Winter 2020

Behavioral and Genetic Mechanisms of Social Evolution: Insights from Incipiently and Facultatively Social Bees

Wyatt Abram Shell
University of New Hampshire, Durham

BEHAVIORAL AND GENETIC MECHANISMS OF SOCIAL EVOLUTION: INSIGHTS FROM INCIPIENTLY AND FACULTATIVELY SOCIAL BEES

BY

WYATT A. SHELL

B. A. Biology, Bard College, 2010

DISSERTATION

Submitted to the University of New Hampshire

in Partial Fulfillment of

the Requirements for the Degree of

Doctor of Philosophy

In

Biological Sciences: Integrative and Organismal Biology

December, 2020

This dissertation was examined and approved in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Integrative and Organismal Biology by:

Dissertation Director, Dr. Sandra Rehan
Associate Professor of Biology, York University

Dr. Jessica Bolker
Professor of Biological Sciences

Dr. James Haney
Professor of Biological Sciences

Dr. Arturo Andrade
Assistant Professor of Biological Sciences

Dr. Adrienne Kovach
Associate Professor of Natural Resources and the Environment

On December 8, 2020

Approval signatures are on file with the University of New Hampshire Graduate School.

ACKNOWLEDGMENTS

I have a ton of people to thank. First and foremost, though, I want to thank my dissertation adviser and research mentor, Sandra Rehan, who for years now has tirelessly challenged and enthusiastically supported me as an academic, a research colleague, and friend. Sandra has provided me with endless opportunities for exploration and growth as a young scientist, and has combined in her advising role a demanding, wholistic, and ceaselessly engaging approach to research with a great sense of humor, care, and patience. Simply put, the years I've worked with Sandra have been the most challenging and rewarding of my career; I could go on, but suffice to say I am very grateful and fortunate!

I want to thank the members of my dissertation committee, Jessica Bolker, Arturo Andrade, Adrianne Kovach, and Jim Haney for providing so much supportive counsel and constructive feedback on my research as this dissertation has evolved. I also want to thank UNH faculty outside of my committee, Leslie Curren, Chris Neefus, Patty Bedker, Don Chandler, Anita Klein, Larry Harris, Win Watson, and István Mikó for their considerate support and affable conversation during the past five years.

I also want to thank the wonderful folks in the UNH Biology offices – Diane Lavalliere, Maria Shortt, and Charlotte Cooper – for their warmth and tireless logistical and moral support for myself and the rest of the department; thanks as well to those among UNH's Graduate School admin, Dovev Levine-Leung, Cari Moorhead, and Laurie Witham, and Michael Thompson from Research Development for all of your help in getting me to where I am now. Along similar lines, thanks very much to Toni Westbrook for helping me out of digital snares when using UNH's server cluster.

Naturally, I also want to thank a large handful of other Bee Lab alumni/ae who have variously supported me through the sharing of research trials, manuscript reviews, good spirits, and – in some cases – outright lasting friendship. Jake, Michael, Robert, Margot, Nick, Minna, Katherine, Erika, Sabine, Celia, Sarah, Sean, Dusty, Liz, Salena, Cullen, Hannah, Stephanie, Kiley, and Amy – thank you all! I also want to thank other graduate students from UNH's Biology Department and other institutions who brightened my grad student experience through their good natures and science savvy: Nate, Annie, Seth, Rory, Eric, Morgan, Jenny, Kate, Haley, Jenna, Ben, Alina, Jessica, Jon, Meghan, Jen, Andrew, Taruna, Mia, Maranda, and Brooke – thank you guys, too!

The community that has supported my dissertation work in the Bee Lab extends to the international level – thank you to Fred and colleagues at Genome Quebec for so amiably helping with RNA library prep and sequencing; Charles Blatti for much-needed assistance in operating the CisMetalysis pipeline; Michael Schwarz, Scott Groom, and Rebecca Dew for your hospitality, scientific insights, and good humor both in Australia and the States; and the Biology Department at York University for helping to host me during BeeCon 2019.

Also I want to thank the academic and professional mentors I've had along the way to my dissertation work: Dr. Philip Johns, who got my feet under me and gave me so much to explore as an undergrad researcher at Bard College; and Dr. Gerald Wilkinson, who gave me the chance to expand my abilities and responsibilities as a research assistant at the University of Maryland. Both of these guys have supported me ever since, and I am so grateful for the strong foundations working with each of them set for me moving forward.

I have had the great fortune and joy of being supported by many friends and academic colleagues during these years, but I want to especially thank those people in my personal life

who have bolstered me with near-daily support: my mom and dad, my girlfriend Hannah, and my dear friends Benjamin and Jake. Collectively, you have helped re-center my thinking and bolster my resilience when things have been particularly tough and have allowed me to more fully acknowledge accomplishments as I've steadily travelled further down this road. For this and so much more, I thank you all very much.

Finally, I want to thank the sources of funding that have helped make this work possible: the University of New Hampshire; the National Science Foundation – Integrative Organismal Systems: Behavioral Systems (1456296); and National Science Foundation Graduate Research Fellowship (1450271).

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Bees are highly charismatic and ecologically valuable organisms, and most popularly represented by the honey bee. Honey bees are well known in no small part because they are eusocial: a single reproductive queen continually lays eggs, and is supported by overlapping generations of thousands of non-reproductive workers, all performing specialized tasks to feed and protect the colony. Despite the ecological edge it seems to confer, eusociality has emerged in relatively few bee lineages; most have instead either remained solitary (the ancestral state for all bees) or demonstrate any of a range of less derived forms of social organization. Researchers have for decades been steadily teasing out the ecological, developmental, and evolutionary factors that may drive the emergence and elaboration of insect social complexity. This dissertation aims to join that effort by offering a handful of additional insights emerging from empirical testing of major social evolutionary hypotheses in bees of facultative and early sociality.

My introductory Chapter 1 elaborates on the question of eusociality in greater detail and lays out the major social evolutionary hypotheses and their syntheses. I argue in support of research among bees of early or facultative sociality as systems in which much-needed empirical testing of evolutionary theory may be performed. In Chapter 2, I use relatedness and demographic data to calculate the inclusive fitness costs and benefits of social nesting in the small carpenter bee (*Ceratina calcarata*) which may rear a single worker-like daughter to aid in brood care. I find that social nesting may be advantageous to social nest mothers rather than daughters in this species, contrary to the expectations of kin selection theory. In Chapter 3, I further investigate sociality in *C. calcarata* using brain transcriptomic data that captures patterns of *cis*-regulation and gene expression associated with female maturation and two well-defined

behavioral states, foraging and guarding, concurrently demonstrated by mothers and daughters in social nests. I find that the early social nest environment may have a strong effect on gene expression; and reveal foraging and guarding behaviors to be underpinned by deeply conserved genes that are differentially expressed within a highly modular gene network. In Chapter 4, I draw on another set of brain transcriptomic data, this time reflecting first and second year solitary females, queens, and workers of the long-lived and facultatively eusocial small carpenter bee, *Ceratina japonica*. I find that queen and worker phenotypes are underpinned by highly divergent gene regulatory pathways. I also show how genes underlying *C. japonica*'s queens and workers are well-conserved and demonstrate strikingly similar patterns of expression in other bees of early eusociality. I also discover that while the social nest environment may induce some shared shifts in lifetime gene expression among queens and workers vs solitary females, the role of oxidative damage reduction may be a proximate mechanism of prolonged longevity regardless of social phenotype.

Appendix A details my development of polymorphic microsatellite markers using the *C. calcarata* genome, which were used in Chapter 2 of this dissertation and are now a publicly available tool for further research; the remaining Appendices provide mainly supplementary methods and figures for Chapters 3 and 4.

CHAPTER 1

INTRODUCTION

BEHAVIORAL AND GENETIC MECHANISMS OF SOCIAL EVOLUTION: INSIGHTS FROM INCIPIENTLY AND FACULTATIVELY SOCIAL BEES

[Shell WA, Rehan SM. 2018 Behavioral and genetic mechanisms of social evolution: insights from incipiently and facultativly social bees. *Apidologie.* **49**, 13-30 (doi: https://doi.org/10.1007/s13592-017-0527-1)]

Phenotypic plasticity, an organism's ability to modify physiological or behavioral traits following environmental cues, has been comprehensively studied over the past four decades (West-Eberhard 1989; Nijhout 2003). In some cases, variation in the regulation of a conserved set of genes can lead to the expression of multiple, highly discrete phenotypes (Whitman and Agrawal 2009). Such polyphenisms are relatively common in the insects (Emlen and Nijhout 2000; Simpson et al. 2011) and are informative targets for addressing questions regarding the origins and elaboration of derived phenomena (e.g. caste determination, Evans and Wheeler 2001; Fjerdingstad and Crozier 2006) including the paradox of eusociality. Eusociality is considered the most complex form of social organization and eusocial species are often ecologically dominant (Michener 1969; Wilson 1971). However, despite its biological prevalence and ecological success (ants alone represent as much as 25 percent of all terrestrial biomass, Schultz 2000), the emergence of eusociality remains a rare and highly derived event (Wilson and Hölldobler 2005; Nowak et al. 2010). Determining the molecular and environmental factors underlying the origins of eusocial behavior has consequently become a central focus in the field of social evolution (Michener 1974), and will be the central focus of this dissertation.

Eusociality has emerged independently only a handful of times in nature, but has done so within bees more than any other lineage (likely four independent origins, **Figure 1.1**; Cardinal and Danforth 2011; Gibbs et al. 2012; Rehan et al. 2012). The repeated evolution of eusociality within so widely diversified a monophylum has made eusocial bees an exemplary group for comparative research (Fischman et al. 2011). Accordingly, studies in socially complex bees have generated well-founded theories regarding the origins of eusociality; and have contributed to the emergence of the field of sociogenomics, which integrates the disciplines of molecular biology and behavioral genetics with insights from phylogenetics and behavioral ecology to investigate the molecular architecture of social behavior and organization (Robinson 2002; Robinson et al. 2005).

Advanced eusocial bees are considered to have transitioned past an evolutionary "point of no return" (Wilson 1971; Wilson and Hölldobler 2005), and thus provide mainly inferential insights regarding the origins of their complex sociality (Schwarz et al. 2007; Boomsma 2009). By contrast, it is thought that the vast majority of the more than 20,000 bee species worldwide (as many as 94 percent, Kocher and Paxton 2014) are either solitary or lack clearly defined castes (Michener 2007). Considered alongside their eusocial relatives, bees can thus be seen to comprise a natural gradient of social complexity (Kocher and Paxton 2014), formally termed the "social spectrum" (Rehan and Toth 2015; **Table 1.1**). There is strong evidence that non-eusocial species are capable of moving along this social spectrum in either direction, "toward" and "away from" social complexity (West-Eberhard 2003; Field et al. 2010; Rehan and Toth 2015). Advanced eusociality thus does not represent a necessary evolutionary eventuality for social species (West-Eberhard 2003), but rather indicates the far extreme of fundamentally flexible social organization. Studies which look across this social spectrum have consequently spurred a

concerted push to better unify theory (Rehan and Toth 2015; Toth and Rehan 2017) and to resolve semantic confusion within the field of social evolution (Dew et al. 2016). To fully appreciate the value of comparative research across the social spectrum, it is necessary to first briefly conceptualize this gradient, and to articulate the defining qualities of its major classes.

Table 1.1. Definitions of social terminology – adapted from Rehan and Toth 2015

Stage	Social class	Cooperative brood care	Overlapping generations	Division of labor	Facultative sociality	Species examples – with genomes
Early	Solitary	No	No	No	No	*Megachile rotundata*, *Dufourea novaeangliae*, *Habropoda laboriosa* (Kapheim et al. 2015)
	Subsocial	No	Some	No	Yes	*Ceratina calcarata* (Rehan et al. 2016)
	Incipiently social	Yes	Some	Some	Yes	*Ceratina australensis* (Rehan et al. 2018)
Late	Primitively eusocial	Yes	Some	Yes	Yes	*Lasioglossum albipes* (Kocher et al. 2013); *Bombus terrestris*; *Bombus impatiens* (Sadd et al. 2015)
	Advanced eusocial	Yes	Yes	Yes	No	*Apis mellifera*, *Apis florea*, *Melipona quadrifasciata* (Kapheim et al. 2015)

In solitary bees, reproductive females independently establish a nest and singly forage for their offspring, which require no care during their maturation (Michener 2007). Once provisioning is complete, females often either abandon their brood (e.g. *Colletes* and *Osmia*; Černá et al. 2013) or die before their offspring reach adulthood (Michener 2007). Though solitary species are not burdened by the increased risk of disease transmission associated with social living (Fu et al. 2015), they are extremely susceptible to parasitism and predation (Wcislo and Cane 1996). Increased parasitism and predation rates are thus considered primary ecological

drivers towards social organization (Lin and Michener 1972; Wcislo et al. 2004; Rehan, Schwarz and Richards 2011).

In a handful of seemingly solitary bees, the reproductive female will remain at her nest to guard and clean the developing brood (e.g. *Ceratina japonica*, Sakagami and Maeta 1984, *Lasioglossum laticeps*, Plateaux-Quénu 2008; and *Ceratina calcarata*, Rehan and Richards 2010); and this extended parental care defines them as subsocial behavior (Michener 2007). Subsociality is the simplest form of social organization along the social spectrum (Rehan and Toth 2015) and is considered a necessary precondition for the evolution of more complex social groups (Wilson 1971; Michener 1974). While subsocial bees are nest-loyal, and at least some are capable of nestmate recognition (Rehan and Richards 2013), they remain effectively casteless and independent as adults.

Like solitary and subsocial species, incipiently social bees remain totipotent throughout adulthood and are able to found nests of their own (Sakagami and Maeta 1989; Rehan, Richards, and Schwarz 2010). They are distinguished from their subsocial cousins by their capacity to nest cooperatively under a simple division of labor to rear a single brood (West-Eberhard 1987). As they represent species at the origins of social organization, subsocial through incipiently social bees may be collectively referred to as early stage social species (**Table 1.1**; Rehan and Toth 2015).

On the other end of the social spectrum are bees representative of late stage sociality: the primitively eusocial species. In these groups, a reproductive female continues to lay eggs after initially founding her nest (either alone or with nestmate workers), and may do little else once the first generation of her offspring reach maturity (Wilson 1971; Michener 2007). Individuals of primitively eusocial species are typically monomorphic, though a division of labor based on age

and body size is common (e.g. *Halictus rubicundus*, Soucy and Danforth 2002; *Megalopta genalis*, Smith et al. 2003; *Halictus scabiosae*, Brand and Chapuisat 2012). Worker-type offspring in primitively eusocial colonies may maintain a capacity for reproductive behavior, though ovary development is often inhibited (Michener 1974). These offspring are rarely limited to a single role, and may switch tasks over their lifetime. For instance, in both sweat bees and small carpenter bees, workers may rapidly develop into reproductives upon removal of a dominant reproductive nestmate (Eickwort 1986; Rehan et al. 2014). The workers of advanced eusocial species, by contrast, develop into functionally sterile castes, often with distinct behaviors and morphologies (e.g. *Apis* and *Melipona* bees, Wilson 1971; Michener 1974, 2007). This distinct caste differentiation is considered a defining feature of advanced eusocial species, which are no longer capable of reverting to simpler forms of social organization (Wilson and Hölldobler 2005).

Intriguingly, early stage social bees may demonstrate 'facultative sociality,' or a capacity to express more than one form of social organization (**Figure 1.1**; Smith et al. 2003; Rehan et al. 2011, 2014). The crepuscular sweat bee (*Megalopta genalis*), for instance, nests either solitarily or in small, primitively eusocial colonies (Smith et al. 2003; Wcislo et al. 2004; Kapheim et al. 2013; Jones et al. 2017). The subsocial small carpenter bee (*Ceratina calcarata*), by comparison, is facultatively incipiently social and may rear a worker-like offspring to aid in brood care and feeding (Rehan and Richards 2010a; 2010b; Rehan et al. 2014). The behaviors characteristic of early stage social bees are theorized to be necessarily antecedent to advanced eusociality, and thus provide a tractable opportunity to empirically test social evolutionary theory (e.g. Wilson 1971; Michener 1974; Seger 1983; Rehan and Toth 2015). Those which are also facultatively social, however, provide particularly informative insights, as the costs, benefits and molecular

mechanisms underlying multiple classes of social organization may be explored concurrently within a single species (West-Eberhard 2003; Rehan and Toth 2015). To make evolutionarily informed conclusions about the molecular origins of complex social behavior across the social spectrum, well-resolved phylogenetic histories and species delineations must first be established (Danforth et al. 2013). Fortunately, considerable efforts have been made to resolve the phylogenies of two socially diverse bee groups: the sweat bees (Halictidae) and carpenter bees (Xylocopinae).

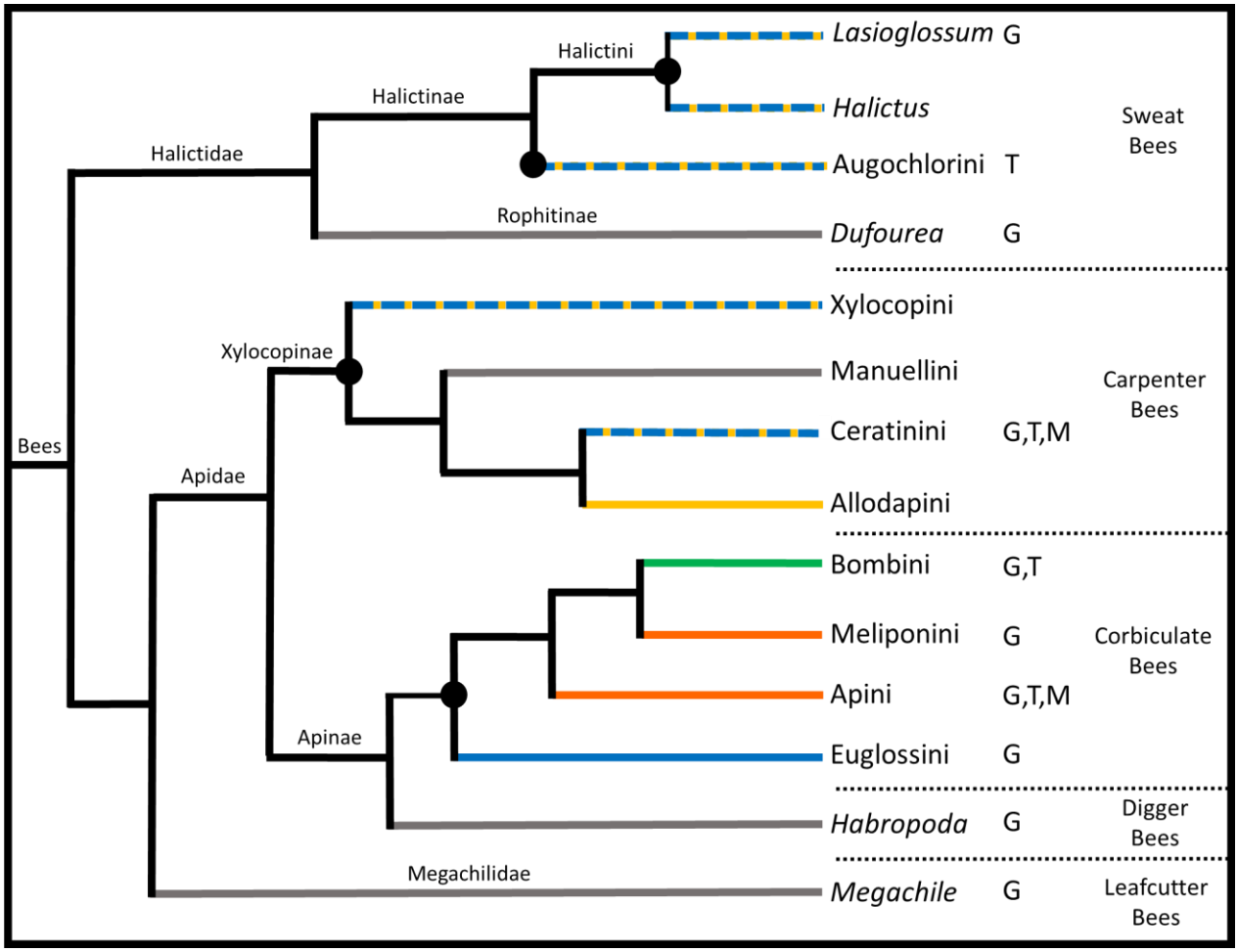

Figure 1.1. Phylogeny of major bee groups and social lineages. Black nodes indicate independent origins of sociality. Branch colors indicate lineages containing species which are: *gray*, solitary; *blue*, solitary through primitively eusocial; *green*, primitive eusocial; *yellow*, primitive through advanced eusocial; *orange*, advanced eusocial. Branches *hashed with yellow* indicate lineages containing facultatively social species. Letters indicate published resources for that lineage (G: genome; T: transcriptome; M: methylome).

Represented by solitary, facultatively social and eusocial species, the wide behavioral diversity of the halictids has made them an informative system for studies of behavioral plasticity and social evolution (Schwarz et al. 2007; Kocher and Paxton 2014). Multiple revisions of halictid phylogeny (Danforth 2002; Brady et al. 2006; Gibbs et al. 2012) have come to posit two origins of eusociality within Halictidae: one in the tribe Augochlorini (Danforth and Eickwort 1997; Danforth et al. 2013) and another for the tribe Halictini (**Figure 1.1**; *Halictus* and *Lasioglossum*, Gibbs et al. 2012). This research has also provided evidence of multiple reversions from social to solitary behaviors in this group (Danforth et al. 2003).

The carpenter bees (subfamily Xylocopinae) collectively represent the entirety of the social spectrum across their four tribes: the solitary Manueliini, incipiently social Xylocopini, solitary to eusocial Ceratinini, and incipiently social to eusocial Allodapini (**Figure 1.1**). This rich diversity, combined with a well-resolved phylogeny (Rehan et al. 2010, 2012; Rehan and Schwarz 2015) and phylogeography (Dew et al. 2016; Shell and Rehan 2016), make the xylocopine bees an exceptional system for sociogenomic research (Rehan and Toth 2015). Further, as the Xylocopinae and Apinae are sister subfamilies (Cardinal and Danforth 2011), research involving carpenter bees may directly inform our understanding of the dynamics underlying advanced eusociality in corbiculate species (**Figure 1.1**; Rehan and Toth 2015; Toth and Rehan 2017). Phylogenetic work within the Xylocopinae has revealed a single origin of sociality followed by as many as four reversions to a solitary lifestyle (Schwarz et al. 2007; Rehan et al. 2012).

Well resolved phylogenies among social lineages provide much-needed evolutionary frameworks by which to contextualize comparative molecular research (Rehan and Toth 2015; Romiguier et al. 2015; Branstetter et al. 2017; Peters et al. 2017). Such research has helped to

reveal that species may evolve along a non-linear trajectory between simple and complex social organization: incipiently social lineages, for instance, frequently revert to solitary life (Wcislo and Danforth 1997; Rehan et al. 2012; Gibbs et al. 2012). Phylogenetic studies also indicate that the evolution of advanced eusociality is rare, and likely checked by prohibitive natural barriers (Rehan et al. 2012). Integrative research which builds off this phylogenetic foundation may now benefit from ongoing advances in genomic and transcriptomic sequencing methods. As such, studies which strive to integrate genomic, transcriptomic, and behavioral data are poised to gain meaningful insights into the molecular signatures of phenotypic plasticity.

In this introductory chapter, we provide a detailed overview of sociogenomic research in bees of early stage and facultative sociality. The rich behavioral and ecological diversity of these species provides a powerful workspace in which to explore the molecular, developmental, and environmental drivers of social transition (West-Eberhard 1987; Rehan et al. 2014; Patalano et al. 2015). Despite their informative value, however, there remains a paucity of research in these groups. We thus argue in support of additional sociogenomic research into bees of early stage facultative sociality, and highlight the socially diversified *Ceratina* small carpenter bees and halictid sweat bees as promising model systems for future studies.

FACULTATIVELY SOCIAL BEES

Facultative sociality is not a distinct social category (like subsociality or eusociality), but rather indicates a species' capacity for plasticity in its social organization. For example, conspecifics of some facultatively social sweat bees either establish a solitary nest, or found a primitively eusocial colony in which a helper generation assists in the rearing of a second brood of reproductive individuals (e.g. *Halictus rubicundus*, Soucy and Danforth 2002; *Megalopta*

genalis, Smith et al. 2003; *Lasioglossum calceatum*, Davison and Field 2016). Facultatively social nesting in some carpenter bees, by contrast, may involve incipiently social brood care by a cooperative pair of age-matched females (e.g. *Ceratina australensis*, Rehan et al. 2014), or production of a single, worker-like offspring alongside future reproductives (e.g. *C. calcarata*, Rehan et al. 2014; C. *japonica*, Sakagami and Maeta 1984). Where advanced eusocial Hymenoptera feature distinctive and often multifaceted reproductive and morphological caste systems, the behavioral phenotypes of facultatively social bees straddle species-specific ranges along the social spectrum. This behavioral diversity, both within and between facultatively social species, provides an exceptional opportunity to explore how phenotypic variation across modes of social organization is reflected at the molecular level (Kocher and Paxton 2014).

Facultatively social bees are particularly tractable models for comparative sociogenomic study because variation in their social phenotype may be reliably elicited through experimental manipulation (Schwarz et al. 2007; Kocher and Paxton 2014; Rehan and Toth 2015). This can be seen, for example, in the induction of cooperative, multi-female nests in otherwise mainly subsocial species (e.g. *Ceratina japonica*; Sakagami and Maeta 1984, 1987; *Lasioglossum* spp., Jeanson et al. 2005; 2008). Sakagami and Maeta (1995) elicited reproductive division of labor in the small carpenter bees, *C. japonica* (Sakagami and Maeta 1984) and *C. okinawana* (Sakagami and Maeta 1989), by constraining conspecific females to cohabitate. In these induced social nests, one female assumed queen-like behavior, laying eggs and defending the nest, while the second primarily foraged. In all cases of social nesting, the reproductive role was reliably adopted by the larger of the two females, in age-matched and even mother-daughter cooperative pairs (Sakagami and Maeta 1984, 1989, 1995). In similar fashion, Holbrook et al. (2009) observed that individuals of the normally solitary sweat bee (*Lasioglossum NDA-1*) assumed

dedicated guarding or digging behaviors when forced to nest in pairs. Though paired individuals secured deeper and better-guarded nests compared to their solitary counterparts. Holbrook et al. (2009) noted that such arrangements rarely occur in wild populations. Considerable phenotypic plasticity may thus be widespread in bees of early stage sociality, even in those species which rarely demonstrate facultative sociality in the wild.

Environmental factors, including local climate, resource availability, predation and parasite pressure, have long been theorized to comprise the primary ecological drivers of social evolution (Lin and Michener 1972; Evans 1977; Strassman and Queller 1989; Kocher et al. 2014). Accordingly, social complexity in facultatively social bees is often reliably predicted by ecological condition (Richards and Packer 1996; reviewed in Purcell 2011). For example, brood parasitism in the Australian small carpenter bee (*Ceratina australensis*, Rehan, Richards, and Schwarz 2010) appears to be a key selective pressure favoring social nesting (Rehan, Schwarz, and Richards 2011). The influence of environment on social organization has also been particularly well-documented in the sweat bees (Plateaux-Quénu et al. 2000; Schwarz et al. 2007; Davison and Field 2016). In *Halictus rubicundus* (Soucy and Danforth 2002), solitary or social reproduction appear to be largely dependent on geographic location: solitary *H. rubicundus* females produce a single brood in cooler regions (Eickwort et al. 1996) and social females produce two broods in warmer regions (Yanega 1989, 1993; Soucy 2002). Field et al. (2010) demonstrated this organizational lability by inducing the offspring of solitary or social populations to switch behavioral type via translocation between cool and warm locations. In other species of sweat bees, social phenotype has been reliably predicted specifically by altitude (*Lasioglossum calceatum*, Sakagami and Munakata 1972) or even micro-habitat composition. For instance, in the sweat bee, *L. balecium*, social nests were consistently observed in warmer,

sunnier sites, while solitary nests were typically found in adjacent cooler, well-shaded locations (Hirata and Higashi 2008). In some cases, however, variation in social organization is neutrally or otherwise not predictably affected by environment (e.g. *Lasioglossum apristum*, Miyanaga et al. 1999; *L. malachurum*, Richards 2000; Wyman and Richards 2003), supporting the role of genetic architecture as primary in determining social phenotype (Soucy and Danforth 2002; Soro et al. 2010).

BEHAVIORAL AND GENETIC MECHANISMS OF SOCIAL EVOLUTION

A suite of major theoretical frameworks has been put forward to explain the mechanisms by which complex social behavior may have evolved. Here, we provide an overview of some of these hypotheses which, though distinct in their viewpoints, are not mutually exclusive, and have been recently brought together in theoretical syntheses (Rehan and Toth 2015; Toth and Rehan 2017). A tendency to reuse a nesting substrate among kin groups is a well-supported precondition for the emergence of social behavior (Wild and Koykka 2014). As it may increase the opportunity for inclusive fitness benefits, reuse or inheritance of a 'family' nest is expected to promote group nesting, even with some nestmate competition. Accordingly, nest reuse has been observed across facultatively social *Ceratina* (Sakagami and Maeta 1987; Rehan et al. 2009, 2014, 2015) and in other basically social bees (e.g. *H. rubicundus*, Yanega 1990). By contrast, those species which are known to frequently disperse prior to the establishment of a new nest (e.g. *Ceratina calcarata* and *C. flavipes*) are less likely to nest socially (Sakagami and Maeta 1987; Rehan and Richards 2010). Along with nest reuse, multi-voltinism, the capacity to produce more than one brood during a reproductive season, has been suggested as antecedent to eusocial behavior (Seger 1983). The ability to produce multiple generations of offspring in a single

breeding season establishes a scenario in which the offspring of mated, overwintering females have an opportunity to remain at the nest as helpers in the following reproductive season. This proposition has found demonstration in primitively eusocial bees (e.g. *Halictus sexcinctus*, Richards 2001) and is supported by research in bivoltine *Ceratina* (e.g. *C. okinawana*, Sakagami and Maeta 1995). The prolonged maternal care behavior often seen in multivoltine species has also been suggested as prerequisite to more complex social structuring (West-Eberhard 1987) and is likely readily reinforced through selection (Wade 2001).

West-Eberhard (1996) also considered the evolutionary necessity of parental care behavior in her proposition of the *ovarian ground-plan hypothesis*, which offers that extended maternal care behaviors seen in solitary and subsocial Hymenoptera may have gradually differentiated into the dedicated forager and egg-layer roles seen in incipiently social and primitively eusocial Hymenoptera (**Figure 1.2**). The *reproductive ground-plan hypothesis* builds on this prediction by suggesting that such specialized roles could gradually lead to the emergence of task-specific worker castes, as seen in advanced eusocial species (Amdam et al. 2004, 2006). In many incipiently social bees, however, reproductive and foraging behavior are expressed in tandem; as such, a decoupling of foraging and reproductive behavior is not expected to have initiated advancements in social complexity in most lineages (Schwarz et al. 2011; Rehan et al. 2014).

It might be expected that the initial co-option of maternal care genes into relatively queen-like and worker-like phenotypes would be hindered, or outright prevented, by strict pleiotropic constraints on those genes (Gadagkar 1997). In consideration of this, Gadagkar (1997) proposed the concept of *genetic release*, in which a gene duplication event or significant variation in the timing of gene expression could free key genes from the effects of stabilizing

selection. Variation in gene expression, in combination with the evolution of separate developmental programs, could then allow for the mutual decoupling of strictly reproductive activities (e.g. mating, egg laying) from brood care and nest maintenance tasks. Directional selection through inclusive fitness could then reinforce the expression of increasingly queen-like and worker-like behaviors and physiologies. The *maternal heterochrony hypothesis* builds on this concept by proposing that a shift in the timing of expression (heterochrony) of maternal care genes may account for the evolution of precocious, sib-social care behavior seen in advanced eusocial species (Linksvayer and Wade 2005). From this approach, the emergence of queen and worker-type behaviors could result from a slight variation in the regulation of an otherwise conserved set of genes.

Rehan et al. (2014) directly assessed the maternal heterochrony hypothesis through transcriptomic analysis of the subsocial small carpenter bee, *Ceratina calcarata*. In *C. calcarata*, mothers may rear a worker-like daughter to aid in brood care and foraging, and thus enter an incipiently social stage (Rehan and Richards 2010b). Rehan et al. (2014) compared brain gene expression of individuals from five focal stages in *C. calcarata*'s reproductive season and, as predicted by the maternal heterochrony hypothesis, gene expression patterns of worker-like daughters most closely mirrored those of post-reproductive mothers, and included upregulation of genes implicated in maternal care. Further support for the underlying role of molecular heterochrony in emergent brood care behavior has also been found through similar transcriptomic works in other Hymenoptera (e.g. *Polistes metricus*, Toth et al. 2007, 2010).

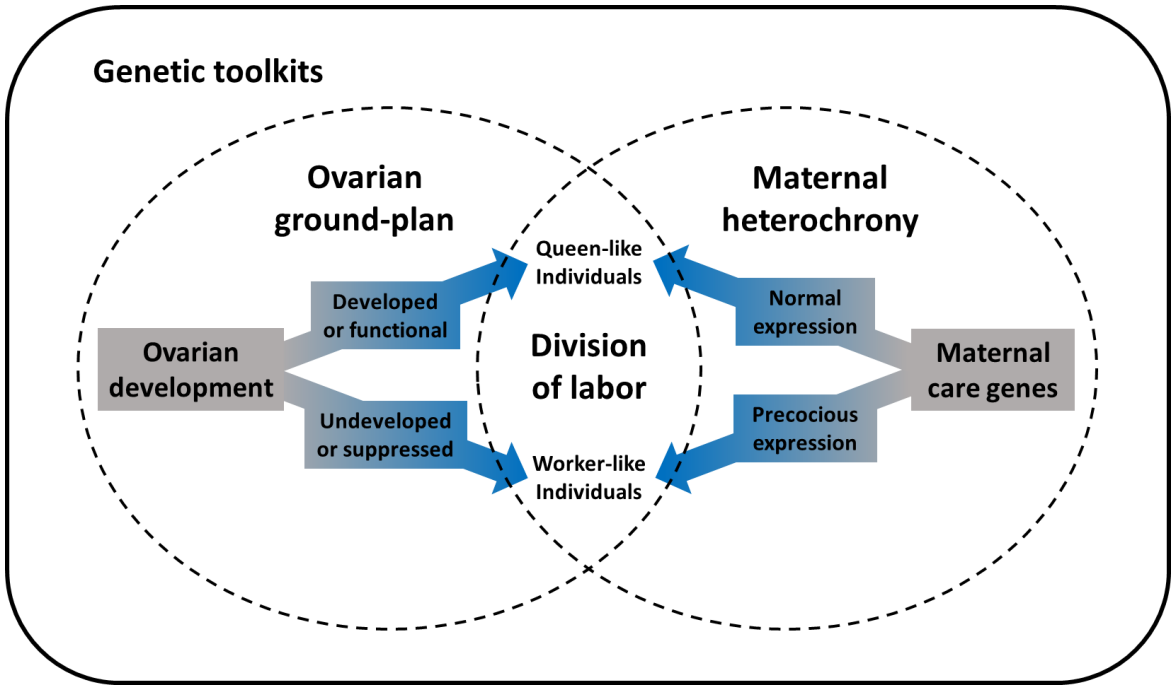

Figure 1.2. Nestedness and interconnectedness of sociogenomic theory. Deeply conserved genes (genetic toolkits) underlie core developmental and behavioral traits. These toolkit genes may become available for social roles following release from pleotropic constraints through relaxed selection. Reproductive polyphenism and division of labor may evolve by differences in hormone titers and gene expression levels associated with ovarian development (ovarian ground-plan) and/or temporal variation in the expression of genes associated with maternal care behavior (maternal heterochrony).

