December 2013

Understanding honey bee worker sterility: a conceptual-empirical framework

Emma K. Mullen
The University of Western Ontario

Supervisor
Dr. Graham Thompson
The University of Western Ontario

Graduate Program in Biology

A thesis submitted in partial fulfillment of the requirements for the degree in Master of Science

© Emma K. Mullen 2013

Follow this and additional works at: https://ir.lib.uwo.ca/etd

Part of the Biology Commons

Recommended Citation
https://ir.lib.uwo.ca/etd/1734

This Dissertation/Thesis is brought to you for free and open access by Scholarship@Western. It has been accepted for inclusion in Electronic Thesis and Dissertation Repository by an authorized administrator of Scholarship@Western. For more information, please contact tadam@uwo.ca, wlswadmin@uwo.ca.
UNDERSTANDING HONEY BEE WORKER STERILITY: A CONCEPTUAL-EMPIRICAL FRAMEWORK

(Thesis format: Integrated Article)

by

Emma Kate Mullen

Graduate Program in Biology

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science

The School of Graduate and Postdoctoral Studies
The University of Western Ontario
London, Ontario Canada

© Emma Kate Mullen 2013
Abstract

Kin selection explains how complex social behaviour can evolve at the gene level, but this theory does not identify which genes are necessary for the expression of altruism. In my first chapter I overview seven criteria for genes for altruism using the honey bee as a model species. In the second chapter I explore one criterion in detail – that altruism genes are differentially expressed between reproductive and sterile workers. I used results from previous microarray studies to reconstruct nine knowledge-based gene networks that describe reproductive altruism by means of ovary activation and de-activation. All networks were enriched for Gene Ontology terms pertaining to reproduction, and the hub genes in each network tend to consist of genes involved in expression and signaling. 138 genes overlap among networks for workers of different ages and tissues. These networks provide testable hypotheses that explain the expression of altruistic sterility in workers.

Keywords: Kin selection; reproductive altruism; Apis mellifera; co-citation network; meta-analysis; ovarian development
Statement of Co-Authorship

Chapter 1: Understanding honey bee worker sterility was written by Emma Kate Mullen with Dr. Graham Thompson as a co-author. Specifically, Dr. Thompson provided scientific oversight and edited the written content. I am intending to submit this chapter as a manuscript for publication.

Chapter 2: Gene co-citation networks associated with worker sterility in honey bees was written by Emma Kate Mullen with Dr. Graham Thompson and Dr. Mark Daley as co-authors. Emma Kate Mullen constructed the gene networks, analyzed the statistics, and wrote the paper. Dr. Graham Thompson helped conceive of the idea, provided scientific oversight, and edited the paper. Dr. Mark Daley provided scientific oversight, assisted in graphical interpretation, and approved of the manuscript before I submitted it for publication.
Acknowledgements

I would like to express my thanks to Dr. Graham Thompson, my supervisor and mentor. Without him this work could not have been completed. He has helped me think critically, push the boundaries, and develop into a true scientist. I would also like to thank my advisors Dr. Bryan Neff and Dr. Susanne Kohalmi for their advice and comments on my project. Furthermore, this thesis could not be complete without the members of the Social Biology Group at Western. Alison Camiletti was a true mentor to me and I can’t thank her enough for all of her knowledge, advice, and warmth throughout my research endeavors. Catherine Gao gave me scientific counsel and reassurance in the lab, and David Scaduto helped me with formatting and to believe in myself in times of frustration. Last, but not least, I am very grateful for my family. Thank you to my parents Dan and Mary Mullen, my sister Sarah Mullen, and my grandmother Margaret Rhude for all your support and love throughout this process. You have always encouraged me to pursue my aspirations.
Table of Contents

Table of Contents

ABSTRACT .................................................................................................................. ii
STATEMENT OF CO-AUTHORSHIP ........................................................................ iii
ACKNOWLEDGEMENTS .............................................................................................. iv
TABLE OF CONTENTS ................................................................................................. v
LIST OF TABLES ........................................................................................................ vii
LIST OF FIGURES ...................................................................................................... viii
LIST OF APPENDICES ............................................................................................... ix
LIST OF ABBREVIATIONS ......................................................................................... x

CHAPTER 1: Understanding honey bee worker sterility: a conceptual-empirical framework

1.1 The origins of social complexity ................................................................. 1
1.2 Honey bees as a model insect society ..................................................... 3
1.3 Gap between gene theory and genes ......................................................... 6
1.4 Towards a molecular understanding of worker sterility ...................... 7
  1.4.1 Prediction 1: Genes for altruism should satisfy Hamilton’s rule ........ 8
  1.4.2 Prediction 2: Genes for altruism should be environmentally sensitive . 9
  1.4.3 Prediction 3: Genes for altruism should increase in number and complexity with social-behavioural sophistication ................. 11
  1.4.4 Prediction 4: Genes for altruism should co-evolve with, or depend on, the previous evolution of genes for kin recognition ............ 12
  1.4.5 Prediction 5: Genes for altruism may reside in regions of low recombination, exhibit co-expression, and show modular genetic architecture .......................................................... 13
  1.4.6 Prediction 6: Genes for altruism should have partially additive effects ................................................................. 15
  1.4.7 Prediction 7: Genes for altruism should show strong pleiotropy .... 16
1.5 Conclusions ...................................................................................................... 17

CHAPTER 2: Gene co-citation networks associated with worker sterility in honey bees ................................................................................................................... 28

2.1 Introduction ..................................................................................................... 28
2.2 Methods .......................................................................................................... 32
  2.2.1 Meta-analysis and network construction ......................................... 32
  2.2.2 Within network analysis ................................................................. 35
  2.2.3 Between network analysis ................................................................. 36
2.3 Results ............................................................................................................. 36
  2.3.1 Networks from brain tissue analysis ................................................. 39
  2.3.2 Network from whole body tissue analysis ........................................ 65
  2.3.3 Genes of functional importance ....................................................... 65
  2.3.4 Network enrichment analysis ............................................................. 67
  2.3.5 Gene overlap among networks ......................................................... 69
  2.3.6 Co-citation network bias ................................................................. 69
2.4 Discussion ....................................................................................................... 70
  2.4.1 Co-citation networks ........................................................................ 70
  2.4.2 Testing candidate pathways for ovary activation ......................... 73
2.4.3 A single conserved network for ovary activation?.................................75
CHAPTER 3: Discussion..................................................................................84
  3.1 Technical limitations with using multiple microarrays.................85
  3.2 Building a network based on Drosophila homologs...............86
  3.3 Advancing gene networks for worker sterility.................................87
APPENDIX A: Chapter 2 Detailed Results..................................................92
CURRICULUM VITAE...............................................................................96
## List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1</td>
<td>Studies included in the meta-analysis.</td>
<td>38</td>
</tr>
<tr>
<td>Table 2</td>
<td>Enrichment analysis of gene networks.</td>
<td>68</td>
</tr>
</tbody>
</table>
## List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1</td>
<td>Co-citation networks from Grozinger et al. (2003) gene lists for 1-, 2-, 3-, and 4-day-old workers (A-D, respectively).</td>
<td>40</td>
</tr>
<tr>
<td>Figure 2</td>
<td>Co-citation networks from Backx et al. (2011) gene lists for 4-, 6-, 8-, and 10-day-old workers (A-D, respectively).</td>
<td>52</td>
</tr>
<tr>
<td>Figure 3.</td>
<td>Co-citation network from Cardoen et al. (2012) gene list for 18-day-old workers.</td>
<td>64</td>
</tr>
<tr>
<td>Figure 4.</td>
<td>The most connected genes of each network and their functions</td>
<td>66</td>
</tr>
</tbody>
</table>
List of Appendices

<table>
<thead>
<tr>
<th>Table/Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure A.1</td>
<td>Degree distributions for each gene network and their relative $R^2$ values.</td>
<td>92</td>
</tr>
<tr>
<td>Table A.1</td>
<td>Gene loss occurring from converting original differentially expressed genes to those that appear connected in a co-citation network.</td>
<td>93</td>
</tr>
<tr>
<td>Table A.2</td>
<td>Edge confirmation by Drosophila Interaction Database</td>
<td>94</td>
</tr>
</tbody>
</table>
List of Abbreviations

egr1: early growth response 1

QTL: Quantitative Trait Loci

MRJP: Major Royal Jelly Proteins

QMP: Queen Mandibular Pheromone

GO: Gene Ontology

DroID: Drosophila Interaction Database

DEGs: Differentially Expressed Genes

miRNA: microRNA

IIS: Insulin/Insulin-like Signaling

HVA: Homovanillyl Alcohol

GRN: Gene Regulatory Network
1. Understanding honey bee worker sterility: a conceptual-empirical framework

1.1 The origins of social complexity

A central goal of evolutionary biology is to understand the origins of biological complexity. While it has long been recognized that biological complexity evolves at the level of the gene, only recently has it become more widely appreciated that fundamental increases to complexity result from changes in the way genes are transmitted from generation to generation (Maynard Smith and Szathmary 1995; Queller 2000). For example, the evolution of eusocial societies, in which individuals form colonies with reproductives who pass on genes and non-reproductives who do not, is one of the most recent ‘major’ evolutionary transitions on Earth. Though this type of transition to group-living is rare, it is not singular. Eusociality, in one form or another, has evolved at least a dozen times across different spectra of life (Choe and Crespi 1997). Moreover, eusociality can be evolutionarily labile and lost through reverse transitions (Wcislo and Danforth 1997) or via extinction of eusocial lineages (Engel 2001).

The most advanced animal societies are found among humans and insects (Wilson 1975). Like humans, eusocial ants, bees, wasps and termites show bewildering complexity in how their societies are structured. Yet for insects, this complexity is not associated with culture but rather is derived from a simple division of labour into reproductive and non-reproductive specialists (Wilson 1971). It is known from kin theory that this reproductive division in labour evolves from selection at the level of
the gene (Hamilton 1964; Crozier and Pamilo 1996; Queller and Strassmann 1998; Keller 2009; Bourke 2011) and therefore studying divisions in labour from a “gene’s-eye-view” (sensu Dawkins 1976) can provide key insights into how complex social systems evolved from more individualistic ancestries.

