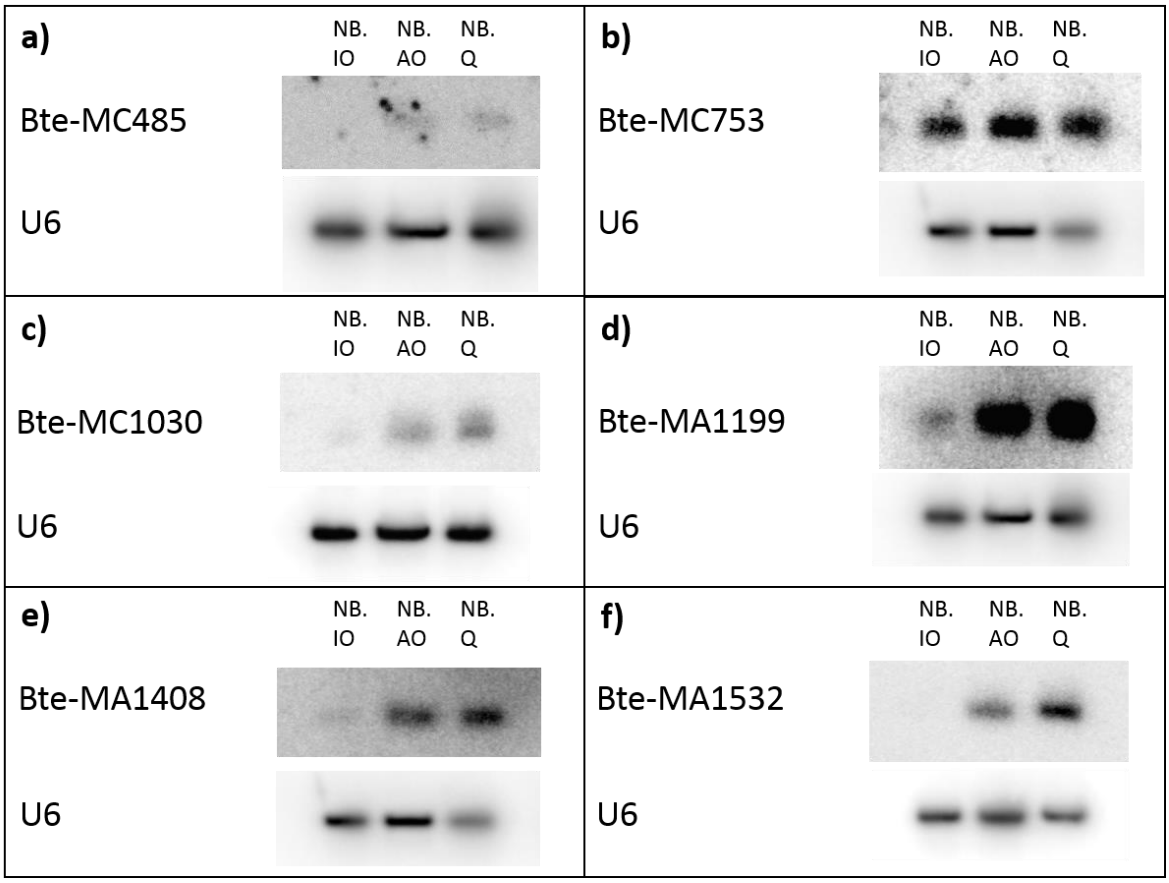


**Figure A10.1: Northern blot expression profiles for six new miRNAs (a-f) that were classified as differentially expressed in ovaries from inactive-ovary workers, active-ovary workers, and queens in *Bombus terrestris* adult females according to miRNA-seq. The U6 panel (control) demonstrates equal loading for each sample. NB.IO, inactive-ovary worker; NB.AO, active-ovary worker; NB.Q, Queen.**