Complementary to both heterochrony and ground-plan hypotheses is the idea of the role of genetic 'toolkits' - highly conserved genes which play a focal role in the developmental pathways of core physiological and behavioral traits across taxa (Toth and Robinson 2007). The *genetic toolkit hypothesis* suggests that modifications to these key genes, or the molecular pathways regulating their expression (i.e. transcription factors), could underlie the emergence of behaviors leading to social organization (**Figure 1.2**; True and Carroll 2002; Carroll 2005). Expanded by Bloch and Grozinger (2011), the functionality of socially co-opted toolkit genes or pathways could become sensitive to signals from the social environment, and thus become socially regulated. The genetic toolkit hypothesis has found support in widely conserved genetic

signaling across both taxonomic and social lineages within bees (Woodard et al. 2011) and among social Hymenoptera more broadly (Toth et al. 2014; Berens et al. 2014; Morandin et al. 2016). For instance, two species of primitively eusocial bumble bee, *Bombus terrestris* and *B. impatiens,* were found to share many genes and gene regulatory elements with both *P. metricus* and *A. mellifera,* including those involved in the regulation of social behaviors (Sadd et al. 2015). In a comparison of transcriptomes across hymenopteran lineages (i.e. bees, ants, and wasps), Berens et al. (2014) determined that, despite high variation in expressed genes, metabolic pathways and molecular functions were relatively well-conserved across distant social lineages. They thus suggested support for a "loose" genetic toolkit: even when relatively few genes are expressed in common, the expression of complex sociality appears to involve highly conserved transcription factor and molecular functional networks.

In contrast to the highly-conserved nature of genetic toolkits, *novel* or *taxonomically restricted genes* (i.e. those unique to clades or lineages) have also gained support as likely to play a key role in the evolution of complex social behavior (Johnson and Tsutsui 2011; Sumner 2014). With the evolution of eusociality, it may be expected that group behavioral and physiological coordination would come to rely on specialized genes and gene expression networks (Johnson and Linksvayer 2010). Evidence of caste-specific expression in genes unique to bees (e.g. *A. mellifera*, Harpur et al. 2014) and other social Hymenoptera (e.g. *Polistes* wasps, Ferreira et al. 2013; *Temnothorax* ants, Feldmeyer et al. 2014) suggests positive selection on novel genes may underlie derived worker phenotypes (Kapheim et al. 2015). Accordingly, detection of lineage-specific variations in gene families, such as a bias towards gustatory chemoreceptors in *Bombus* (Sadd et al. 2015) or expansion in odorant receptor genes in the clonal raider ant (*Ooceraea biroi*; McKenzie et al. 2016), have helped to reveal how changes at

the molecular level may dictate the evolutionary trajectory of social Hymenoptera. Despite many novel genes between eusocial lineages, a large body of research in toolkit genes indicates that complex gene expression networks and metabolic pathways remain highly conserved (Simola et al. 2013; Berens et al. 2014; Morandin et al. 2016). Therefore, in theory, i) eusociality may evolve from a highly diverse suite of antecedent genetic profiles; but ii) the composition of genes and functional regulatory networks underlying social phenotypes may be increasingly constrained as species approach advanced forms of social complexity. It remains unclear, however, whether the apparent acceleration in rates of positive selection detected among eusocial lineages may instead indicate relaxed selection operating on relatively reduced effective population sizes in these taxa (Harpur and Zayed 2013; Romiguier et al. 2014).

SOCIOGENOMICS: A COMPARATIVE APPROACH TO SOCIAL EVOLUTION

Broadly, sociogenomic research addresses the feedback circuit between genetics and environmental cues: i) how molecular functions and expression affect behavior and social organization, and ii) how social and environmental cues further modify signaling pathways, development, physiology, and behavior (Robinson 2002; Robinson et al. 2005). The publication of the *A. mellifera* genome represented a critical foundation for concerted sociogenomic research (The Honeybee Genome Sequencing Consortium 2006) and remains a valuable resource for comparative studies of behavioral plasticity and social evolution (Fischman et al. 2011; Dolezal and Toth 2014). As genes are generally conserved across genera, particularly within taxonomic clades, the expanding wealth of honey bee sociogenomic research can act as a guide for explorations into the molecular architecture of species considered antecedent to the eusocial form (Page and Amdam 2007; Fischman 2011; Woodard et al. 2011).

The determination of the molecular signals underlying the origins of caste determination in advanced eusocial Hymenoptera represents a central endeavor in the field of sociogenomics (Toth et al. 2007; Toth and Robinson 2009), and has been astutely explored (Barchuk et al. 2007; reviewed in Berens et al. 2014). As with other complex traits of interest, caste determination can be traced to suites of genes whose functions are tied to key developmental (e.g. vitellogenin), hormonal (e.g. juvenile hormone), and metabolic pathways (e.g. insulin pathway genes; reviewed in Page and Amdam 2007; Corona et al. 2016). As the expression of these critical pathways remains highly sensitive to external cues (such as pheromone signaling, Grozinger et al. 2003; and nutritional intake, Ament et al. 2008), they can be considered as highly responsive to factors within the social environment. It has thus been theorized that gene pathways in eusocial taxa, which are involved in the ontogenetic determination of caste fate during the earliest stages of development, may have arisen through the co-opting of genes previously involved in core physiological processes (i.e. genetic toolkit hypothesis, see also Toth et al. 2007; Toth and Robinson 2009). We can use the power of comparative sociogenomic research across the social spectrum to trace the function of candidate genes across species of varying degrees of sociality (reviewed in Robinson et al. 2005; Smith et al. 2008).

SOCIOGENOMICS ACROSS THE SOCIAL SPECTRUM

Continued advancements in the accuracy and accessibility of sequencing technologies has made the acquisition and analysis of genomic and transcriptomic data an increasingly achievable undertaking; and has greatly expanded the molecular resources available for social bee lineages (**Table 1.1**; **Figure 1.1**). Comparisons of differentially expressed genes (DEGs) within *A. mellifera*, for instance, have revealed clear distinctions in brain gene expression between scout,

non-scout (Liang et al. 2012), and recruit worker behavioral variants (Southey et al. 2016). As genomic and transcriptomic resources are developed for species of early stage and facultative sociality (e.g. *Lasioglossum albipes*; Kocher et al. 2013; *Ceratina calcarata*; Rehan et al. 2014, 2016) extended comparative studies within and between Hymenoptera of distinct social organizations and lineages will be made possible (reviewed in Berens et al. 2014; Kapheim 2016). In the genome of *C. calcarata*, for instance, key transcription factor and chemosensory gene groups were identified as significantly expanded (Rehan et al. 2016), directly in line with expectations from observed gene family expansions in *A. mellifera* (Kapheim et al. 2015).

Insights into the molecular signals of ontogenetic variation between social lineages have also been gained through DEG research in facultatively social bees. Analysis of transcriptomic data recently developed for *M. genalis* revealed that while both functional gene groups and associated expression levels varied greatly over the course of an offspring's development, there were consistent shifts in primary functional processes during each stage of maturation (Jones et al. 2015). As female *M. genalis* develop from egg to adult, expression patterns consistently shift from predominantly cellular development and differentiation, to establishment of neurological systems and specialized metabolic pathways at adulthood (Jones et al. 2015). These results contrast with similar DEG analyses of offspring caste determination in eusocial species, in which early nutritional signals lead to increasingly distinct developmental gene expression pathways towards adulthood as either a sterile worker or reproductive queen (as seen in *A. mellifera*, Barchuk et al. 2007;*Polistes* wasps, Hunt et al. 2010; and *Formica* ants, Morandin et al. 2015). More recently, however, research investigating differences in brain and abdominal gene expression among solitary and social *M. genalis* revealed stark differences between reproductive and non-reproductive roles. The abdominal gene expression patterns of workers that became

replacement queens mirrored those of reproductive foundresses, while those of non-reproductive worker daughters stood apart. Differences in brain gene expression patterns among castes were not nearly as strong, however, suggesting that changes in reproductive functionality may precede changes in behavior (Jones et al. 2017). The caste-biased differences in abdominal gene expression detected in *M. genalis* were found to overlap with those of obligately eusocial taxa, suggesting there may be considerable conservation of the regulatory mechanisms underlying eusocial organization across evolutionary lineages. Further research comparing the transcriptomic profiles of offspring developmental stages and adult reproductive strategies (solitary vs. social) in other bees of early stage sociality more broadly, is a necessary next step.

Genomic and transcriptomic resources have also enabled the exploration of evolutionary developmental (evo-devo) theories as to how highly complex behavioral traits may emerge from a genetic functional architecture, which can be at once widely conserved (Rittschof et al. 2014) and pleiotropically constrained (Carroll 2008; Chen et al. 2013). Transcriptional regulatory networks (TRNs) are the highly complex and interconnected molecular signaling pathways (between transcription factors (TFs) and their target genes) underlying both gene expression and environmentally responsive regulation (Luscombe et al. 2004). TRN analysis in bees was pioneered by Chandrasekaran et al. (2011), who combined the brain transcriptomic profiles from 853 honey bees, representative of 48 distinct behavioral phenotypes, to assemble a brain TRN for *A. mellifera*. Their work indicated that brain gene expression would reliably predict behavioral phenotype, and that disparate behavioral phenotypes frequently rely on both shared and targeted TFs. A more recent population genomic study of the *A. mellifera* brain TRN revealed that regulatory and coding sequences throughout the TRN are under purifying selective pressure (Molodtsova et al. 2014). Proteins on the periphery of this regulatory network were most likely

to be free of pleiotropic constraint and thus available for adaptive selection. Accordingly, this work provided additional support for the genetic toolkit theory within the evo-devo framework, highlighting the role of alterations to TFs regulating gene expression, rather than changes in the genes themselves, as likely critical in the emergence of complex social behaviors.

The importance of TF functionality in social evolution was further highlighted through comparative genomic studies at the phylogenetic scale. Kapheim et al. (2015) compared the genomic data of 10 bee species (representing three families and containing two independent origins of eusociality) to investigate conserved molecular dynamics across social lineages. In contrast to species of early stage sociality, those of greater social complexity featured an increased frequency of transcription factor binding sites and putatively methylated genes. Additional evidence of accelerated evolution among eusocial taxa was also seen in similar studies (Woodard et al. 2011; Simola et al. 2013; Harpur and Zayed 2013; Harpur et al. 2014; Romiguier et al. 2014), and suggests that a capacity to regulate gene expression may become increasingly essential in the later stages of social organization. However, large sets of distinct and often lineage-specific genes were also detected in these studies (Woodard et al. 2011; Simola et al. 2013; Kapheim et al. 2015). Thus, despite considerable convergence in social organization and gene network functionality among eusocial taxa, there appears to be great variability in the possible evolutionary routes to complex social organization.

The increased achievability of sociogenomic methods opens doors to the integrative study of previously under-researched taxa. In the past four years, the *A. mellifera* genome has been joined by the draft genomes of two advanced eusocial (*Apis florea* and *Melipona quadrifasciata*; Kapheim et al. 2015) and two primitively eusocial bees (*Bombus terrestris* and *Bombus impatiens*, Sadd et al. 2015); three solitary (*Megachile rotundata*, *Dufourea*

novaeangliae, and *Habropoda laboriosa*; Kapheim et al. 2015) and two facultatively social bees (*Lasioglossum albipes*, Kocher et al. 2013; *Ceratina calcarata*, Rehan et al. 2016); as well as one putatively facultatively social bee (*Eufriesea mexicana*, Kapheim et al. 2015; **Table 1.1**, **Figure 1.1**). As genomic and transcriptomic data continue to be developed in additional social and solitary bees, we will continue to gain more comprehensive insights into how novel genes, genetic release from pleiotropic constraints, and convergent gene regulation underlie the foundational transitions in social form (Rehan and Toth 2015).

EPIGENETICS AND PHENOTYPIC PLASTICITY

Variation in gene expression via methylation is a well-appreciated mechanism of phenotypic plasticity across many taxa (Goldberg et al. 2007; Glastad et al. 2011). DNA methylation data from social Hymenoptera (Yan et al. 2015) indicate that a capacity for epigenetic response to ecological or social environment may play a role in the emergence of some complex traits (Weiner and Toth 2012; Kapheim et al. 2015). Studies investigating the epigenetics of phenotypic plasticity in *A. mellifera* (Lyko et al. 2010; Foret et al. 2012; Li-Byarlay et al. 2013) and bees of early stage sociality (Kocher et al. 2013; Rehan et al. 2016) have detected conserved expansions in possible methylation sites (CpG densities), and have consequently proposed this expanded capacity for DNA methylation as putatively central in the evolution of social complexity (Glastad et al. 2011; Weiner and Toth 2012). Though CpG densities are indicators for the capacity of methylation (Weiner and Toth 2012; Bewick et al. 2017), only by examining an organism's total methyl-modified genome can one empirically explore changes in epigenetic architecture across behavioral states and taxa (e.g. bees, *A. mellifera*, Lyko et al. 2010; *C. calcarata*, Rehan et al. 2016; *Polistes* wasps, Patalano et al. 2015;

Standage et al. 2016; and the clonal raider ant, *Cerapachys biroi*, Libbrecht et al 2016; reviewed in Hunt et al. 2013). Nevertheless, studies exploring methylation activity in primitively eusocial *Polistes* wasps found little to no evidence of gene methylation in this group (*P. canadensis*, Patalano et al. 2015; *P. dominula*, Standage et al. 2016). Thus, though some form of epigenetic regulation of gene expression via DNA methylation appears widespread in many species, its exact function in the evolution of social organization remains irregular and unclear (Bewick et al. 2017). Despite a large body of literature regarding its implications in Hymenoptera, published methylome data in bees is currently limited to two species: *A. mellifera* (Lyko et al. 2010) and *C. calcarata* (Rehan et al. 2016). Additional resources for bees of early stage and facultative sociality could help to further illuminate how methylation affects phenotype, and whether epigenetic modifications influence cooperative behavior at the origins of social organization (Weiner and Toth 2012).

THE ECO-EVO-DEVO APPROACH

As genomic and transcriptomic data provide unprecedented insights into the molecular drivers of phenotypic plasticity, the role of environmental factors in affecting gene expression pathways is becoming increasingly well appreciated (Gilbert 2012; Weiner and Toth 2012; Schlichting and Wund 2014). Social species are affected by signals from both their ecological habitat (e.g. temperature or predator pressure) and their social environment (Bloch and Grozinger 2011; Toth and Rehan 2017). Social environmental signals include: pheromones (Le Conte and Hefetz 2008), maternal manipulation of provisions for developing larvae (Rehan and Richards 2010b; Lawson et al. 2016), or reproductive policing by nestmates (Olejarz et al. 2016), any of which may affect changes in an individual's development or behavior. By integrating

measurements in the social and ecological environment with ontogenetic, phylogenetic and genomic datasets, researchers now have the chance to capture an unprecedentedly holistic portrait of behavioral plasticity and social evolution (Toth and Rehan 2017). This revelation has formed the basis of the emerging "eco-evo-devo" approach, under which the flexible molecular architecture underlying an individual's developmental pathways is continuously influenced by environmental signal factors (reviewed in Abouheif et al. 2014; Gilbert et al. 2015; Toth and Rehan 2017).

Central to the eco-evo-devo framework is the concept of genetic assimilation, in which an initially plastic phenotypic response to local environmental pressure may gradually, through accommodating mutations and selective reinforcement, lead to canalization for discrete phenotypes (Waddington 1961). Evidence for genetic assimilation has been clearly demonstrated in the tobacco hornworm (*Manduca sexta*, Suzuki and Nijhout 2006), in which experimental alteration of environmental factors was shown to lead to novel phenotypic responses, which may then be readily reinforced at the genetic level through accommodating mutation. To date, comparative research in *Apis* (Woodard et al. 2011; Harpur et al. 2014) and among social bee lineages (Kapheim et al. 2015) has suggested that genetic assimilation may play a similar role in the evolution of insect sociality.

Recent sociogenomic research in *M. genalis*, however, offers the first empirical assessment of this idea in a facultatively eusocial species (Jones et al. 2017). Jones et al. (2017) found that genes with worker-biased expression in *M. genalis* overlapped with genes shown to be rapidly evolving in ten other bee species. Similarities in the patterns of evolutionary rate and caste-biased gene expression between *M. genalis* and advanced eusocial species (e.g. *Apis*) thus suggest that strong positive selection through genetic accommodation may have driven the

elaboration of social phenotype (i.e. a worker caste) in the ancestors of modern eusocial taxa (Jones et al. 2017). The developmentally constrained caste systems of advanced eusocial bees may thus be the result of sustained environmental pressures operating on the flexible "substrate for selection" seen in bees of early and facultative sociality (Kapheim et al. 2015; Toth and Rehan 2017).

Approaching from the eco-evo-devo framework, the work presented in this dissertation integrates comparative transcriptomic, molecular and behavioral ecological methods to study bees of early and facultative sociality. Focusing on genus *Ceratina*, this work aims to expand our empirical understanding of the diverse mechanisms underlying phenotypic plasticity in general, and the evolution of social complexity in particular.

CHAPTER 2

THE PRICE OF INSURANCE: COSTS AND BENEFITS OF WORKER PRODUCTION IN A FACULTATIVELY SOCIAL BEE

[Shell WA, Rehan SM. 2018 The price of insurance: costs and benefits of worker production in a facultatively social bee. *Behavioral Ecology.* **29**, 204-211
(doi: https://doi.org/10.1093/beheco/arx146)]

ABSTRACT

Kin selection theory is foundational in helping to explain the evolution of sociality; however, the degree to which indirect fitness benefits may underlie helping behavior in species of early stage sociality has received relatively little empirical attention. Facultatively social bees, which demonstrate multiple forms of social organization, provide prime systems in which to empirically test hypotheses regarding the evolutionary origins of sociality. The subsocial small carpenter bee, *Ceratina calcarata*, may establish a social nest by manipulating brood provisions to rear a worker daughter, which then assists in critical late-season alloparental care. In this chapter, we combine nest demographic and behavioral data with genetic relatedness estimates to calculate the relative inclusive fitness of both subsocial and social reproductive strategies in *C. calcarata*. Social mothers benefit from improved likelihood of brood survivorship and have higher fitness than subsocial mothers. Worker daughters have low indirect fitness on average, and will not produce their own offspring. Among-sibling relatedness is significantly higher in social nests than subsocial nests, though mothers of either reproductive strategy may mate multiply. Though this study corroborates the ultimate role of indirect fitness and assured fitness returns in the evolution of social traits, it also offers additional support for maternal manipulation as the proximate mechanism underlying evolutionary transitions in early stage insect societies.

INTRODUCTION

Eusociality is one of the most complex forms of social organization in nature (Wilson 1971). Although eusocial organisms are represented by a diverse suite of taxa (e.g. naked mole rats, Jarvis 1981; thrips, Crespi 1992; shrimp, Duffy 1996; termites, Thorne 1997), Hymenoptera collectively contain more eusocial species than any other group (Wilson 1971). Obligately eusocial bees (e.g. *Apis mellifera*) demonstrate complex reproductive division of labor. Each individual's role within the colony is irreversibly determined during development, and a reproductive queen's lifetime fitness depends on the collective effort of thousands of sterile workers (Wilson 1971; Wilson and Hölldobler 2005; Michener 2007). Despite their established ecological dominance, however, eusocial bees represent relatively few species. Most of the more than 20,000 bee species worldwide are solitary (Michener 2007) and the remainder demonstrate forms of non-eusocial organization (Rehan and Toth 2015). Solitary nesting is ancestral in bees; but evidence suggests that lineages may undergo continuous evolutionary gains or losses in their social complexity (Szathmáry and Smith 1995; Danforth 2002; Rehan and Toth 2015). The evolutionary origins of an obligate and sterile worker caste thus appear paradoxical: why would an individual sacrifice its direct fitness to assist in rearing another's offspring? Further, how might such a seemingly altruistic behavioral phenotype be selectively reinforced?

Inclusive fitness theory suggests that indirect fitness benefits to the altruist may be enough to account for the origin and elaboration of the advanced eusocial worker caste (Hamilton 1964; West-Eberhard 1975; Trivers and Hare 1976; Foster et al. 2006). As formalized by Hamilton (1964), if an altruist's helping behavior were to contribute to the direct fitness of close genetic relatives, its indirect fitness gains could outweigh the incurred costs of forgoing some or all of its own reproduction. Kin selection is thus considered a plausible explanation for

the evolution of sociality in Hymenoptera, in which female siblings are expected to share significantly more of their genetic identity with each other compared to their mother or brothers (Hamilton 1972; Lin and Michener 1972). Though the multiple mating of advanced eusocial species appears to confound these expectations by reducing among-sibling relatedness (Palmer and Oldroyd 2000), monandry is thought to be ancestral to Hymenoptera and suggests indirect fitness could have facilitated the emergence of early stage social traits (Hughes et al. 2008).

The biological applicability of kin selection has been the subject of heated debate within the field of social evolution (West-Eberhard 1975; Wilson 2005; Gadagkar 2010; Nowak et al. 2010; Marshall 2011). While the maintenance and elaboration of social traits by inclusive fitness remains a widely accepted theory (Foster et al. 2006; Hughes et al. 2008; Bourke 2011; Quiñones and Pen 2017), relevant examinations within early stage social systems remain limited (Leadbeater et al. 2011; Rehan et al. 2014b; Kapheim et al. 2015). Similarly, though many insights into the evolution of sociality have been gained through the study of advanced eusocial taxa, reproductive divisions of labor are often obligate in such species, and provide only inferential insights regarding their evolutionary origins (Queller and Strassmann 1998; Wilson and Hölldobler 2005; Toth et al. 2007). Facultatively social species, by contrast, demonstrate a capacity for multiple degrees of social organization, and exhibit at least two distinct nesting phenotypes (e.g. solitary and eusocial nesting). Accordingly, such species provide a unique opportunity to empirically investigate whether theoretically predicted mechanisms of evolution have a biologically realized influence on gains or losses in social complexity across the social spectrum (Rehan and Toth 2015; Shell and Rehan 2017; Toth and Rehan 2017; Quiñones and Pen 2017).

Divisions of labor among Hymenoptera are frequently based on differences in age or body size. For instance, larger and older daughters in *Polistes* paper wasp colonies are more likely to assert reproductive dominance than their smaller or younger siblings if given a viable opportunity to do so (Hughes and Strassmann 1988; reviewed in Jandt et al. 2014). Among facultatively eusocial bees, foundresses are the eldest in the family unit, and often able to manipulate the sex, size, and behavior of their offspring through selective fertilization, provision investment, and physical coercion, respectively (Sakagami and Maeta 1984; Yanega 1989; Aneson and Wcislo 2003; Smith et al. 2003; Rehan and Richards 2010b; Kapheim et al. 2011, 2015). In this way, mothers maximize their reproductive investment while minimizing potential conflict or competition from their offspring. Maternal manipulation of brood has consequently been proposed as a proximate mechanism of early divisions of labor across multiple lineages (Alexander 1974; Craig 1979; Ratnieks and Wenseleers 2008), and evidence shows that, in combination with high intracolonial relatedness, it likely reinforces early stage social organization (Crespi and Ragsdale 2000; Richards et al. 2005; Kapheim et al. 2015).

Cost-benefit analyses are an effective means of evaluating different life histories (Hamilton 1964; Trivers 1971) and have been applied in facultatively social bees to compare the relative fitness of alternative reproductive strategies (Augusto and Garófalo 2004; Pech et al. 2008; Rehan et al. 2014b; Kapheim et al. 2015). Where brood loss to predation or parasitism may represent a significant natural problem for solitary nests, social nesting may provide a means of increasing fitness through improved resistance to these pressures (Rehan et al. 2011, 2014b; Yagi and Hasegawa 2012). For example, social nests of the facultatively eusocial sweat bee, *Lasioglossum baleicum*, were better able to preserve developing larvae compared to solitary nests when under pressure from ant predation (Yagi and Hasegawa 2012). Consequently, higher

fitness for social reproductive females and their sterile first-brood workers compared to solitary conspecifics supported the role of inclusive fitness in maintaining social traits (Yagi and Hasegawa 2012). Inclusive fitness benefits also likely contribute to the maintenance of division of labor in facultatively social orchid bees (Pech et al. 2008). However, persistent agonistic interactions among siblings or between a mother and her brood indicate that social behavior may be prompted by physical aggression rather than sibling relatedness in this group (Augusto and Garófalo 2004; Pech et al. 2008). Taken together, studies suggest that while ecological pressures and genetic identity among family groups likely help to maintain and even reinforce social traits through inclusive fitness, i) social nesting may not be an advantageous strategy for all individuals involved; and ii) physical and/or aggressive interactions among nestmates may be required to elicit and maintain sib-social care behaviors (Ratnieks and Wenseleers 2008). Cost-benefit analyses of facultatively social species are few, yet critical to understanding the underlying mechanisms for social organization in early stage societies.

The small carpenter bee, *Ceratina calcarata* demonstrates a form of facultative incipient sociality across its range in eastern North America (Rehan and Sheffield 2011; Shell and Rehan 2016a) where it produces one brood per year (Johnson 1988; Rehan and Richards 2010b). All reproductively active female *C. calcarata* nest subsocially by providing extended parental care for their maturing brood. Some mothers, however, establish social nests by also producing a worker daughter to assist with late-season brood feeding and defense (**Figure 2.1**; Rehan et al. 2014a; Lawson et al. 2016). Maternal manipulation is thought to play an important role in *C. calcarata*'s nesting biology (Rehan and Richards 2013). Specifically, a reproductive female can choose whether to fertilize her eggs and, with carefully controlled pollen provisioning, can determine both the sex and body size of her developing brood (Rehan et al. 2014b; Lawson et al.

2016). Socially nesting mothers initiate their nest by provisioning a fertilized egg with a relatively small amount of pollen (Lawson et al. 2016), which subsequently develops into a dwarf eldest daughter (Rehan and Richards 2010b). As differences in body size govern dominance hierarchies in *C. calcarata*, this particular daughter's small stature at adulthood is thought to facilitate her mother's ability to coerce her into a worker role (Rehan and Richards 2013; Rehan et al. 2014a; Withee and Rehan 2016).

Adult *C. calcarata* offspring must be fed in late summer to ensure their overwintering survival (Durant et al. 2016; Lewis and Richards 2017; Mikát et al. 2017). It is during this late summer feeding that the worker daughter may contribute to her siblings' survivorship by acting as a secondary forager: either gathering pollen and nectar alongside her mother, or acting as the sole forager in orphaned nests (**Figure 2.1**; Rehan et al. 2014a). Outside of its social context, the worker daughter's diminutive size effectively negates her chances of surviving the winter season to reproduce in the following year (Rehan and Richards 2010b; Mikát et al. 2017). Thus, though the decision to establish a brood with a dwarf daughter may initially appear to represent a needlessly high maternal cost, worker daughters are thought to represent an investment in late season brood insurance for some social mothers (Mikát et al. 2017). By contrast, though subsocial mothers rear only reproductively viable brood, their nests may fail when left without feeding services in the likely event of late-season maternal mortality. *Ceratina calcarata* nests thus provide a natural system in which to empirically assess the inclusive fitness of an incipient form of social nesting. Here, we estimate genetic relatedness within and between sympatric subsocial and social *C. calcarata* colonies. Next, we assess the potential influences of mate frequency and maternal body size on social phenotype. We then combine genetic and nest

demographic data to calculate the relative costs and benefits of subsocial and social reproduction in a species capable of both reproductive strategies in sympatry.

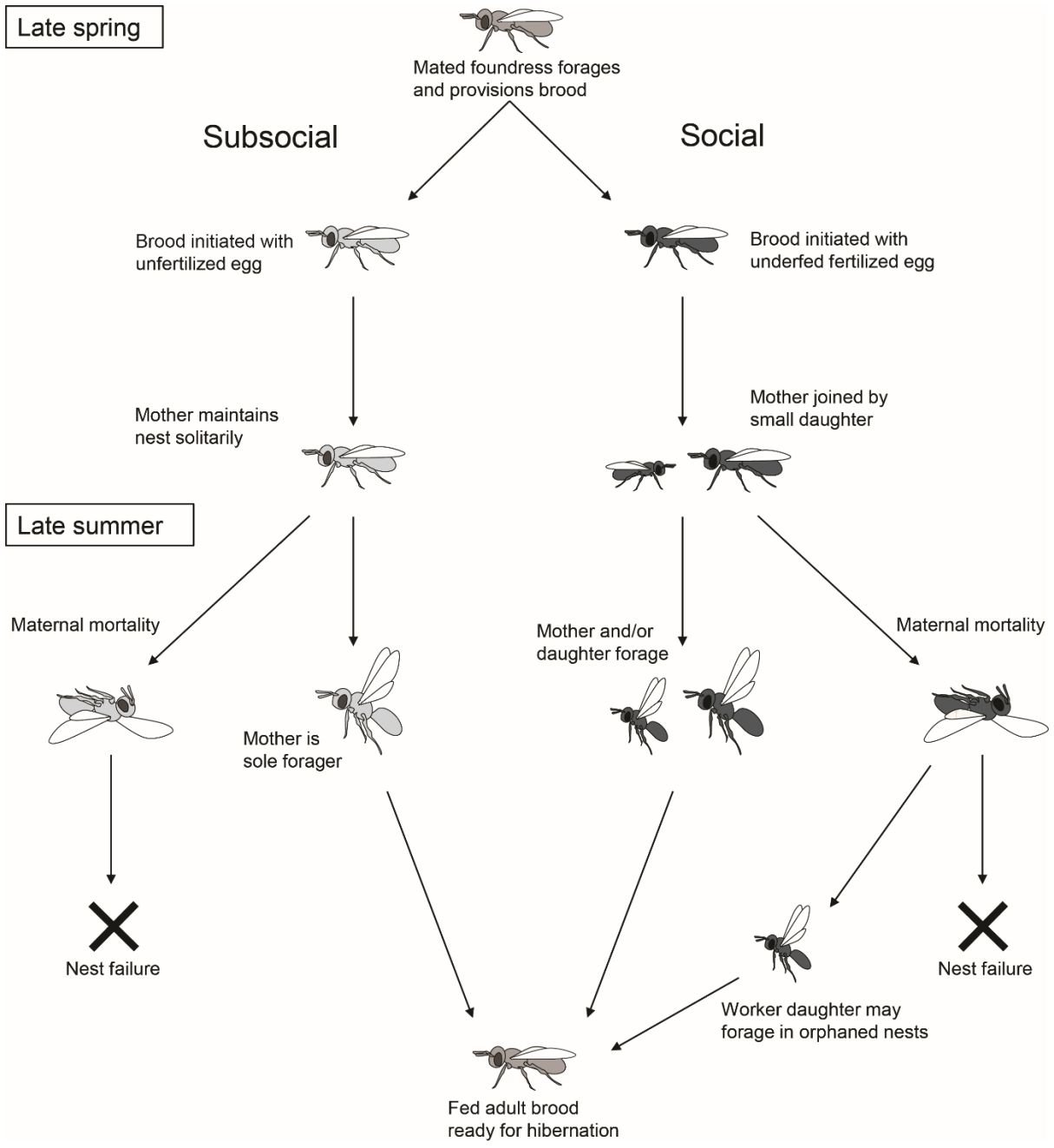

Figure 2.1. Annual life cycle and reproductive strategies of *Ceratina calcarata*. In late spring, each female disperses, mates, and establishes a new nest either subsocially (light gray) or socially (dark gray). During the late summer adult brood feeding period, social mothers and/or their worker daughters forage. A worker daughter may save her orphaned nest from failure in the event of maternal mortality. Sufficiently fed adult brood of either nest type survive a lengthy winter hibernation to disperse the following spring.

METHODS

Nest collections and assessment

A total of 167 *Ceratina calcarata* nests were collected from Durham, New Hampshire during the 2014 through 2016 summer field seasons. *Ceratina* nests were identified in the field using burrow entrance holes in dead broken stems of staghorn sumac (*Rhus typhina*) and berry brambles (*Rubus* spp.). Nests were gathered at dawn, while adults were still dormant, as this ensured collection of foraging individuals (i.e. mother and worker daughter) and prevented offspring escape. Nests were refrigerated to sedate individuals, and were then dissected lengthwise to reveal nest architecture and brood composition. Total brood cells, pollen provisions, brood parasitism and mortality were recorded, along with the developmental stage, brood cell position, and sex of each offspring (assessable at the pupal developmental stage or later); mothers were then measured for overall body size using head width as an accurate proxy (Rehan and Richards 2010b). As worker daughters are reared in the first brood cell, social nests were defined as those which contained a female offspring in the first brood cell; and subsocial nests were defined as those with a male offspring in the first position (Johnson 1988; Rehan and Richards 2010b; Lawson et al. 2016; Lewis and Richards 2017). Sex ratio (male and female brood count), social category (subsocial vs social), total clutch size (all provisioned brood cells), live brood (offspring alive at time of measurement), and maternal body size were assessed in all nests for which relevant data was available.

DNA extraction, amplification and allelic profiles

Nests used for genotyping were collected during the full brood stage, wherein the reproductive female has finished laying eggs and has assumed a brood guarding and cleaning

role. Gathering nests at this stage thus ensures that both the reproductive female and her complete brood are collected. Twenty-seven full brood nests of mixed-sex brood, containing 257 individuals in total, were selected for DNA extraction and genotyping. Of the 27 nests, 19 were social and the remaining eight were subsocial. A modified Phenol-Chloroform Isolation protocol (Kirby 1956) was used to extract DNA from the abdomen and three legs of each adult, and from the full body of each late-stage pupa. Each individual was then screened at eight polymorphic microsatellite loci (**APPENDIX A**; Shell and Rehan 2016b) using the fluorescent M13-tail methodology described in Schuelke (2000). PCR reactions were mixed to a volume of 11µl as follows: 5.45µl ddiH_20; 2.0µl 5x HF Buffer (Thermo Scientific); 0.2µl [10mM] dNTPs; 0.1µl Phusion HF Taq Polymerase (Thermo Scientific); 0.25µl [10mM] forward primer; 0.5µl Fluorescent M13 oligo [10mM], 0.5µl [10mM] reverse primer; 2.0µl DNA template. Thermocycler programs were run following primer annealing specifications from Shell and Rehan (2016b). After amplification was confirmed via gel electrophoresis (1% agarose gel) PCR product was mixed with Hi-Di Formamide (Applied Biosystems, Foster City, CA, USA) and submitted to the DNA Analysis Facility at Yale University for fragment analysis on a 3730xl Analyzer (Applied Biosystems, USA).

Individual allelic profiles were called via manual inspection of peaks in Peak Scanner 2 (Applied Biosystems). Intra- and inter-colony maximum likelihood relatedness scores between mothers and their offspring, as well as among all siblings and non-nestmates, were then assessed using ML-Relate (Kalinowski et al. 2006). Pairwise statistical tests of relatedness were then performed using one randomly sampled female per nest to account for variation in brood counts among nests (Rehan et al. 2014b). Variation among female sibling allelic profiles also allowed for assignment of an estimated sire count for each genotyped nest (as in Richards et al. 2005).

Statistical analyses of demography and relatedness were then performed in JMP Pro 13 (SAS Institute Inc., Cary, NC).