Comparative gene level studies especially have potential to reveal the cryptic genes due to shared ancestry that underlie the evolution of seemingly disparate or convergent social traits (Crespi 1996), and there may well be a common genetic architecture (i.e., the genes involved, their effects, their location on the chromosome, etc.) that underpins this diversity (e.g., Johnson and Linksvayer 2010). For example, much like conserved genes that underpin diversity in form (Giudice 2001; Carroll et al. 2005), similarly conserved genes may underpin diversity in social behaviour (Toth and Robinson 2007). One such gene is early growth response 1 (egr1), a brain transcription factor whose expression was initially found to be socially responsive in songbirds (Mello et al. 1992). Songbirds up-regulate this gene in response to song, but a homologue of this same gene is up-regulated in cichlid fish with changes in social status (Burmeister et al. 2005), and in rats upon receipt of maternal care (Hellstrom et al. 2012). The egr1 gene thus demonstrates how expression of a single gene can underpin variation in behavioural traits among taxa. This and other examples (e.g., the neuropeptides oxytocin and arginine vasopressin; Carter et al. 2008) show the potential for comparative genomics applied to social phenotypes (de Bono 2003; Robinson et al. 2008; Liang et al. 2012).
The conceptual premise for expecting conserved sets of genes to underlie basic social evolution could therefore be strong (Reaume and Sokolowski 2011). Insect species provide a good subject to study the evolution of the genes that underlie the expression of sociality (Beshers and Fewell 2001; Boomsma and Franks 2006). For one, social insects are often regarded as pinnacles of social success, whereby the depth of division in labour and degree of interconnectedness far exceeds that which is typical of vertebrate social systems (Alexander et al. 1991; Holldobler and Wilson 2009). Second, owing to their phylogenetic diversity, whereby sociality has arisen repeatedly, to different levels and in different clades (Hines et al. 2007; Schwarz et al. 2007; Cardinal and Danforth 2011; Woodard et al. 2011), social insects lend themselves well to comparative analyses of phenotypic (Thompson and Oldroyd 2004) or genotypic (e.g., Fischman et al. 2011; Woodard et al. 2011) characters. Finally, unlike most vertebrates, some insect species are amenable to behavioural and genetic manipulation, and therefore provide tractable models for studying the interplay between genes, the environment and social versus asocial life (LeBoeuf et al. 2013).

1.2 Honey bees as a model insect society

Among the insects, the honey bee *Apis mellifera* has emerged as the pre-eminent model system for understanding the origins of social complexity (Weinstock et al. 2006). Like all eusocial taxa, honey bees are characterized by specialized castes (Michener 1974). For *A. mellifera* this division is so pronounced that the non-reproductive worker caste is sterile (Seeley 1985; as it is for other *Apis* spp.). This
sterility is maintained by indirect selection that promotes reproductive self-sacrifice and alloparental care from the worker caste (Foster et al. 2006a). This altruism is stimulated at the proximate level by a collection of social and environmental factors, with a key factor for honey bees being the presence of a queen (Backx et al. 2012). The queen’s pheromones, as well as those from her brood, serve as an honest signal to which workers attend and inactivate their ovaries (Kocher et al. 2009). In her presence workers remain more-or-less sterile, but otherwise some proportion of queenless workers may activate their ovaries and lay eggs (Miller and Ratnieks 2001).

The precise threshold beyond which workers ‘switch’ their own reproductive machinery off in order to assist the queen depends on the social and ecological context, as predicted by response threshold models (Barron and Robinson 2008), and even from Hamilton’s Rule: $rb > c$. This behavioural rule-of-thumb shows that, despite a cost $c$ to the altruist’s individual fitness, genes for reproductive altruism can nonetheless spread if their effect sufficiently benefits $b$ relatives of degree $r$ who – by definition – are likely to carry copies of that gene. As recapped by Crozier (2008), the condition under which the inequality is met varies with socio-genetic and ecological context. Even for a given set of conditions the threshold response is likely to be genetically variable (Oldroyd and Fewell 2007); workers with a genetically low threshold will be sensitive to queen signal and need only a low dose of pheromone to inactivate their ovaries, whereas workers with a genetically high threshold will be relatively insensitive to queen signal. So-called ‘anarchist’ workers provide a good example of heritable variation in worker sterility; workers of the anarchist strain have
such a high threshold response that they readily activate their ovaries and lay eggs despite the presence of the queen (Oldroyd et al. 1994). The evidence for heritable variation underlying worker ovary activation – essentially, underlying worker sterility – suggests that it may be possible to identify the loci and genes responsible. In effect, genes that explain variation in ovary activation are examples of genes that explain this particular form of reproductive altruism (Oldroyd and Thompson 2007).

With the increasing power and accessibility of tools that facilitate the interrogation of honey bee genomes (Aase et al. 2005; Weinstock et al. 2006; Solignac et al. 2007; Munoz-Torres et al. 2011; Schulte et al. 2013), it has become increasingly feasible to map, isolate and characterize genes associated with social traits. For example, Quantitative Trait Loci (QTLs) that regulate worker sterility in anarchist workers are now mapped to specific linkage groups (Oxley et al 2008), as are other segregating loci related to social life (Hunt et al. 1995; Lattorff et al. 2007; Shorter et al. 2012). From such mapping studies, other socially relevant genes have been isolated (e.g., foraging genes by Rueppell et al. (2004); reproductive behaviour genes by Jarosch et al. (2011); and fighting genes by Kiya et al. (2012)) and a growing number are becoming carefully characterized (Fussnecker and Grozinger 2008; Wang et al. 2009).

Based on these initial findings, there is growing interest in studying the evolutionary genetics of honey bee social life, with this new field often termed ‘sociogenomics’ (Robinson et al. 2008; Smith et al. 2008; Fischman et al. 2011). Yet, despite this
enthusiasm advances to date have been limited, with sociogenomics often focusing on idiosyncratic aspects of honey bee social life (e.g., genetic basis of pollen hoarding; Hunt et al. 2007; Linksvayer et al. 2009) and not explicitly focusing on genes that underlie reproductive altruism per se. This lack of focus may be due in part to sociogenomics still emerging from its infancy (it has not yet been 10 years since the first honey bee genome), but may likewise stem from discord between molecular biologists and bee (or other) sociobiologists. If honey bee workers are considered reproductively altruistic, and this altruism has evolved via kin selection on genes for sacrificing personal reproduction in order to help the queen, then it is simply expected that workers to carry these genes for altruism (Thompson et al. 2013). Despite this expectation, there remains a considerable gap between the knowledge of gene theory and the genes themselves.

1.3 Gap between gene theory and genes

If kin theory is so gene-centric – literally, ‘The Genetical Evolution of Social Behaviour’ (Hamilton 1964), then why have so few genes been uncovered? I propose that there are two reasons for this gap between sociobiology and molecular biology. First, kin theory enthusiasts – who are most likely to understand the genetic basis of sociality – have tended not to embrace the tools of molecular biology. On the other hand, molecular biologists – who are best equipped to map and characterize genes – tend not to adopt social species as models of gene discovery. So, there simply has not yet been a coordinated effort to screen, prioritize and verify candidate genes relevant to sociobiology in the same way that these efforts have revolutionized the human
behavioural and health sciences (Ebstein et al. 2010) or, alternatively, have revolutionized the economically-driven agricultural (Dennis et al. 2008) or medical sciences (Haines and Pericak-Vance 2006).

Second, even with the initial success of sociogenomics there has not yet been a robust framework developed for guiding gene discovery. After all, ‘genes for altruism’, as Hamilton would put it (Hamilton 1963), are not just any genes, they are genes whose transmission to subsequent generations is dependent on indirect effects (Wolf et al. 1998; Linksvayer and Wade 2005), and this quality can be exploited to aid in their identification. For example, social genes are likely to be differentially expressed (Seger 1981; Queller and Strassmann 1998; Bourke 2011), may be more variable (Linksvayer and Wade 2009) and have a higher rate of molecular evolution (Bromham and Leys 2005; Hall and Goodisman 2012; Kent et al. 2012) than their homologues in non-social linages, or relative to other genes in the same genome. It should therefore be possible to harness these and other qualities of social genes to focus a renewed search for their identification and, beyond that, assemble these genes into functional pathways that provide a more integrated description of social behaviour, as it is expressed within natural social environments (Chapter 2).

1.4 Towards a molecular understanding of worker sterility

In the 50 years since Hamilton’s insight, kin theory has become sophisticated (Frank 1998) and is sometimes misunderstood (Dawkins 1979; Foster et al. 2006b; Thompson 2006; Abbot et al. 2011) – yet the theory remains elegant at its core, and
can be interpreted to generate several straightforward predictions regarding the nature of genes for altruism. In a recent review commemorating the semi-centennial anniversary of Hamilton’s landmark 1964 papers, Thompson et al. (2013) outline what they consider to be the seven most likely characteristics of genes underlying altruistic traits, in any organism. Using a kin-theoretic framework, they describe how to identify and analyze genes for altruism in humans and social insects. This framework stands to accelerate social gene discovery by drawing attention from molecular biologists to the abstract ‘genes’ that are so central to kin theory, but that for the most part have yet to be isolated or characterized. These characteristics are outlined below, but despite the potential utility of Thompson et al.’s (2013) framework it remains untested against any social genome. In this review I apply this newly developed scheme to the honey bee genome. My goal is to relate all seven predictions to what is known from published honey bee genomic studies, review the molecular evidence for each characteristic and, where it is lacking, suggest how any gaps might be addressed.

1.4.1 Prediction 1: Genes for altruism should satisfy Hamilton’s rule

Worker sterility in honey bees is one of the best known examples of reproductive altruism (Oldroyd and Thompson 2007). Textbook references to this case are illustrative because even first-time students to kin thinking can envision the familiar workings of a domesticated beehive, or may have first-hand experience of a worker’s selfless sting. The concept of working, or sacrificing one’s self, on behalf of the whole is thus made, and the very teaching of social evolution via Hamilton’s Rule has
become closely associated with this one common Hymenopteran insect. But even if honey bees were rare or unknown to most, their matriarchal, ancestrally monogamous families (Hughes et al. 2008), among other demographic or ecological predispositions to influence the reproductive success of kin (Michener 1974; Winston 1987; Crespi 1994), suggest that honey bee workers evolved, as for other social taxa, via kin mediated selection on genes controlling the timing and target of parental care (Crozier and Pamilo 1996; Queller and Strassmann 1998; Linksvayer and Wade 2005).

1.4.2 Prediction 2: Genes for altruism should be environmentally sensitive

In the non-social ancestor of the honey bee tribe (Apini) there was no division of reproductive labour among females (Thompson and Oldroyd 2004) – no queen or worker caste. Presumably there was, however, genetic variation among pro-social females for when, and toward whom, they would direct maternal care. Indirect selection might thus have favoured genetic variants that directed care towards related, but not descendent, brood. If so, the relevant ‘gene for altruism’ associated with a shift from parental to alloparental care could increase in frequency. This evolutionary scenario assumes, however, that only *some* of the females make this shift in the timing (tending to early) and direction (tending towards non-descendent kin) of care, while the remaining females (those tending to receive help) remained firmly reproductive. To the extent that this or any other shift towards altruistic behaviour is mediated by genes, then these genes must have been sensitive to the presence of conspecifics comprising their social environment – that is, they switch *on* in altruistic helpers but switch *off* in selfish reproductives (Thompson et al. 2006). As reviewed in
Bourke (2011) this conditional expression of altruism genes is a logical requirement for their evolution.