Fitness calculations

Mature brood which are not fed before the end of the blooming season do not survive the winter (Durant et al. 2016), and orphaned nests with no additional foraging activity will fail (Lewis and Richards 2017). Foraging activity by a mother or worker daughter indicates successful brood feeding and can therefore be used as a proxy for expected brood overwintering survivorship. Empirical frequencies of nest orphanage, and foraging activity by mothers and worker daughters, were made available by a comprehensive study of foraging behavior (Mikát et al. 2017). These values were used to calculate average expected survivorship for subsocial and social foundresses, which were then combined with relatedness and demography datasets to calculate the relative fitness of subsocial and social nesting strategies.

On average, *C. calcarata* nests have a 70% chance of being orphaned during the mature brood feeding stage (Mikát et al. 2017). Therefore, likelihood of survivorship in a subsocial nest was calculated as one minus the average probability of a mother dying or abandoning her brood ($S_{sub} = 1 - 0.70$). A worker daughter was observed to adopt foraging responsibilities in 18% of orphaned social nests (Mikát et al. 2017). Thus, likelihood of survivorship in social nests was calculated as one minus the probability that a nest is orphaned and not then fed by a worker daughter ($S_{soc} = 1 - (0.70 * 0.82)$). Average subsocial and social maternal fitness values were then calculated as a mother's relatedness to her offspring (r_m) multiplied by the average number of live brood for her nesting type ($subsocial = N_{sub}; social = N_{soc}$), then multiplied by the probability of offspring survival for that nesting strategy (S_{sub} or S_{soc}). The average live brood

count for social mothers was penalized by 1 to represent the cost of rearing a non-reproductive daughter:

Eq. 1 *Maternal inclusive fitness (IF)*: *subsocial*

$$IF_{sub} = [(r_m) * (N_{sub})] * (S_{sub})$$

Eq. 2 *Maternal inclusive fitness (IF)*: *social*

$$IF_{soc} = [(r_m) * (N_{soc} - 1)] * (S_{soc})$$

A worker daughter is active in 29% of social nests, either foraging alongside her mother (11%) or operating as the sole remaining forager in an orphaned nest (18%) (worker daughters do little to no foraging in the remaining social nests, Mikát et al. 2017). Mothers that survive to the end of the blooming season are expected to be able to provide sufficient late-season feeding for their brood's overwintering survival. Interestingly, in the 11% of cases where a worker daughter aids her mother in foraging, both individuals forage at half the rate of a lone forager (Mikát et al. 2017). As such, a worker daughter may be expected to receive a half the return on her total potential indirect fitness when she aids her mother in a non-orphaned nest $(wD_{coop} = 0.30 * 0.11)$, and her full indirect fitness in the 18% of cases where she assumes the role of sole forager for her siblings in the event of nest orphanage ($wD_{solo} = 0.70 * 0.18$). The average inclusive fitness for the worker daughter (IF_{wD}) is thus calculated as i) her total expected average indirect fitness ($IndF_{wD}$) multiplied by the likelihood of her contributing as the sole remaining forager (wD_{solo}), plus ii) half her expected average indirect fitness ($IndF_{wD} * 0.5$), multiplied by the likelihood of her foraging alongside her mother (wD_{coop}).

Eq. 3 *Worker daughter inclusive fitness (IF$_{wD}$)*

$$IF_{wD} = [(IndF_{wD}) * (wD_{solo})] + [(IndF_{wD} * 0.5) * (wD_{coop})]$$

RESULTS

Demography

Genotyped nests ranged in clutch size from 4 to 15 individuals, with female offspring comprising between 18 and 89% of the total brood. In the population-wide dataset of 167 nests there were 122 social and 45 subsocial nests. Social nests contained significantly larger clutch sizes on average (t-test, t = 2.10, df = 80, p = 0.04), however, the percentage of live brood was not significantly different between nest types (t = -0.068, df = 71, p = 0.95; **Figure 2.2**), nor was the frequency of brood parasitism (social = 0.11 ± 0.54; subsocial = 0.27 ± 1.03; t = -1.29, df = 53, p = 0.33). Brood sex ratios were significantly more female biased in social than subsocial nests (mean female $\%_{Social}$ = 57 ± 0.02; $\%_{subsocial}$ = 36 ± 0.038; t = -4.47, df = 31, p < 0.0001). There was a significant positive correlation between maternal body size and clutch size in social nests (social; $F_{1, 95} = 3.93$, $r^2 = 0.04$, p = 0.05), but not for subsocial nests (subsocial; $F_{1, 28} = 0.001$, $r^2 < 0.001$, p = 0.99). Similarly, a significant positive correlation between maternal body size and percent female sex investment was found across social nests ($F_{1, 94} = 5.3492$, $r^2 = 0.05$, p = 0.02), but not subsocial nests ($F_{1, 28} = 0.0222$, $r^2 < 0.001$, p = 0.88). The frequency of social nest formation was found to increase with maternal body size (Logistic regression, $\chi^2 = 4.04$, df = 1; p = 0.04) such that for each 0.1 mm increase in maternal head width the likelihood of forming a social nest increased by four percent.

Table 2.1. Nest demography and relatedness estimates from genotyped colonies. Average nest demographics reported here diverge from population-wide values, as these nests represent a subset of the 167 used for population-wide demography and fitness analyses. ID = sample nest identifier label; Class = social organizational status of genotyped nest; Clutch = total provisioned brood; Sex ratio = the number of female brood divided by the total brood; Mating = estimated mate count from genotypic data; M:O = maximum likelihood relatedness (MLR) between a mother and her total offspring; S:S = MLR among two female siblings; M:W = MLR between mother and her worker daughter; W:S = MLR between worker daughter and one female sibling; W:B = MLR between worker daughter and one male sibling.

	Nest demographics				Average relatedness scores				
ID	Class	Clutch	Sex ratio	Mating	M:O	S:S	M:W	W:S	W:B
L28	Social	8	0.63	Mono	0.50	0.72	0.55	0.25	0.25
L38	Social	5	0.40	Mono	0.50	0.65	0.50	0.74	0.25
L49	Social	6	0.83	Mono	0.57	0.61	0.66	0.59	0.25
L50	Social	8	0.50	Mono	0.60	0.90	0.50	0.69	0. 25
L70	Social	5	0.80	Mono	0.62	0.88	0.50	0.65	0. 25
L73	Social	12	0.58	Mono	0.66	0.91	0.62	1.00	0.25
L75	Social	4	0.75	Mono	0.50	0.64	0.66	0.79	0.25
L76	Social	10	0.60	Mono	0.62	1.00	0.50	0.91	0.25
L103	Social	8	0.75	Mono	0.56	0.51	0.50	0.79	0.25
N60	Social	8	0.75	Mono	0.58	0.63	0.50	0.70	0.25
Q33	Social	10	0.40	Mono	0.50	1.00	0.70	0.65	0.25
Q39	Social	10	0.80	Mono	0.43	0.84	0.50	0.73	0.25
R09	Social	10	0.60	Mono	0.61	0.74	0.65	0.76	0.25
R11	Social	12	0.67	Mono	0.53	0.54	0.50	0.50	0.25
R49	Social	8	0.63	Mono	0.51	0.57	0.50	0.69	0.25
S04	Social	6	0.50	Mono	0.50	0.83	0.57	0.67	0.25
L30	Social	7	0.71	Multi	0.52	0.30	0.50	0.60	0.25
L37	Social	11	0.36	Multi	0.54	0.78	0.62	0.69	0.25
L72	Social	9	0.89	Multi	0.53	0.30	0.50	0.56	0.25
Social average		8.3	0.64		0.55	0.70	0.55	0.68	0.25
L54	Subsocial	7	0.57	Mono	0.49	0.52	-	-	-
L97	Subsocial	7	0.43	Mono	0.71	0.78	-	-	-
Q21	Subsocial	8	0.63	Mono	0.50	0.48	-	-	-
R17	Subsocial	11	0.18	Mono	0.58	0.53	-	-	-
R20	Subsocial	13	0.62	Mono	0.50	0.36	-	-	-
R25	Subsocial	15	0.47	Mono	0.58	0.52	-	-	-
N41	Subsocial	8	0.38	Multi	0.53	0.28	-	-	-
R43	Subsocial	4	0.50	Multi	0.50	0.38	-	-	-
Subsocial average		9.1	0.47		0.55	0.48	-	-	-

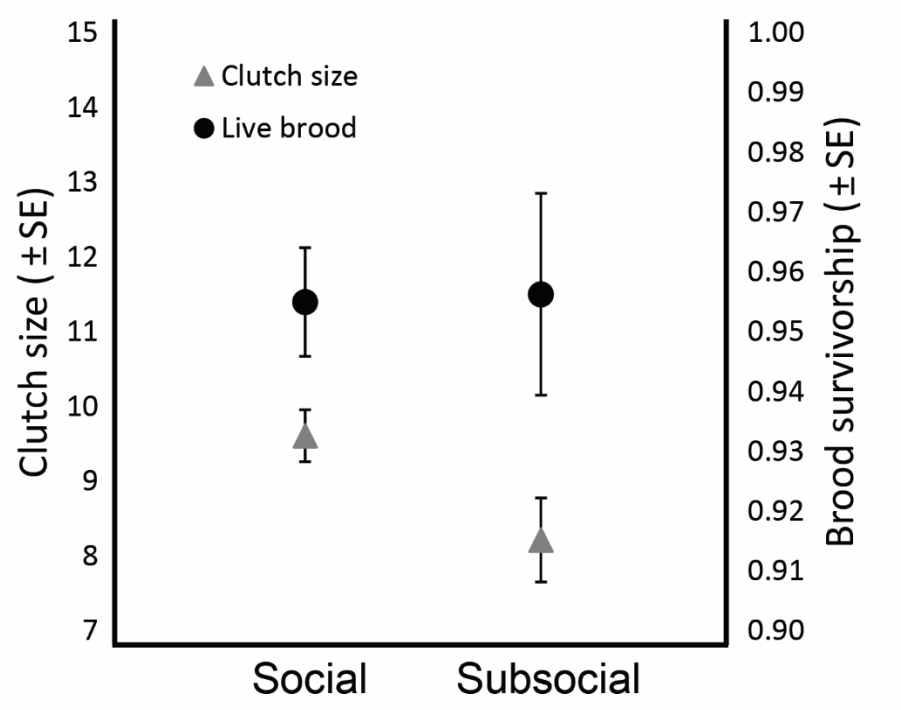

Figure 2.2. Mean clutch size and live brood in social and subsocial nests of *Ceratina calcarata*. Though social mothers produce significantly more brood (t = 2.10, df = 80, p = 0.04), average brood survivorship to adulthood is nearly identical for social and subsocial nest types (t = -0.068, df = 71, p = 0.95).

Relatedness and inclusive fitness estimates

Average intercolony relatedness was low (R = 0.07), though average relatedness between mothers and their brood was identical in both social and subsocial nests (R = 0.55; t test, t = 0.08 df = 10, p = 0.94; **Table 2.1**). Average relatedness among female siblings in social nests (R = 0.70), however, was significantly higher than subsocial nests (R = 0.48) (t = -3.11, df = 18, p = 0.0061). Within social nests, average worker daughter relatedness to sisters (R = 0.698) was not significantly different from the relatedness among non-worker sisters (R = 0.701) (t = -0.045, df = 34, p = 0.96). Across all genotyped nests, relatedness values between female offspring and their male siblings was 0.25. Genotyping colonies also revealed that 22 of the 27 nests were singly sired. The relatedness among female siblings of singly mated nests (R = 0.71) was

significantly higher than multiply mated nests (R = 0.47) (t = -4.158; df = 5.34; p = 0.008). Intracolony allelic profiles indicate that 16% (3/19) of social nest mothers and 25% (2/8) of subsocial nest mothers were multiply mated, though the frequency of multiple mating was not significantly different between nest types (Fisher's Exact Test, p = 0.62; **Table 2.1**).

Though social mothers provisioned larger clutches than subsocial on average, accounting for the rearing of a non-reproductive worker daughter resulted in similar direct fitness values for both reproductive strategies (subsocial = 4.29, social = 4.51; Wilcoxon, $\chi^2 = 0.2218$, df = 1, p = 0.64). Higher expected offspring survivorship in social colonies ($S_{soc} = 0.426$) compared to subsocial nests ($S_{sub} = 0.30$) resulted in higher inclusive fitness for social mothers ($IF_{soc} = 1.92$; vs $IF_{sub} = 1.29$). Though worker daughters do not produce offspring, their average indirect fitness was high ($IndF_{wD} = 4.06$). The likelihood that a worker daughter becomes solely responsible for the survival of the brood, however, is relatively low ($wD_{solo} = 0.126$). Further, her indirect fitness is halved in the few cases where she forages alongside her mother ($wD_{coop} = 0.033$; $Eq.3$). This low overall probability of indirect fitness gains through helping, combined with her lack of direct fitness, results in low average inclusive fitness for worker daughters ($IF_{wD} = 0.58$). Comparing fitness values calculated for the three reproductive strategies across the total assessed population, social mothers have significantly higher fitness than subsocial mothers, and subsocial mothers significantly higher fitness than worker daughters on average (Kruskal-Wallis, df = 2, $\chi^2 = 154.53$, $p < 0.0001$; **Figure 2.3**).

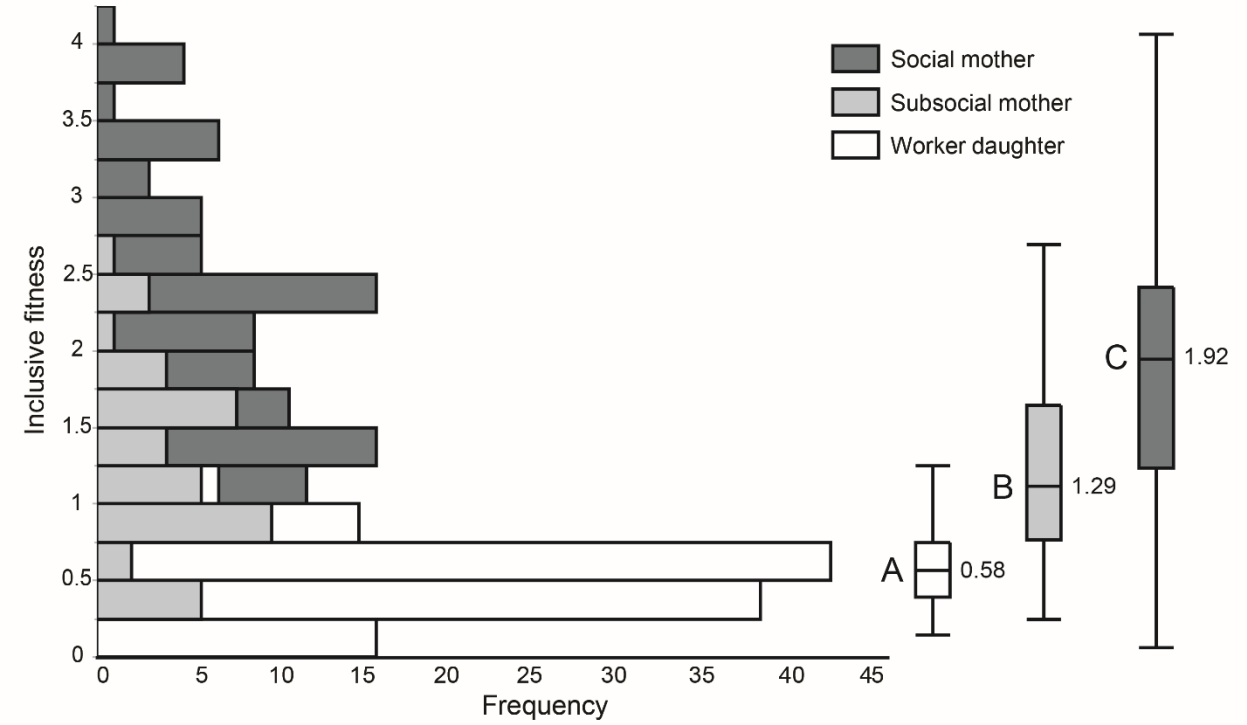

Figure 2.3. Histogram and box plots of population-wide distribution of estimated inclusive fitness values for social mothers (dark gray; N = 122), subsocial mothers (light gray; N = 45), and worker daughters (white; N = 122). Average inclusive fitness values for worker daughters (A), subsocial mothers (B), and social mothers (C) (Kruskall-Wallis, df = 2, $\chi^2 = 154.53$, $p < 0.0001$; Wilcoxon for Each Pair, social mother vs subsocial mother $p < 0.0001$, social mother vs worker daughter $p < 0.0001$, subsocial mother vs worker daughter $p < 0.0001$). All values were calculated using nest demography data and imputed relatedness scores.

DISCUSSION

Ceratina calcarata demonstrates a relatively simple form of facultative social organization, and one that is characteristic of bees in the earliest stages of social evolution (Rehan and Richards 2010b; Rehan et al. 2014a; Rehan and Toth 2015). These data represent the first ever empirical assessment of inter- and intra-colony relatedness in *C. calcarata*, and a cost-benefit analysis of a bee representative of both subsocial and incipient social reproductive strategies, in which a single worker daughter may be produced. Demographic analyses confirmed positive relationships between maternal body size and brood composition (Johnson 1988; Rehan

and Richards 2010b; Rehan et al. 2014a), and revealed that socially nesting females provision larger clutch sizes and invest in a significantly greater proportion of female offspring (**Figure 2.2**). As the probability of social nesting increased with female body size, this study provides further support for the role of maternal body size in determining division of labor in *C. calcarata* (Rehan and Richards 2010b; Withee and Rehan 2016), other *Ceratina* small carpenter bees (e.g. *C. japonica*, Sakagami and Maeta 1984), and among facultatively social bees more broadly (e.g. *Halictus rubicundus*, Field et al. 2012; *Megalopta genalis*, Kapheim et al. 2011). Screening at eight polymorphic microsatellite loci revealed higher average intracolony relatedness among female siblings in social nests compared to subsocial, and indicates that *C. calcarata* is capable of limited polyandrous mating. Inclusive fitness calculations reveal that while social nesting is advantageous in *C. calcarata*, worker daughters receive few fitness benefits for their role. Social nests are initiated though careful control of pollen provisions and sib-social care is likely enforced through differences in body size. *Ceratina calcarata* provides a clear example of how an early division of labor may be initiated by maternal manipulation and consequently maintained through maternal inclusive fitness benefits.

Variations in *Ceratina calcarata* relatedness and implications of multiple mating

Microsatellite screening revealed no difference in maternal relatedness to brood between social and subsocial nests. The higher average relatedness among social female siblings compared to subsocial, however, represents a notable asymmetry, and one which has been detected in other facultatively social bees (e.g. *Ceratina australensis*, Rehan et al. 2014b; and *Megalopta genalis*, Kapheim et al. 2015). High intracolony relatedness, particularly within female-biased broods, is thought to play an important role in the emergence of social

organization (Trivers and Hare 1976; Hughes et al. 2008). However, the degree to which social nesting may be heritable or influenced by individual condition remains an aim for further investigation (Hamilton 1964; Crozier and Pamilo 1996).

Colony relatedness data additionally indicate *C. calcarata* females are primarily singly mated (22/27 mothers), with a capacity for occasional multiple mating with two or more males (5/27 mothers). Most species of bees are thought to be monandrous (Strassmann 2001; Wilson 2005), though many demonstrate a facultative capacity for polyandry (e.g. *Bombus* species with two to four sires, Estoup et al. 1995; *Lasioglossum malachurum* reporting up to three sires, Paxton et al. 2002). Polyandry is taken to an extreme degree in some advanced eusocial bees (e.g. *Apis mellifera* with over 17 sires on average, Laidlaw and Page 1984), but may be disadvantageous in non-*Apis* bees, in which intermediate levels of colony genetic heterogeneity may incur a fitness cost. For instance, *Bombus terrestris* females mated to two males had reduced reproductive output compared to females mated to either one or four males (Bae and Schmid-Hempel 2001). Limited polyandry among social nesting *C. calcarata* suggests that strict monandry may not be necessary for the persistence of early social traits. Additional studies in *C. calcarata* and other early stage social bees are needed to better understand the degree to which mating frequency may affect the emergence and maintenance of social phenotypes.

The costs and benefits of social and subsocial nesting in *Ceratina calcarata*

According to kin selection theory, genes underlying social care behavior are more likely to be passed on when an altruist's gains in indirect fitness outweigh its direct fitness losses incurred by forgoing reproduction (Hamilton 1964). The low average inclusive fitness of worker daughters compared to both social and subsocial mothers thus seems to contradict theoretical

expectation (**Figure 2.3**). Although a worker daughter's foraging behavior is comparable to that of her mother (Mikát et al. 2017), surviving late-season mothers can sufficiently feed their mature brood alone (Lewis and Richards 2017). For this reason, on the occasion a worker daughter provisions alongside her mother, her contributions may represent a supplemental rather than necessary resource for brood survival. Additionally, although sib-social provisioning may be critical to preventing brood mortality, worker daughters assume their role of sole forager with relatively low frequency in orphaned nests (Mikát et al. 2017). Therefore, despite high relatedness between worker daughters and their siblings, reinforcement of sib-social care behavior in *C. calcarata* does not appear to be fully explained by indirect fitness benefits to the helping individual (Hamilton 1964). Although she may be making the best of a bad situation by foraging for her siblings, in the absence of inclusive fitness benefits to the worker daughter maintenance of division of labor in *C. calcarata* appears best explained by maternal manipulation (Johnson 1988; Crespi and Ragsdale 2000; Rehan and Richards 2013; Rehan et al. 2014a; Withee and Rehan 2016; Mikát et al. 2017).

Maternal manipulation of brood has long been theorized as a proximate mechanism underlying the emergence of social traits (Alexander 1974; Craig 1979) and has been suggested as an explanation for facultative sociality in other species (e.g. *Lasioglossum malachurum*, Richards et al. 2005; *M. genalis*, Kapheim et al. 2016). For instance, *M. genalis* queens manipulate pollen provisions to reduce the body size of their offspring (Kapheim et al. 2011). As observed in *C. calcarata*, *M. genalis* workers have low inclusive fitness, and it is thought that their small adult body size facilitates physical coercion into their worker roles (Arneson and Wcislo 2003; Kapheim et al. 2016). Body size affects behavioral roles in *C. calcarata* (Withee and Rehan 2016), and appears to predict the probability of social nesting. Larger *C. calcarata*

mothers may thus be better-equipped to manipulate pollen during the mass provisioning period (Rehan and Ricahrds 2010b; Lawson et al. 2016), and may be more capable of establishing a social hierarchy during the adult offspring feeding stage (Withee and Rehan 2016; Mikát et al. 2017).

While mothers of either strategy are capable of independently feeding their brood for a successful overwintering period, all mothers are also equally likely to orphan their adult offspring at the end of the brood rearing season (Mikát et al. 2017). In the event of late season orphanage, a social mother's direct fitness may be insured through sib-social brood provisioning. Similar insurance-based advantages to social nesting have also been detected in the facultatively social hover wasp, *Lieostenogaster flavolineata*, in which the fitness of brood rearing adults is likely to be assured through the collective effort of multiple helpers (Field et al. 2000). Assured fitness returns have been proposed as an important ultimate mechanism for the maintenance of social reproduction in facultatively eusocial bees (Smith et al. 2007), and among social allodapine bees (Schwarz et al. 2010) and wasps (Gadagkar 1990, 2001). The fitness advantage observed in social nesting *C. calcarata* thus suggests assured fitness returns may operate in this and other bee lineages in selecting for group living and early divisions of labor (Gadagkar 1990).

CHAPTER 3

SOCIAL MODULARITY: CONSERVED GENES AND REGULATORY ELEMENTS UNDERLIE CASTE-ANTECEDENT BEHAVIORAL STATES IN AN INCIPIENTLY SOCIAL BEE

[Shell WA, Rehan SM. 2019 Social modularity: conserved genes and regulatory elements underlie caste-antecedent behavioural states in an incipiently social bee. *Proceedings of the Royal Society B.* **286**, 20191815 (doi: https://doi.org/10.1098/rspb.2019.1815)]

ABSTRACT

The evolutionary origins of advanced eusociality, one of the most complex forms of phenotypic plasticity in nature, has long been a focus within the field of sociobiology. Although eusocial insects are known to have evolved from solitary ancestors, sociogenomic research among incipiently social taxa has only recently provided empirical evidence supporting theories that modular regulation and deeply conserved genes may play important roles in both the evolutionary emergence and elaboration of insect sociality. There remains, however, a paucity of data to further test the biological reality of these and other evolutionary theories among taxa in the earliest stages of social evolution. Here, we present brain transcriptomic data from the incipiently social small carpenter bee, *Ceratina calcarata*, which captures patterns of *cis*-regulation and gene expression associated with female maturation, and underlying two well-defined behavioural states, foraging and guarding, concurrently demonstrated by mothers and daughters during early autumn. We find that an incipiently social nest environment may dramatically affect gene expression. We further reveal foraging and guarding behaviours to be putatively caste-antecedent states in *C. calcarata*, and offer strong empirical support for the

operation of modular regulation, deeply conserved and differentially expressed genes in the expression of early social forms.

INTRODUCTION

An organism's biological whole can be conceived of as the sum of many discrete (yet ultimately interconnected) parts (e.g. gene regulatory networks; metabolic states; organs), any of which may experience evolutionary pressures that are unique to its component identity (West-Eberhard 2003). In this way, observable phenotypic plasticity – the ability of a single genotype to produce multiple phenotypes under the influence of varying environmental conditions (Simpson et al. 2011) – can be understood as a product of the organismal system's essential modularity (West-Eberhard 2003). This is conspicuously demonstrated by the developmental plasticity of eusocial insects, such as honey bees (*Apis mellifera*), in which female offspring develop into either a reproductive queen or sterile worker depending on which nutritional cue is received early in life (Evans and Wheeler 1999). Honey bee phenotypic plasticity extends further through a complex form of age polyethism: as a worker matures, her behavioural tendencies shift several times, from nursing brood when young, to potentially guarding the nest and eventually foraging towards the end of her life (Seeley 1985; Whitfield et al. 2003). Advanced eusocial Hymenoptera, which are defined by ontogenetically canalized divisions of labour between a reproductive queen and her sterile workers, have consequently allowed for revolutionary explorations into the molecular biology of insect behaviour within complex social environments (Toth and Robinson 2009; Abouheif et al. 2014).

Accordingly, extensive genomic and transcriptomic research exploring caste determination and behavior across highly eusocial Hymenoptera has illuminated an array of important genetic

and regulatory elements (Corona et al. 2016; Weitekamp et al. 2017). It is now widely appreciated that eusociality's ontogenetic and behavioral variety may be rooted in the modular nature of transcription factors and regulatory networks (Khamis et al. 2015; Chandrasekaran et al. 2011; Molodtsova et al. 2014) and their combined influence on differential gene expression (Gospocic et al. 2017). As such, eusociality's evolutionary origins may be tied to a process of regulatory network differentiation: simultaneously defining new regulatory subunits and co-opting pre-existing genes into increasingly diverse and/or novel functions (West-Eberhard 2003; Espinosa-Soto and Wagner 2010). Concurrently, comparative sociogenomic research has shown strong support for the theory that eusocial behavioral phenotypes are underpinned in part by 'toolkit' genes (Toth and Robinson 2007; Rittschof and Robinson 2016), which are genes that are deeply conserved across widely diverged taxa and appear to demonstrate consistent underlying roles in the expression of complex social traits (Morandin et al. 2016; Rehan et al. 2018).

It is widely accepted that advanced eusocial Hymenoptera evolved from solitary ancestors, with lineages gradually gaining traits of social complexity as an increasingly well-defined division of labour is established (Rehan and Toth 2015). Ongoing comparative research involving Hymenoptera which demonstrate non-eusocial forms of social organization (i.e. incipient sociality) is rapidly advancing our understanding of the molecular and environmental factors that may have contributed to the emergence and elaboration of sociality (Jones et al. 2017; Kocher et al. 2018; Saleh and Ramírez 2019; reviewed in Rehan and Toth 2015; Shell and Rehan 2018). These works have revealed that, similar to what has been observed among advanced eusocial Hymenoptera, transcriptomic rifts may be forming within these less socially derived species, dividing nestmates along behavioral and ontogenetic lines (e.g. *Ceratina*

australensis, Rehan et al. 2018; *Megalopta genalis*, Jones et al. 2017; *Polistes canadensis*, Ferreira et al. 2013). These are critical and evolutionarily consequential delineations, as they signify a regulatory separation of gene sets indicative of distinct phenotypic states and/or ontogenetic trajectories; a process of molecular compartmentalization necessary for the reduction (and eventual removal) of ancestral pleiotropic constraints putatively antecedent to a developmentally canalized division of labour (West-Eberhard 2003; Gadagkar 1997).

Recent evidence has also provided support for the social ladder hypothesis, which states that differentially expressed and deeply conserved genes likely play an important role in both the early and later stages of social evolution (Rehan and Toth 2015; Jones et al. 2017; Toth and Rehan 2017; Doganitz et al. 2018). For example, many of the same genes and regulatory elements that underlie advanced eusocial division of labour in *A. mellifera* play a conserved role in the incipiently social Australian small carpenter bee (*C. australensis*, Rehan et al. 2018). There remains, however, a paucity of transcriptomic datasets which capture gene expression and regulatory patterns underlying ontogeny and behavioral plasticity among incipiently social species; data which are necessary for further empirical examination of these hypotheses within a pan-social comparative framework (Weitekamp et al. 2017; Shell and Rehan 2018).

The incipiently social small carpenter bee, *Ceratina calcarata*, is an emerging model organism for studies of early insect sociality (e.g. Glastad et al. 2017; Rubin et al. 2019). After establishing a nest and provisioning her brood, a *C. calcarata* mother guards and cleans her maturing offspring through adulthood (Rehan and Richards 2010a, 2010b). Once her brood has matured, the mother resumes foraging to feed her adult offspring, ensuring their survival during the long winter diapause (Mikát et al. 2017). It is during this temporary but highly interactive autumn nest phase that *C. calcarata* nests may become incipiently social when the mother is

joined by one of her daughters in guarding and foraging for the nest (Mikát et al. 2017; Rehan et al. 2014). In these social nests, the mother and her daughter forage and guard at roughly equivalent rates (Mikát et al. 2017; Rehan et al. 2014); despite roughly a year's difference in age, both females demonstrate conspicuous and phenotypically similar behaviors in tandem. In this way, *C. calcarata* provides a prime natural experiment to disentangle how brain gene expression and regulation may vary both with age and socially-mediated behavioral phenotype in a transiently incipiently social bee. Here, we investigate *C. calcarata* transcriptomic data within a comparative framework to: i) examine time course variation in gene expression across the colony cycle; ii) identify patterns of differential gene expression and *cis*-regulatory enrichment associated with foraging and guarding behavior in mothers and daughters; and iii) determine whether deeply conserved genes appear to play a role in a species demonstrating emergent social traits.

METHODS

Sample collection and RNA sequencing

Ceratina calcarata nests were collected from *Rubus* and *Rhus spp.* branches around Durham, NH, USA over the course of the 2016 active season (May through August). Branches were dissected lengthwise to secure females and to assess nest developmental stage (following Rehan and Richards 2010a, 2010b). In May, founding nest (FN) mothers excavate or reoccupy a nesting burrow in preparation for brood rearing (Rehan and Richards 2010b). By June, mothers have mated and begin actively brooding (AB), provisioning brood cells with pollen balls and laying a single egg on each. By July, mothers of full brood (FB) nests guard and clean their offspring as the young mature. When the brood have reached maturity in August, autumn

mothers (AM) must forage again to ensure their survival during a lengthy overwintering period (September through April). Adult females identified at each of these five life history stages (see **Figure 3.1**) were immediately flash frozen in liquid nitrogen following nest dissection. During the autumn nest stage, social nest mothers and daughters both guard and forage for the nest. To examine brain gene expression patterns underlying these behaviors, we targeted foraging mothers, guarding mothers, foraging daughters and guarding daughters (further details on sample collection protocol can be found in Supplementary Methods, Supplementary Material online).

A total of 33 individuals were selected for whole head RNA extraction (**Table S3.1**; this and all remaining supplementary tables can be found at UNH DropBox: https://unh.box.com/s/4ss3kyw9e5j2vir4uvbx05qnoo38poxb). Heads removed on dry ice and immediately processed using the QIAGEN RNeasy Kit and protocol (Cat # 73404). RNA sample quality was then confirmed on an Agilent Tape Station 2200 and submitted for library prep and paired-end Illumina HiSeq 2500 sequencing by Genome Quebec. Read data were then aligned to the *C. calcarata* genome (Rehan et al. 2016) before being used for analysis (data accessible via NCBI SRA PRJNA434715).

Gene expression and *cis*-regulatory element enrichment analyses

Brain gene expression data was used to perform several analyses of gene expression (**Table S3.1**). First, a time course analysis of gene expression was performed using maSigPro v3.7 (Conesa and Nueda 2018) in R v3.4.3 (R Core Team 2013) to assess consistencies in gene expression variation during female aging. Analyses of differentially expressed genes (DEGs) using DESeq2 (Love et al. 2014) were then performed in R to explore foraging and guarding behavior in mothers and daughters. A complementary weighted gene co-expression network

analysis (WGCNA, Langfelder and Horvath 2008) was also run to further inspect these data. Gene ontology (GO) term enrichment was then determined for gene lists from maSigPro and DESeq2 analyses using topGO v3.7 (Alexa and Rahenfuhrer 2016), and significantly enriched GO term lists ($p < 0.05$) were further reduced using Revigo to select terms which featured a dispensability rating ≤ 0.5 (Supek et al. 2011). Transcription factor binding site (TFBS) motif enrichment among all DEGs determined for each of our focal conditions was then identified using the programs Stubb (Sinha et al. 2006) and cis-Metalysis (Ament et al. 2012) searching against the JASPAR Insect and Vertebrate databases for motif reference (Khan et al. 2018). (Full details on maSigPro, DESeq2, WGCNA, and cis-Metalysis analyses can be found in **APPENDIX B**).

Comparative analysis

BLASTn (Camacho et al. 2009) was used to assess gene homology between the *C. calcarata* genome (Rehan et al. 2016) and publicly available genomic and transcriptomic datasets from 5 bees, 3 ants, 2 wasps, and the fruit fly using cut-offs of >65% shared ID and p-values <1.0E-5 (**Table S3.2**). OrthoFinder v2.3.2 (Emms and Kelly 2015) was then run using default settings to compare *C. calcarata* amino acid sequence data to publicly available sets from some of these same species (i.e. 4 bees, one ant, and one wasp) to further examine orthology. Homologous genes were then compared between *C. calcarata* and a total of 32 additional studies collectively examining variations in gene expression by reproductive development, age, caste, and/or behavioral state in 5 bees, 5 ants, 2 wasps, the house mouse, stickleback fish, and fruit fly (**Table S3.3**). Where possible, these studies were also used to compare significantly enriched GO terms and TFBS motifs.

RESULTS

Read mapping and time course analyses

Paired-end Illumina sequencing generated an average of 31.8 MB of raw sequence data for each of the 33 samples (1.05 GB in total; **Table S3.4**). Better than 99% of the raw data for each sample met cutoff criteria for alignment, allowing the sequence data to map back to an average of 10,855 genes at 32 x read coverage across all samples.

Time course analysis identified 141 genes which collectively demonstrated one of four focal patterns in expression (at FDR < 0.01; **Figure 3.1; Tables S3.5, S3.6**). Three of these patterns appeared to be associated with processes involved in maturation from pre-reproductive (i.e. month-old) to post-reproductive (i.e. year-old) females. A total of 64 of these genes were most upregulated in month-old females before steadily declining in expression with age. GO term enrichment for this set (**Table S3.6**) identified multiple metabolic processes associated with energy production (e.g. phosphorus metabolic process). Six genes were upregulated as females reach their tenth month (e.g. *meiosis arrest female protein 1*) which featured functions associated with reproduction and neurotransmitter transport. Following this, a total of 28 genes were found to steadily increase to a most elevated expression in year-old females. GO term enrichment for this set included production of cyclic and aromatic compounds (**Table S3.6**).

The remaining 43 genes were upregulated in both pre- and post-reproductive females, suggesting this set is not so closely regulated by ontogeny (**Figure 3.1; Table S3.5**). Notable genes in this set included *histone h2a* and *pro-corazonin preproprotein*; and enriched GO terms (**Table S3.6**) included neuropeptide hormone activity and protein transmembrane transport.

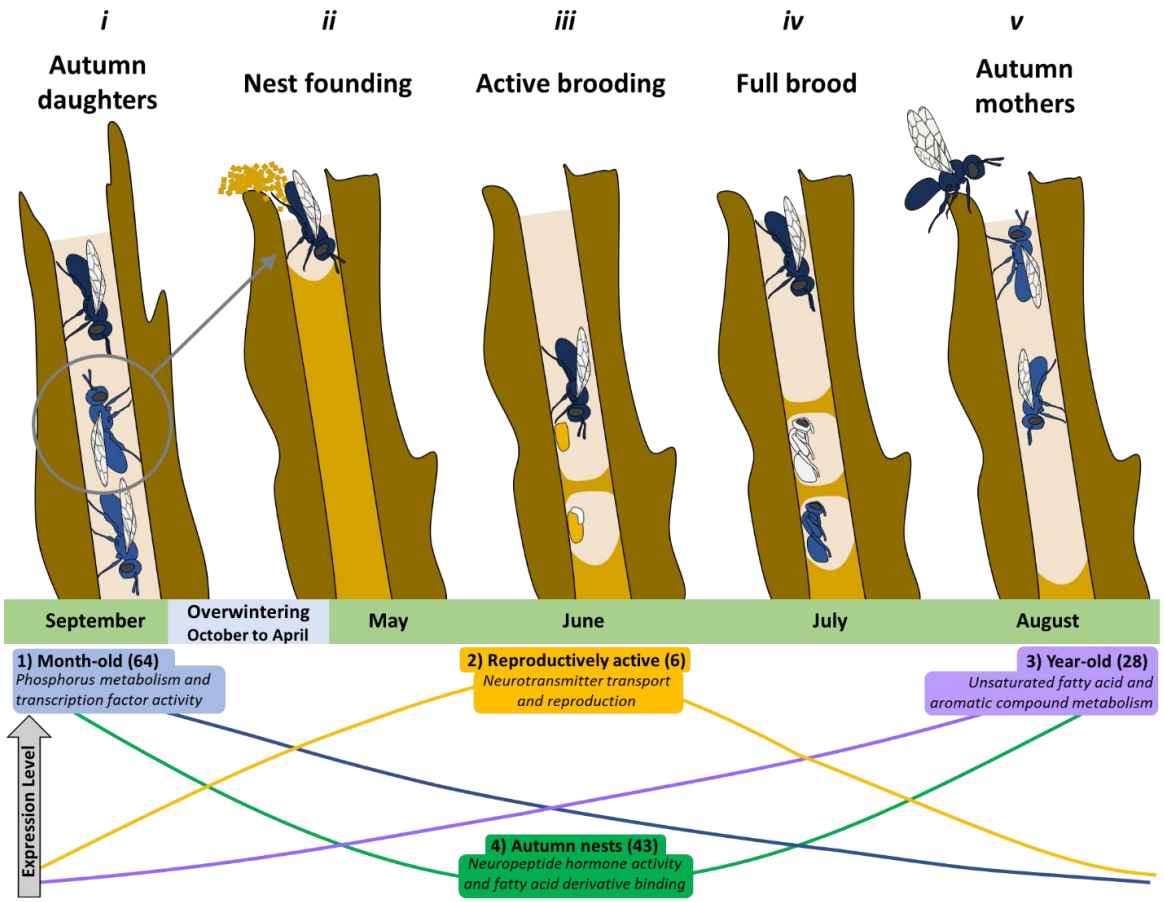

Figure 3.1. Timeline of *Ceratina calcarata* adulthood illustrating five major waypoints associated with reproductive maturation and senescence: *i*) a recently eclosed autumn daughter (circled) occupies her natal nest wherein she may assist her mother (dark blue) in foraging and guarding; *ii*) after overwintering, unmated females disperse to establish their own nest; *iii*) now reproductively active, females gather pollen and nectar to provision their own brood; *iv*) mothers remain in the nest to guard and clean their brood as the young mature; *v*) in autumn, mothers (dark blue, on nest entrance) resume foraging to feed their adult offspring, ensuring their overwintering survival. Depicted below the timeline are the four major patterns of overall gene expression identified through time course analysis. Sets 1, 2, and 3 capture genes demonstrating age-associated variations in expression; set 4 captures genes upregulated in both autumn mothers and daughters. Gene counts (in parentheses) and representative GO terms enriched for each set is provided. For full list of genes and GO terms see Tables S5, S6, S8, and S10.

Differential gene expression by behavioral state and age

Analyses of gene expression among behavioral states (i.e. foraging vs guarding vs nesting individuals) captured 2,017 distinct significantly differentially expressed genes overall (at FDR <0.05; **Table S3.7**). Comparing genes uniquely upregulated among phenotypes, 292 genes were upregulated specifically in nesting females, compared to 483 genes associated with foragers, and

1,104 genes upregulated in guards (**Figure 3.2**). Forager-associated genes were conserved across bees (*C. australensis*; *A. mellifera*), wasps (*Polistes metricus*), and ants (*Temnothorax longispinosus* and *Solenopsis invicta*; **Figure 3.3**) and were enriched for gene silencing and immune related processes (**Tables S3.7**, **S3.10**). Guard-associated genes were well-conserved among the limited datasets that allowed for comparison, which included *C. australensis* and *A. mellifera* (**Tables S3.8**, **S3.9**); this set was enriched for processes involved in carbohydrate derivative metabolism, regulation of defense response, and neurotransmitter transport (**Tables S3.7**, **S3.10**).

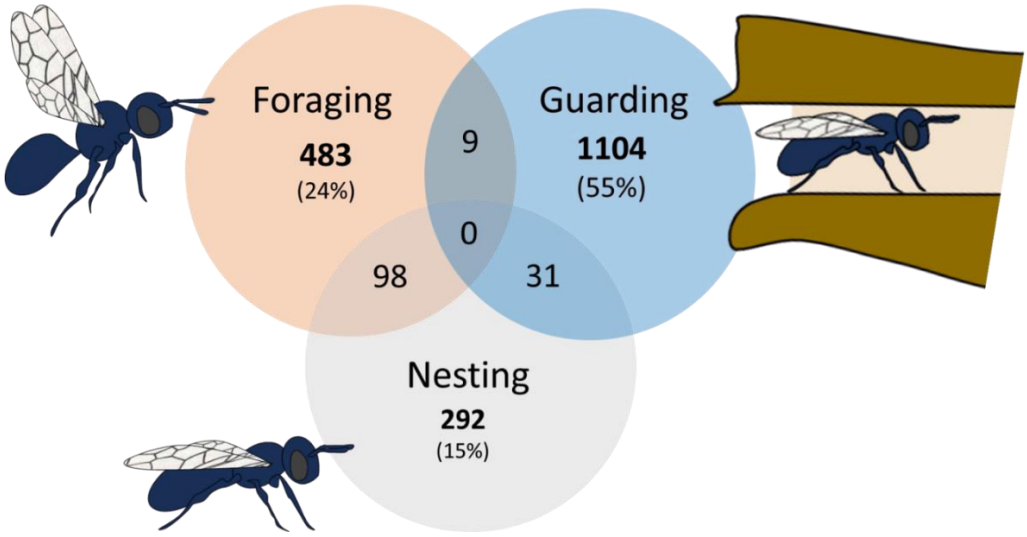

Figure 3.2. Venn diagram of upregulated genes differentiating guarding and foraging individuals in comparison to non-foraging, non-guarding 'nesting' individuals. Unique gene counts by category are provided in bold, and respective percentages of total genes are indicated in parentheses (6% of DEGs are shared among phenotypes). Compared to nesting controls, foraging and guarding individuals feature a roughly 2-and 4-fold increase in number of upregulated genes respectively.

A total of 5527 significantly differentially expressed genes were identified across all behavior by age comparisons (at FDR < 0.05; **Table S3.11**). Comparing the effects of age on either behavioral phenotype, 616 genes were found to distinguish foraging mothers (N_{DEGs} =

198) from foraging daughters (N_{DEGs} = 418, **Figure S3.1D; Table S3.11**). GO enrichment indicates that foraging mothers rely more on organic cyclic compound metabolism and lipid catabolic process, whereas daughters may require generation of energy, including carbohydrate metabolic processes. A total of 2488 DEGs distinguished guarding behavior in mothers (N_{DEGs} = 1292) from daughters (N_{DEGs} = 1196, **Figure S3.1E**). Notable genes upregulated in guarding daughters included *syntaxin 1a*, *glutamate* and many glutamate-associated genes (**Figure 3.4**). Mother guard enrichment indicated roles for positive regulation of translation and histone deacetylase activity (**Table S3.11**). Daughter guards were enriched for demethylation, lipid metabolic process, and neurotransmitter transport.

A total of 114 genes distinguished guarding from foraging behavior among mothers, most of which were upregulated in guarding individuals (N_{DEGs} = 73, **Figure S3.1A; Table S3.11**). While guarding mothers were enriched for methylation and neurological system process, foraging mothers featured response to stress. A total of 3031 genes separated guarding (N=1451) from foraging behavior in daughters (N=1581, 52%; **Figure S3.1B; Table S3.11**). Notable conserved genes upregulated in guarding daughters over foraging included many glutamate-associated genes and *syntaxin 1a*. Where foraging daughters were enriched for biological processes involved in lipid and carbohydrate metabolism and torso signaling pathways (**Figure 3.4**; **Table S3.11**), guarding daughters featured greater enrichment for neurotransmitter transport and regulation of synapse structure or activity. Genes and GO terms associated with both foraging and guarding in daughters overlapped among younger females of other social taxa most consistently with pre-dispersal females in *C. australensis* (N_{DEGs}=279), and to a lesser extent among *A. mellifera* (N_{DEGs}=17), *T. longispinosus* (N_{DEGs}=6), and *S. invicta* (N_{DEGs}=16; **Tables S3.8-S3.10**). Gene ontology enrichment for methylation and histone modification were generally

well-conserved among guarding or foraging mothers and older individuals of other social taxa (bees, *C. australensis* and *A. mellifera*; and ants, *Formica exsecta*; **Table S3.9**).

WGCNA identified a total of 35 distinct and statistically supported modules of co-expressed genes, each positively or negatively correlated with foraging, guarding, nesting, mothers, and daughters (**Tables S3.12-S3.16**; **Figures S3.5-S3.10**). Overall, this analysis strongly corroborated results of our DEG analysis: e.g. daughters were again shown to have much higher counts of significantly associated genes ($N_{daughters}$ = 695; p = 4.8e-75) compared to mothers ($N_{mothers}$ = 296; p = 8.4e-25); and the glutamate family of genes was again found to be expanded specifically among guarding daughters (see **Tables S3.13**, **S3.15**).

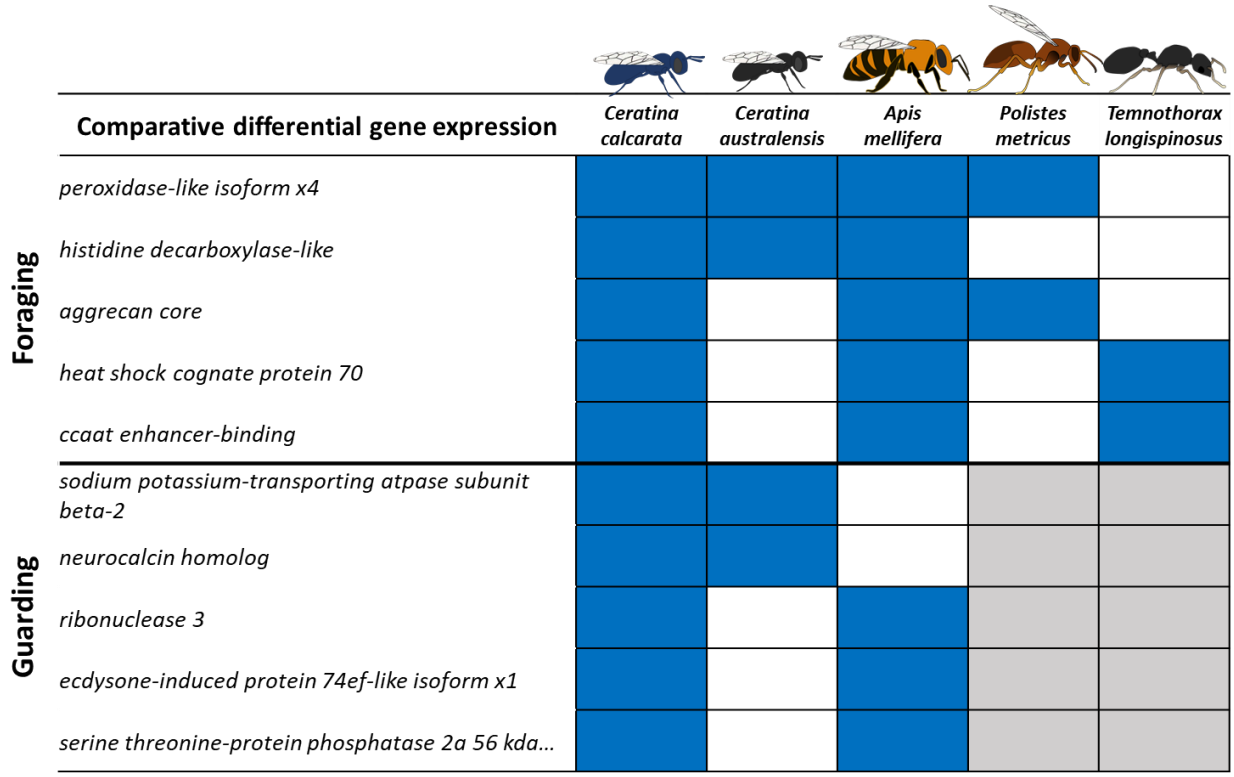

Figure 3.3. An illustrative subset of all genes associated with foraging or guarding behavior in *C. calcarata* which matched with strong statistical support to genes in *C. australensis*, *A. mellifera*, *P. metricus*, and/or *T. longispinosus* (for full gene lists and references see Table S2.8). Blue boxes indicate shared genes and similar regulatory contexts between *C. calcarata* and each species; white boxes indicate a lack of contextual or regulatory overlap. Grey boxes indicate no applicable comparison. Overall, genes associated with foraging behavior in *C. calcarata* appear to be deeply conserved across Hymenoptera. Fewer studies examine the transcriptomics of guarding behavior, but existing bee literature also reveals important conserved candidate genes.

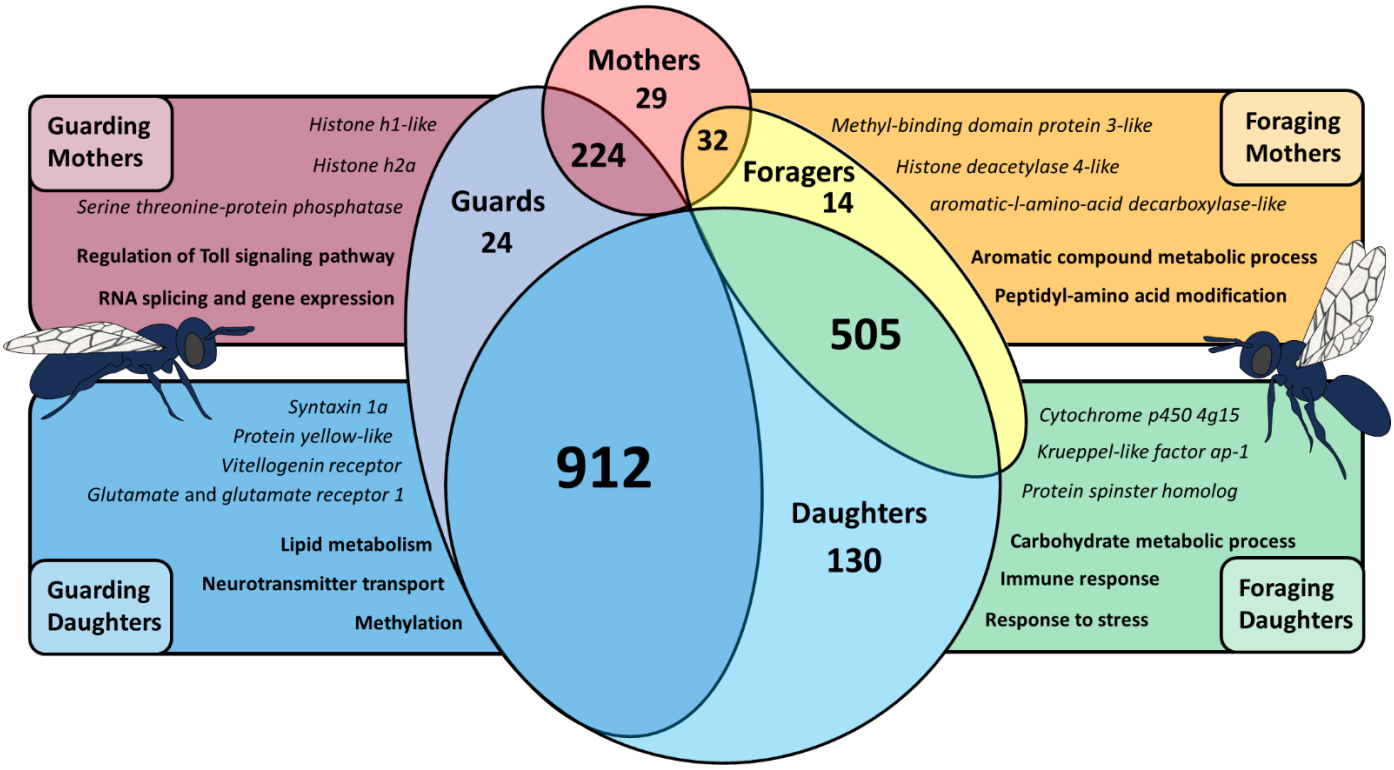

Figure 3.4. Summary of unique upregulated genes and enriched GO terms identified in each focal age-behavioral state, namely foraging and guarding mothers and daughters. Numbers represent counts of DEGs uniquely upregulated by category (circle sizes and spans of overlap are relative); a selection of notable genes (in italics) and enriched GO terms (bold) for each role are provided in four corner panels. In all cases, greater counts of upregulated genes are consistently found in daughters over mothers and guards over foragers.

Cis-regulatory enrichment

Significant enrichment of transcription factor binding site (TFBS) motifs associated with behavioral states revealed a total of 705 functionally unique motifs (**Tables S3.17, S3.18**). Motifs accommodating transcription factors with known neural function (e.g. memory, learning, circadian rhythm), enriched upstream of genes upregulated in foragers and guards, were largely unique to behavior (**Figure 3.5**). These neural-associated transcription factors include some which were, themselves, differentially regulated (**Table S3.8**). Comparisons of TFBS motifs enriched among foragers in publicly available datasets (e.g. Khamis et al. 2015; **Table S3.18**) revealed BarH1 to have a conserved role underlying differential gene expression in *C.*

australensis and foraging behavior in *A. mellifera*. The guard-associated TFBS motif for CTCF was also previously identified as regulating social conflict response in *C. calcarata* and in caste-associated behaviors in *A. mellifera.*

A total of 308 TFBS motifs were enriched across up-regulated genes among foraging mothers and daughters (**Table S3.17; Figure S3.18**). Foraging daughters were enriched for many TFBS motifs (N = 193) which bind pairs of TFs upregulating immune response (e.g. *Kruppel*) and neuronal development (e.g. *scalloped*). Compared to daughters, foraging mothers were enriched for fewer motifs given the DEG count (N_{TFBS} = 115; chi-squared test, χ^2=2.47, df=1, p=0.116), binding many multifunctional TFs which likely regulate large suites of processes (e.g. *IRF2*), as well as those involved in learning (e.g. *CREB1*; **Table S3.18**). Conserved TFBS motifs associated with foraging mothers (e.g. *SP1* and *USF1*) or daughters (e.g. *slbo*) were also enriched among DEGs in *C. australensis* or foragers in *A. mellifera* (**Table S3.18**).

A total of 172 TFBS motifs were enriched among mother and daughter guard-associated DEGs (**Table S3.17**), including those binding TFs involved in methylation, neural function, and circadian rhythmicity (**Table S3.18; Figure S3.11**). TFBS motifs specific to guarding daughters (N=124) bind TFs involved in neurological development and activity as well as immune response and functionality (**Table S3.18**). Compared to daughters, guarding mothers featured significantly reduced TFBS motif enrichment given DEG count (N_{TFBS} = 48; chi-squared test, χ^2=37.14, df=1, p<0.0001), with sites binding TFs involved in reproductive maturity and a diverse set of regulatory roles. Guarding mother and daughter-associated transcription factors, such as *opa*, *utraspiracle* and *repo*, also regulate foraging and nursing behavior in *A. mellifera* [11] or social polyphenism in *C. australensis* ([38]; **Table S3.18**).

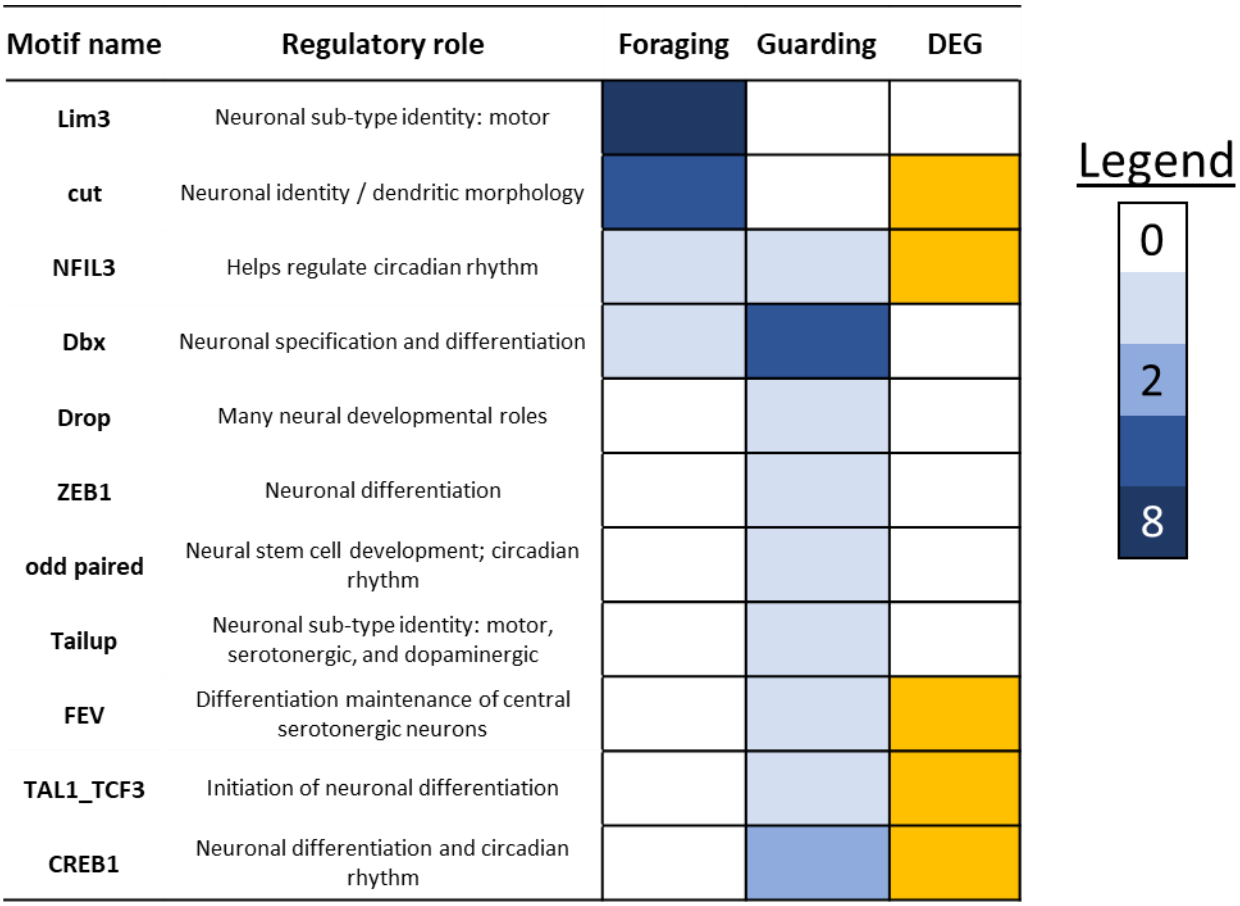

Figure 3.5. Heat map highlighting transcription factor binding site motifs with known neural regulatory roles, significantly enriched in the promoter regions of genes associated with foraging and guarding individuals regardless of age (for full list see **Figure S3.11**; **Table S3.18**). Motif names are presented in order of phenotypic affiliation, along with a summary description of regulatory roles. Enrichment counts for each motif in the upregulation of genes associated with each biological context is then indicated by color (legend: white = no enrichment; blue = enriched, with darker blues indicating greater counts). The rightmost column indicates in yellow whether the associated transcription factor is also differentially expressed in this study.

DISCUSSION

This brain transcriptomic dataset captures variations in gene expression underlying both a phenological time course and conspicuous behavioral states in *Ceratina calcarata*, an incipiently social small carpenter bee. In this discussion, we consider the strong and systemic influence of a relatively simple and transient social environment; and examine how differences in behavioral phenotype and age reveal a high degree of modularity and variation in regulation and expression

of associated genes. We also discover that some of the same genes and regulatory elements associated with foraging and guarding behavioral states in *C. calcarata* include those which may play conserved roles in phenotypic plasticity and social complexity across social taxa.

Time course variations in gene expression

Time course analysis revealed two major biological events which unfold over the course of *C. calcarata*'s lifetime. The first involves a major shift in metabolism associated with maturation: from generation of energy when young to the production of aromatic and cyclic compounds when old. Greater energy and metabolic requirements among young adults have been observed across eusocial Hymenoptera (e.g. *A. mellifera*), attributable to developmental processes and caste roles (Ament et al. 2008). By contrast, enrichment among year-old mothers for processes involving organic and aromatic compound metabolism may suggest that mothers increasingly rely on chemical signaling as they begin interacting with their adult brood. Chemical signaling is widely employed among Hymenoptera as a means of nestmate communication and may reinforce divisions of labour (e.g. *A. mellifera*, Leonhardt et al. 2016). Accordingly, chemical signaling may play a role, alongside aggression and social experience, in affecting gene expression and behavior in *C. calcarata*'s autumn nests (Withee and Rehan 2017).

The second major event is reflected in the relatively large portion of genes upregulated simultaneously in both autumn nest mothers and daughters, a pattern in overall gene expression that is highly consistent with the findings of previous brain transcriptomic work in *C. calcarata* (Rehan et al. 2014). Mothers and daughters are otherwise separated by nearly a year of physiological maturation and experience (Rehan and Richards 2010a, 2010b) indicating that upregulation in this set of genes is more likely associated with the conditions of the autumn nest

rather than ontological stage. Considered from an evo-devo approach, continuous variations in phenotype (e.g. behavior) are expected to be organized by critical 'switches' (e.g. particular life stages or environmental circumstance) that demark modular points of segregation among ontogenetic stages or regulatory states (West-Eberhard 2003). Among eusocial taxa, changes in nest social environment, such as during the shift between reproductive and brood caring phases in the clonal raider ant (*Ooceraea biroi*, Libbrecht et al. 2018), have been shown to induce substantial variation in underlying gene expression. Reproductively successful *C. calcarata* females can arguably be seen to experience a similar series of regulatory switch points: departing the incipiently social environment of their natal nest to establish and maintain their own brood as a subsocial parent, and later reentering the incipiently social environment to care for their own adult brood.

Considering this series of shifts, one particularly notable gene identified within this set is *pro-corazonin preproprotein*, which encodes a necessary precursor to *corazonin*, a widely conserved cardiac and neuropeptide (Veenstra 1991). *Corazonin* contributes to social and behavioral phenotypic plasticity in other species – facilitating shifts between solitary and gregarious phases in two locusts (*Locusta migratoria* and *Schistocerca gregaria*, Sugahara et al. 2015) and reproductive and working states in a ponerine ant (*Harpegnathos saltator*, Gospocic et al. 2017) – and may play a comparable role in *C. caclarata*. Overall, lifetime variation in *C. calcarata*'s brain gene expression appears to be strongly affected by the nest social environment, potentially involving global gene regulatory networks with distinct neural and metabolic states (Cardoso et al. 2015). Further, the molecular dynamics underlying *C. calcarata*'s shifts in nest social environment suggest additional support for the role of certain conserved genes in the operation of social plasticity (Toth and Robinson 2007; Rehan and Toth 2015).

Extensive differential regulation underlies behavioral states

Differential gene expression analysis revealed two- and four-fold expansions in gene upregulation among foraging and guarding individuals compared to the nesting (i.e. non-foraging, non-guarding) state. This result was also reflected by gene cluster analyses, which assigned the largest number of positively and significantly correlated genes to the most strongly guarding-associated module. Our results are thus consistent with other transcriptomic works typifying gene expression among narrowly defined behavioral states in advanced eusocial insects (e.g. ants, Kohlmeier et al. 2019; and honey bees, Whitfield et al. 2003). For example, in the acorn ant (*T. longispinosus*), four times more variation in gene expression was detected among behavioral states than in comparisons involving age or fertility status (Kohlmeier et al. 2019). Although some studies suggest little variation in gene expression underlying tasks performed by age-matched individuals (e.g. guarding vs undertaking in *A. mellifera*), this may also be attributable to the strongly age-associated polyethism of these species (Cash et al. 2005).

As hypothesized, comparative analyses revealed that both foraging and guarding associated genes in *C. calcarata* are generally well-conserved across taxa, including its incipiently social congener *C. australensis* (Rehan et al. 2018), and among eusocial bees (Khamis et al. 2015), wasps (Toth et al. 2010), and ants (Kohlmeier et al. 2019; Feldmeyer et al. 2014) (**Tables S3.8**, **S3.9**; **Figure 3.3**). Although fewer studies allowed for comparison specifically against guard-type roles, our study still offers additional support for the operation of deeply conserved and differentially expressed genes across social taxa (Rehan et al. 2018; Rehan and Toth 2015; Saleh and Ramírez 2019).

Notably, 517 of the DEGs associated with foraging or guarding behavior in this study collectively correspond to 58% of all genes previously identified with signatures of positive selection in the *C. calcarata* genome (N_{total} = 877 genes; Rehan et al. 2016). Theory suggests

that, as insects gain traits of social complexity, genetic release from pleiotropic constraint should lead to elevated rates of DNA recombination among genes associated with potentially caste-antecedent behavioral phenotypes (Rehan and Toth 2015; Gadagkar 1997; Rehan et al. 2016; Kent et al. 2012). As modular subunits (e.g. regulatory networks underlying behavioral states) are established and selectively reinforced, previous genetic correlations (i.e. pleiotropic constraints) are expected to relax, allowing associated genes to be rapidly co-opted into new functional roles (West-Eberhard 2003). The elevated rates of molecular evolution in genes specifically upregulated within narrowly defined and socially mediated behavioral states in *C. calcarata* thus suggest support for these theoretical predictions. Across divergent lineages with well-defined divisions of labour, overall rates of accelerated molecular evolution appear to be correlated with degrees of social complexity (Doganitz et al. 2018)); and recent comparative studies among both facultative and obligate eusocial species have consistently detected elevated rates of recombination specifically among genes associated with the worker caste (e.g. *M. genalis*, Jones et al. 2017; *A. mellifera*, Harpur et al. 2014) or working behavior (Kent et al. 2012). In much the same way, it appears to be primarily the genes associated with socially-mediated behavioral states that are experiencing genetic release in *C. calcarata* (Gadagkar 1997; Rehan et al. 2016).

Distinct neural regulatory pathways underlie guarding and foraging behaviors

Neural-associated TFBS motifs enriched in the promoter regions of genes upregulated in foraging or guarding individuals were largely specific to behavior and included binding sites for transcription factors that were themselves differentially regulated. Foragers were consistently enriched for sites binding the TFs *cut* (also a DEG) and *Lim3*, whereas guards were mainly

enriched for motifs binding the TFs *Dbx* and *CREB1* (also a DEG), which may also play a role in *C. calcarata* aggression (Withee and Rehan 2017). The behavioral regulatory patterns in *C. calcarata* thus resemble those detected in the brain transcriptional regulatory network of *A. mellifera*, in which behavioral states were found to be underpinned by role-specific TF modules (Chandrasekaran et al. 2011). As such, the foraging and guarding behavioral phenotypes may represent distinct neurogenomic states, each subject to its own suites of state-specific selective pressures (West-Eberhard 2003; Chandrasekaran et al. 2011; Cardoson et al. 2015). Over time, these distinctions could conceivably drive a molecular wedge between foraging and guarding behavioral phenotypes in a *C. calcarata*-like lineage putatively antecedent to more canalized caste roles (West-Eberhard 2003).

Age effects on behavioral states

Of the phenotypes examined, guarding daughters were consistently underpinned by the largest and most distinct set of genes and regulatory elements, including many which have been strongly associated with social plasticity in other species (e.g. *syntaxin 1a*, Kocher et al. 2018). Perhaps most intriguingly, though, guarding daughters were repeatedly associated with the greatest number of glutamate-related genes. Glutamate and its receptors have been found to play a key role in task specialization within divisions of labour in other social insects (e.g. *C. australensis*, Rehan et al. 2018; *A. mellifera*, Liang et al. 2014; and leaf-cutting ants, *Atta vollenweideri*, Koch et al. 2013); and in the adoption and/or expression of social traits among vertebrates (including mice, Xiao et al. 2017; dogs and humans, O'Rourke and Boeckx 2019). Glutamate-related genes thus represent a particularly promising deeply conserved candidate suite

for further functional and comparative genomic investigations into the emergence and elaboration of sociality (Toth and Robinson 2007; Rittschof and Robinson 2016).

Differences in age accounted for the majority of quantitative and qualitative variation in both DEGs and enriched TFBS motifs: daughters in either foraging or guarding roles consistently featured greater numbers of DEGs and TFBS motifs compared to mothers. For example, while age-associated DEGs split roughly evenly between guarding mothers and daughters, guarding mothers featured significantly fewer TFBS motifs than expected. Potentially owing to major ontogenetic and metabolic differences, behavior-specific switches triggered by the conditions of the incipiently social autumn nest environment may be causing a considerable degree of dissociation in gene regulation and expression along highly age-specific lines (West-Eberhard 2003). While expression of either behavioral state among mothers may involve comparatively constrained regulatory pathways, potentially pleiotropically tied to the ontogenetic aspects of reproductive maturation and activity (Gadagkar 1997), these same states appear more dynamically and expansively regulated among daughters, incorporating many more TFs and genes. Consequently, while foraging and guarding behaviors may yet appear phenotypically similar between mothers and daughters, the associated underlying regulatory pathways and gene expression patterns appear to be undergoing substantial differentiation and elaboration; a process strongly suggestive of a lineage actively experiencing an appreciable augmentation in social complexity.