If so, then one opportunity for identifying genes for altruism is via expression-based genomic screens – for example, microarray scans comparing queen and worker transcriptomes. These screens have been performed and typically do yield widespread expression differences (Evans and Wheeler 2001; Cristino et al. 2006; Barchuk et al. 2007; Grozinger et al. 2007). Yet, because queens and workers are differentiated with respect to a large number of traits, it would be difficult to pinpoint the precise genes associated with any one trait, altruistic or otherwise. The most experimentally relevant studies are therefore those that control for caste, and thus compare gene expression differences between selfishly reproductive and altruistically non-reproductive members of a single caste. For example, Thompson et al. (2006, 2008) used microarrays to compare the gene expression profiles of mutant ‘anarchist’ workers that activate their ovaries and lay eggs, and wild type workers that were typically sterile. This precise screen reveals a small number of interesting genes, including genes that encode the egg-yolk protein vitellogenin and that encode several of the major royal jelly proteins. Similarly, Cardoen et al. (2011) used arrays to compare sterile and reproductive workers and reveal hundreds of genes differentially expressed. To the extent that these experimental designs do capture genes that explain the functional expression of sterility, then these types of studies do identify candidate genes for worker reproductive altruism.
1.4.3 Prediction 3: Genes for altruism should increase in number and complexity with social-behavioral sophistication

Mutations that predispose essentially selfish individuals toward altruism may initially involve a single gene, such as a gene that renders mutants less reproductive or more likely to help, defend or provision beyond their own brood. If selection promotes this initial gene-based division in labour, then selection may subsequently embellish this division for evermore specialized or efficient divisions that involve larger numbers of downstream and interacting genes. This co-evolutionary scenario between social and genomic complexity is consistent with the ‘genetic release’ hypothesis of Gadagkar (1997). He hypothesized that because the solitary ancestors of social Hymenoptera would have had to perform the full suite of reproductive and non-reproductive tasks necessary to raise brood, the genes involved may be constrained in their response to selection via antagonistic pleiotropy. The evolution of task specialization in which different females perform different tasks therefore releases this hypothetical constraint and allows genes to respond to directional selection. If Gadagkar’s idea is correct, then I expect well differentiated ‘advanced’ eusocial bee species to have genomes enriched for genes related to task specialization, relative to ‘primitively’ eusocial species.

Hunt et al. (2010) found that rates of molecular evolution were indeed faster at loci related to queen function in Apis mellifera than were their orthologues in non-social Hymenoptera or solitary non-Hymenopteran insects, providing some support to the ‘genetic release’ hypothesis. However, this pattern was not detected at loci related to
worker function, and it remains equivocal whether task specialization co-evolves with particular gene family expansions. Until the number of fully sequenced bee genomes increase to include species from different grades of sociality – from solitary to primitively eusocial to advanced eusocial – then a provisional test of this hypothesis might come from comparative genomic studies that measure caste biased gene expression. Because workers and queens are specialized for different behavioural repertoires, then worker-biased (or queen-biased) gene expression provides a crude test of this idea (Evans and Wheeler 2001; Grozinger et al. 2007). I can determine whether directional selection is acting on caste-biased genes by comparing their substitution rates at non-synonymous and synonymous sites to the substitution rates of non caste-biased genes. Finally, evidence for molecular co-evolution between genes and sociality may be found on a smaller scale, within particular gene families. Evidence for this is seen in major royal jelly proteins that are derived from the Drosophila Yellow gene family but are unique to the (eusocial) Apis clade. Major royal jelly proteins are thought to have co-evolved with honey bee sociality and apparently radiated within that lineage (Drapeau et al 2006).

1.4.4 Prediction 4: Genes for altruism should co-evolve with, or depend on, the previous evolution of genes for kin recognition

In order for any gene for altruism to evolve its effect must be directed towards other carriers in which the gene is typically not expressed. Hence kin selection should favour kin recognition, which, in social insects, involves perception of shared genetic or environmental cues (Ratnieks 1991). For honey bees kin recognition has been
relatively well studied, with semi-heritable cuticular hydrocarbons playing a major role (e.g., Fletcher et al. 1987; Page et al. 1989; Arnold et al. 1996; Arnold et al. 2000; Le Conte and Hefetz 2008). Although the exact mechanism for hydrocarbon matching remains unknown, the molecular identities of some recognition proteins have been uncovered (Kamikouchi et al. 2004; Calvello et al. 2005; Fang et al. 2012).

In polyandrous hives where queens are multiply mated, kin selection further predicts that workers will preferentially feed larvae that are their full sisters over their half sisters (Page et al. 1989). Arnold et al. (2000) have shown that *A. mellifera* workers are able to discriminate between full- and half-sisters, suggesting that advanced kin recognition systems have evolved and may enable workers to strategically maximize their inclusive fitness. To determine whether the evolution of altruism genes preceded or occurred alongside the evolution of kin recognition genes, a type of comparative analysis can be conducted through a phylogeny (Harvey and Pagel 1991). Eusocial species, like the honey bee, would possess both traits, whereas subsocial species, like the small carpenter bee in the same subfamily (Apinae), do not exhibit reproductive altruism but have evolved kin recognition (Rehan and Richards 2013). One could score the presence or absence of ‘kin recognition’ and ‘altruism’ as discreet variables in related bee or Hymenopteran species to determine whether these traits arose together or independently.

**1.4.5 Prediction 5: Genes for altruism reside in regions of low recombination, exhibit co-expression, and show modular genetic architecture**
If selection promotes genetic task specialization, then it may likewise favour gene linkage into co-expressed functional groups. The evolution of sociality may in time select for genetic loci for worker traits to be grouped closer together on a chromosome. For example, if genes a, b and c are essential to queen-worker dimorphism, then genetic recombination separating these genes may prove maladaptive and selection should favour their joint segregation and joint transmission.

I therefore predict genes co-adapted to a social complex to not only be linked but to map to areas of low recombination (Linksvayer et al. 2013). There is emerging support for this prediction in the eusocial fire ant *Solenopsis invicta*, where a tight linkage of genes that experience no recombination are responsible for several traits involved in social organization (Wang et al. 2013), arguing for the existence of a ‘social chromosome’

The support for this in honey bees has yet to be seen. Recent genetic maps of the honey bee genome show that there are very high recombination rates distributed throughout the genome (Beye et al. 2006; Weinstock et al. 2006), and genes specifically related to behaviour and worker phenotype are among those that experience the highest rates of molecular evolution (Kent et al. 2012), although this observation may depend on the precise calculation performed (Hunt et al. 2013).

Efforts to map genes associated with worker sterility have already begun (Oxley et al. 2008), and further progress with mapping loci involved in worker- vs. queen-biased traits will help us better understand their degree of genetic linkage. For example, certain QTLs and genes associated with worker behaviour such as foraging and
defense (Hunt et al. 2007) and alloparental care (mrjps; Drapeau et al. 2006) have been physically mapped, but have not yet been compared to QTLs for other worker traits to determine their proximity in the genome. Furthermore, I can indirectly test this prediction by examining the expression profiles of worker-trait genes. If two genes are expressed together, they may be similarly regulated and may therefore be found in close genetic proximity (Stuart et al. 2003). Gene co-expression networks (that cluster together genes expressed similarly) could be constructed to show how genes underlying different worker traits occur together in network modules. Chandrasekaran et al. (2011) constructed a network based on co-expression data and transcriptional binding data on 48 behavioural worker phenotypes and noticed most genes converged onto three unique modules that represent ‘foraging’, ‘maturation’ and ‘aggression’. Furthermore, all three distinct worker behaviour modules were controlled by four transcription factors (broad, dl, lilli, and GB13780), showing that the genes that underlie various worker behaviours depend on shared transcriptional regulators. From here it would be possible to scan the genome to determine if the multifunctioning transcription factors or the worker-trait genes that comprise this modular genetic architecture do occur in areas of low recombination.

1.4.6 Prediction 6: Genes for altruism should have partially additive effects

Natural selection for any trait requires heritable variation, and to be heritable genes must exhibit additive effects. The genes for altruism that allow queen and worker phenotypes to evolve must therefore have an additive genetic effect. One way to test for additive effects is to examine whether the development of a larva into a queen or a
worker is at least partially influenced by her genotype. To date, genetic components to caste differentiation are known from sixteen species of wasps, ants, bees and termites (Schwander et al. 2010), but in many of these cases the genetic effect may be weak, epistatic (i.e., non-additive), conditional on environmental context (i.e., GxE dependent) or stem from maladaptive interspecific hybridization (Schwander et al. 2010). In honey bees experimental evidence for heritable effects on reproductive fate can be seen in queen-worker caste differentiation (Tilley and Oldroyd 1997), worker egg-laying behaviour (Robinson et al. 1990) and in ovary activation (reviewed in Backx et al. 2012). To directly estimate additive effects on aspects of worker sterility it should be possible to employ breeding designs that partition the genetic variance into additive and non-additive components. One such estimate of the narrow sense heritability of ovariole number in a population of cape honey bees (A. mellifera capensis) is 19% – 57% (Goudie et al. 2012).

1.4.7 Prediction 7: Genes for altruism should show strong pleiotropy

Social behaviours are complex in that they are the outcome of perceiving, interpreting and responding to cues from conspecifics, and as such can involve several processes (neuronal, hormonal, motor, etc.). Genes for social behaviours, including altruism, may therefore exhibit strong pleiotropy, as for other complex traits (e.g., Featherstone and Broadie 2002; Solovieff et al. 2013). From socio-genetic studies to date, genetic pleiotropy is known from eusocial ants, social slime moulds and social vertebrates (Foster et al. 2007; Keller 2009), but the evidence from honey bees is scarce. In A. mellifera the workers are noticeably different from queens in their physiology,
morphology and behaviour, implying a high level of molecular coordination within each caste. For example, workers with inactive ovaries work faster (Dampney et al. 2004), forage earlier (Oldroyd and Beekman 2008) and accelerate their metabolism relative to ovary activated workers of the same age. One of the best examples of social pleiotropy is the yolk protein *vitellogenin* that has evolved multiple social roles in the honey bee (Nelson et al. 2007). In queens, the gene is expressed in the ovaries as a food source to the developing embryo, but in barren workers, vitellogenin is expressed in the hypopharyngeal gland (head) where it is incorporated into royal jelly and fed to larvae as a form of alloparental care (Amdam et al. 2003). This dual social role for vitellogenin – in parental and alloparental care – demonstrates how a single gene can affect more than one social trait.

1.5 Conclusions

In this review I have outlined the potential to find genes associated with altruism, a social behaviour that enabled the evolutionary transition to eusociality multiple times throughout the tree of life. These genes have been predicted to exist since the 1960s, but the disconnect between theoretical advances of kin theory and empirical evidence has prevented their discovery. In this review, I applied a recently adopted conceptual framework for finding these genes for altruism to an important model of social behaviour. Predictions 2 (genes for altruism are environmentally sensitive) and 6 (genes for altruism should be additive) have some support, with expression data revealing novel altruistic genes to examine further and evidence to support that altruistic genes are in fact heritable. The least evidence exists with regards to
Prediction 4 (genes for altruism should depend on the previous evolution of genes for kin recognition) and 5 (genes for altruism may show modular genetic architecture), with limited information suggesting altruistic genes evolved with or after kin recognition genes and that these genes reside in areas of low recombination in the genome.

I have also suggested how these shortcomings may be addressed. For example, it should be possible to use expression profiling and network modeling to identify even more genes associated with altruistic social behaviour. Genes identified in these networks can then be tested by manipulating their expression through RNA interference and observing the phenotype outcome of silencing specific genes (Gadau and Hunt 2009). These and previously identified genes can then be tested for other predictions such as whether it is co-expressed with other worker related genes or exhibits multiple functions related to social behaviour. Although these characteristics have been specifically applied to the honey bee, they may also pertain to other social organisms and even the commonly studied behaviours of humans that lead to our complex and cooperative societies. Furthermore, understanding the genetic underpinnings of social structure for insects like honey bees, ants and termites has economic advantages with regards to agriculture and pest management. I hope that following this conceptual framework will, in time, enable characterization of all the genes that underlie altruism and other complex behaviours in social organisms that Hamilton so long ago predicted.
References


Linksvayer TA, Wade MJ. 2009. Genes with social effects are expected to harbor more sequence variation within and between species. Evolution 63:1685-1696.


2. Gene co-citation networks associated with worker sterility in honey bees

2.1 Introduction

The well-understood theory of kin selection explains how complex social behaviour can evolve at the gene level (Hamilton 1964; Lehmann and Keller 2006; Bourke 2011), yet the theory is incomplete because it does not predict which genes promote the expression of reproductive altruism. The recent genome sequencing of the honey bee *Apis mellifera* (Weinstock et al. 2006) and of other social organisms (e.g., *Dictyostelium discoideum* by Eichinger et al. 2005; several ant species by Gadau et al. 2012) is creating new opportunities to identify genes involved in reproductive regulation and social coordination. For example, *vitellogenin* (Amdam et al. 2004; Thompson et al. 2008), *major royal jelly proteins* (Thompson et al. 2006), insulin signaling genes (Grozinger et al. 2007; Cardoen et al. 2011b), and ecdysteroids (Paul et al. 2006; Wang et al. 2012) are among a growing set of genes implicated in reproductive regulation. Despite these advances from microarray and quantitative PCR studies, there has not yet been an attempt to link these genes into functional pathways that explain the phenotypic expression of worker sterility.

Honey bees are a model system for studying the sociogenomic basis of worker reproductive altruism and sterility (Sumner 2006; Weinstock et al. 2006; Robinson et al. 2008). Like other highly social taxa, eusociality in honey bees is characterized by a reproductive division in labour between reproductive and non-reproductive specialists (Seeley 1985). The queen caste is sexual and highly fecund, with well-developed ovaries that each contain ~150-180 ovarioles. The worker caste, by contrast, is non-
sexual and has only rudimentary ovaries with few ovarioles (Velthuis 1970). Workers are effectively sterile in the presence of a functional queen, and though this trait has many physiological components, sterility is most commonly measured as a function of ovary in-activation (Backx et al. 2012). One approach to identifying genes integral to the expression of worker reproductive altruism and sterility is therefore to screen for genes that control ovary activation (Thompson et al. 2006).

For workers, sterility from ovary inactivation is not obligate but rather is conditional on social context. As predicted from kin theory, workers refrain from activating their ovaries to lay eggs when the indirect fitness pay-off surpasses a conditional threshold (Barron and Robinson 2008). For individual workers this threshold is in part dependent on queen fecundity, and is communicated to workers by the queen’s pheromonal signal (Hoover et al. 2003; Le Conte and Hefetz 2008). When a queen is healthy and fecund, her daughter workers will generally refrain from activating their ovaries, but when she is weak or absent, a proportion of workers may activate their normally dormant ovaries to lay unfertilized eggs (Visscher 1989). Because worker sterility is conditional on the strength of queen signal, I likewise expect genes regulating ovary activation to be conditionally expressed – in particular, in response to queen mandibular pheromone (QMP).

Previous studies have begun to identify genes differentially expressed as a function of pheromones (Grozinger et al. 2003; Thompson et al. 2006, 2008; Alaux et al. 2009; Kocher et al. 2010; Cardoen et al. 2011b), but as yet no study has systematically
compared these gene lists or compiled them into a network of potentially interacting genes that collectively function to turn worker ovaries on and off (Bloch and Grozinger 2011; Wang et al. 2012). Inferring a gene network for the control of worker ovary activation will help determine how worker sterility is regulated at the molecular level, and will represent our best example yet of how genes interact with each other and with their environment to coordinate one of the best-known forms of reproductive altruism.

Using a network biological approach (Barabasi and Oltvai 2004), I first collect studies from the literature that identify genes differentially expressed by workers as a function of queen signal. Second, using data from comparable studies I infer, for the first time, the functional relationship among candidate genes using co-citation networks. A co-citation network is a graphical representation of how genes might interact with each other to functionally affect a phenotype. The networks infer pairwise interactions between genes if they are mentioned within the same sentence of a written abstract published in PubMed – the co-citation being used to suggest a functional relationship between them (Jensen et al. 2006). Despite how new the co-citation network approach to describe molecular interactions is, several studies have validated them experimentally to show they do indeed predict meaningful biological relationships among genes and proteins (e.g., Jenssen et al. 2001; Stephens et al. 2001; Chen and Sharp 2004; Schmelzer et al. 2008; Wiggins et al. 2010).
Candidate genes identified from microarray studies alone are typically those with the highest or most consistent expression differences. Network analysis, by contrast, builds upon these gene-list outputs to identify genes of importance via a different criterion – namely, those with the highest connectivity (Newman 2010). Identifying well-connected ‘hub’ genes within networks can help pinpoint the crucial junctures that enable network function (Jeong et al. 2001). Given the flurry of gene expression analyses that followed the Honey Bee Genome Sequencing Project (e.g., Cristino et al. 2006; Drapeau et al. 2006; Wheeler et al. 2006), there is now worldwide interest in converting the data generated from these analyses into provisional networks that describe how worker sterility is regulated within eusocial bee colonies. Moreover, the as-yet-unknown network is potentially related to the networks that regulate other aspects of honey bee social coordination, such as a tendency to specialize on pollen vs. nectar among foraging workers (Amdam et al. 2006; Hunt et al. 2007), or the tendency for individual workers to specialize on within-colony vs. out-of-colony tasks (Wang et al. 2010; Bloch and Grozinger 2011).

For honey bees, several studies have suggested a single, conserved pathway that regulates ovaries in response to pheromonal cues (Thompson et al. 2006, 2008; Abbot et al. 2011; Cardoen et al. 2011b). In this study I test this single-pathway hypothesis by generating co-citation networks from genes previously implicated in the regulation of worker ovaries. First, I identify suitable microarray experiments that derive gene sets related to ovary activation. I then use the computer software suite GENOMATIX PATHWAY SYSTEM (GENOMATIX, Munich) to evaluate whether co-citation networks
can adequately explain variation in this trait. From the networks inferred, I test whether worker ovary activation is best explained by a single, conserved pathway that is retrieved by different studies, or whether variation in this trait is better explained by multiple networks that vary with regard to the age, population or pheromone treatment of workers. This latter scenario would suggest that no single pathway explains the conditional expression of worker sterility, and that multiple pathways are utilized by workers under different circumstances, in different populations or at different phases in a worker’s life. Finally, our analysis will allow us to test the extent to which any inferred networks show homology to those known from Drosophila or other insects, as predicted from recent sociogenomic theories (Amdam et al. 2006; Toth et al. 2007).

2.2 Methods

2.2.1 Meta-analysis and network construction

In October 2012, I compiled microarray data from the literature by searching the Web of Science using the following search criteria: [TOPIC = honey bee OR honeybee OR Apis mellifera] coupled with [TOPIC = gene expression OR microarray] and [TOPIC = steril* OR ovar*], whereby the latter terms capture topics such as sterile, sterility, ovary, ovarian, etc. I also searched for analyzed microarray data directly using the search function of ArrayExpress online databases (www.ebi.ac.uk/arrayexpress) with the filter [Species = Apis mellifera].
To circumscribe studies that most closely identify genes that regulate ovary activation, I included data sets from studies that met the following criteria. Studies must have i) reported normalized gene-expression differences between ovary-active and ovary-inactive adult workers, ii) controlled for genetic and environmental variation, iii) used queen mandibular pheromone (Plettner et al. 1993) as the principle cue for manipulating ovaries, and iv) quantified the level of ovary activation via an explicit scoring scheme. Studies that were generally on-topic but that did not have a pheromone-untreated control group (Kocher et al. 2010), did not use queen mandibular pheromone (Alaux et al. 2009; Wang et al. 2012) or did not explicitly score ovaries (Alaux et al. 2009) are valuable in their own right but were excluded from our meta-analysis.

Prior to up-loading acquired microarray data into GENOMATIX PATHWAY SYSTEM, I first generated standardized gene lists. This pre-processing step enabled comparison between studies that used different sets of microarray probes. For studies using the complimentary DNA (cDNA) platform described in Whitfield et al. (2002), I converted expressed sequence tag (EST) accessions to the corresponding gene accession from Version 2.0 of the Official Gene Set (Munoz-Torres et al. 2011). I then manually curated and identified the single best significant ($E$-value < $10^{-5}$) BLASTp match in *Drosophila melanogaster* (Version 5.10; Jasny and Bloom 2000). Bee ESTs that did not correspond to a coding sequence in the fly were, out of necessity, excluded from downstream analysis (a minority of genes, see Results). For studies using the honey bee oligonucleotide microarray, in which probes are already
linked to the Official Gene Set (array described at ArrayExpress under accession A-MEXP-755), I simply used BLASTp to directly assign the most likely *D. melanogaster* homologue. The meta-data matrices that I used as input for pathway analysis therefore consisted of fruit fly homologues that correspond to the differentially expressed bee genes. The database and pathway analysis algorithms of Pathway System software are optimized for the fly and I simply transferred the direction and magnitude of bee gene expression changes to the fly homologues. I uploaded gene lists and expression profiles to Genomatix Pathway System. The algorithm uses a gene recognition strategy described by Frisch et al. (2009) to scan the PubMed database for genes mentioned together and it subsequently builds the network by adding interactions with the highest number of co-citations first. To minimize falsely implied connections between genes and increase the reliability of the networks, I applied the ‘function word’ filter recommended by Jensen et al. (2006) in which interactions are only drawn between genes if the sentence linking them explicitly implies a functional interaction recognized by the built-in search tool. For example, gene x ‘inhibits’, ‘phosphorylates’, ‘is the target of’, gene y, etc. Finally, because gene-data loss occurs when converting ESTs to official bee genes, genes to fly homologs, and finally at the level of co-citation in the literature, I applied a series of Chi-square tests for independence to determine whether the gene composition of networks at each stage of analysis were an unbiased sample of the original gene expression dataset. Specifically, I calculated whether or not the proportions of up- and down-regulated genes at each stage were similar to or significantly different ($\alpha = 0.05$) from the proportions initially reported by the studies.
2.2.2 Within network analysis

For each network I first used the Universal Protein Resource (UniProt; Bairoch et al. 2007) to assign each network gene a cellular function (e.g., kinase, cofactor, etc.). Specifically, I queried the UniProtKB database (gene name AND organism: “Drosophila melanogaster [7227]”) for all genes and distinguished these types of protein products visually using graphical symbols based on what is assigned under “General Annotation” of each gene name.