CHAPTER 4

MOLECULAR PATHWAYS UNDERLYING CASTES AND LONGEVITY IN A FACULTATIVELY EUSOCIAL SMALL CARPENTER BEE

ABSTRACT

Unravelling the evolutionary origins of eusocial life, one of the most complex forms of social organization known in nature, is a longstanding endeavor in the field of evolutionary-developmental biology. Although descended from solitary ancestors, eusocial insects such as honeybees have evolved an ontogenetic division of labor between a reproductive queen and non-reproductive worker caste. Workers perform a variety of age-associated tasks over their roughly month-long life, while queens continuously produce brood for up to five years. It has been proposed that i) the pronounced phenotypic plasticity of eusocial caste systems may have evolved through the co-option of deeply conserved and differentially expressed genes; and ii) caste longevity may be tied to differences in both oxidative damage and mitigation capacity. Although these hypotheses have received illuminating examination among highly eusocial corbiculate bees, there remains a paucity of empirical data from other social bee lineages. Here we present brain transcriptomic data from a Japanese small carpenter bee, *Ceratina japonica* (Apidae: Xylocopinae), which demonstrates both solitary and eusocial nesting in sympatry and lives two or more years in the wild. These data capture patterns of gene expression and *cis*-regulation associated with first- and second-year solitary females, queens and workers, providing an unprecedented opportunity to explore the molecular mechanisms underlying caste-antecedent

phenotypes in a long-lived and facultatively eusocial bee. We find that *C. japonica*'s queen and worker classes are underpinned by highly divergent and modular gene regulatory and co-expression pathways, involving many differentially expressed genes that are strikingly well-conserved among other eusocial bee lineages. We also discover that the social nest environment may induce shared variations in queen and worker lifetime gene expression versus solitary females. Finally, these data offer support for the role of oxidative damage reduction as a proximate mechanism of prolonged longevity in insects.

INTRODUCTION

The emergence of obligate eusocial life, in which a reproductive individual is supported by many non-reproductive relatives (Michener 1969; Wilson 1971), is considered one of the major evolutionary transitions in biological complexity (Szarthmary and Maynard Smith 1995). A defining feature of eusocial organization is pronounced phenotypic plasticity, the capacity for a single genotype to produce multiple phenotypes under different environmental conditions (West-Eberhard 2003; Simpson et al. 2011). For example, in honey bees, a female will develop either into a reproductive queen or a non-reproductive worker depending on the dietary and environmental conditions of her infancy (Engels and Imperatriz-Fonseca 1990). This ontogenetic split is highly consequential, as queens and workers go on to exhibit substantially different physiologies and behavioral states. Where a queen spends most of her life inside the nest, continually producing brood and interacting with her colony, her sterile worker offspring perform a variety of complex, age-related tasks to support and protect the nest (Seeley 1985; Whitfield et al. 2003). This rich phenotypic variation by caste extends to life expectancy: active workers live for two to six weeks (Remolina et al. 2007), but queens may spend up to five years continuously producing additional brood (Blacher et al. 2017).

Research exploring caste-associated variations in honey bees has provided a wealth of insights into the molecular mechanisms underpinning differences in behavior and longevity (Evans and Wheeler 1999; Corona et al. 2005). For example, we now appreciate that both the developmental and behavioral plasticity of honey bees emerges from a highly directional regulatory network of modular suites of co-expressed genes (Chandrasekaran et al. 2011; Molodtsova et al. 2014). Further, studies have shown that the variations in longevity seen in honey bees and bumblebees may be tied to individual reproductive capacity (Blacher et al. 2017; Lockett et al. 2016), metabolic requirements, and oxidation reduction activity (Keller and Jemielity 2005, Williams et al. 2008; Li-Byarlay and Cleare 2020). The relatively abbreviated honey bee worker lifespan, for instance, may partly be explained by the high metabolic costs of frequent flight, which causes increasing harm as oxidation reduction activity fades with worker age (Remolina et al. 2007; Williams et al. 2008). By contrast, not only do long-lived queens rarely fly, there is clear evidence that reproduction-associated genes, such as *vitellogenin*, may also contribute to mitigating oxidative damage (Seehuus et al. 2006).

Despite their extreme phenotypic plasticity, advanced eusocial bee lineages are known to have evolved from solitary ancestors (Rehan and Toth 2015). Research among bee lineages of less derived sociality thus presents an exciting opportunity to empirically test major evolutionary hypotheses regarding the emergence of social organizational systems (Rehan and Toth 2015; Toth and Rehan 2017; Shell and Rehan 2018). For instance, recent works have indicated that similar behavioral states and reproductive roles among independent social bee lineages may be underpinned by the consistent differential expression of deeply conserved genes (Rehan et al. 2018; Saleh and Ramirez 2019; Kapheim et al. 2020). Research has also revealed that worker-like roles often feature expanded regulatory elements and evidence of positive protein evolution

(Jones et al. 2017; Shell and Rehan 2019), as predicted by genetic release (Gadagkar 1997). Additionally, comparative research is finding growing support for the role of sociality itself as exerting considerable influence on gene expression pathways associated with behavioral states (Rehan and Toth 2015; Rubinstein et al. 2019) and lifespan among bee lineages (Lucas and Keller 2019). There remains, however, a paucity of datasets which capture transcriptomics of emergent social traits and aging among facultatively eusocial bee taxa, which would allow for empirical testing of hypotheses that seek to bridge the effects of developmental time and social environment on phenotypic plasticity (West-Eberhard 2003; Toth and Rehan 2017).

Ceratina japonica is a long-lived species of small carpenter bee capable of forming either solitary or eusocial nests in sympatry across its native range in Japan (Sakagami and Maeta 1977, 1984, 1987). As common to all social species of small carpenter bees around the globe (Rehan 2020; Rehan et al. 2010; Udayakumar and Shivalingaswamy 2019; Sakagami and Maeta 1995), *C. japonica* forms or reuses a linear burrow within the stems of pithy plants in which it rears around eight offspring a year (Sakagami and Maeta 1977, 1984). In solitary nests, brood cell provisioning, guarding, and rearing are accomplished by a single reproductive female. By contrast, social nests typically contain two adult females: the larger mother is reproductively dominant and guards the nest, while the smaller daughter forages (Sakagami and Maeta 1984, 1987). *Ceratina japonica* also lives and reproduces for two years (Sakagami and Maeta 1984). In this way, *C. japonica* presents a novel opportunity to simultaneously test hypotheses regarding the evolution of social phenotypic plasticity and longevity in a species demonstrating facultative eusociality. Here we investigate *C. japonica* brain transcriptomic data to: i) identify patterns of differential gene expression and *cis*-regulatory enrichment among this species' three naturally co-occurring classes (i.e. queens, workers, and solitary females); ii) explore molecular signatures

of aging within and among these classes; and iii) assess the degree to which molecular elements associated with sociality and longevity may be conserved among this and other social insect lineages.

METHODS

Sample collection and sequencing

Ceratina japonica queens, workers, and solitary females were collected from nests primarily found in dead, broken stems of *Hydrangea sp.* around Sapporo, Japan in July 2015. Social nests are most often established in pre-established burrows; workers are primarily foragers and are capable of reproduction, though most eggs are consumed by queens; queens are the dominant egg layers and primary guards; both females regularly engage in trophallaxis (Sakagami and Maeta 1984, 1987). Solitary nests are typically established in newly dug burrows, in which the reproductive female lays eggs, forages for her brood, and guards her brood as she is able. Social and solitary nest statuses were determined during nest dissection: those containing eggs and/or larvae along with two adult females were deemed social and those with only one adult female solitary. Individual ages were determined using nest condition (clean nest walls year 1, and soiled reused nests year 2) which corresponded to adult female wing wear (0-3 first and 4-5 second brood rearing season). To assign classes within social nests, individuals were assessed for combined metrics of both body size (intertegular diameter and wing length) and ovarian development (sum of lengths of three largest ovarioles); in accordance with known biology (Sakagami and Maeta 1984, 1987, 1989) the larger and more reproductively established of the two individuals was assigned to the queen class (Sakagami and Maeta 1984). Following class determination, a total of 18 individuals (three each for old and young, totaling six queens, six

workers, and six solitary females). Bees were dissected on dry ice for brain RNA extraction using the QIAGEN RNeasy Kit and protocol (Cat # 73404). Acceptability of RNA sample quality was confirmed on an Agilent Tape Station 2200 before the set was submitted to Genome Quebec for library prep and 150 base pair (bp) paired-end (PE) Illumina HiSeq 2500 sequencing. Read data were then aligned to the *C. japonica* genome (unpub. data) before being used for further analysis (data are accessible via NCBI PRJNA 413373).

Gene expression and network analyses

Differentially expressed genes (DEGs) were determined using DESeq (Anders and Huber 2012) with corroboration by DESeq2 (Love, Huber, Anders 2014), comparing individuals by class (e.g. queens vs workers) and age (e.g. old vs young queens, **Table S4.1**; this and all remaining supplementary tables can be found at UNH DropBox: https://unh.box.com/s/4ss3kyw9e5j2vir4uvbx05qnoo38poxb). A weighted gene co-expression network analysis (WGCNA) was then run to further examine patterns of expression across all genes (Langfelder and Horvath 2008). Normalized expression data from all sampled individuals (N=18) was filtered to remove any genes that featured few or no expression values in at least one biological group. Remaining samples were then assigned corresponding biological trait data (i.e. queen, worker, solitary, young, and old) and further inspected using the default hclust distance method to produce a first-pass dendrogram (Zhang and Horvath 2005), allowing us to identify and remove outliers (N=2, Cjap2, young worker, and Cjap6, old worker, **Figure S4.4**); the remaining 16 samples were retained for further analysis. Following analysis of soft thresholding model powers we selected a power of 4 – the lowest value for which the scale free model achieved a plateau R^2 of 0.80 – for the full WGCNA analysis (**Figure S4.5**), which was then

performed as instructed by the authors' guide (Langfelder and Horvath 2014). Minimum module size was set to 30 and module merge cut height set to 20 to ensure assembly of sizeable and biologically realistic modules. Sets of significantly distinct gene co-expression modules were defined, and each individual gene was tested both for membership within each module and for association with each trait group. Each co-expression module was then tested for trait association. WGCNA assigned the overall gene set to a total of 37 statistically supported modules of co-expressed genes (**Figures S4.5, S4.6**), each positively or negatively correlated with queens, workers, solitary, old, and young females (**Figure S4.7;** See **Tables S4.8-4.13**). Ranked, hierarchical tables of most- to least-associated gene modules were exported for each trait of interest. A summary network of the three most positively and significantly correlated modules for each class was then exported, setting the edge connectivity threshold to > 0.10, to visually explore hub genes and inspect overall network connectivity among classes in Gephi (Bastian et al. 2009). The ForceAtlas2 rendering function was used in Gephi to achieve an optimal visual interpretation of the structural qualities of the gene network: specifically, ForceAtlas2 highlighted node (i.e. gene) communities and thus facilitated examination of their essential modularity (Jacomy et al. 2014). Hub genes were then defined as those genes with a module membership score (i.e. interconnectivity) greater than 0.9 and both positive gene significance (i.e. trait association) and p value < 0.05. The top five most interconnected and annotated hub genes for each trait in the network were then labeled. Gene ontology (GO) term enrichment was then determined for DEG and WGCNA module gene lists using togGO v3.7 (Alexa and Rahnenfuhrer 2016); lists of significantly enriched GO terms ($p < 0.05$) were further condensed using dispensability ratings ≤ 0.5 in Revigo (Supek et al. 2011).

Cis-regulatory analysis

Analyses of regulatory enrichment upstream of differentially expressed genes was performed using Stubb (Sinha et al. 2006) and cis-Metalysis (Ament et al. 2012) referencing the JASPAR insect and vertebrate databases to identify transcription factor binding site (TFBS) motifs. Fist, relative regulatory direction (i.e. up or down) was determined for all genes with complete expression data by calculating Z-scores across all biological conditions. Z-scores provide an accurate measure of positive or negative standard deviations away from average expression: a score near 0 indicates little difference from average expression, and positive and negative values indicate up- and down-regulation respectively. Using this logic, DEGs with Z-scores less than or equal to -0.333 were considered as down-regulated; greater than or equal to -0.333 but less than or equal to 0.333 non-differentially regulated; and those with a Z-score greater than 0.333 up-regulated in the given condition. Calculated regulatory directions among DEGs were then confirmed for consistency against results of DESeq analyses. Regulatory directional values were then taken together with promoter region sequence data drawn from 5kb windows upstream of all genes for which sequence and expression data were available (N = 8,936 genes) and fed into Stubb to identify TFBS motif enrichment. Stubb scores enrichment of both single and compound TFBS motifs across each biological condition based on both whether the motif is present in associated gene promoter regions and the z-score defined regulatory directions of each gene. Stubb outputs were then fed through cis-Metalysis, which selects the top 1% most significant portion of the results to produce a list of best-supported TFBS motif enrichment associations for up- and down-regulated genes in each condition. Given the exhaustive permutation testing of the cis-Metalysis pipeline (i.e. for all motifs A and B, tests are performed for the presence of just motif A, just B, and all logical motif pairs: A and B, A and not B, not A and B, and not A and not B) raw output lists were extensive. We therefore consolidated

these results for further analysis by removing functionally redundant outputs (e.g. while the outputs “not A and B” and “just B” are technically distinct, they are effectively both functionally reducible to “just B;” whereas the motif outputs “A and B” and “A and C” are both distinct pairs, and would thus both be kept for further analyses).

Comparative Analyses

BLASTn was used to explore homology between the *C. japonica* genome (unpub. data) and publicly available genomic and transcriptomic datasets from nine bees, four ants, two wasps, and the fruit fly using cut-offs of >70% shared ID and p-values <1.0e-5 (**Table S4.3**). OrthoFinder v2.3.2 (Emms and Kelly 2015) was then employed using default settings to compare amino acid sequence data between *C. japonica* and these same species where data were available (five bees, two ants, and one wasp) to corroborate and further assess orthology (**Table S4.7**). These analyses allowed for comparison of *C. japonica* to a total of 35 other studies ranging from variation in gene expression by reproductive development, age, and/or caste in 15 bees, 5 ants, 2 wasps, the fruit fly, mouse, and stickleback fish (**Table S4.3**). These studies were also used to compare significantly enriched GO terms and TFBS motifs where possible.

Rank-rank hypergeometric overlap analysis

Rank-rank hypergeometric overlap analysis (RRHO, Plaisier et al. 2010) was used to assess degrees of correlation in gene expression variation between queens and workers of *C. japonica* to comparable biological roles in six additional bee species, one wasp, and one ant. To prepare for this RRHO analysis, results of differential gene expression analyses comparing a queen or reproductively dominant role to a worker or non-reproductive role were compiled.

BLASTn results were then used to pair *C. japonica*'s queen vs worker results with those of each other species, pruning any non-homologous genes from each set. These lists were then used as input for RRHO analysis, which uses species-specific log2 fold change values to first rank each list of genes by trait-association, and then assess significance of gene expression rank correlation between both species. RRHO analysis concludes with the generation of a summary heat map matrix of hypergeometric p-values and an output of a list of genes of most significant correlation in expression for further inspection.

RESULTS

Read mapping and differential gene expression by age and class

Illumina 150bp PE sequencing produced an average of 39.9MB of raw sequence data for each of our 18 whole head samples (718.4 MB in total; **Table S4.2**). On average, 93% of raw sequence data for each sample passed initial quality checks for further alignment, enabling the mapping of an average of 9231 genes at 35x read coverage across all samples. Overview assessment of gene expression via principal components analysis revealed 60% of the total variation in the data was explainable by the first (3421 genes, 38%) and second components (1961 genes, 22%; **Figure S4.1**) which appeared to reflect age and reproductive status respectively.

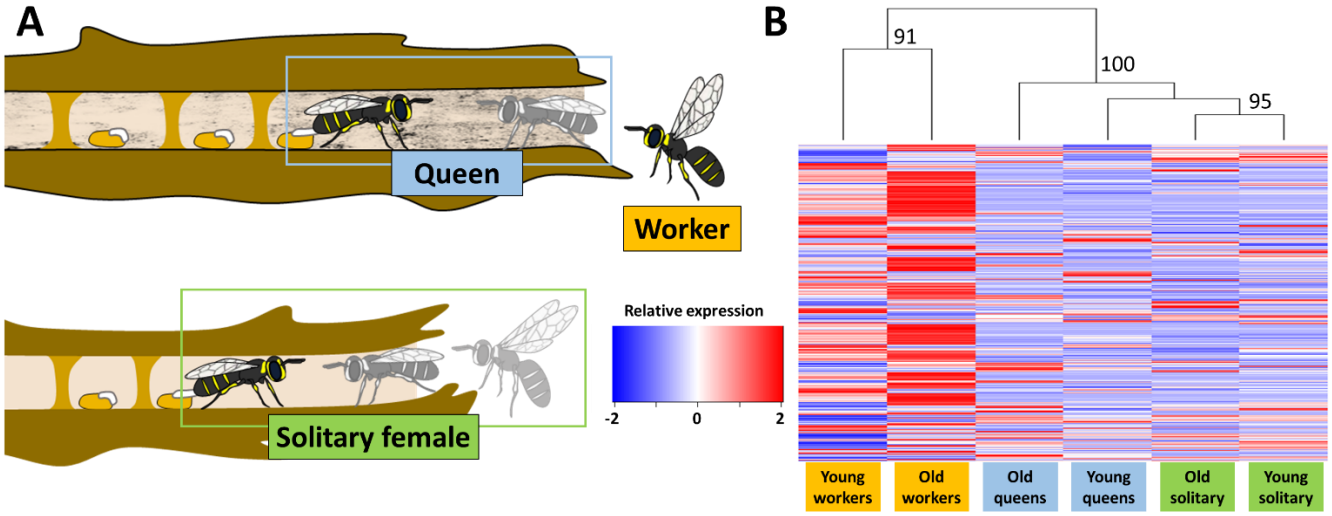

Figure 4.1. A) Summary illustrations of *Ceratina japonica* nesting biology. In social nests, the queen acts as the primary reproductive and nest guard while worker forages; in solitary nests, the lone reproductive female produces, forages for, and guards her brood. B) Heat map of significantly differentially expressed genes (FDR corrected P values< 0.05; N = 471) identified among young (first year) and old (second year) queens, workers, and solitary females (three samples per group; relative expression values are in $\log_2$fold change). Hierarchical cluster analysis reveals strong support for three focal categories: i) non-reproductive females (workers), ii) social reproductive females (queens), and iii) solitary reproductive females.

Effects of class regardless of age

Analysis of gene expression variation by class (i.e. queens vs workers vs solitary females) revealed an additional 349 significantly differentially expressed genes (DEGs, at FDR < 0.05; **Table S4.4**; **Figure 4.1, Figure 4.2**) of which 50 were unique to this analysis. Most of this set was upregulated specifically in workers (N_{unique} = 230, 69%) which also featured the greatest numbers of strongly DEGs ($N_{log2fc > 2}$ = 97, 85%) and significantly more than expected even given total DEG counts ($\chi 2$ = 30.04, d.f. = 6, p = 0.000039; **Figure S4.2; Table S4.17**). Some notable worker class-associated DEGs included *cytochromes P450 6B1* and *6k1*, *Troponin C*, and two copies of *General odorant binding protein 69a*. GO term enrichment for workers revealed a wide array of at least 27 metabolic processes (including carbohydrate, fatty acid, and hormone metabolism) along with evidence of immune function (regulation of cellular defense)

and oviduct development (**Table S4.4**). A total of 35 DEGs were uniquely upregulated in queens, including *cytochrome P450 4g15, 10 kDa heat shock protein mitochondrial,* and *Heat shock 70 kDA protein*. Enrichment among queens highlighted core biological processes and molecular functions, such as protein processing and folding. Although just five GO terms were enriched in social females, both queens and workers, this set notably included taste receptor activity and glucocorticoid binding. A total of 40 DEGs were uniquely upregulated in solitary females, including *vitellogenin*, *major royal jelly protein 2*, and *cytochrome P450 4C1*. GO term enrichment captured clear evidence of both core and metabolic processes (**Table S4.4**), including 15 terms also enriched in queens (e.g. protein deglutamylation, supramolecular fiber organization) and nine terms also enriched among workers (e.g. polysaccharide, and carbohydrate metabolic processes).

Effects of aging within class

A total of 421 DEGs were associated with age differences among queens, workers, and solitary females (**Figures 4.1, S4.3; Table S4.5**). A total of 47 DEGs separated young (N = 18) from old queens (N = 29); old queens notably featured upregulation of five copies of *protein lethal(2)essential for life*, three *cytochrome P450*s, *major royal jelly protein2*, and both *heat shock 70 kDa protein* and *oxidoreductase YrbE*. GO term enrichment revealed that queens are enriched for immune processes and responses to biotic and chemical stimuli when young, but were enriched for oxidation reduction and integrin-mediated cell adhesion, when old. Workers featured the greatest differences by age of any class, with a relatively large set of 89 DEGs splitting young (N = 24) from old (N = 65). As workers age, upregulated gene count expanded to include an array of notable DEGs including, *general odorant binding protein 69a*, two

*Cytochrome P450*s, *Troponin C*, and five copies of *myosin heavy chain muscle*. Worker functional profile also shifted dramatically with age, from positive regulation of gene expression when young, towards evidently high locomotor activity when old (e.g. flight and locomotion) underpinned by metabolic (e.g. lipid metabolism) and immune-associated processes (e.g. hemocyte proliferation). Regardless of age, workers were enriched for structural constituent of muscle, taste receptor activity, and activity of nutrient reservoirs and oxidoreductase (**Table S4.5**). Solitary females featured just 28 DEGs associated with changes in age, dropping from 17 genes upregulated when young to just eleven when old. Functional enrichment revealed that while both young and old solitary females are underpinned by suites of metabolic processes (including of galatolipid, icosanoid, and unsaturated fatty acid), overall functional profile shifts towards increased immune responsiveness, protein methylation, and oxidoreductase activity with age (**Table S4.5**).

Effects of aging among classes

Examining shared age effects among classes, queens and workers featured more significantly upregulated genes in common than either did with solitary females, both among young (N = 5, 10%) and old individuals (N = 10, 11%); notable DEGs included *putative gustatory receptor 23a* among young workers and queens, and both *cytochrome p450 6A1* and *major royal jelly protein 2* among old (**Table S4.5**). At the functional level, young queens and workers were enriched for 11 GO terms in common (8%), including taste receptor activity, positive regulation of DNA binding, and detection of chemical stimulus. By comparison, old queens and workers shared just three enriched terms in common (2%) which represents a significant reduction in shared terms with age ($\chi 2 = 9.9159$, d.f. = 4, $p = 0.0419$; **Table S4.17**).

Queens and solitary females shared just one DEG and three enriched GO terms at each age stage, notably including upregulation of *Cytochrome P450 4C1* and enrichment for oxidation reduction processes among old females. Just one gene in the deeply conserved Band 7 protein family was upregulated across all classes when young, and one unannotated gene, Cjapo_01990, was consistently upregulated when old. Functional enrichment also revealed shared activity of oxidoreductase among all classes when old (**Table S4.5**).

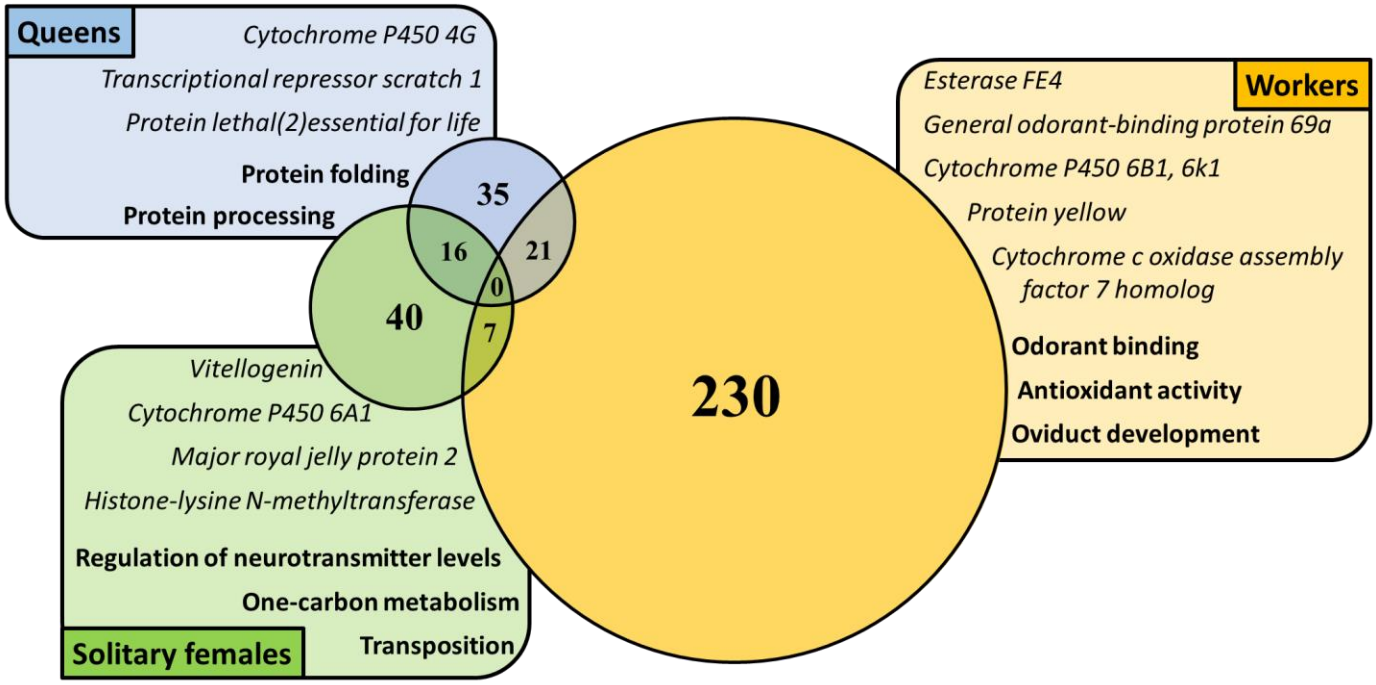

Figure 4.2. Summary Venn-diagram of unique significantly upregulated genes and enriched GO terms identified between queens (blue), workers (yellow), and solitary females (green) regardless of age (circle size and spans of overlap are relative); a selection of notable genes (in italics) and enriched GO terms (bold) are provided for each class. Workers featured nearly six times as many uniquely upregulated genes as queens or solitary females. For full lists of DEGs and GO terms see electronic supplementary material (**Table S4.4**).

Network analysis

All 8936 genes and 16 of our 18 samples passed filtering checks for measures of expression among samples and for use in weighted gene coexpression network analysis (WGCNA). Network topographic rendering revealed that genes underlying queens and solitary females cluster tightly by trait, and are separate from but well interconnected with each other (**Figure 4.3**). By comparison, while some worker-associated gene modules (especially 'salmon') were distinctly disconnected from the core network, others (including the most positively worker trait-associated module, dark red, correlation= 0.75, p = 2.7e-17) were interconnected with mostly solitary female-associated genes (**Figs. 4.3, S4.8**). Notable hub genes from across the top three queen-associated modules (black, pink, and dark green; **Figs. S4.9-4.11**) included *transcription factor GAGA, calcyclin-binding protein*, *cysteine and histidine-rich domain-containing protein* and *97 kDa heat shock protein* (**Tables S4.8, S4.13**). GO term enrichment from these sets indicates that regulation of hormone levels, chromatin organization, and oxidoreductase activity are among highly queen-associated processes. Solitary female-associated modules (midnight blue, light green, and pale turquoise; **Figs. S4.12-4.14**) included *dopamine D2-like receptor*, *broad-complex core protein isoforma 1-5*, and *neurogenic locus notch homolog protein 1* as hub genes (**Tables S4.10, S4.13**). Core solitary female-associated processes include sexual reproduction, oxidation-reduction, and polysaccharide metabolic process. Hub genes identified within worker-associated modules (dark red, white, and salmon; **Figs. S4.15-4.17**) included *cytochrome b561*, *esterase FE4*, and *x-linked retinitis pigmentosa GTPase regulator* (**Tables S4.9, S4.13**). Worker hub genes also included *catalase*, which was notably located roughly in the center of the overall WGCN. Strongly worker-associated GO enrichment suggested a role for many immune (e.g. Toll signaling pathway), metabolic (e.g. oligosaccharide metabolism), and both muscular and neuronal activity (**Table S4.13**).

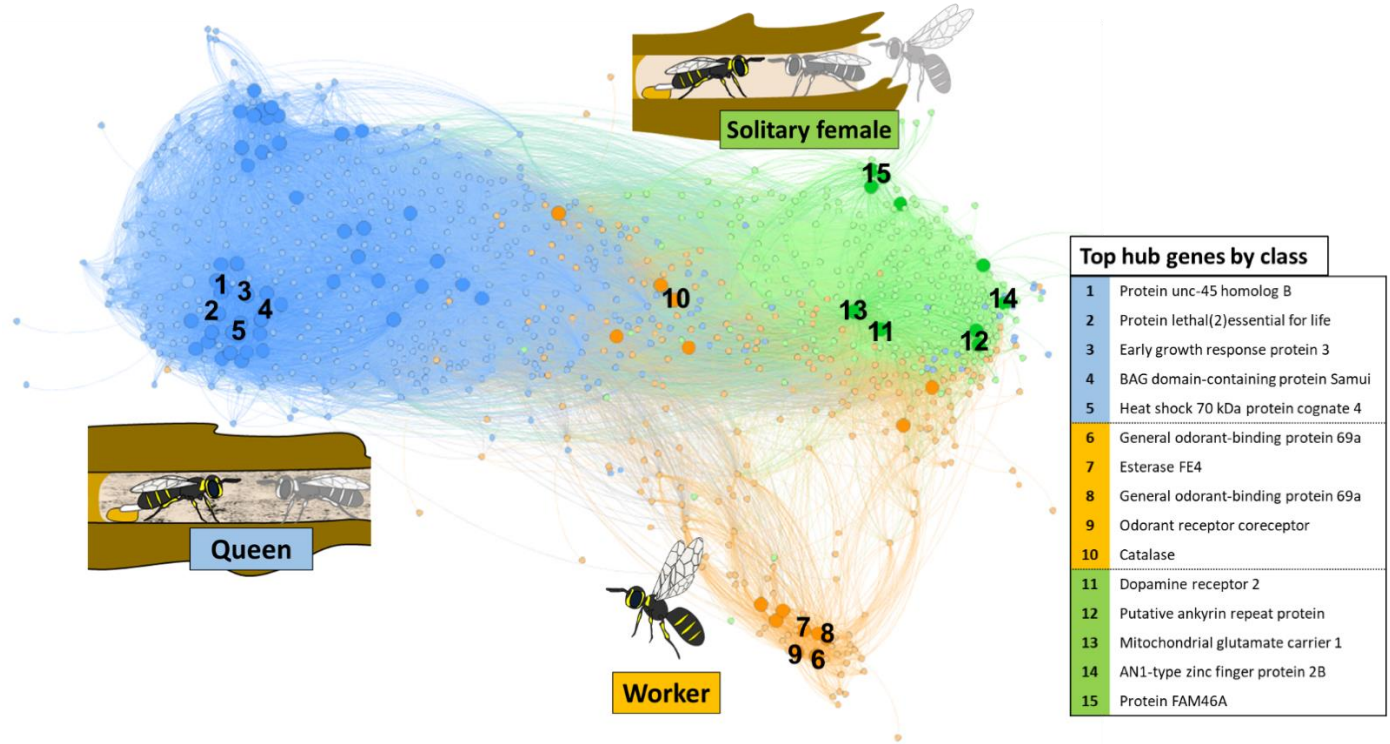

Figure 4.3. Weighted gene co-expression network rendering only the top three most significantly and positively correlated gene modules for each class for summary visualization (Queens, blue; solitary females, green; workers, orange; $N_{totalNodes}$ = 903 genes, $N_{totalEdges}$ = 40,174; for full module details see supplementary **Table S4.13**); top five annotated hub genes by class are indicated by number and named in legend. Queen- and solitary female-associated genes occupy generally distinct regions of an otherwise well-interconnected overall network; although some worker-associated genes are interconnected more tightly with those of solitary females, a considerable number cluster tightly on the periphery of the network, suggesting class-specific expression patterns. For full WGCNA results across all 37 modules see electronic supplementary materials (**Tables S4.8-S4.13**).

Cis-regulatory enrichment

Transcription factor (TF) binding site enrichment by role identified a total of 429 significantly enriched and functionally unique up- or down-regulating TFs (**Tables S4.15, S4.16**). Workers featured many more upregulating TFs than solitary females or queens ($N_{workers}$ = 223, vs N_{queens} = 30, $N_{solitary}$ = 90). By contrast, queens featured significantly fewer and solitary females significantly greater TF counts than expected given respective DEG counts ($\chi 2 = 21.34$, d.f. = 2, p = 0.00002; **Table S4.17**). Considering enrichment only for those TFs with known

neural, immune, and reproduction regulatory associations, workers featured significantly greater neural-associated TFs ($\chi 2 = 6.85$, d.f. = 2, $p = 0.0325$), but significantly fewer reproduction-associated TFs than expected given total TF count ($\chi 2 = 11.12$, d.f. = 2, $p = 0.00385$). By contrast, although solitary females and queens featured significantly fewer neural-associated TFs, solitary females were enriched for significantly more reproduction-associated TFs than expected (**Table S4.17**). Counts of immune-associated TFs varied by role but were not significantly different than expected given respective total TF counts ($\chi 2 = 0.157$, d.f. = 2, $p = 0.924$). Notably, a few TFs expected to bind to significantly enriched site motifs were themselves differentially expressed (i.e. BH1, eve, RREB1) or identified during WGCNA analysis (**Figure 4.4**).

Comparative analyses of gene expression

Rank-rank hypergeometric overlap (RRHO) analyses of gene expression variation between queens and workers of *C. japonica* and comparable roles in eight additional hymenopteran species collectively identified a total of 3328 genes as strongly and significantly correlated in all comparisons except vs *Temnothorax longispinosus* ($p = 0.7$; **Table S4.14**, **Figure 4.5**). The strongest correlations were detected between *C. japonica* and other *Ceratina* species (i.e. *C. calcarata* and *C. australensis*), followed by those of biologically comparable phenotype (i.e. primitive eusociality; *M. genalis* and *E. robusta*), followed by advanced eusocial Hymenoptera. Among primitively eusocial taxa, strongly correlated genes included the *broad complex* protein family, *dopamine receptor 1* and six *zinc finger* proteins among queens and *glutamate decarboxylase* among workers. Overall, RRHO analysis detected much greater and stronger conservation of gene expression patterns among queens and reproductive dominants in

primitively social bees and more orthologous genes expression among workers or subordinates in advanced eusocial Hymenoptera.

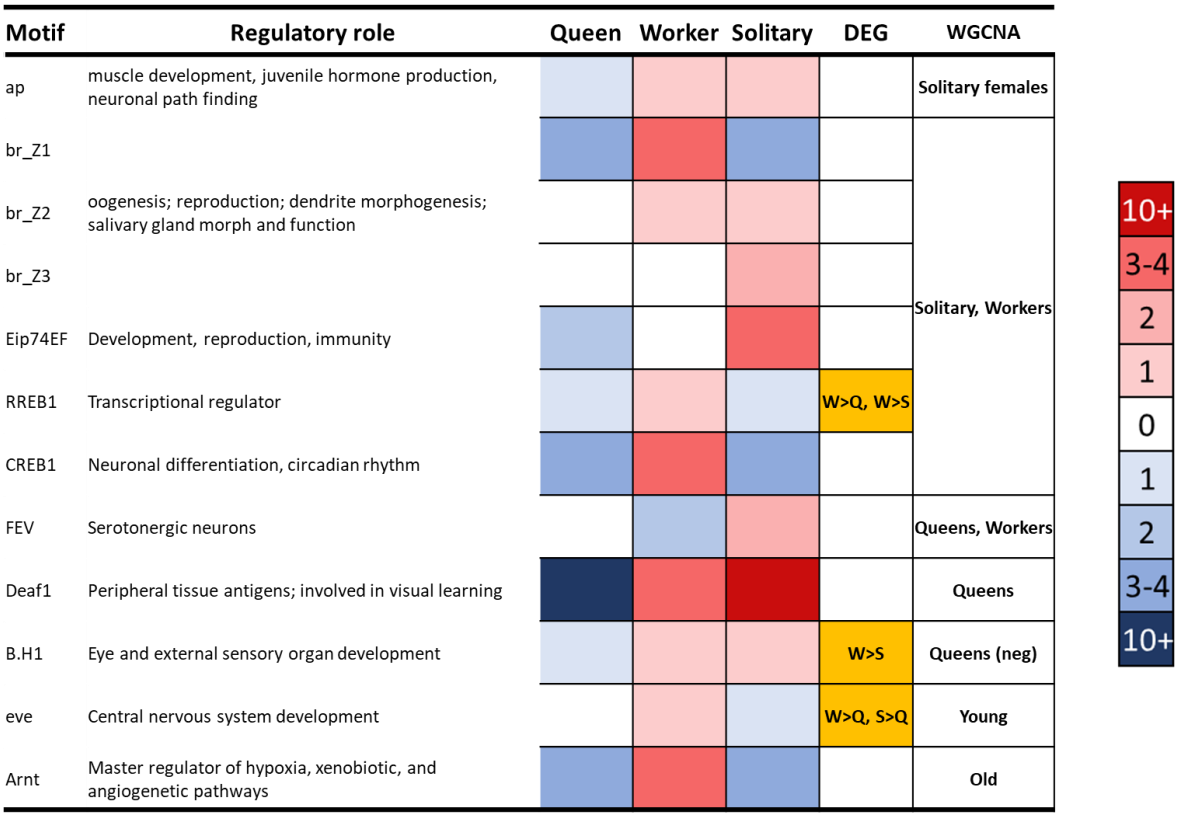

Figure 4.4. Heat map highlighting TFBS motifs with known neural, immune, or reproductive regulatory roles, significantly enriched in the promoter regions of genes associated with queens, workers, or solitary females (for full list, see **Table S4.16**). Motif names are presented in order of phenotypic association (WGCNA, **Tables S4.8-S4.13**), followed by a summary description of regulatory role, their enrichment and regulatory status among queens, workers, and solitary females, and their status and context as differentially expressed in this study (yellow, **Table S4.4, S4.5**). The rightmost column contains phenotype-association and module information for all motifs from WGCNA. Where queens are generally down-regulated and workers generally up-regulated, and the regulatory enrichment and direction of solitary females falls somewhere in between.

Notably, genes identified during RRHO analysis included those that were significantly differentially expressed in biologically comparable contexts in this and other studies (**Figure 4.6**; **Table S4.6**). For example, genes upregulated in *C. japonica* queens included those upregulated

across social Hymenoptera, such *TBC1 domain family member 1*, previously associated with queens or social reproductive dominants (*C. calcarata*, Shell and Rehan 2019; *A. mellifera*, Grozinger et al. 2007; *T. longispinosus*, Feldmeyer et al. 2014). Two other queen-associated genes, *muscle LIM protein Mlp84B* and the long form of *paramyosin*, were upregulated in mated honey bee queens (Manfredini et al. 2015) and guarding honey bees, mice, and sticklebacks (Rittschof et al. 2014). Genes upregulated in *C. japonica* workers were also well-conserved across taxa. For example, *cadherin-89D*, was found to be significantly upregulated among workers or social reproductive subordinates (e.g. *C. calcarata*, Shell and Rehan 2019; *M. genalis*, Jones et al. 2017; *E. dilemma*, Saleh and Ramirez 2019; and *Polistes metricus*, Berens et al. 2014) and *GTP cyclohydrolase 1* was upregulated in the forager or reproductive subordinate role in ants (*T. longispinosus*, Kohlmeier et al. 2019) and bees (e.g. *C. australensis*, Rehan et al. 2018; and *C. calcarata* Shell and Rehan 2019).

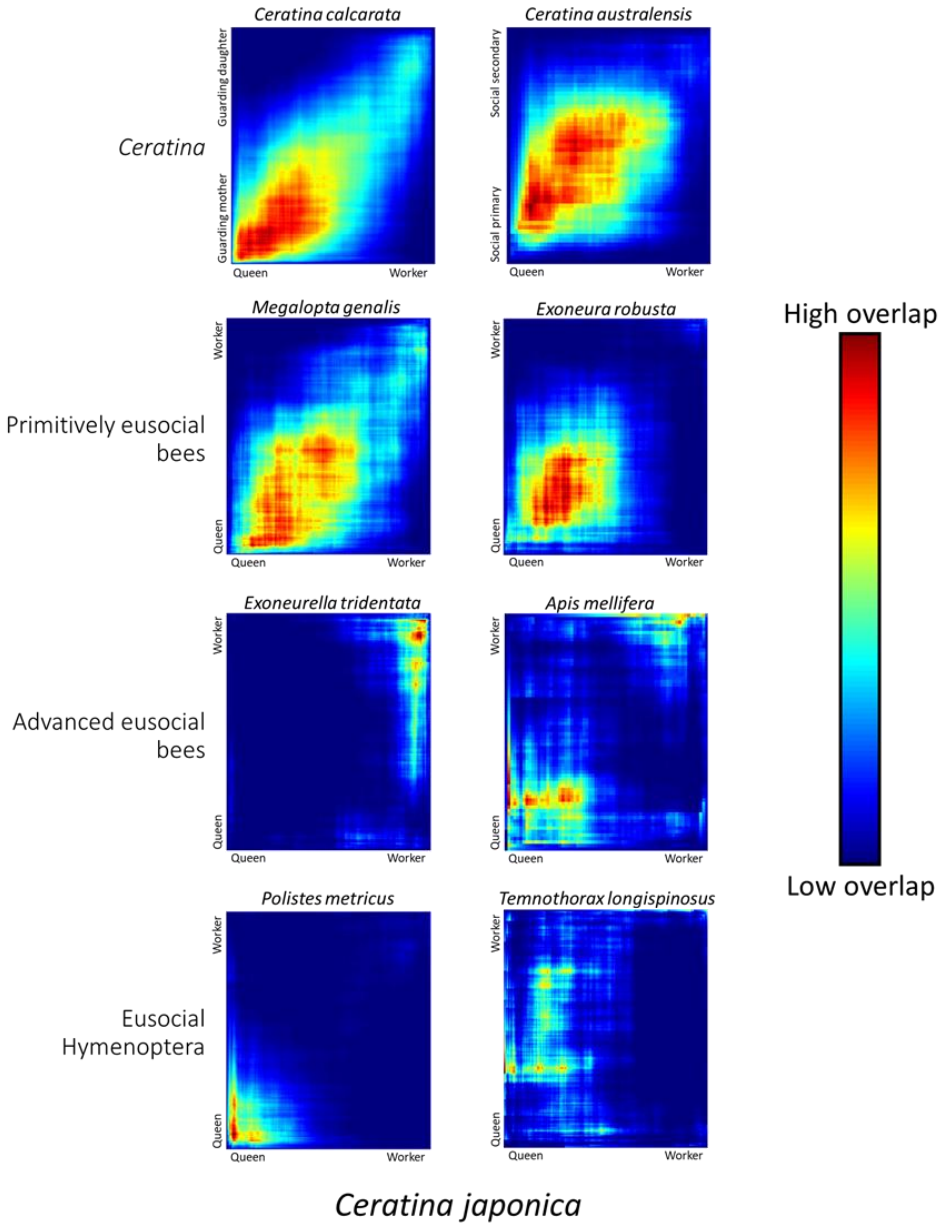

Figure 4.5. Rank-rank hypergeometric overlap (RRHO) plots portraying correlations between genes ranked by association with queens through workers in *C. japonica* (all x-axes) to homologous genes associated with comparable roles in eight additional hymenopteran species (y-axes). Overall, all comparisons are highly significant except vs *Temnothorax longispinosus* ($p = 0.7$; **Table S4.14**). The most positively and significantly correlated comparisons were between *C. japonica* and i) other ceratinine species, despite variation in complexity of social phenotype; and ii) other primitively eusocial bees, *Megalopta genalis* and *Exoneura robusta*, despite phylogenetic distance among lineages. Although correlations are also significant among advanced eusocial bees and wasps, range of overlap becomes much more localized by queen or worker phenotype at this level.

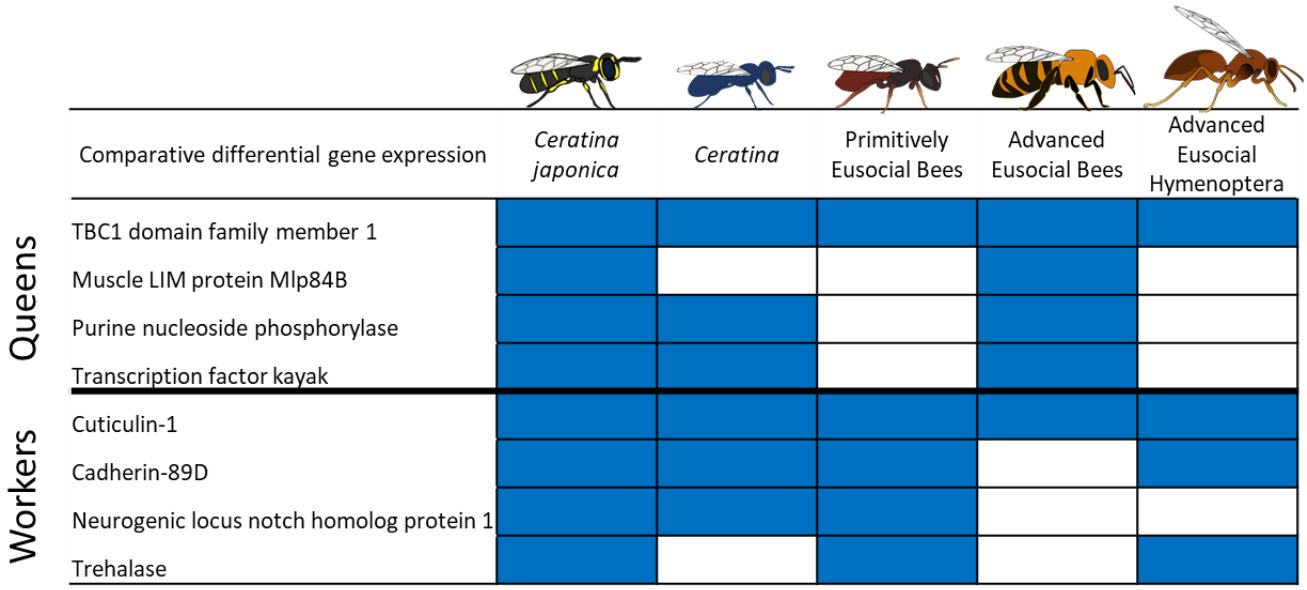

Figure 4.6. A representative subset of all genes associated with queen or worker phenotypes in *C. japonica* which were identified during both orthofinder and RRHO analyses and matched with strong statistical support to genes in species of *Ceratina* (i.e. *C. calcarata* and/or *C. australensis*), primitively eusocial bees (e.g. *E. robusta*, *M. genalis*, or *E. dilemma*), advanced eusocial bees (e.g. *A. mellifera*), and/or advanced eusocial Hymenoptera (e.g. *Polistes metricus*, *Temnothorax longispinosus*). Blue boxes indicate shared genes and similar regulatory contexts between *C. japonica* and at least one species from the indicated category; white boxes indicate a lack of contextual or regulatory overlap. Overall, genes associated with queen and worker phenotypes in *C. japonica* are very well conserved among similar roles in other ceratinine taxa, both primitively and advanced eusocial bee lineages, and to a lesser extent among other advanced eusocial Hymenoptera (**Tables S4.6**, **S4.7**, **S4.14**).

DISCUSSION

Here we provide insights into the molecular dynamics underlying both aging and caste-antecedent behavioral classes of *Ceratina japonica*, a long-lived species of small carpenter bee capable of eusocial nesting. We uncover considerable genetic variation among queens, workers, and solitary females, and highlight how regulatory enrichment and a highly modular co-expression networks may play especially important roles in *C. japonica*'s social phenotypes. We also consider how differentially expressed genes underlying *C. japonica*'s form of eusociality correlate significantly with those observed across other social taxa, providing empirical support

for the genetic toolkit hypothesis (Berens et al. 2015; Rehan and Toth 2015). Finally, we examine how gene expression patterns shift over *C. japonica*'s two-year lifespan; and identify a potential proximate mechanism for this relatively prolonged longevity in a small carpenter bee.

Social effects

Workers are highly distinct from queens and solitary females

Analyses revealed clear distinctions among our three phenotypic classes, the most notable being that workers are highly distinct from queens or solitary females. Workers expressed almost six times as many differentially expressed genes as queens or solitary females. Workers also featured extended enrichment for neural-associated TFs, some of which are themselves differentially expressed in workers (e.g. *homeobox protein Barh1*). Other worker enriched TFs, such as *apterous*, *CTCF*, and *Lim3*, have been previously identified as likely directing worker-like phenotypes including foraging across other incipiently social (e.g. *C. australensis*, Rehan et al. 2018; *C. calcarata*, Shell and Rehan 2019) and eusocial bees (e.g. *A. mellifera*, Khamis et al. 2015; *E. robusta*, *Exoneurella tridentata*, Shell et al. *in revision*). Worker uniqueness was further corroborated by our weighted gene co-expression network (WGCN) analysis, which reveals that many key worker-associated genes lie on the periphery of the overall network. This disconnect indicates that genes underpinning the worker phenotype are distinct from solitary female or queen-associated pathways (Zhang and Horvath 2005; Langfelder and Horvath 2008). Previous network analyses among social insects have shown that caste phenotypes can often be tied to highly modular sets of co-expressed genes (Chandrasekaran et al. 2011; Shell and Rehan 2019) and have suggested that genes on the periphery of a regulatory network are likely targets of strong directional selection (Molodtsova et al. 2014). The apparent expansion of worker-associated TFs and DEGs in *C. japonica* echoes evidence of positive selection or regulatory

expansion seen in worker phenotypes across both ant and bee lineages, collectively representing emergent through derived social phenotypes (Harpur et al. 2014; Feldmeyer et al. 2014; Jones et al. 2017; Shell and Rehan 2019). As such, this work supports the hypothesis that ancestral pleiotropic constraints are more likely to loosen in the non-reproductive role of a lineage experiencing evolutionary gains in social complexity (Gadagkar 1997).

Notably, three genes – *odorant receptor coreceptor* and two copies of *general odorant binding protein 69a* (*Obp69a*) – were identified as both highly expressed worker-associated DEGs and functionally important worker hub genes on the network periphery. Odorant binding proteins (OBPs) and receptors (ORs) underpin insect olfactory behavior and pheromone binding (Fan et al. 2011), and *Obp69a* has been shown to help moderate social interactions among fruit flies (Bentzur et al. 2018). The role of ORs and OBPs in chemical communication is relatively well-appreciated among honeybees (*Apis mellifera*, Forêt and Maleszka 2006; Iovinella et al. 2011) and other advanced eusocial Hymenoptera (i.e. ants, McKenzie et al. 2016; wasps, Jandt et al. 2014), as they likely contribute to nestmate recognition and the direction of caste behavior and physiology (Dani 2009; Shah and Renthal 2020). OBPs have also been found to play an important role in other primitively eusocial bees (e.g. *Bombus terrestris*, Colgan et al. 2011) and still-flexible reproductive dominance hierarchies of incipiently social small carpenter bees (Rehan et al. 2014, 2018). Taken together, while their exact physiological function within *C. japonica* remains a question for future research, the evident involvement of ORs and OBPs in *C. japonica*'s worker class highlights chemical communication as of potential importance for the behavior and social dynamics of this species, and among bees of early eusociality generally (Woodard et al. 2011; Wittwer et al. 2017).

Social phenotypes have divergent gene regulatory and co-expression pathways

Regulatory enrichment suggests that expression pathways underlying *C. japonica*'s queen and worker phenotypes have opposing directional expression and are class specific. For example, while almost half of all TFBS motifs enriched in queens or workers (N = 128 of 297; 41%) upregulated workers while downregulating queens, just four TFBS motifs were identified as upregulation both roles (1.3%). Notably, extensive downregulation in queens is directly in line with behavioral observations by Sakagami and colleagues (1993) who wrote that the queen phenotype was at least partly "characterized by two *negative* key tasks – continuous resting and abandonment of foraging." Further, although *C. japonica* demonstrates a form of facultative eusociality, these stark variations by class are more typical of those seen among advanced eusocial hymenopteran lineages (e.g. ants, Feldmeyer et al. 2014; *A. mellifera*, Grozinger et al. 2007). By comparison, solitary female regulatory enrichment and directionality overlaps with that of both queens *and* workers. For example, solitary females are enriched for 125 binding sites accommodating down-regulating TFs also found in queens, and another 28 binding sites accommodating upregulating TFs also found in workers. WGCN analysis partially corroborates this apparent phenotypic directionality, while highlighting the divergence of social from solitary phenotypes. Summary network modeling suggests that while queen and worker associated modules are partly isolated from each other, both feature greater interconnection to each other (N = 3633) than either does to solitary females ($N_{ToWorkers}$ = 2781; $N_{ToQueens}$ = 2186; **Figure 4.3**). In sum, while the gene regulatory and co-expressive pathways underpinning queen and worker phenotypes reveal clear functional distinctions between roles, both remain at least partially interconnected, and associated with those of solitary females. Evolutionary-developmental hypotheses regarding the emergence of social traits propose that eusocial nesting emerged from

the gradual co-option of key genes underlying ancestrally maternal behaviors and reproductive physiology (West-Eberhard 1996; Toth and Robinson 2007; Rehan and Toth 2015; Shell and Rehan 2018). This study provides empirical support for this prediction by qualifying a transcriptomic roadmap between the expression of ancestral solitary reproduction and foraging, and the social nest roles of reproductive queen and foraging worker.

Deeply conserved and differentially expressed genes underlie primitive eusociality

Despite substantial phylogenetic divergence and an independent origin of eusociality, *C. japonica*'s queen and worker classes demonstrate many traits that are phenotypically consistent with those of other primitively eusocial bees (e.g. *M. genalis*, Wcislo and Gonzalez 2006; *Exoneura robusta*, Schwarz 1986, 1987) including (though not limited to) division of reproductive labor by age and body size, a reproductively viable but dedicated foraging class, and regular trophallaxis among nestmates. The social ladder framework predicts that the earlier stages of social evolution are rooted in the differential expression of otherwise deeply conserved genes, with sustained selective pressure gradually driving change at the level of proteins and taxonomically restricted genes (Rehan and Toth 2015). Comparative RRHO analysis supports this prediction, revealing that genes associated with queen and worker phenotypes in *C. japonica* are conserved across other *Ceratina*, among primitively eusocial bees and – to a lesser but still significant extent – among advanced eusocial bees and wasps. Both *C. calcarata* and *C. australensis* regularly demonstrate incipiently social phenotypes (Withee and Rehan 2017; Steffen and Rehan 2020) and, as relatively close cousins of *C. japonica*, share many of the life history traits consistent across the small carpenter bee genus (e.g. stem nesting; Sakagami and Maeta 1977). As such, observed correlations may be explained as much by broad similarities in

social and ecological phenotypes as they are by phylogenetic proximity among ceratinine lineages (Rehan et al. 2010). By comparison, *C. japonica* is a distant cousin of primitively eusocial *E. robusta* (Xylocopinae; Rehan et al. 2012) and of even further remove from *M. genalis* (Halictinae, Cardinal and Danforth 2011). Strongly positive and significant correlations in gene expression detected among these lineages and two independent origins of sociality is thus better explained by shared phenotypic traits than by shared phylogenetic ancestry. Correlations between *C. japonica* and advanced eusocial taxa rapidly thin out, localizing to just those genes that are either strongly queen- or worker-associated. Our results thus offer empirical support for the predictions of the social ladder framework: deeply conserved and differentially expressed genes appear to play a relatively major and consistent role during the evolution of early social traits across lineages; however, subsequent changes past the evolutionary "point of no return" into advanced eusociality (Wilson and Hölldobler 2005) appear more lineage-specific, increasingly involving taxonomically restricted genes (Toth and Rehan 2015).

Aging Effects

Gene expression and regulation by age

There were considerable shifts in gene expression and regulatory enrichment by age across classes. Of any phenotype, workers evidently undergo the most dramatic changes with age: upregulating motif enrichment increases by over four-fold, and functional enrichment expands from terms associated with metabolism and immunity to those of locomotion, flight, and odorant binding. By comparison, and despite some regulatory shifts of their own, queens and solitary females largely maintain signals of metabolism and immunity with age. Dramatic age-associated changes in worker regulatory and molecular profile, including functional shifts towards elevated activity and odorant binding, have also been detected among other eusocial bees (Whitfield et al.

2003; Colgan et al. 2011). For example, *A. mellifera* workers experience changes in brain gene expression over their lifetime, precipitating a major shift from intranidal nurse to extranidal foraging behaviors (Whitfield et al. 2003). Previous work among incipiently social (e.g. *C. calcarata*, Shell and Rehan 2019) and advanced eusocial Hymenoptera (e.g. *Harpegnathos saltator*, Gospocic et al. 2017), indicates that the social environment may induce highly novel neurogenomic states (Cardoso et al. 2015). Although second year *C. japonica* workers and queens have likely experienced a year or more in a social nest environment, dramatic brain gene expression changes are primarily seen in workers, as seen in honey bee and ant workers (Whitfield et al. 2003; Morandin et al. 2015).

Despite the relatively unique trajectory with aging seen in workers (**Fig 4.1B**), there were notable consistencies among classes as well. Most notably, social queens and workers shared the greatest number of age associated DEGs, both among young and old individuals (**Table S4.5**). Immune-associated genes, such as *cytochrome P450 6A1* and *major royal jelly protein 2* are both upregulated among older workers and queens, suggesting that a measure of expanded immunocompetency is required among individuals living in a social nest environment. Complexity in immune responsiveness is well known among other eusocial bees, especially *A. mellifera*, in which extranidal foragers (Vannette et al. 2015) and their colony must persistently resist pathogenic and oxidative damages (Evans et al. 2006; Wilson-Rich et al. 2008). Detection in *C. japonica* of an increase in immune-associated gene expression with senescence in both social nest classes, however, is atypical (Doums et al. 2002; Schmid et al. 2008; Moret and Schmid-Hempel 2009). Ultimately, the influence of the social nest environment may best explain why more DEGs are shared between queens and workers than with solitary females, despite

dissimilar physiological and behavioral traits (i.e. reproduction vs foraging; Shell and Rehan 2019; Cardoso et al. 2015).

Molecular signatures of longevity in C. japonica

Despite distinct life histories, queens, workers, and solitary females of *C. japonica* are all equally capable of living up to two years or more (Sakagami and Maeta 1989). Examination of aging in *C. japonica* thus provides a unique opportunity to identify potential molecular components underlying this insect's longevity. Very notably, just one GO term, oxidoreductase activity, was enriched among older individuals of each class (**Table S4.5**). This was subsequently joined by additional class-specific enrichment for various redox-associated genes and processes. As a byproduct of aerobic metabolism, an ability to mitigate oxidative damage has been consistently and independently implicated as critical to the longevity of many taxa, from bivalves (Ungvari et al. 2011) to mammals and birds (Munshi-South and Wilkinson 2010), and among social insects (Hsu and Hsieh 2014; Negroni et al. 2019; Li-Byarlay and Cleare 2020). Studies among advanced eusocial bees have shown that ageing workers may offset expected oxidative damages through increases in expression of redox-related proteins, such as *vitellogenin* (Seehuus et al. 2006) or *catalase* (Hsu and Hsieh 2014). Notably, *catalase* and four copies of *esterase fe4* were found to be among the most highly interconnected worker module genes and strongly worker associated DEGs. Like *catalase*, *esterase fe4* plays an essential role in mitigating damage by oxidation or environmental stress such as insecticides across many insects, including social bees (*A. mellifera*, Dussaubat et al. 2016; and *A. cerana*, Ma et al. 2018). Any increased environmental and metabolic costs associated with continuous foraging by *C. japonica*'s worker class may thus be counterbalanced by greatly intensified redox activity.

Additionally, among other primitive eusocial bees, such as *B. terrestris*, worker longevity is positively correlated with reproductive opportunity (Lockett et al. 2016). The observed longevity across *C. japonica*'s classes may thus be partly explained by the fact that – despite policing by queens – workers remain reproductively viable and routinely attempt to lay eggs in their nests (Sakagami and Maeta 1984, 1987). Overall, transcriptomic variation in *C. japonica* thus suggests additional support for the proposition that individual longevity among social insect lineages may be tied to ancestrally reproductive pathways (Amdam et al. 2004) and oxidation-reduction activity (Harman 1992; Li- Byarlay and Cleare 2020).

CONCLUSIONS

Phenotypic plasticity is a multifaceted and highly dynamic phenomenon, particularly within the context of social evolution. The field has rapidly advanced in its understanding of the molecular dynamics underlying highly elaborated insect societies by studying advanced eusocial species with rigidly defined social castes; however, studying these advanced species does not allow us to directly explore the evolutionary origins of division of labor. Bees of early stage sociality, on the other hand, with less defined castes that are shaped by a wide range of social, biotic, and abiotic factors, provide a unique opportunity to test major social evolutionary hypotheses. However, few studies have explored the genetic changes underlying sociality in these species. In this dissertation, I address this knowledge gap by applying integrative research methods to species of early stage and facultative sociality. My focus on these species will advance understanding of the ecological and molecular factors driving the evolution of behavioral plasticity and social organization in insects. Below, I summarize the major conclusions from each chapter and provide future directions.

COSTS AND BENEFITS OF SOCIAL NESTING

In Chapter 2, I combined relatedness and demography data to compare the average inclusive fitness of subsocial mothers, social mothers, and worker daughters in the small carpenter bee, *Ceratina calcarata*. *Ceratina calcarata*'s social and subsocial nesting polyphenism has gone previously unreported across its range (Rau 1928; Johnson 1988; Rehan and Richards 2010a; Lawson et al. 2016; Lewis and Richards 2017); though this could be attributable to the subtle differences between nest types. This sympatric variation in social phenotype thus provides the perfect opportunity to examine the overall costs and benefits of a

very early form of social nesting, and to inspect for evidence of inclusive fitness advantages hypothesized to accompany the emergent worker daughter role (Hamilton 1964). As revealed in Chapter 2, both social and subsocial mothers produce similar clutch sizes, and relatively infrequent brood losses to parasitism do not appear to affect one nesting type more than the other. Further, contrary to the expectations of kin selection theory, relatedness and demography data suggests that worker daughters may not actually receive much of an inclusive fitness return for their helping behavior. It therefore appears that, while the emergence of social traits likely does require high relatedness and inclusive fitness benefits overall, selection on physiological and behavioral traits that maximize maternal rather than worker daughter fitness may represent a more recurrent proximate mechanism for evolutionary transitions towards early social organization.

TRANSCRIPTOMICS OF INCIPIENT SOCIALITY

As expanded in Chapter 3, *Ceratina calcarata*'s social autumn nest stage provides a prime natural experiment to disentangle the effects of age from behavioral state on gene expression and regulation in a species of incipient sociality. In this chapter, I analyzed brain transcriptomic data which captured both a maturational time course and clearly defined behavioral states observed in social nest mothers and worker daughters during the Autumn nest stage. Results from Chapter 3 suggest that *C. calcarata*'s behavioral plasticity may be underpinned by conserved genes differentially expressed within a highly modular network. Further, I found evidence that even a transiently incipiently social nest environment may have major influence on patterns of gene regulation and expression. Overall, these results lend important empirical support to the functional role of multiple mechanisms theorized to contribute

to the evolution of social complexity (toolkit genes (Toth and Robinson 2007) and the social ladder hypothesis (Rehan and Toth 2015)). Moreover, *C. calcarata*'s highly dynamic and socially-responsive gene regulatory network speaks to the pressing need for similar studies in additional incipiently and facultatively social taxa as outlined in Chapter 1 (Shell and Rehan 2018).

MOLECULAR PATHWAYS OF CASTE AND LONGEVITY

In Chapter 4, I examined brain transcriptomic data from a long-lived small carpenter bee capable of eusocial nesting (*Ceratina japonica*, Sakagami and Maeta 1984). The facultative eusociality and long lifespan of *C. japonica* offered an unprecedented opportunity to simultaneously explore the molecular mechanisms underpinning behavioral castes and insect longevity. My Chapter 4 results indicate that the transcriptomic profile of *C. japonica*'s workers is highly distinct from those of queens and solitary females and may be undergoing a considerable measure of genetic release (Gadagkar 1997). I also discover that division of labour in *C. japonica* may rely on chemical communication and distinct gene regulatory and co-expression pathways separate queens and workers, despite a shared genome and reproductive potential among all nestmates (Sakagami and Maeta 1984). Despite great phylogenetic distance, differentially expressed genes associated with the queen and worker classes in *C. japonica* are deeply conserved and remarkably similar in expression across taxa representative of independent origins of primitively and advanced eusociality. Further, longevity in social and solitary *C. japonica* females appears to be tied to a capacity to mitigate oxidative damage; though social nesting females may undergo shared changes with age that are unique to the social environment. Chapter 4 thus provides empirical support for a highly modular gene co-expression network

(West-Eberhard 2003; Molodtsova et al. 2014). This Chapter also highlights changes in behavior and age-associated genotypes and phenotypes as a response to the strong and persistent influences of social organization itself (Rubenstein et al. 2019); and, finally, offers additional evidence that oxidation reduction activity may play an important role in insect longevity, regardless of solitary or social phenotypes (Harman 1992; Li-Byarlay and Cleare 2020).

FUTURE DIRECTIONS

The phenomenon of eusociality is perhaps one of nature's greatest and most bizarre achievements: a puzzling and fascinating "evolutionary cul-de-sac" (Gadagkar 1991) that manages to be at once ecologically dominant and disarmingly fragile. And, while the 'routes' from solitary life towards the "point of no return" (Wilson and Hölldobler 2005) may take generations to fully map, collective research efforts are revealing or confirming major patterns in social evolution. For example, rather than a punctuated equilibrium, social evolution appears by and large to be a process of gradual and indeterminate gains (or losses) of social traits (Rehan and Toth 2015). In this light, the emergence of early sociality in no way necessitates further elaboration for a given lineage. That said, in whatever lineage they may emerge, social phenotypes themselves evidently appear to influence subsequent genotypic change (Toth and Rehan 2017; Rubenstein et al. 2019). Future research is needed within this puzzling evolutionary paradigm. Comparative studies which sample widely across solitary and facultatively social species will help clarify if there are consistent ecological and molecular factors accompanying phenotypic divergence in the earliest stages of social evolution.

APPENDIX A

DEVELOPMENT OF MULTIPLE POLYMORPHIC MICROSATELLITE MARKERS FOR THE SMALL CARPENTER BEE, *CERATINA CALCARATA*, USING GENOME-WIDE ANALYSIS

[Shell WA, Rehan SM. 2016 Development of multiple polymorphic microsatellite markers for the small carpentr bee, *Ceratina calcarata*, using genome-wide analysis. *Journal of Insect Science*. **16**, 57 (doi: https://doi.org/10.1093/jisesa/iew042)]

INTRODUCTION

Microsatellite markers are popular and frequently employed for studies of relatedness and population genetics. Owing to their high mutation rate and population variability, microsatellite loci can be targeted to reveal subtle changes in population structure and composition, kinship, patterns of paternity, and heritability (reviewed in Powell, Machray, and Provan 1996; Sunnucks 2000). The design and optimization of these powerful molecular tools has been improved and greatly expedited by next-generation sequencing (Grover and Sharma 2016). Given their multiscalar capacity to reveal patterns of both population and family structure, the development of microsatellite markers remains an informative and important endeavor.

Bees are represented by over 20,000 described species and occur on all continents except Antarctica (Michener 2007). Highly efficient pollinators, bees make a significant contribution to the productivity of both agricultural and natural systems (Kremen, Williams, and Thorp 2002; Klein et al. 2007; Brittain et al. 2012). Fine scale population research has revealed a great deal about how bees affect and are affected by ecological conditions (Cameron et al. 2011; Bartomeus et al. 2013). Such studies have detailed species composition and population distribution, (e.g. *Melipona* spp. Tavares et al. 2013; *Bombus* spp. Geib, Strange, and Galen 2015), effects of land use practices (Dreier et al. 2014), as well as conflicts between managed and wild bee populations (Moreira et al. 2015). The development of microsatellite markers thus allows for the

comprehensive study of bee biology and demography at a macroscopic scale, and informs our ability to implement biologically meaningful pollinator conservation practices.

The small carpenter bee, *Ceratina calcarata*, is one of five very recently diverged and largely sympatric species of *Ceratina* found across eastern North America (Rehan and Sheffield 2011; Shell and Rehan 2016). *Ceratina calcarata* appears to be a generalist pollinator (McFrederick and Rehan 2016) and, given its broad range and high abundance, contributes to the productivity of a large number of ecological and agricultural systems. *Ceratina calcarata* is also subsocial: females provide extended maternal care to their brood, and defend and clean their offspring into adulthood (Rehan and Richards 2010). Subsocial behavior is considered foundational to the evolution of more complex social forms (reviewed in Rehan and Toth 2015); as such, *C. calcarata* is also emerging as a model organism for studies of social evolution (Rehan, Berens and Toth 2014).

Significant molecular resources are available for *C. calcarata* in the form of an annotated transcriptome (Rehan, Berens, and Toth 2014), methylome and genome (Rehan et al. 2016). Among many powerful and practical applications, such data avails the rapid and reliable development of molecular markers. Here, we isolated microsatellite loci from the *C. calcarata* genome, and optimized a suite of 21 polymorphic markers in 39 individuals from across the species' range. These primers make available multiscalar population genetics studies, and will allow researchers to investigate relatedness and patterns of parentage in this unique subsocial pollinator.

Materials and methods

Microsatellite loci were isolated from the recently published *C. calcarata* genome (Rehan et al. 2016) using the Microsatellite Identification Tool (MISA; Thiel et al. 2003) interfaced with an executable version of Primer 3 (Untergasser et al. 2012; Koressaar and Remm 2007). MISA was used to trim genomic reads to lengths of 20 kb to facilitate scans. To ensure microsatellite quality, MISA was configured to select loci based on strict minimum motif repeat requirements (mononucleotides = 15; dinucleotides = 7; tri- through hexanucleotides = 5). The resulting 2,010 putative microsatellite loci and flanking regions were sorted in descending order of motif length and motif repeat count. Flanking regions were then visually inspected for self-complementarity to prevent hairpin or excessive primer dimer formation. Putative primer pairs were also screened to ensure less than 1°C difference in melting temperature between forward and reverse oligos.