Second, I analyzed our networks above the level of the gene, via enrichment analysis as implemented in FUNCASSOCIATE software (Berriz et al. 2003). Here, I used a Fisher’s exact test (with a Monte Carlo False Discovery Rate simulation; adjusted family-wise error rate $\alpha=0.05$) to determine the most common functions and pathways of each network. To do this, I first calculated the number of genes expected to have particular Gene Ontology (GO) functions for a random network of the same size, assuming the random network samples genes relative to their true frequency in the (Drosophila) genome. I then compared this null expectation to the actual number of genes observed for the same functions in each of our networks. If my inferred networks are biologically functional, then I expect an over-representation of genes for that GO function.

For each network I also identified the highest connected (‘hub’) genes, and plotted the degree distribution (see Appendix Figure A.1), where the degree is the number of connections per gene (Newman 2010). For the single most connected gene in each
network, I verified its implied interactions by querying genes against the *Drosophila* Interaction Database (DroID version 2013_02 database, http://www.droidb.org). This database is searchable for experimental evidence from protein-protein interaction studies, genetic interaction studies, transcription factor-gene interaction studies, and miRNA-gene interaction studies, if any.

2.2.3 Between-network analysis

Because the input studies to our meta-analysis are variable with respect to populations of bees, experimental detail, and even array platform (Table 1), I expect our co-citation networks to vary. As a proxy for topographical convergence between inferred networks, either between studies for a given worker age or within studies for a different bee age, I determined the number of genes found to occur in more than one network. I then noted whether these recurring genes comprised the highest connected genes of any networks.

2.3 Results

I included six studies that represent 14 different microarray experiments in our meta-analysis (Table 1). These studies are comparable in that they screen for genes differentially expressed as a function of worker ovary activation in the presence or absence of queen pheromone, and therefore generate data that is suitable input for our proposed network analysis. From the set of input studies I can now potentially construct networks that describe the interactomes within brain, abdominal or whole body tissues, and do so across a range of worker ages from 1-day to 18-days post
eclosion. On average, 52% (1884 of 3625) of ESTs identified from cDNA microarrays corresponded to Official Gene Set bee genes. Of all bee genes, including those from oligo arrays, roughly 78% (4409 of 5675) had unambiguous fruit fly homologs, and 19% (830 of 4409) of these genes had sufficient co-citation data to be incorporated into networks. Summary statistics for these genomic data are provided in Appendix Table A.1.
Table 1. Studies included in the meta-analysis. I included a total of n = 14 meta-datasets that I sourced from the independently published studies listed. For each meta-dataset I provide summary information on the experimental platform, the type of tissue and the age of workers. I also provide the number of differentially expressed genes (DEGs) that I converted to fly homologs prior to network construction.

<table>
<thead>
<tr>
<th>Study</th>
<th>Experimental platform</th>
<th>Tissue type</th>
<th>Age of workers(^2)</th>
<th>DEGs</th>
<th>DEGs after conversion</th>
<th>Number of potential networks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grozinger et al. 2003</td>
<td>Wild type bees in cages with QMP(^1)</td>
<td>Brain</td>
<td>1</td>
<td>287</td>
<td>181</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>1080</td>
<td>469</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td>1242</td>
<td>540</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4</td>
<td>391</td>
<td>334</td>
<td></td>
</tr>
<tr>
<td>Thompson et al. 2006</td>
<td>Anarchist vs. wild type bees in hive with queen</td>
<td>Brain, Abdomen</td>
<td>4</td>
<td>20</td>
<td>13</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>20</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>Grozinger et al. 2007</td>
<td>Wild type bees in cages with QMP</td>
<td>Brain</td>
<td>10</td>
<td>221</td>
<td>103</td>
<td>1</td>
</tr>
<tr>
<td>Thompson et al. 2008</td>
<td>Anarchist vs. wild type bees in hive with queen</td>
<td>Brain, Abdomen</td>
<td>4</td>
<td>7</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Cardoen et al. 2011</td>
<td>Wild type bees in hive without queen</td>
<td>Whole body</td>
<td>18</td>
<td>1292</td>
<td>1077</td>
<td>1</td>
</tr>
<tr>
<td>Backx et al. 2012</td>
<td>Wild type bees in cages with QMP</td>
<td>Brain</td>
<td>4</td>
<td>564</td>
<td>338</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>6</td>
<td>782</td>
<td>527</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>8</td>
<td>623</td>
<td>428</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>10</td>
<td>534</td>
<td>387</td>
<td></td>
</tr>
</tbody>
</table>

\(^1\) QMP, queen mandibular pheromone; \(^2\) Days post-eclosion
2.3.1 Networks from brain tissue analysis

From the 14 different data sets, I successfully generated 9 networks. The remaining five data sets from Thompson et al. (2006), Thompson et al. (2008), and Grozinger et al. (2007) (Table 1) were not amenable to network analysis due to either the small number of DEGs identified (≤ 20 per experiment) and the even smaller subset that were suitable for downstream analysis via homology to the fly, or due to the small number of co-citations found in the literature. Eight out of nine networks were derived from worker brain tissue. Figure 1 shows the set of networks inferred from the DEG sets of Grozinger et al. (2003), which correspond to workers of different ages. In each data set, I infer a single main network that incorporates a majority of genes, within only a minority of genes excluded from the main network to form minor connections among themselves, or to remain unconnected as singletons. Some networks reveal genes that are potentially of functional importance – for example, \textit{dlg1} in Network 1C or \textit{arm} in Network 1D are particularly well connected. From the Grozinger et al. (2003) study, the networks I infer also vary in size, in this case ranging from \( n = 24 \) - 135 genes. The networks I infer from other brain tissue data sets showed comparable topologies. Figure 2 shows the networks derived from 4-, 6-, 8-, and 10-day old bees, as inferred from the DEG sets identified by Backx (2011). These networks vary in size from 34 to 63 genes and reveal additional highly connected genes: \textit{Rel} in Network 2A, \textit{bsk} in Network 2B, and \textit{abd-A} in Network 2C.
Figure 1A. Co-citation network from Grozinger et al. (2003) gene list for 1-day-old workers. Each node is a gene and each edge a potential interaction between two genes. Pink genes are up-regulated in workers exposed to queen mandibular pheromone (as determined in the original study), and blue genes are correspondingly down-regulated. Genes highlighted with a circle function in the biological processes pertaining to ‘reproduction’ (GO:0000003).
Figure 1B. Co-citation network from Grozinger et al. (2003) gene list for 2-day-old workers. Notation is the same as for Figure 1A.
Figure 1C. Co-citation network from Grozinger et al. (2003) gene list for 3-day-old workers. Notation is the same as for Figure 1A.
Figure 1D. Co-citation network from Grozinger et al. (2003) gene list for 4-day-old workers. Notation is the same as for Figure 1A.
Figure 2A. Co-citation networks from Backx et al. (2011) gene list for 4-day-old workers. Description of symbols and topography is as for Figure 1, with the exception of a colour code change: genes ranging from yellow to pink are up-regulated in ovary in-active bees and genes ranging from green to blue are down-regulated in ovary active bees.
Figure 2B. Co-citation networks from Backx et al. (2011) gene list for 6-day-old workers. Notation is the same as for Figure 2A.
Figure 2C. Co-citation networks from Backx et al. (2011) gene list for 8-day-old workers. Notation is the same as for Figure 2A.
Figure 2D. Co-citation networks from Backx et al. (2011) gene list for 10-day-old workers. Notation is the same as for Figure 2A.
Figure 3. Co-citation network from Cardoen et al. (2011) gene list for 18-day-old workers. Description of symbols and topography is as for Figure 1, with the exception of a colour code change: genes ranging from green to blue are up-regulated in ovary in-active bees and genes ranging from yellow to pink are down-regulated in ovary active bees. Inset shows a close-up of a region of interest.
2.3.2 Network from whole body tissue analysis

The single experimental data set that was derived from whole body tissue (head + thorax + abdomen) yielded an expansive co-citation network, as expected given the tissue heterogeneity (Figure 3). This network corresponds to workers that are 18-days of age, and is inferred from the DEG set identified by Cardoen et al. (2011b). This dataset corresponds to the oldest aged workers included in our meta-analysis, and the inferred main network consists of 323 genes with only three genes that remain disconnected. A Histone 2A variant gene (His2Av) is shown to have as many as 24 functional connections.

2.3.3 Genes of functional importance

Eight out of nine networks (all except Network 2D) contain highly connected genes that show between four (Hsp83; Network 1A) to twenty-four (His2Av; Network 3) interactions (Figure 4). GO analysis suggests that these so-called hub genes function to regulate gene expression (Rel, abd-a, arm, and His2Av), are involved in signaling (dlg1, bsk, Rho1), or are molecular chaperones (Hsp83). Six out of eight of these functionally important genes (bsk, abd-A, Hsp83, Rho1, dlg1, and arm) also have GO functions related to reproduction.

In addition to co-citation support in the literature, I found experimental support for several of the hub gene interactions. DroID analysis confirmed that fully 32% of co-citation interactions (24 of 74) are associated with protein-protein interactions, transcription factor-gene interactions, genetic interactions, or combinations thereof, in
Figure 4. The most connected genes of each network and their functions. Black hubs are from networks created with brain gene expression and grey hubs are from networks created with whole body gene expression. Hubs are grouped based on their degree (number of genetic interactions) and also on their functions in gene expression, cell signaling, or other processes.
fruit flies and other model organisms (Summary of DroID analysis is available as Appendix Table A.2). All hub genes had at least one confirmed interaction, and one gene (Hsp83) had all of its four interactions experimentally confirmed. This level of cross-validation suggests that connections I have inferred from PATHWAY SYSTEM analysis are biologically robust.

2.3.4 Network enrichment analysis
The networks inferred here show evidence for functional enrichment for genes related to multiple biological processes. Sixty terms were enriched in all networks, including oogenesis (GO:0048477), neuron differentiation (GO:0030182), and response to chemical stimulus (GO:0042221), among others (Table 2). These networks were also enriched for genes related to ‘reproduction’ (GO:0000003), but even here the number and identity of these genes depended on the age of the bee and the type of tissue (Table 1). From this analysis, I reveal a total of 170 different genes involved in reproduction in our networks. It is worth noting that the highest connected genes in Networks 1A-D, as well as Networks 2B and 2C, are implicated by GO analysis to function in reproduction. These highest connected genes are clearly relevant to our trait of interest, worker sterility.

Several pathways were also enriched across multiple networks. All but one network (Network 2C) were enriched for functional constituents of the ‘cell surface receptor signaling pathway’ (GO:0007166). Three (Network 2A, 2B, 3) were enriched for ‘insulin receptor signaling pathway’ (GO:0008286), one (Network 2C) for ‘dopamine
Table 2. Enrichment analysis of gene networks. For each network that I inferred (Networks 1-3) I found the GO terms that are most likely to represent the network’s biological functions. Sixty biological processes were significantly enriched among all nine networks ($P < 0.01$). Nine selected processes, the number of genes associated with these processes, as well as the observed/expected (O/E) ratios averaged over all networks, are presented.

<table>
<thead>
<tr>
<th>Category</th>
<th>GO term</th>
<th>GO number</th>
<th>NG</th>
<th>Average O/E</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ovary Activation</td>
<td>Reproduction</td>
<td>GO:0000003</td>
<td>250</td>
<td>3.06</td>
</tr>
<tr>
<td></td>
<td>Female gamete generation</td>
<td>GO:0007292</td>
<td>171</td>
<td>4.40</td>
</tr>
<tr>
<td></td>
<td>Oogenesis</td>
<td>GO:0048477</td>
<td>171</td>
<td>4.46</td>
</tr>
<tr>
<td>Brain and Behaviour</td>
<td>Neuron differentiation</td>
<td>GO:0030182</td>
<td>208</td>
<td>6.35</td>
</tr>
<tr>
<td></td>
<td>Response to chemical stimulus</td>
<td>GO:0042221</td>
<td>194</td>
<td>4.55</td>
</tr>
<tr>
<td>Signaling</td>
<td>Signal transduction</td>
<td>GO:0007165</td>
<td>253</td>
<td>3.17</td>
</tr>
<tr>
<td></td>
<td>Cell communication</td>
<td>GO:0007154</td>
<td>332</td>
<td>3.05</td>
</tr>
<tr>
<td>Foraging/Flight Related</td>
<td>Compound eye development</td>
<td>GO:0048749</td>
<td>109</td>
<td>4.83</td>
</tr>
<tr>
<td></td>
<td>Locomotion</td>
<td>GO:0040011</td>
<td>141</td>
<td>5.02</td>
</tr>
</tbody>
</table>

1. The total number of genes found in all networks associated with the GO term
receptor signaling pathway’ (GO:0007212), and one (Network 1B) for ‘steroid hormone-mediated signaling pathway’ (GO:0043401).

2.3.5 Gene overlap among networks
There was visually little gene overlap among the networks, and no single gene was found in all nine networks. Yet, a noteworthy proportion of genes (136 of 824, or 17%) were found in more than one network, with 10 of these genes found to span all four networks inferred from the Backx (2011) study. Moreover, I found 34 genes to span all four networks inferred from the Grozinger et al. (2003) study. In total, I found 96 genes to span at least two networks across all of the different studies. The most recurring genes, Src42A and Hsp83, were found in five of the nine networks. The hub genes of four networks were also re-occurring genes; Hsp83 is found in five networks, Rho1 and dlg1 in four, and Abl in two.

2.3.6 Co-citation network bias
A majority of networks (6 of 10) showed no statistical bias with respect to proportion of genes up- or down-regulated, compared to the original data sets from which the networks were inferred ($\chi^2$ tests, $P > 0.05$ in each case). The remaining networks did however include a disproportionate number of genes up-regulated ($\chi^2 = 13.58-24.4$, d.f. = 3, $P <0.001$ in all cases), either upon conversion to fly homologues (Networks 1B) or upon retrieving co-citation links from PubMed (Networks 1D, 3 and 4). This biased sampling of some DEG sets simply reflects the information currently available within the PubMed database.
2.4 Discussion

In this study I have generated the first gene networks that describe variation in ovary activation among honey bee workers. As such, these networks may be useful for understanding how worker reproductive altruism and sterility is regulated at the physiological and molecular level. The networks presented are derived from a functional analysis of gene sets previously identified from microarray studies (Table 1). One pattern to emerge from this meta-analysis is that each network is generally inclusive, with a majority of annotated genes forming linkages into single, main networks (Figures 1-3). This level of connectivity, together with the substantial size of some networks, suggests that the underlying DEG sets are biologically informative, and that the networks do reflect functionally interacting genes (Grigorov 2005). Moreover, each network, though highly variable in terms of gene membership, is enriched for biological processes related to reproduction, including oogenesis and other functional terms that are consistent with reproduction and reproductive regulation (Table 2). Our meta-analysis has therefore yielded a set of graphical hypotheses that potentially describe adaptively complex and biological functional networks that worker bees use to regulate personal reproduction within a social context.

2.4.1 Co-citation networks

I included six studies and 14 datasets in our meta-analysis, nine of which were suitable for network analysis (Table 1). The eight networks derived from brain tissue varied substantially in gene membership, size and topology. Figure 1 shows the four
networks inferred through co-citation analysis of genes identified by Grozinger et al. (2003). The smallest network (n = 24 genes) describes how workers that are very young, only 1-day post eclosion, respond to queen pheromone at the molecular level (Figure 1A). This network is neither fully connected nor very large, but does show balanced expression between genes up- or down-regulated, and does identify a single most-connected gene that encodes a heat-shock protein (Hsp83). Bees this young may therefore show insufficient gene activity in response to pheromone to meaningfully assemble this activity into a functional network. Nonetheless, our identification of Hsp83 is significant because this gene has previously been singled-out in relation to honey bee caste differentiation (Evans and Wheeler 2001; Barchuk et al. 2007), oogenesis (Wilson and Dearden 2013), and is a marker for changes in queen reproductive status (Grozinger et al. 2007; Cardoen et al. 2012). All of these implied changes reflect a potential role in reproduction and reproductive divisions in labour.

Bees just one and two days older, however, show evidence of massively coordinated expression (Figures 1B, 1C). These co-citation networks are markedly more complex, and thus are more informative with respect to inferring function and identifying hub genes. The Figure 1B network differed in the degree of up and down gene regulation from that of the fly homolog gene list with 49% (54/110) up-regulated vs. 59% (227/469) up-regulated, respectively. The bias against up-regulated genes in this network suggests that the original gene set contained important elements to reproduction, but that have no GenBank homologue in fly. The highest connected gene in Network 1B is Rho1, a protein that regulates signaling pathways in
development (Magie and Parkhurst 2005). Another highly connected gene in this network is Abl tyrosine kinase, which functions in neural development in honey bees and influences behavioural maturation (Fu and Whitfield 2012). Finally, caged 4-day old workers, at the age just prior to ovary activation (Thompson et al. 2006), yield a relatively small co-citation network (Figure 1D). Again, the bias against genes up-regulated suggests that some proportion of genes associated with reproductive regulation in bees have not previously been mentioned in the literature and is consistent with PubMed simply having more information on certain genes than others. In this network, the highest connected gene arm encodes a transcriptional activator in the wnt pathway to regulate the expression of many genes (Brembeck et al. 2006).

Figure 2 shows the four networks inferred through co-citation analysis of genes identified by Backx et al. (2011). These gene sets are comparable to those identified in Grozinger et al. (2003) in that they are disregulated in groups of caged workers in response to queen pheromone, but measure gene expression over an older age period (Table 1). The 4-day old bee co-citation network is comparable in size to that inferred from Grozinger et al.’s (2003) study (50 vs. 35 genes), but shared only one gene in common (sima) and differed in pattern of gene expression (43 vs. 27 % of genes up-regulated). The highest connected genes in these networks were the immune-related transcription factor Relish (Evans et al. 2006), the oogenesis-related signaling protein basket (Wilson et al. 2011), and abdominal-A, a transcription factor implicated in abdomen and gonad development (Warrior 1994).
Figure 3 shows the single network inferred through co-citation analysis of genes identified by Cardoen et al. (2011b). This network depicts a putative molecular mechanism through which the oldest workers included in our meta-analysis turn their ovaries ‘on’ and ‘off’ in response to pheromonal cues. This is the only co-citation network included in our meta-analysis that is derived from whole body tissue, as opposed to brain tissue (Table 1). This network is well connected, with only three genes separate from the main component, suggesting that the majority of these genes are indeed functioning together. This network showed some bias towards up-regulated genes in comparison to the initial DEG list, again during the detection of co-cited genes. The gene His2Av is not obviously related to reproduction but I have identified this gene silencer as a very well connected gene (24 interactions) and thus may be important to sterility. This gene is connected to those directly implicated in reproduction, including eggless. The gene eggless has a role in Drosophila oogenesis (Clough et al. 2007), so the identity of His2Av and its neighbours within our co-citation network for worker ovary activation and sterility warrants further attention.

2.4.2 Testing candidate pathways for ovary activation

The networks generated in this study provide an opportunity to test previous ideas on the make-up of pathways for ovary activation in honey bees. One pathway that has been implicated in reproductive regulation is the insulin/insulin-like signaling (IIS) pathway. The IIS pathway acts upstream of ecdysteroid and juvenile hormone regulation to control solitary insect reproduction (Flatt et al. 2005) and is required for insect vitellogenesis (LaFever and Drummond-Barbosa 2005). In social
Hymenoptera, it appears to have a critical influence in the evolution of eusociality, as it has been specifically implicated in both reproductive division of labour between castes (Wheeler et al. 2006; Toth et al. 2007) and age-related division of labour within the worker caste (Ament et al. 2008). Our enrichment analysis has identified significant elements of the IIS pathway in some of our networks. Network 2A (adjusted $P<6.00E-03$), Network 2D ($P<1.00E-03$), and Network 3 ($P<7.00E-03$) were enriched for ‘insulin receptor signaling pathway’ (GO:0008286). Six key genes involved in this pathway and its regulation were present in the three networks, including the ligand *Ecdysone-inducible gene L2*, the receptor *chico*, the signaling molecules *Pten*, *dock*, and *Tsc1*, and the target transcription factor *foxo* (Brody 1999). This representation of genes involved throughout the entire pathway, as well as the enrichment across networks from different studies suggests the IIS pathway is implicated in the control of worker sterility.