Candidate primers were then assessed for amplification performance and polymorphism in 16 females from across *C. calcarata*'s range (*C. calcarata* is haplodiploid, thus primers must be screened in females to accurately assess heterozygosity). We followed the methodology of Schuelke (2000) and designed forward primers modified with a partial M13 tail. This M13 oligo extension allowed fluorescently dyed universal probes to be incorporated during PCR. A set of universal primers was labeled with three dyes from the DS-33 set (FAM, PET, and VIC) to allow for downstream multiplexing. PCR reactions were executed in an 11 µl volume (5.45 µl ddiH20; 2.0µl 5x HF Buffer (Thermo Scientific); 0.2µl [10mM] dNTPs; 0.1µl Phusion HF Taq Polymerase (Thermo Scientific); 0.25µl [10mM] forward primer; 0.5 µl Fluorescent M13 oligo [10mM], 0.5µl [10mM] reverse primer; 2.0µl DNA template) using an Eppendorf Mastercycler gradient thermocycler. PCR reactions involved five stages: i) initial denaturing at 98 °C for 40s, ii) a touchdown series of 10 cycles at 98 °C for 10s, 72 °C for 15s (cooling incrementally to

primer-specific T_a), and 72 °C for 15s, iii) 20 cycles at 98 °C for 10s, primer T_a for 15s, and 72 °C for 15s, iv) 8 cycles at 98 °C for 10s, 62 °C for 15s, and 72 °C for 15s, v) final extension at 72 °C for 10 min. PCR products were mixed with HiDi Formamide (Applied Biosystems, Foster City, CA, USA) before being sent to the DNA Analysis Facility at Yale University for fragment analysis on a 3730xl Analyzer (Applied Biosystems, USA). Alleles were scored using Peak Scanner 2 (Applied Biosystems).

After initial PCR, 21 loci were further screened for polymorphism and performance in an additional 23 females (total screening panel N = 39), following the above methods. Loci were tested for Hardy-Weinberg and linkage disequilibrium using GenePop 4.2 (Raymond and Rousset 1995; Rousset 2008). Each locus was then assessed for expected and observed heterozygosity (H_e and H_o respectively) and total alleles (k) using GenAlEx 6.502 (Peakall and Smouse 2006; Peakall and Smouse 2012). All 21 microsatellite loci were uploaded to GenBank under accession numbers KU945359-KU945379.

Results and discussion

None of the 21 loci were found to be in linkage disequilibrium, nor did they diverge significantly from Hardy-Weinberg equilibrium following sequential Bonferroni correction. Observed and expected heterozygosities ranged from 0.08 to 0.82 (mean 0.47) and 0.26 to 0.88 (mean 0.56) respectively (**Table A1**). Our results are thus very similar to those of other recent hymenopteran microsatellite development projects (Rabeling et al. 2014; Vickruck 2015; Chen et al. 2015).

Table A.1. Primer sequences and locus characteristics of twenty-one microsatellite loci developed for *C. calcarata*. Information presented includes primer sequence; GenBank Accession Number; repeat motif; allele count (k); allele size range; number of individuals successfully screened (N); observed (H_o) and expected (H_e) heterozygosities.

Locus	Primer Sequence 5'-3'	GenBank	Repeat Motif	k	Allele size range	N	H_o	H_e
Ccal01	F: ACAAAACAAAAGCGCGGACA R:GGATTGTCATGACGGGGGAG	KU945359	GGTGAC	8	263-311	38	0.658	0.752
Ccal02	F:AAATCAACCCTAGCCCCAGC R:TACACACAGGTCGTCACGTG	KU945360	CAGCTC	5	226-280	39	0.308	0.409
Ccal03	F: AATAGACGGAGAGCAGCAGC R: TTGTTTCATCTTCGCACGCG	KU945361	AGGCAG	7	152-188	38	0.526	0.664
Ccal04	F: GGAGAACCGAGATACCAGAGG R: TCCCACTTTTTACGGCTCCC	KU945362	ACCGA	2	91-96	36	0.083	0.579
Ccal08	F: TCGATTCACGCAGACCTGAC R: GGATATGCGCCCGTCACTAA	KU945363	CTGA	11	235-287	38	0.579	0.854
Ccal11	F: ATAGGGAGCGAGCTGTTTCG R: TCGTCCGCAGCCATAACAAT	KU945364	AGGTT	6	243-278	37	0.633	0.698
Ccal14	F: GGCGTAGTTCCATCTGTCGT R: TTGCACCGACGATTCTCGAA	KU945365	AACCT	3	164-174	39	0.308	0.406
Ccal16	F: CAGGGAAGGCGGGTATCTTT R: GGCGGTGAAATTGCGACTTT	KU945366	AGGTT	3	252-262	38	0.658	0.502
Ccal17	F: GTGCGCGTAGAACAACCAAG R: AGCCTCGTGCAGCTTACAAT	KU945367	GGCGA	5	195-215	39	0.179	0.295
Ccal18	F: GTTTCATTCGGTTCCGCACC R: CTGAGCCGCGTATCTGCATA	KU945368	GTTCT	2	235-240	39	0.256	0.26
Ccal19	F: TCATTAATTCGGGCGCCTGT R: CTGCCTTTCTCGTCCCTCTG	KU945369	GAACA	3	261-271	38	0.316	0.508
Ccal23	F: AATTCGGCCAAGCTCGTACA R: GGAAACTTGGTTTTCGGCCC	KU945370	GTGCG	6	163-188	39	0.795	0.585
Ccal25	F: AAACGGCGGACTGAAAAACG R: ACTTCGAGTGCGGATTTCGT	KU945371	CCGCA	7	188-218	39	0.436	0.645
Ccal29	F: ACGTTGGACGAACACTGACA R: CCGTGGCTCTCCCTAATCAC	KU945372	AACCT	3	270-280	39	0.359	0.331
Ccal30	F: TACTATGTGATGCGTGCCGT R: CACGAGTGGGTCCCGAATAC	KU945373	ATCAT	5	267-302	39	0.436	0.656
Ccal37	F: CGTCTCGCAGTAACGGTACA R: AGAACAGTCGTGTCCGGTTC	KU945374	AGAA	12	148-196	39	0.769	0.868
Ccal39	F: CAAAGAAATGGCGGGGAACA R: GCGACGGTAATGACTTACAACG	KU945375	TTAT	7	253-277	38	0.553	0.765
Ccal44	F: TTCCCAACACGCTTCGTACA R: TACGTGGATGCATTCGTCCC	KU945376	GTCT	6	201-297	38	0.289	0.319
Ccal48	F: CGATTCCGGTGAAACGCAAG R: CTTCCTTCCTTCCCATGCGT	KU945377	GGAA	5	106-126	39	0.513	0.584
Ccal49	F: CTGCCGTATCCTCTCTCCCT R: GAGAGGCACGCGGGTAATAA	KU945378	GCAC	8	234-262	39	0.744	0.787
Ccal50	F: CCGACCTTTCTCGCAAAACG R: TCTCTGTTTCTTCCCACCGC	KU945379	TGTA	14	227-283	39	0.821	0.879

This suite of microsatellite loci are the first primers developed for a member of the New World *Ceratina*, and the first set developed for the genus based on genomic data. These loci will be informative in exploring the population structure, patterns of parentage, and kinship dynamics in *C. calcarata*. Microsatellite loci developed for other Hymenoptera (*Ceratina flavipes*, Azuma et al. 2005; and *Halictus rubicundus*, Soro and Paxton 2009) demonstrated significant cross-

amplification in closely related species. As *C. calcarata* is closely related to four other native *Ceratina* species (Rehan and Sheffield 2011; Shell and Rehan 2016), these loci will also likely cross-amplify, allowing for genus-wide research and conservation.

Additionally, these microsatellite loci can be powerful markers for understanding the evolution of social structure. Microsatellite markers have been used to reveal variation in queen-worker dynamics by geography (Richards, French, and Paxton 2005), and deviations from expected worker kinship in a eusocial sweat bee (*Lasioglossum malachurum*, Soro et al. 2009). Such markers have also been used to reveal mating structure in a communal bee (*Andrena jacobi*, Paxton et al. 1996) and polymorphism in reproductive strategy and sociality among populations of *Halictus scabiosae* (Ulrich, Perrin, and Chapuisat 2009). As *C. calcarata* is subsocial (Rehan and Richards 2010) it likely represents a foundational stage in the evolution of eusociality (Rehan and Toth 2015). By employing even a subset of our 21 markers in a study of *C. calcarata* inter- and intracolony relatedness, we may be able to uncover similarly cryptic social dynamics in this native pollinator.

Protocols for the discovery and optimization of microsatellite loci are numerous and have evolved over decades of population genetics studies (e.g. Glenn and Schable 2005; reviewed in Zane, Bargelloni, and Patarnello 2002). To secure even one useful microsatellite was originally a laborious procedure with few guarantees: isolated loci were random, or limited to sites complimentary to specially designed probes (Ostrander et al. 1992; Queller, Strassmann, and Hughes 1993; Kijas et al. 1994). Whole genome sequencing technologies have made great advancements in quality and accessibility of genetic resources over the past decade (Hudson 2008; Ekblom and Galindo 2011; vanDijk et al. 2014). Improved availability of such powerful resources has greatly expedited and improved the generation of microsatellite primers for many

novel species (e.g. copperhead snakes, Castoe et al. 2010; water striders, Perry and Rowe 2011). The generation and analysis of whole genomes is still an expensive and bioinformatically exacting endeavor (reviewed in Lemmon and Lemmon 2013); by contrast, microsatellite markers are affordable and adaptable tools, suited to a wide range of research.

APPENDIX B

SUPPLEMENTARY MATERIAL AND METHODS

FOR

CHAPTER 3 – SOCIAL MODULARITY: CONSERVED GENES AND REGULATORY ELEMENTS UNDERLIE CASTE-ANTECEDENT BEHVIORAL STATES IN AN INCIPIENTLY SOCIAL BEE

[Shell WA, Rehan SM. 2019 Social modularity: conserved genes and regulatory elements underlie caste-antecedent behavioural states in an incipiently social bee. *Proceedings of the Royal Society B*. **286**, 20191815 (doi: https://doi.org/10.1098/rspb.2019.1815)]

Supplementary data tables are in UNH DropBox:
https://unh.box.com/s/4ss3kyw9e5j2vir4uvbx05qnoo38poxb

SUPPLEMENTARY METHODS

Sample collection continued

Foraging mothers and daughters were collected on the wing, following extended observation of paint marked individuals during warm, sunny days (Mikát et al. 2017). After an individual was observed departing the nest to forage, the entrance was capped to prevent reentry. The foraging individual was then captured upon return, identified as the mother or a daughter of the nest, and immediately flash frozen. As seen in other *Ceratina* (Sakagami and Maeta 1984), nest guards in *C. calcarata* position themselves at the nest entrance with their abdomens entirely blocking the passageway. To collect guarding mothers and daughters, we thus clipped off the upper part of the stem of an observed nest and immediately froze this section in liquid nitrogen. This method ensured collection of guarding individuals, which were later identified as the mother or a daughter of the nest prior to dissection for RNA extraction.

Further details on time course analysis (maSigPro)

During a standard run, maSigPro takes inputs of normalized gene count data, along with a user-defined matrix of time point conditions. During our run, we provided the normalized count data for females collected from each nest stage (i.e. not including foraging or guarding females). Autumn daughters represent the youngest adult stage in a female's lifetime and were thus used to initiate a natural five-point chronology (i.e. 1 – AD, 2 – FN, 3 – AB, 4 – FB, 5 – AM). In the following lines of maSigPro script, we then assigned our sample replicates to these time points (N=3 replicates per time point) using either a 1 (in) or 0 (out). maSigPro was then run using the "backward" time step (tstep) method with an alpha cutoff for significance = 0.05. We then used the see.genes function with default settings to inspect clusters of genes which maSigPro identified as demonstrating consistent trends in expression over the defined timeline. By default, maSigPro identifies the nine best-supported clusters of genes. We opted to consolidate gene lists from these resulting nine clusters using similarities in overall expression patterns (i.e. while total expression levels varied among clusters, clusters could still be combined using consistencies in gene up- and down-regulation over the defined timeline).

Further details on gene differential expression analysis (DESeq2)

Tests to identify differentially expressed genes were run as two, independent general linear models (GLMs) in R using DESeq2 to identify genes associated with i) *behavioral state*, i.e. comparing all foragers, guards, and autumn nest females to each other in one model, regardless of age; and ii) *behavior by age*, i.e. comparing foraging and guarding mothers and daughters both to each other and to autumn mothers and daughters. In both runs, autumn nest

females (initially assessed during time course analysis) were used as non-foraging/non-guarding 'nesting' controls. Genes identified as significantly differentially expressed through GLM analyses were then pruned using a false discovery rate (FDR) cutoff of FDR < 0.05.

Further details on Weighted Gene Co-expression Network Analysis (WGCNA)

Rather than using differences in gene expression to determine relative up- and down-regulation of genes among defined groups, WGCNA identifies modules of co-expressed genes and then determines their associations with input samples bearing user-defined traits of interest. During a standard run, the WGCNA pipeline takes as input normalized gene expression data along with a data matrix of traits of interest for all samples considered. Following the publicly available user guide and tutorials (https://horvath.genetics.ucla.edu/html/CoexpressionNetwork/Rpackages/WGCNA/Tutorials/index.html) we input normalized gene expression data from females at the autumn nest stage and assigned each sample appropriate traits (i.e. behavioral status as foraging, guarding, nesting, and age status as either a mother or daughter). This total gene expression dataset was then trimmed of genes which featured no expression data for three or more samples. The remaining dataset, representative of 15,106 genes, was then used to cluster our samples using the default scripted hclust distance method, producing a first-pass dendrogram to inspect for outlier samples (**Figure S3.2A**). After removal of a single outlier, re-clustering revealed three focal groups: i) guarding daughters, ii) foraging and guarding mothers, and iii) a mixed group of nesting mothers and daughters with some guarding mothers, and foraging daughters and mothers (**Figure S3.2B**). The full WGCNA analysis was then performed as instructed by the authors' guide, using a soft-thresholding power of 5, as supported by model testing (**Figure S3.3**). Our total gene input data

were subsequently assigned to 35 gene co-expression modules (**Figures S3.4, S3.5**), with between 65 and 2078 genes per module ($N_{average}$ = 432). After gene modules were defined (**Figure S3.4**), further statistical analyses included tests of membership of each individual gene within each module (module membership), tests of association between genes and traits of interest (i.e. foraging, guarding, nesting, daughter, or mother association), and tests of module-trait associations. These analyses concluded with outputs of ranked, hierarchical results tables containing the most statistically supported to least-supported modules and their corresponding genes for each trait of interest (**Tables S3.12-S3.16**). We then used these tables to explore the results of WGCNA, to generate illustrative scatterplots for each trait's best-supported, positively correlated module (**Figures S3.5-S3.10**), and to compare these results to those of DESeq2 analyses.

Cis-regulatory element enrichment analysis continued

To prepare for this analysis, relative regulatory direction was re-determined for each DEG by calculating Z-scores across each biological condition. Z-scores provide an accurate measure of positive or negative standard deviations away from average expression: a score near 0 indicates little difference from average expression, and positive and negative values respectively indicate up- and down-regulation relative to the average. Using this logic, DEGs with Z-scores less than or equal to -0.333 were considered as down-regulated; greater than or equal to -0.333 but less than or equal to 0.333 were considered non-differentially regulated from the average; and those with a Z-score greater than 0.333 were considered up-regulated in that condition. These calculated regulatory directions were then confirmed for consistency against results of DESeq2 analyses. These regulatory directional values were taken together with

promoter region sequence data drawn from 5kb windows upstream of all genes for which sequence and expression data were available (N = 13,468 genes) and fed into Stubb to identify TFBS motif enrichment. Stubb scores both individual and compound TFBS motifs across each biological condition based on both that motif's presence in each associated gene's promoter region and the defined regulatory direction of each gene. Total Stubb outputs were then fed through cis-Metalysis, which selects the top 1% most significant portion of the results to produce a list of best-supported TFBS motif enrichment associations for up- and down-regulated genes in each condition. Given the exhaustive permutation testing of the cis-Metalysis pipeline (i.e. for all motifs A and B, tests are performed for the presence of just motif A, just B, and all logical motif pairs: A and B, A and not B, not A and B, and not A and not B) final output TFBS motif lists were still quite extensive. We therefore consolidated these results for further analysis by removing functionally redundant outputs (e.g. while the outputs "not A and B" and "just B" are technically distinct, they are effectively both functionally reducible to "just B;" whereas the motif outputs "A and B" and "A and C" are both distinct pairs, and would thus both be kept for further analyses).

SUPPLEMENTARY FIGURES

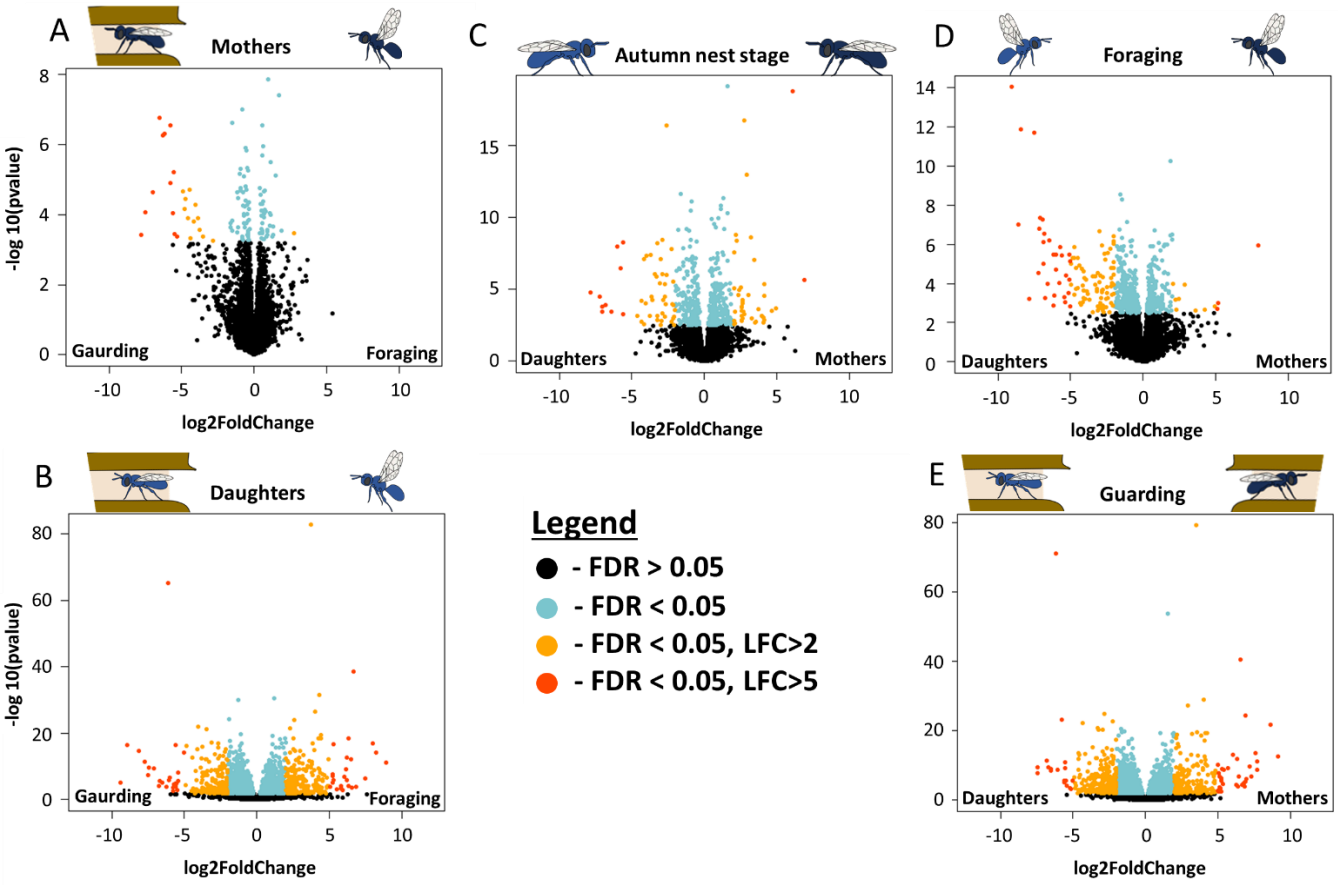

Figure S3.1. Volcano plots presenting results from general linear model analysis of gene expression associated with behavior by age. A) guarding vs foraging behavior in age-matched mothers; B) guarding vs foraging behavior in age-matched daughters; C) autumn 'nesting' daughters vs mothers; D) foraging behavior in daughters vs mothers; E) guarding behavior in daughters vs mothers. In each plot, each dot represents a gene for which a false discovery rate (FDR) value could be determined. Dots are colored by significance (at FDR < 0.05) and log fold change (LFC) value (see legend): black – non-significant; cyan – significant, LFC less than 2; orange – significant, LFC greater than 2; red – significant, LFC greater than 5. Vertical axis displays -log 10 of calculated p-value (note, axis scale varies substantially among plots). Age matched comparisons of behavior reveal much less variation in gene expression underlying behavioral states in mothers (A) compared to daughters (B). Additionally, while there is a fair amount of variation in gene expression associated with the effects of age in both autumn nest (C) and foraging females (D), guarding appears to be underpinned by distinct sets of genes depending on individual age (E). Overall, daughters demonstrate much greater variation in gene expression by behavioral role (B), and guarding behavior appears to be most strongly affected by age (E).

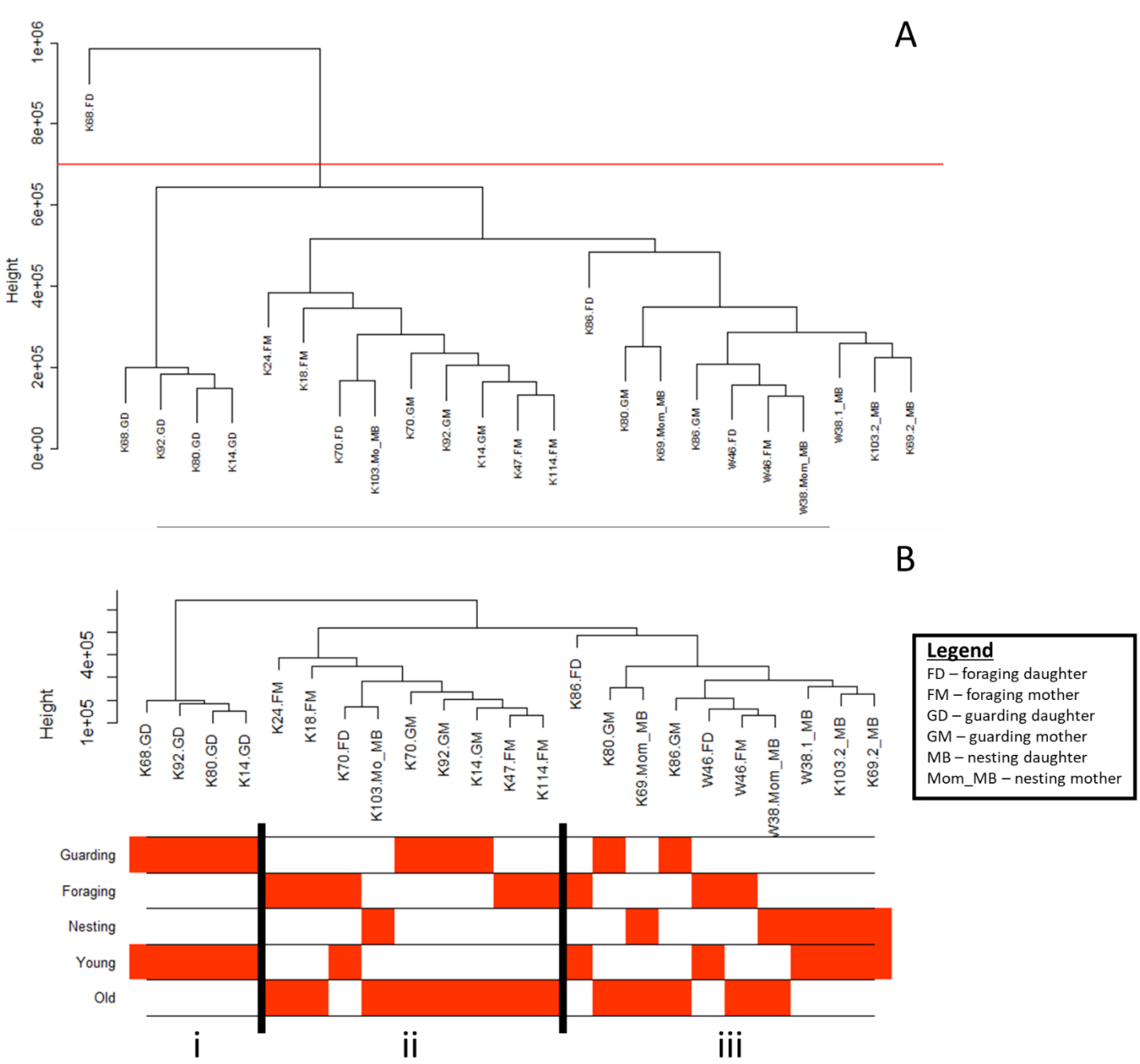

Figure S3.2. A) Sample clustering to detect any outliers based on whole transcriptome expression data. Y-axis displays Height as Euclidean distance. Sample K88FD, one of the four sampled foraging daughters, was found to be a clear outlier and removed for further WGCNA analysis. **B)** Sample dendrogram and trait presence/absence for all samples, minus K88FD. Y-axis displays Height as Euclidean distance. Below clustered branches, behavioral and age-associated traits for each sample is indicated in red. Clustering sorted all samples into three general groups: i) guarding daughters, ii) guarding and foraging mothers, iii) nesting daughters, nesting mothers, and some foraging or guarding mothers. Sample ID numbers (e.g. K68) and phenotypic coding (see legend) are provided.

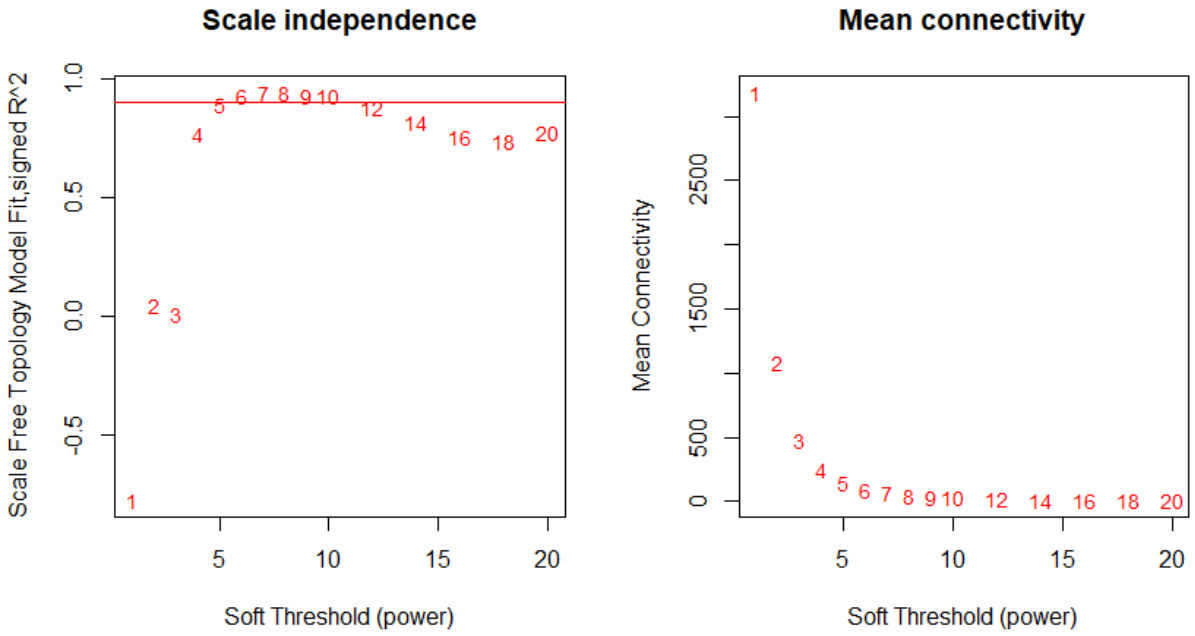

Figure S3.3. Results of automated analysis of network topology for a range of soft-thresholding powers. The y-axis of the left panel shows the scale-free topology fit index curve as a function of soft-thresholding power (x-axis). The red line indicates an R^2 cut-off value of 0.90. The right panel shows the mean model connectivity (y-axis) as a function of soft-thresholding power (x-axis). As it signifies the inflection point of model fit, we selected a soft-thresholding value of 5 for further analyses.

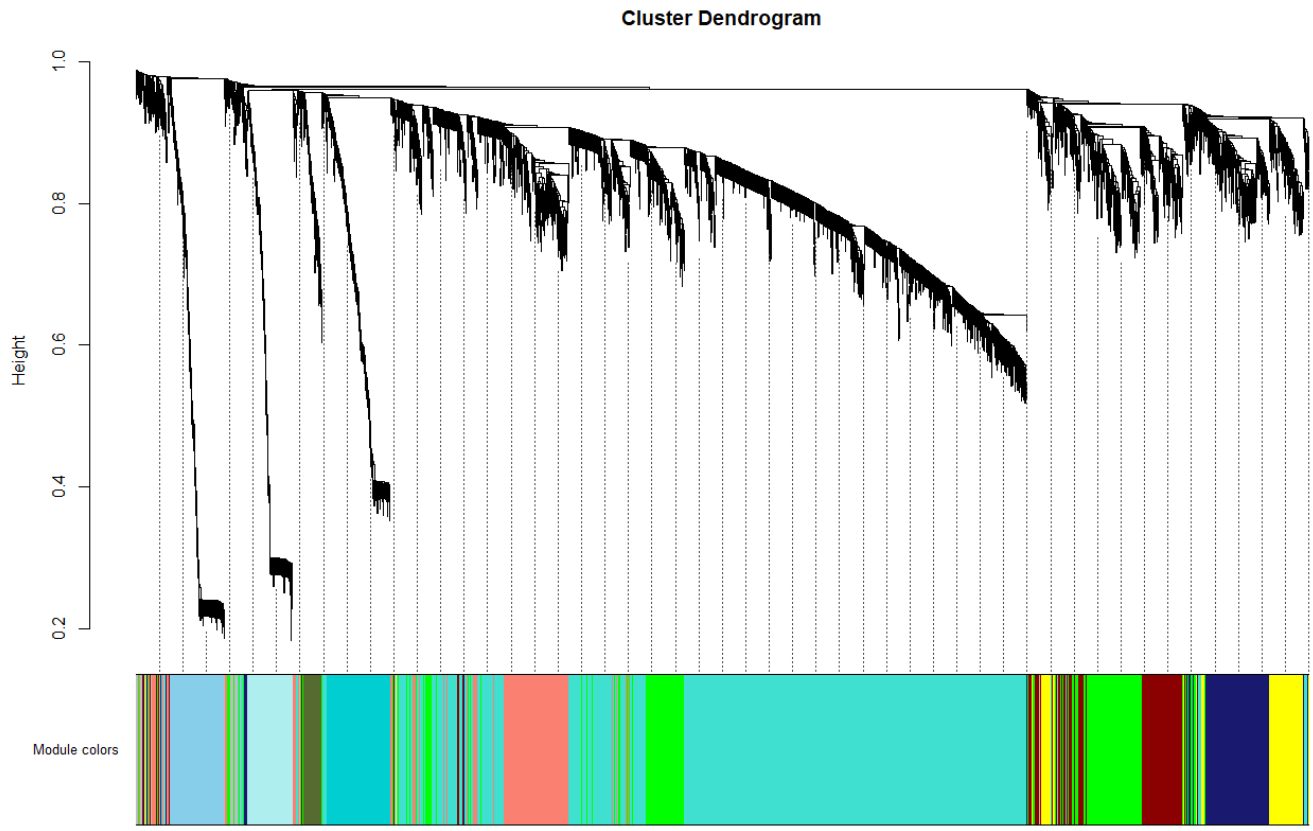

Figure S3.4. Clustering dendrogram of all genes differentially analyzed for age and behavior at the autumn nest stage (i.e. nesting, guarding, and foraging mothers and daughters), with dissimilarity based on topological overlap (i.e. gene network similarity), together with assigned module colors.

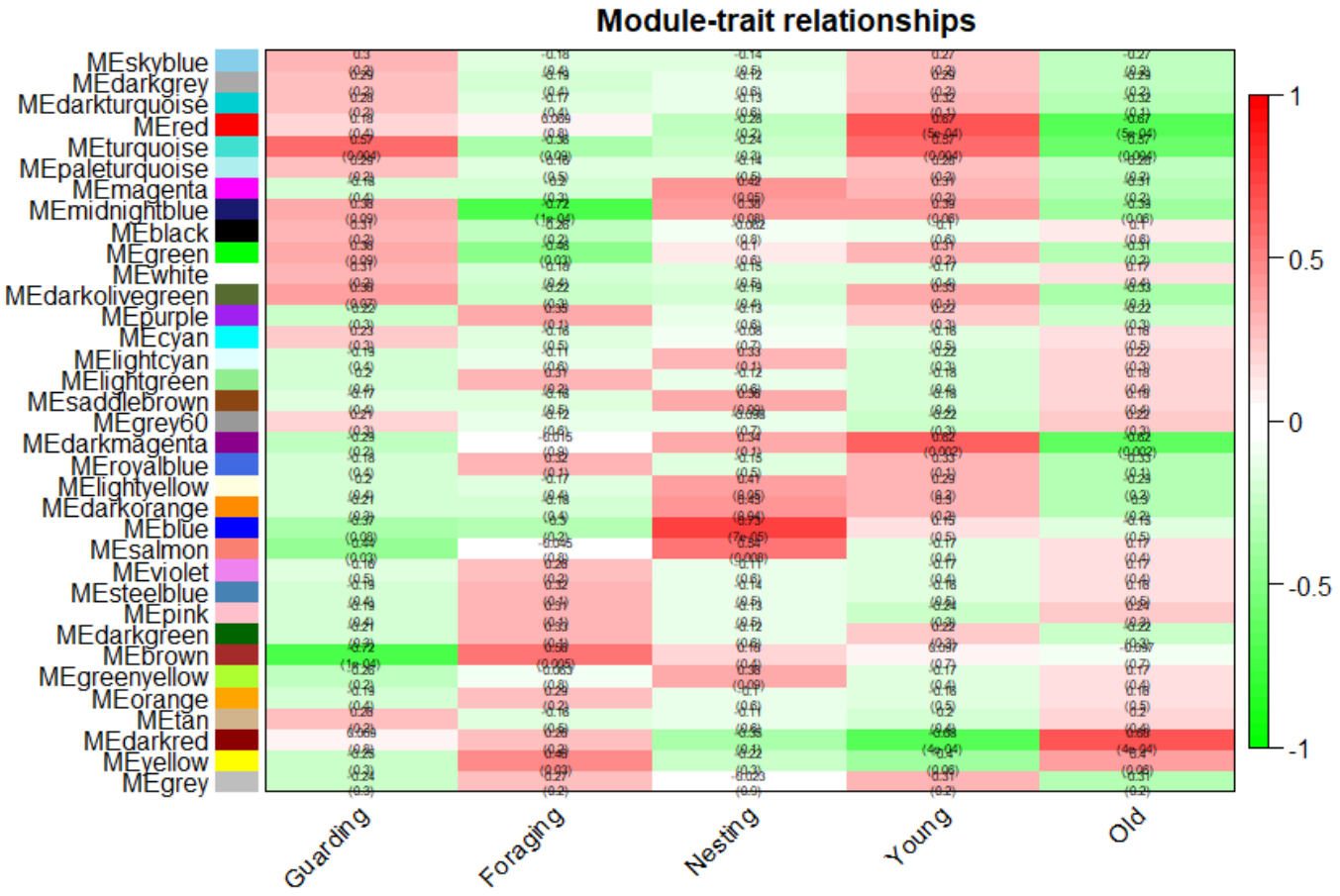

Figure S3.5. Module-trait associations across all behavior and age categories in autumn nests. Each row corresponds to a module's eigengene (most representative gene for the set), and each column specifies each trait (i.e. guarding, foraging, nesting, young – daughters, and old – mothers). Each cell contains the corresponding correlation and p-value for that eigengene-module-trait. The table is color-coded by correlation, according to the color legend on the right (-1, bright green, negative correlation; through +1, bright red, positive correlation). All module-trait correlation and significance values can be found in supplementary data **Tables S3.1-3.6**. Scatterplots were generated for the most statistically significant positively correlated modules for each trait as follows: turquoise - guarding (N_{genes}=2078; cor = 0.57, p = 2.3e-179); brown - foraging (N_{genes}=763; cor = 0.5, p = 1.7e-49); blue - nesting (N_{genes}=1191; cor = 0.72, p=7.3e-191); red - daughters (N_{genes}=695; cor = 0.62, p = 4.8e-75); and dark red - mothers (N_{genes}=296; cor = 0.55, p=8.4e-25).