In addition I have identified elements of two other pathways, the dopamine receptor signaling pathway (GO:0007212, adjusted $P<0.001$) and the steroid hormone mediated signaling pathway (GO:0043401, adjusted $P<0.01$). The dopamine pathway has been implicated in caste differentiation (Hartfelder et al. 1995) and ovary inactivation in the presence of QMP (Vergoz et al. 2012). Honey bees have three dopamine receptors, *Amdop1*, *Amdop2*, and *Amdop3* expressed in their brains and ovaries. The genes *DopR* and *DopR2* were present in Network 2B and correspond to the *Amdop1* and *Amdop2* dopamine receptors. Workers of different ages and behavioural repertoires vary in their expression of all three dopamine receptors in
response to queen pheromone (McQuillan et al. 2012). The homovanillyl alcohol (HVA) component of QMP, one of the principal cues that induced worker sterility via ovary inactivation, binds to the *Amdop3* receptor (Beggs and Mercer 2009), but QMP also modulates the expression of the other two receptors indirectly (Vergoz et al. 2012). Moreover, HVA uptake is accelerated as workers transition between reproductive and non-reproductive states (Sasaki and Nagao 2001; Beggs et al. 2007), suggesting that dopamine signaling is involved in reproductive response thresholds.

Finally, ecdysteroids involved in the steroid hormone mediated pathway have been implicated in bee brain function and oogenesis (Paul et al. 2006). Network 1B contains an ecdysone receptor (*Ecr*) and the ecdysone-induced proteins *Eip78C* and *Hr46* (Wang et al. 2009). *Ecdysone receptor* is needed for ovarian differentiation in *Drosophila* (Hodin and Riddiford 1998). Ecdysteroids are therefore also potentially important to ovary signaling, consistent with Wang et al. (2012) who demonstrated that *Ecr* is expressed in queen and worker ovaries and *Hr46* affects female ovary size. Network 1B is derived from some of the youngest workers included in our analysis (2-days old). If steroid hormone pathways are important to ovary activation, then they would appear to act very early, prior to the visual development of ovaries.

### 2.4.3 A single conserved network for ovary activation?

At first glance, the networks derived here appear to be variable in gene composition, hub gene identity (Figure 4), and the overall interactions they describe (Figures 1-3). This apparent lack-of-convergence onto a single co-citation network, despite them
being inferred from essentially similar datasets (Table 1) suggests that honey bee workers can use different networks to control personal reproduction, perhaps as a function of age, environmental circumstance, or both. Given that workers use different sensory modalities to interact with and respond to changes in their environment, including their social environment (Tanaka and Hartfelder 2004; Cardoen et al. 2011a), it is conceivable that workers may use alternate or redundant pathways to control aspects of reproduction within colonies. Alternatively, a single pathway governing ovary activation in response to social cues seems more parsimonious (Amdam et al. 2006; Toth and Robinson 2007). If so, the multiple networks inferred here may represent segments of the larger, complete network that is still unknown. Different social and environmental signals may activate different suites of genes within this comprehensive network, explaining why the studies that vary in age, pheromone treatment, social structure, and environmental conditions have all captured different gene sets, with some degree of overlap.

The networks identified here represent hypotheses for how workers regulate their ovaries to control reproduction. Within the context of their eusocial colonies, this reproductive machinery is of direct significance to sociogenomic theory that postulates the existence of ‘genes for altruism’ (Hamilton 1963; Dawkins 1976). These genes have rarely been found (Thompson et al. 2013) but the networks presented here, together with other efforts to describe how genes interact with each other and with their cellular, physical and social environment (Bloch and Grozinger 2011; Wang et al. 2012), provide a starting point from which I can begin to test ideas
on the role of specific genes on reproductive phenotypes, including the prospect of finding genes for worker sterility. One approach might be to use functional genomic experiments that perturb hub genes or their neighbours, and then monitor worker phenotype. In particular it will be useful to measure how knock-outs affect worker phenotypes related to reproduction, including response to queen mandibular pheromone, ovary activation and egg laying, as well as other measures of social divisions in labour such as nurse-to-forager transitions or forager specializations. Finally, I suggest that future studies incorporate co-expression information and honey bee protein interactions, so that networks can then be made from *Apis* genes directly without the need for conversion to *Drosophila*, and would not rely on the somewhat haphazard availability of co-citation data in PubMed. The presence of the dopamine, insulin, and ecdysteroid signaling pathways suggests that gene interaction studies will play an important role in uncovering the main endocrine and neuronal control mechanisms that ultimately regulate ovarian physiology and altruistic behaviour in honey bee workers.
References


3. Discussion

The goal of this thesis was to take the well-founded conceptual framework of kin selection and apply it empirically to the honey bee in a case study for reproductive altruism. In Chapter 1, I outlined seven predicted characteristics of ‘genes for altruism’ and illustrated the progress with evidence in the honey bee to date. In Chapter 2, I addressed one prediction specifically – that genes for altruism should be environmentally sensitive – by generating gene networks from published microarray data. Specifically, these networks were constructed using worker gene expression differences upon exposure to queen pheromone, in an effort to better understand the genetic interactions underlying worker sterility and uncover genes involved in reproductive altruism. In this final discussion, I would like to outline some of the challenges, caveats, and future directions with using a brand new co-citation network approach to understanding honey bee sterility.

I predicted my meta-analysis approach of using multiple studies, each slightly varying in experimental design, would yield a set of recurring gene interactions that reveal the main pathway for reproductive regulation. Instead, I found that the gene networks had little genetic overlap, yet they all retained similar functions. If these low levels of gene overlap were due to strictly biological reasons, then this finding will lead to new theories of why and how workers evolved multiple gene networks to govern their reproduction. However, this overlap was likely due to a combination of biological and technical reasons, and now I would like to evaluate the approach of building co-citation networks from several microarrays using fruit fly homologs.
3.1 Technical limitations with using multiple microarrays

The microarrays that were used in this meta-analysis all examined worker gene expression in the presence or absence of queen pheromone, but they were technically variable. Microarrays from Grozinger et al. (2003) were the first expressed sequence tag arrays produced before the honey bee genome was sequenced. The microarrays from Backx (2011) were whole-genome oligonucleotide arrays and included all predicted genes from the genome with additional ESTs. Cardoen et al., (2011) also used a whole-genome microarray, but one that contained an additional 2000 more probes than the ones used by Backx. These studies therefore varied in the original genes that were available on the microarrays, and I created the networks under the assumption that they shared a significant proportion of genes and also that these genes and ESTs reflect a large proportion of the honey bee genome.

For this thesis, caution was exercised when comparing networks made from different microarray datasets to deduce biological meaning. Using different microarray platforms, different statistics, and different laboratories can vary results (Irizarry et al. 2005), and these technical differences likely partially contributed to the low gene overlap I found among networks. The recently introduced BeeSpace project compares worker brain transcriptomes using the same platform and statistical techniques (Zayed and Robinson 2012) in an effort to minimize these differences, but these standards do not yet exist with regards to whole body or abdominal screens used to uncover genes expressed in various ovary states.
3.2 Building a network based on Drosophila homologs

Because network biology is still in its infancy, and reliable modeling programs are limited, I was constrained to exploring *Apis mellifera*-related model organisms. *Drosophila melanogaster* shares a high proportion of single-copy orthologues with the honey bee (Weinstock et al. 2006), and ovarian development between fruit flies and species in the order Hymenoptera are highly conserved (Khila and Abouheif 2010). However, I ideally would have liked to generate networks showing interactions between honey bee genes or proteins directly instead of *Drosophila* homologues, but the majority of the proteome and known genetic interactions for the honey bee have not yet been characterized. Although it is estimated that the genome contains more than 10,000 genes (Weinstock et al. 2006), only a few hundred proteins corresponding to these genes are known (Chan et al. 2011). If I were to construct networks using honey bee genes directly, the networks would reveal few connections between genes due to the limited number of published articles describing interactions between bee genes, proteins, or transcription factors. Using fruit fly homologs that are believed to be similar in structure, function, and evolutionary origin allows us to find potential interactions in the bee that were previously unknown. As fruit fly homologs have been used to predict and annotate the honey bee genome and proteome (Kaplan and Linial 2006; Weinstock et al. 2006), and previous networks based on fly homologues have been capable of predicting important genes for honey bee development (Zheng et al. 2011), our present networks based on these homologs are still biologically informative, although they are provisional.
The bee genes that do not have corresponding homologs in the flies may include key genes for reproductive altruism. Because honey bees are eusocial and fruit flies are solitary, bees likely evolved different genes for social regulation that do not exist in the fly since they diverged from a common ancestor 300 million years ago (Dearden et al. 2006). The co-citation networks that I generated therefore serve as entry-level networks that can be adapted as more knowledge of the honey bee proteome becomes uncovered and honey bee-specific genes can be added to their respective places in the networks.

3.3 Advancing gene networks for worker sterility

Regardless of the technical and species differences, each microarray experiment used in the study generated data to be interpreted in a biologically meaningful way. Two types of entry-level networks that can be constructed from microarray data are co-citation networks and co-expression networks (Shatkay et al. 2000). For this thesis, I chose co-citation networks that provided gene function information and enabled me to perform functional enrichment analyses. Because connections are created when genes are co-mentioned in an abstract, some may not accurately reflect in vivo interactions. However, these networks serve as an entry point for investigating gene relationships (Stapley and Benoit 2000). The main limitation for co-citation networks is that gene interactions are confounded by publications for which genes are not equally represented (Krollinger et al. 2008). Genomatix Pathway System was able to find ~19% of my genes co-cited in the literature. Although I was still able to find interesting putative hubs and biological functions with this amount of genes, I can
only imagine the amount of information contained in a gene network with closer to 100% coverage. Networks based on literature mining have been performed for several organisms including humans (Jenssen et al. 2001) and yeast (Shatkay et al. 2000), and text-mining software to determine accurate gene associations has improved substantially since its beginnings (Jensen et al. 2006; Zhou and He 2008; Steele et al. 2009). Still, in order to improve the gene networks for worker sterility introduced in this thesis, I recommend later studies add co-expression information.

Co-expression networks require gene expression data from multiple time points and cluster genes together that are expressed at similar levels. Although co-expression networks alone do not convey the functions of genes or the nature of their interactions (Shatkay et al. 2000), they would provide more reliability to my networks if two genes are both cited together and expressed similarly (Li et al. 2006). The final step to completing these networks would be if gene sequences were scanned for cis-regulatory elements in their promoter regions so that the transcription factors controlling expression could be determined (Banerjee and Zhang 2002). From here, the optimal gene network – a gene regulatory network (GRN) – can be constructed for the bee, placing the molecular understandings of this model organism on par with *Saccharomyces cerevisiae* (Harbison et al. 2004), *Drosophila melanogaster* (Sandmann et al. 2007), *Escherichia coli* (Resendis-Antonio et al. 2005), *Caenorhabditis elegans* (Arda et al. 2010), and sea urchins (Davidson et al. 2002), for which GRNs already exist.
This thesis was the first study to generate gene networks for worker sterility, the central trait that defines reproductive altruism in the honey bee. Although both honey bee transcriptomics and network biology are still relatively new, I was nevertheless able to identify potential gene interactions, hubs of interest, and key signaling pathways whose importance were previously unknown. With further advances in genomics and network approaches, researchers can expand these entry-level networks and work toward generating the complete gene regulatory network for the honey bee. Identifying this network would be an incredible achievement for both molecular biologists and social biologists, and it would unite our knowledge of the molecular and genetic processes involved in social living with our understanding of kin theory.
References


Appendix A: Chapter 2 Detailed Results

Figure A.1 Degree distributions for each gene network and their relative $R^2$ values.
Table A.1 Gene loss occurring from converting original differentially expressed genes to those that appear connected in a co-citation network. Expressed sequence tags (ESTs) from microarrays are first converted to official bee genes. These genes are then converted to fruit fly homologs, which are entered into the co-citation analysis.

<table>
<thead>
<tr>
<th>Study</th>
<th>ESTs</th>
<th>% of ESTs that code for genes</th>
<th># of bee genes</th>
<th>% of bee genes with fly homologs</th>
<th># of fly homologs</th>
<th>% of homologs in network</th>
<th># of genes in network</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grozinger 2003 Day 1</td>
<td>268</td>
<td>79</td>
<td>213</td>
<td>85</td>
<td>181</td>
<td>13</td>
<td>24</td>
</tr>
<tr>
<td>Grozinger 2003 Day 2</td>
<td>1125</td>
<td>50</td>
<td>560</td>
<td>84</td>
<td>469</td>
<td>23</td>
<td>110</td>
</tr>
<tr>
<td>Grozinger 2003 Day 3</td>
<td>1224</td>
<td>48</td>
<td>590</td>
<td>92</td>
<td>540</td>
<td>25</td>
<td>135</td>
</tr>
<tr>
<td>Grozinger 2003 Day 4</td>
<td>747</td>
<td>94</td>
<td>367</td>
<td>96</td>
<td>334</td>
<td>10</td>
<td>35</td>
</tr>
<tr>
<td>Thompson 2006 - H¹</td>
<td>20</td>
<td>65</td>
<td>13</td>
<td>100</td>
<td>13</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Thompson 2006 - A²</td>
<td>20</td>
<td>65</td>
<td>13</td>
<td>92</td>
<td>12</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Grozinger 2007</td>
<td>221</td>
<td>57</td>
<td>126</td>
<td>82</td>
<td>103</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Cardoen 2011</td>
<td>N/A</td>
<td>N/A</td>
<td>1293</td>
<td>83</td>
<td>1077</td>
<td>30</td>
<td>326</td>
</tr>
<tr>
<td>Backx 2012 Day 4</td>
<td>N/A</td>
<td>N/A</td>
<td>564</td>
<td>60</td>
<td>338</td>
<td>14</td>
<td>49</td>
</tr>
<tr>
<td>Backx 2012 Day 6</td>
<td>N/A</td>
<td>N/A</td>
<td>782</td>
<td>67</td>
<td>527</td>
<td>12</td>
<td>63</td>
</tr>
<tr>
<td>Backx 2012 Day 8</td>
<td>N/A</td>
<td>N/A</td>
<td>622</td>
<td>69</td>
<td>428</td>
<td>10</td>
<td>44</td>
</tr>
<tr>
<td>Backx 2012 Day 10</td>
<td>N/A</td>
<td>N/A</td>
<td>532</td>
<td>73</td>
<td>387</td>
<td>8</td>
<td>34</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td>N/A</td>
<td><strong>52%</strong></td>
<td>N/A</td>
<td><strong>78%</strong></td>
<td>N/A</td>
<td><strong>19%</strong></td>
<td>N/A</td>
</tr>
</tbody>
</table>

¹ Head gene expression, ² Abdominal gene expression
Table A.2 Edge confirmation by Drosophila Interaction Database. Confidence scores (where available) and evidence for Protein-Protein Interactions (PPI), Transcription Factor-Gene (PDI), and Genetic Interactions (GI) occurring between hub genes and their connected gene products in the various networks. Confidence scores range from 0-1. Interactions with higher values are more likely to be biologically relevant than interactions with lower values.

<table>
<thead>
<tr>
<th>Network</th>
<th>Hub Gene</th>
<th>Degree in Networks</th>
<th>Number of Interactions in DroID</th>
<th>% of interactions confirmed by DroID</th>
<th>Associated Gene</th>
<th>Confidence Score</th>
<th>Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Network 1A</td>
<td>Hsp83</td>
<td>4</td>
<td>4</td>
<td>100%</td>
<td>Appl</td>
<td>0.315</td>
<td>PPI</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Hsf</td>
<td>-</td>
<td>PPI</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>eIF-2alpha</td>
<td>-</td>
<td>PPI</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Hsc70-4</td>
<td>0.602</td>
<td>PPI</td>
</tr>
<tr>
<td>Network 1B</td>
<td>Rho1</td>
<td>9</td>
<td>4</td>
<td>44%</td>
<td>tum</td>
<td>-</td>
<td>PPI</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>spir</td>
<td>-</td>
<td>PPI</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Pkn</td>
<td>-</td>
<td>PPI</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Gdi</td>
<td>-</td>
<td>PPI</td>
</tr>
<tr>
<td>Network 1C</td>
<td>dgl1</td>
<td>13</td>
<td>4</td>
<td>31%</td>
<td>Pten</td>
<td>0.34</td>
<td>PPI</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>l(2)gl</td>
<td>-</td>
<td>PPI</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Caki</td>
<td>0.627</td>
<td>PPI</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Cam</td>
<td>0.299</td>
<td>PPI</td>
</tr>
<tr>
<td>Network</td>
<td>Hub Gene</td>
<td>Degree in Networks</td>
<td>Number of Interactions in DroID</td>
<td>% of interactions confirmed by DroID</td>
<td>Associated Gene</td>
<td>Confidence Score</td>
<td>Evidence</td>
</tr>
<tr>
<td>---------</td>
<td>----------</td>
<td>-------------------</td>
<td>-------------------------------</td>
<td>--------------------------------------</td>
<td>----------------</td>
<td>-----------------</td>
<td>----------</td>
</tr>
<tr>
<td>Network 2A</td>
<td>Rel</td>
<td>7</td>
<td>1</td>
<td>14%</td>
<td>cact</td>
<td>0.622</td>
<td>PPI</td>
</tr>
<tr>
<td>Network 2B</td>
<td>Rac1</td>
<td>7</td>
<td>3</td>
<td>43%</td>
<td>Hsp83</td>
<td>-</td>
<td>PPI</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Rho1</td>
<td>-</td>
<td>PPI</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>bsk</td>
<td>0.498</td>
<td>PPI</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GI</td>
</tr>
<tr>
<td>Network 2C</td>
<td>abd-A</td>
<td>4</td>
<td>3</td>
<td>75%</td>
<td>cad</td>
<td>-</td>
<td>PDI</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>trx</td>
<td>-</td>
<td>PDI</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>hth</td>
<td>0.412</td>
<td>PPI</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GI</td>
</tr>
<tr>
<td>Network 3</td>
<td>His2AV</td>
<td>24</td>
<td>4</td>
<td>17%</td>
<td>dom</td>
<td>0.376</td>
<td>PPI</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Pc</td>
<td>-</td>
<td>GI</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Parp</td>
<td>-</td>
<td>PPI</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Top2</td>
<td>-</td>
<td>PPI</td>
</tr>
<tr>
<td>Average confirmed interactions</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>32%</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
CURRICULUM VITAE

Emma Kate Mullen

Education

<table>
<thead>
<tr>
<th>Year Range</th>
<th>Institution</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>2011 - Present</td>
<td>The University of Western Ontario</td>
<td>M.Sc. Candidate</td>
</tr>
<tr>
<td>2007 - 2011</td>
<td>The University of Western Ontario</td>
<td>B.Sc. Honours Specialization in Biology Graduated with distinction</td>
</tr>
</tbody>
</table>

Honours and Awards

<table>
<thead>
<tr>
<th>Year Range</th>
<th>Award Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>2007</td>
<td>Western Continuing Scholarship $10 000</td>
</tr>
<tr>
<td>2007</td>
<td>Allstate Foundation Scholarship $500</td>
</tr>
<tr>
<td>2009 - 2010</td>
<td>Edward Barrow and Ida Hodgins Battle Scholarship $2000</td>
</tr>
<tr>
<td>2010 - 2011</td>
<td>The Albert O. Jeffery Scholarship in Honours Zoology $500</td>
</tr>
<tr>
<td>2011</td>
<td>Helen I. Battle Medal and Scholarship in Zoology $500 and Gold Medal</td>
</tr>
<tr>
<td>2012 - 2013</td>
<td>Natural Sciences and Engineering Research Council of Canada Alexander Graham Bell Canada Graduate Scholarship $17 500</td>
</tr>
<tr>
<td>2013</td>
<td>Best Student Talk Ecology and Evolution Seminar Series</td>
</tr>
<tr>
<td>2013</td>
<td>Western's Best Tutor Students Offering Support London, Ontario Canada</td>
</tr>
</tbody>
</table>
Research Experience

2011 - Present
Graduate Student and Research Assistant
The University of Western Ontario
Biology Department, Graham Thompson PhD

2010 - 2011
Volunteer Research Assistant
The University of Western Ontario
Biology Department, Greg Thorn PhD

Teaching Experience

2011 - 2013
Graduate Teaching Assistant
2290F/G - Scientific Methods in Biology
The University of Western Ontario

2012 - 2013
Biology 1001A/1002B Tutor
Students Offering Support
London, Ontario Canada

2013
Guest Lecturer
3598B - Behavioural Genetics
The University of Western Ontario

2012 - 2013
Guest Lecturer
Grade 8 Biology, Holy Family Catholic School
London, Ontario Canada

Conferences and Presentations


February 2012. The evolution of altruism. BiPED: The Biology-Philosophy Evolution Discussion group

June 2012. A gene network for worker sterility in the honey bee. Arthropod Genomics Symposium, Kansas City MO

July 2012. A gene network for worker sterility in the honey bee. First Joint Congress on Evolutionary Biology, Ottawa ON