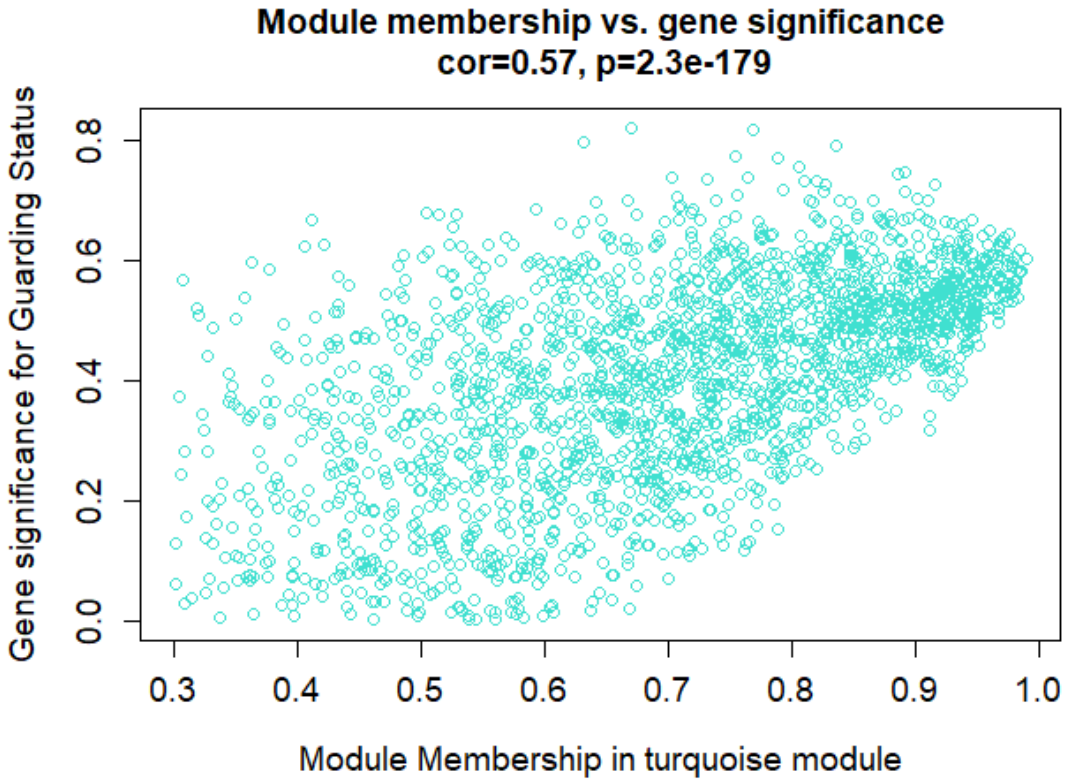

Figure S3.6. A scatterplot of gene significance for association with guarding behavioral state vs. module membership in the turquoise module. The highly significant positive correlation between significant association with guarding behavior and membership in the turquoise module suggests expression of genes in this set is closely tied to the guarding behavioral phenotype in *C. calcarata*.

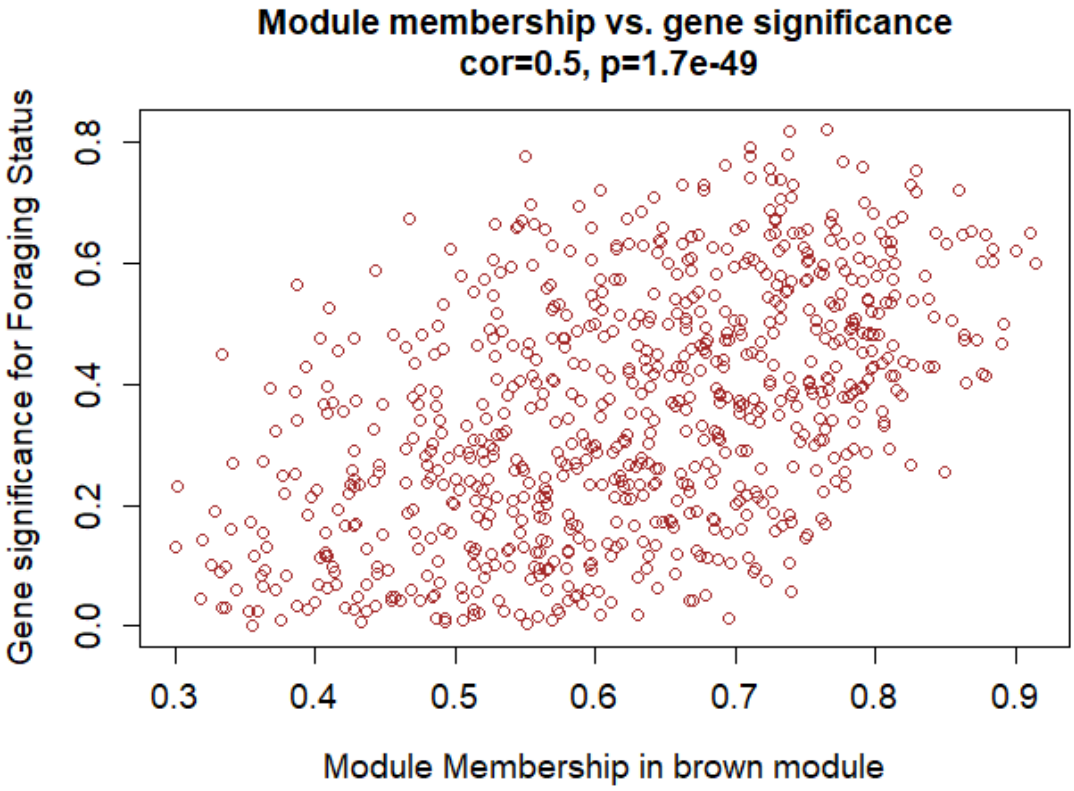

Figure S3.7. A scatterplot of gene significance for association with foraging behavioral state vs. module membership in the brown module. The highly significant positive correlation between significant association with foraging behavior and membership in the brown module suggests expression of genes in this set is closely tied to the foraging behavioral phenotype in *C. calcarata*.

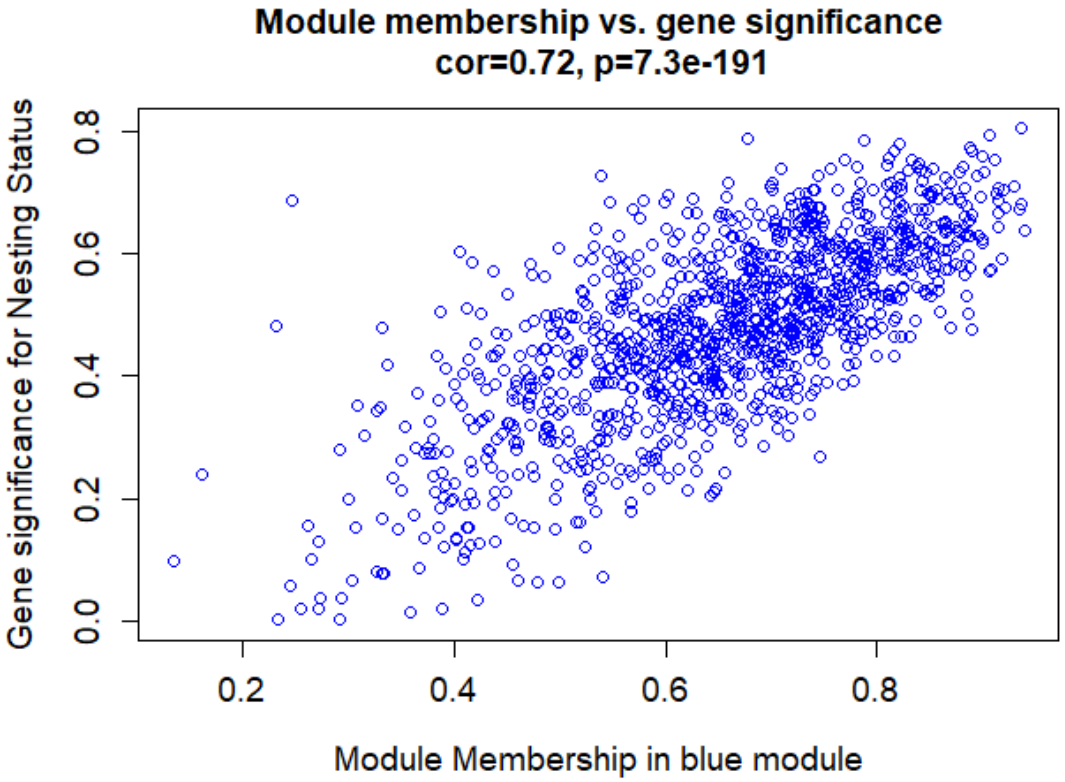

Figure S3.8. A scatterplot of gene significance for association with nesting behavioral state vs. module membership in the blue module. The highly significant positive correlation between significant association with nesting behavior and membership in the blue module suggests expression of genes in this set is closely tied to the nesting behavioral phenotype in *C. calcarata*.

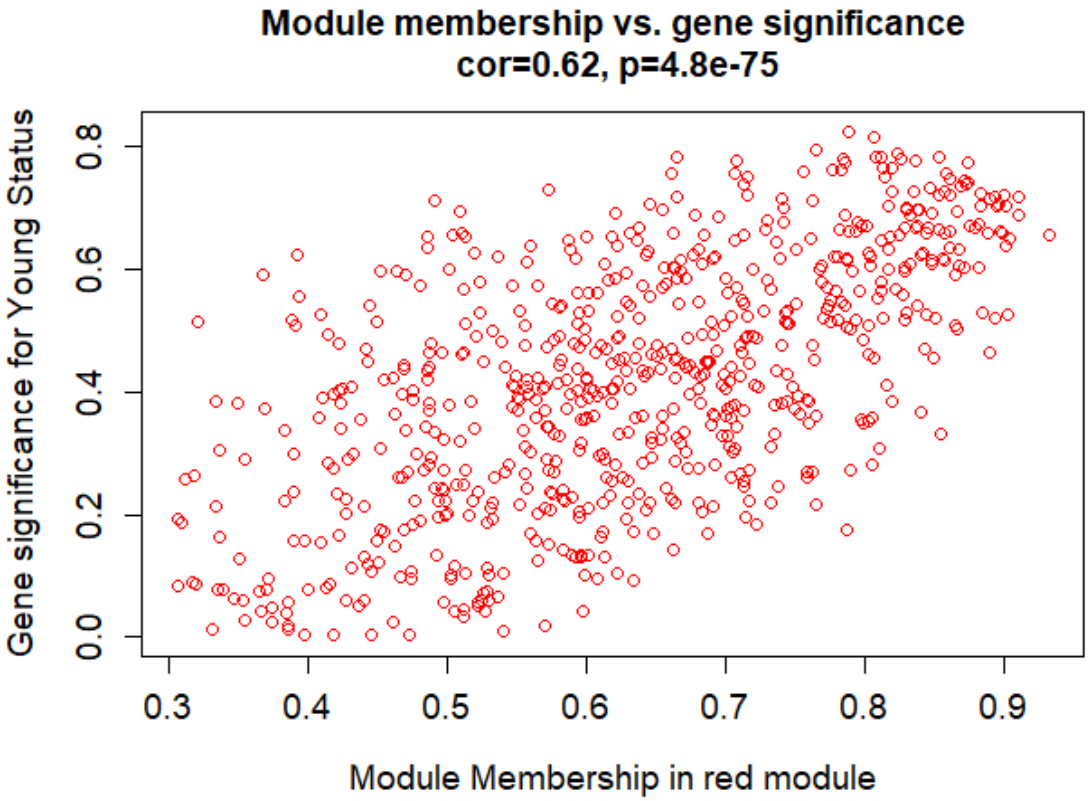

Figure S3.9. A scatterplot of gene significance for association with daughters vs. module membership in the red module. The highly significant positive correlation between significant association with daughter status and membership in the red module suggests expression of genes in this set is closely tied to daughters in *C. calcarata*.

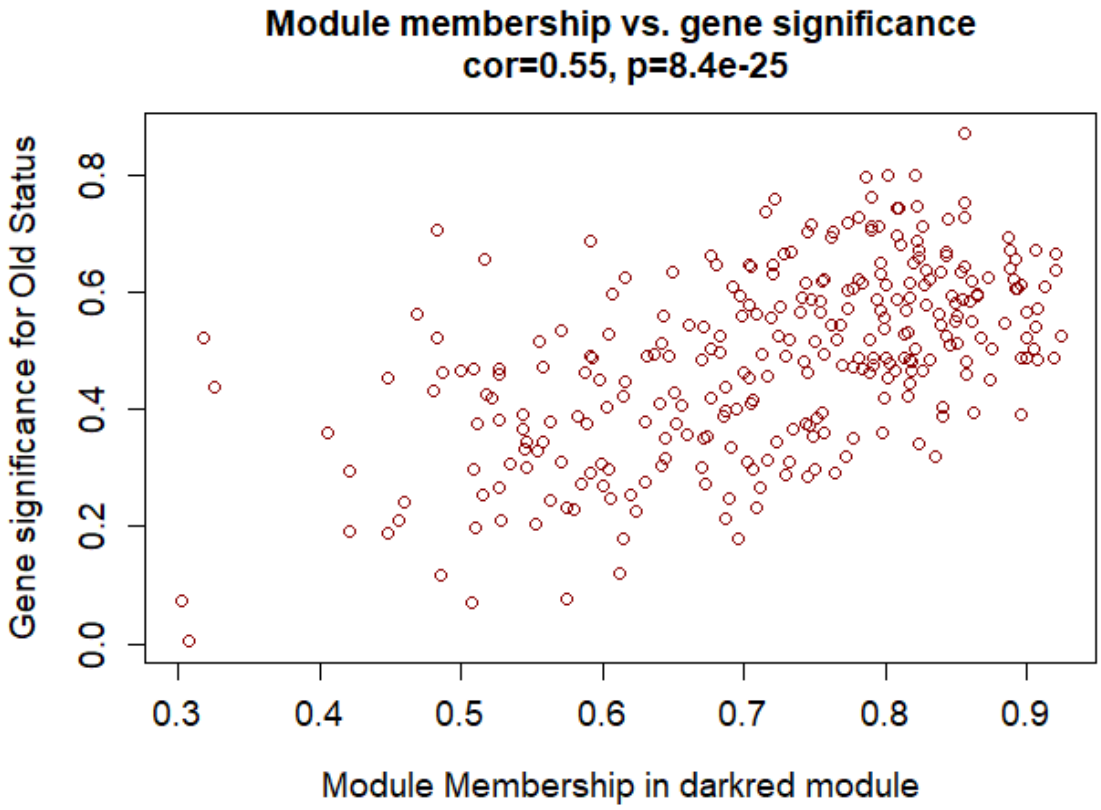

Figure S3.10. A scatterplot of gene significance for association with mothers vs. module membership in the dark red module. The highly significant positive correlation between significant association with mother status and membership in the dark red module suggests expression of genes in this set is closely tied to mothers in *C. calcarata*.

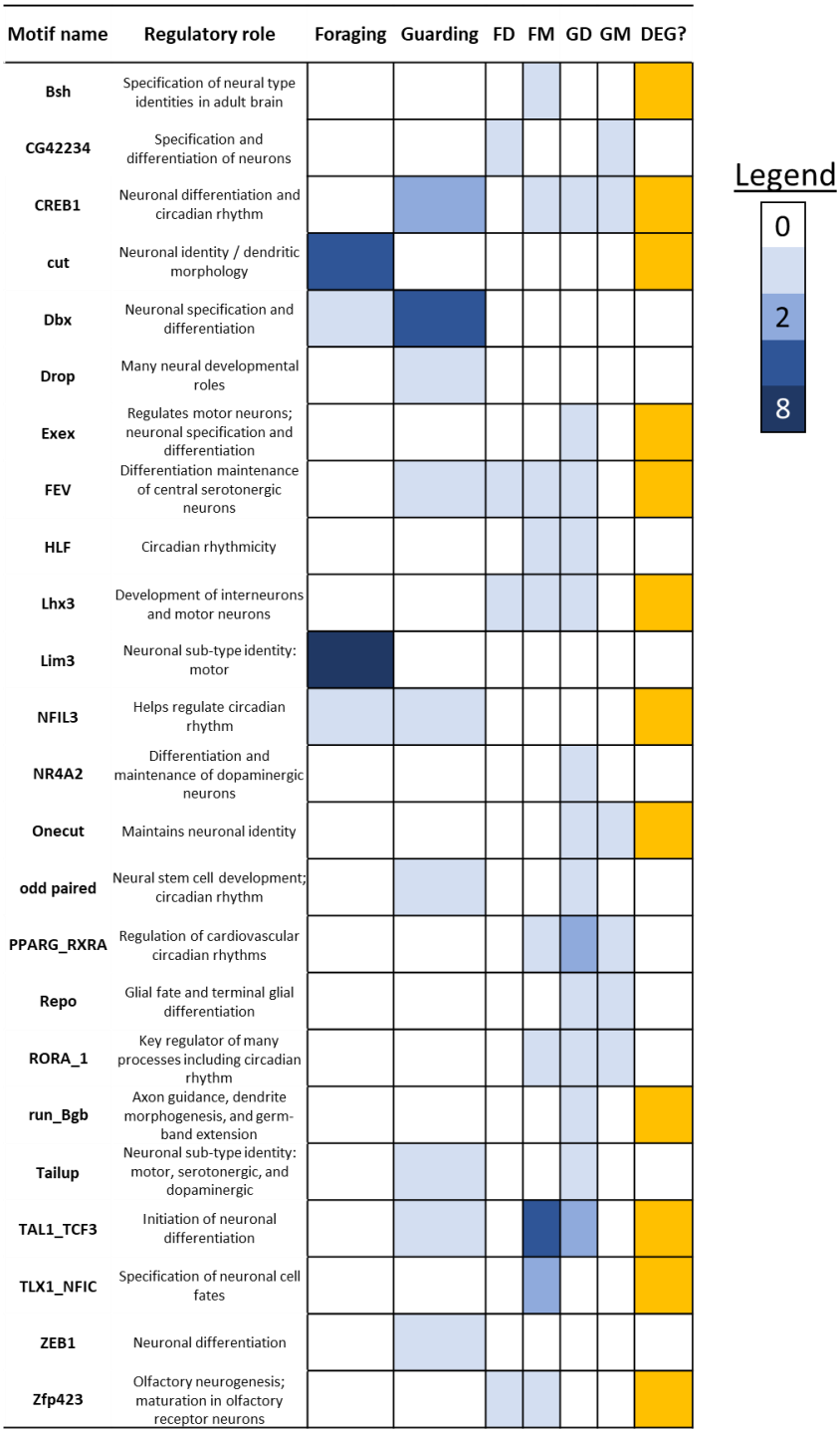

Figure S3.11. Heat map of transcription factor binding site motifs with neural regulatory roles, significantly enriched in the promoter regions of genes associated with behavioral states. Motif names are presented in alphabetical order along with their regulatory roles. Relative importance of each motif in the upregulation of genes associated with each biological context is then indicated by color (see legend): white = no enrichment; blue = enriched, with darker blues indicating greater enrichment frequency. The rightmost column indicates in yellow whether the associated transcription factor is also differentially expressed in this study.

APPENDIX C

SUPPLEMENTARY MATERIAL

FOR

CHAPTER 4 – MOLECULAR PATHWAYS UNDRLYING CASTES AND LONGEVITY IN A FACULTATIVELY EUSOCIAL SMALL CARPENTER BEE

Supplementary data tables can be found at UNH DropBox:
https://unh.box.com/s/4ss3kyw9e5j2vir4uvbx05qnoo38poxb

SUPPLEMENTARY FIGURES

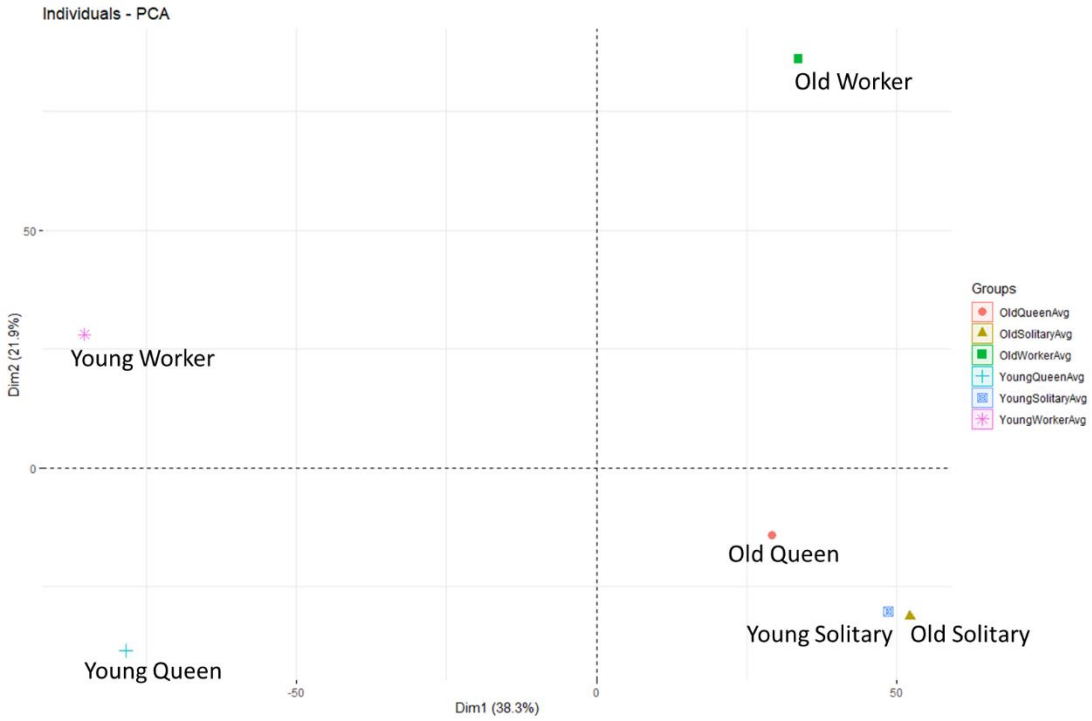

Figure S4.1. Principal components analysis (PCA) plot of gene expression variation among all age-class sample sets. A little more than 60% of total variation in the data is explained by the first two components. Regardless of age, workers are highly distinct from queens and solitary females along the y-axis; while queens and workers both feature comparable separation within class along the x-axis, solitary females are clustered very tightly together. Overall, workers are highlighted as the most distinct of the three biological roles examined in this study.

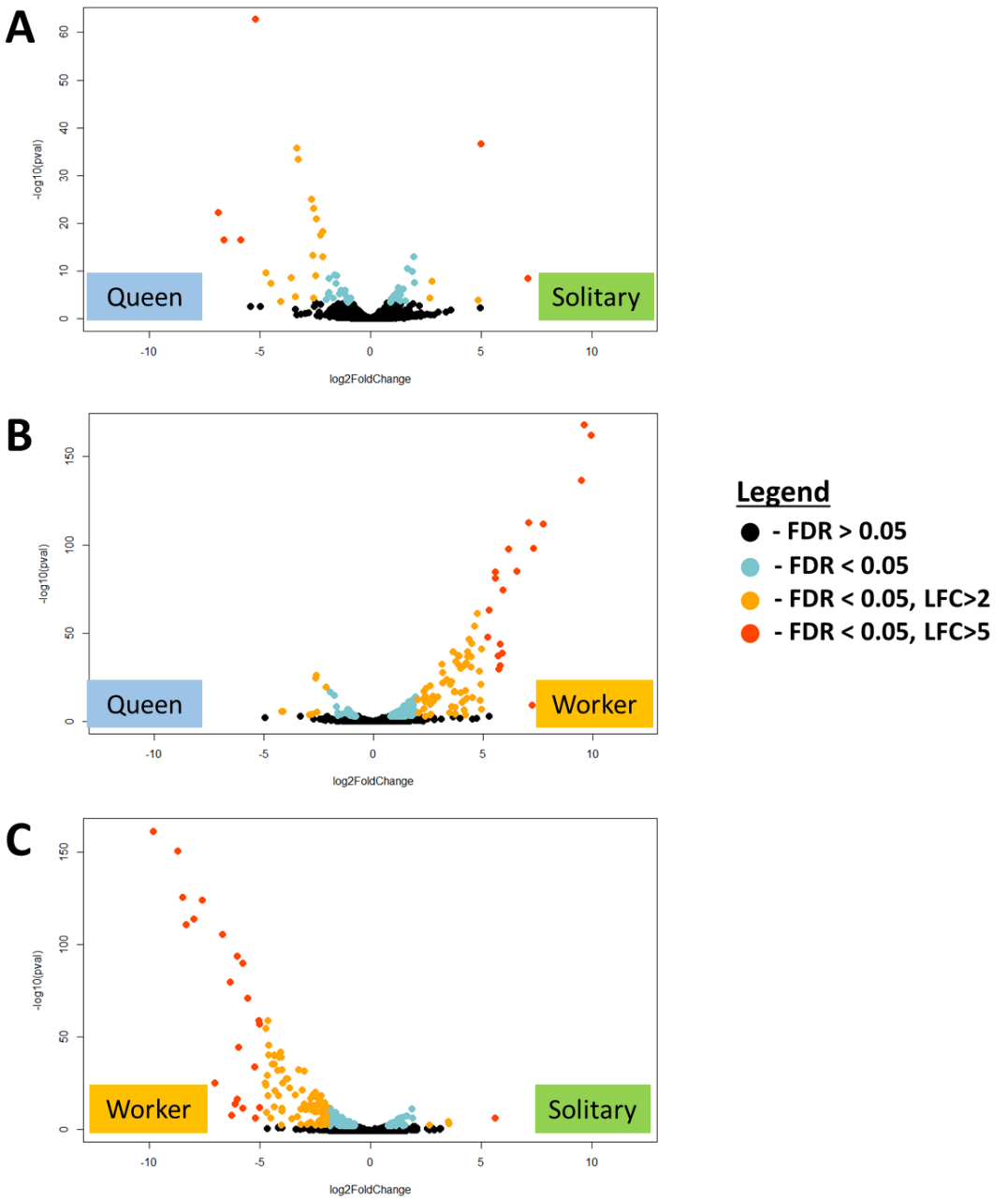

Figure S4.2. Volcano plots illustrating results of gene differential expression analysis across *C. japonica*'s three core biological roles regardless of age. A) queens vs solitary females; B) queens vs workers; and C) workers vs solitary females. In each plot, each dot represents a gene for which a false discover rate (FDR) corrected significance value could be determined. Dots are colored by significance and log fold change (LFC) values (see legend): black – non-significant; cyan – significant, LFC less than 2; orange – significant, LFC greater than 2 but less than 5; red – significant, LFC greater than 5. Vertical axis displays -log10 transformation of p-value (axis scale varies by plot). Queens and solitary females are distinguished by a handful of strongly and significantly differentially expressed genes (A); strongly worker associated genes are more numerous and feature over twice the magnitude in significance, revealing workers as highly distinct from both queens (B) and solitary females (C).

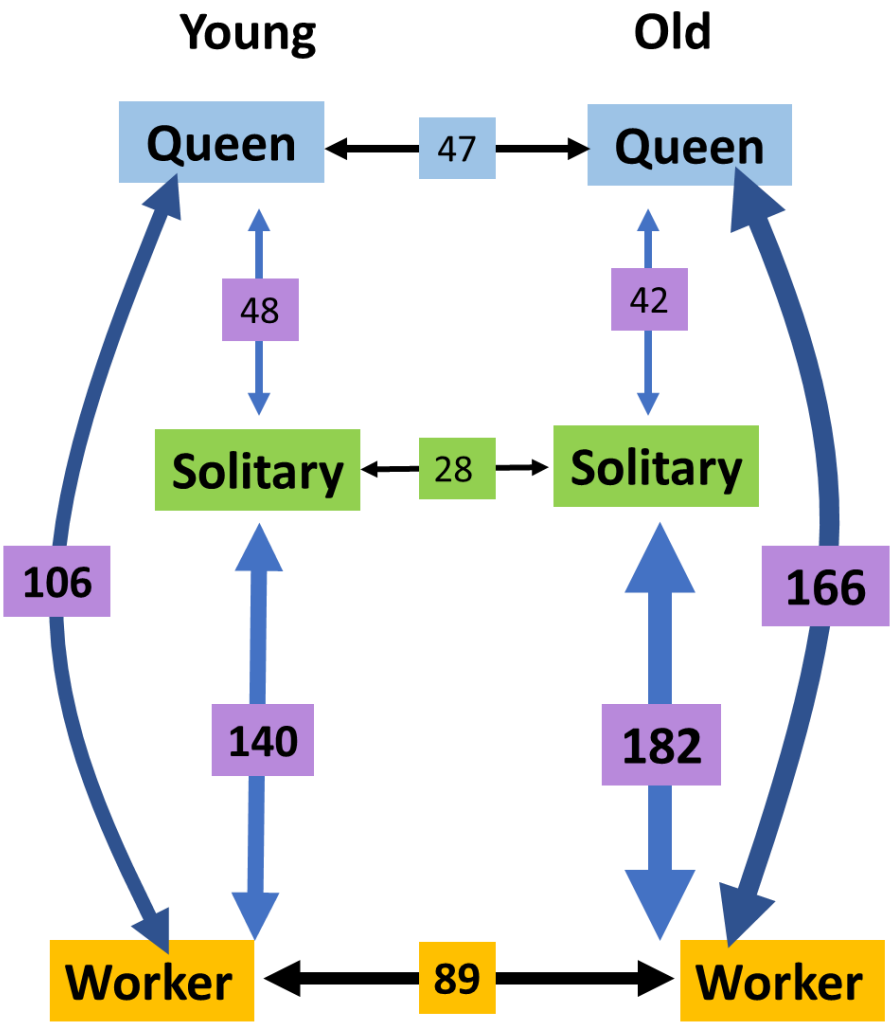

Figure S4.3. Summary schematic comparing total counts of differentially expressed genes (DEGs) separating groups i) by age within classes (class colors) and ii) by class within age brackets (purple); arrows are roughly scaled according to DEG count. Relatively few DEGs separate queens and solitary females regardless of age, but both are distinguished from workers by many DEGs among both young and old individuals. Within classes, age accounts for dramatically more DEGs in workers than in queens or solitary females.

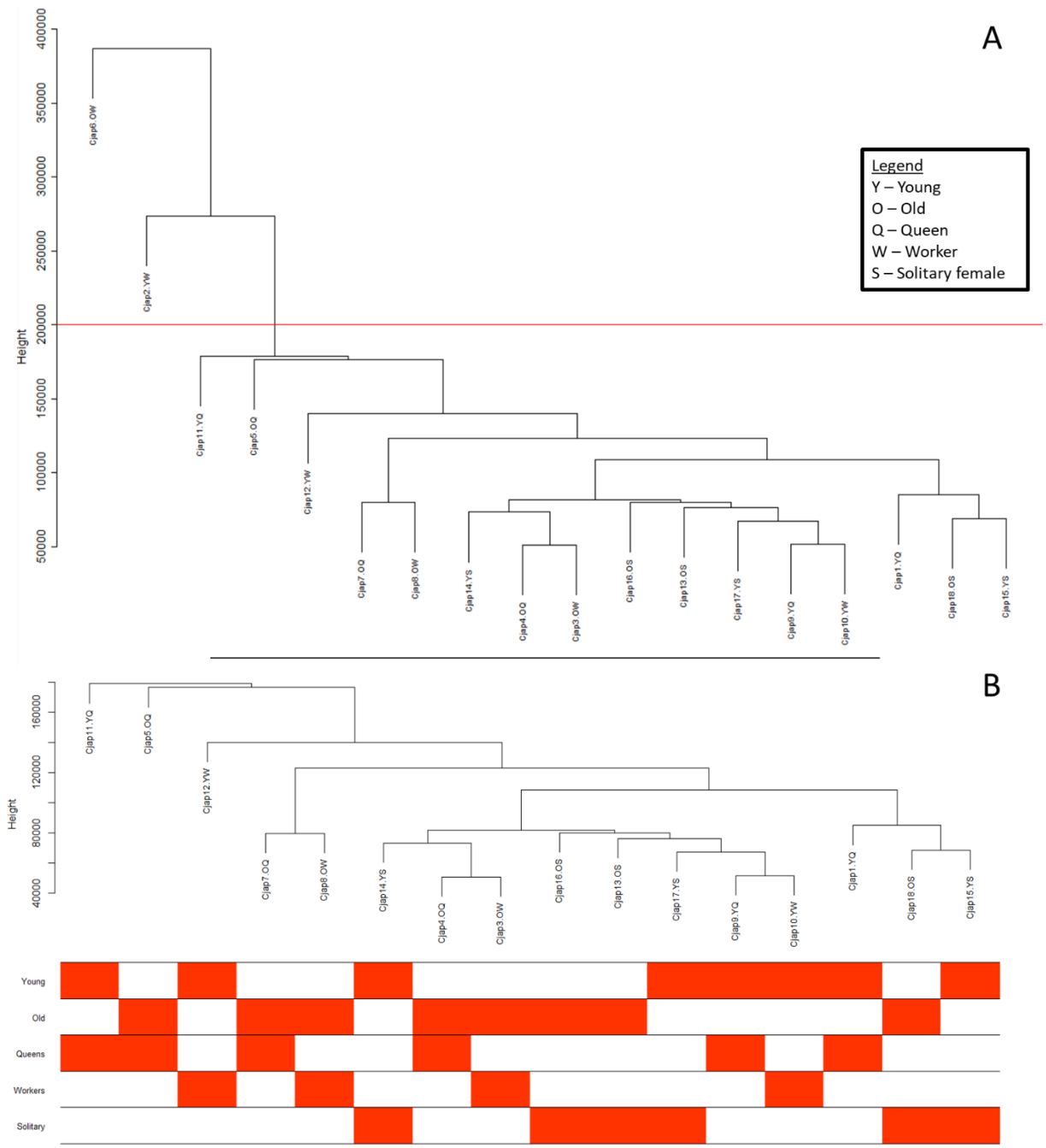

Figure S4.4. A) Weighted gene co-expression network analysis (WGCNA) preparatory sample clustering for outlier detection based on whole transcriptome expression data. Y-axis displays height as Euclidean distance; red line demarks reasonable cut point between core set and outliers. Samples Cjap2.YW and Cjap6.OW, a young and an old worker female, were both identified as clear outliers and removed from further WGCNA. B) Sample dendrogram and trait presence/absence coding for all remaining samples following outlier removal. Clustering sorted samples variably by class and age overall, but notably found support for close pairs between social nest females that physically shared a nest (e.g. Cjap3.OW + Cjap4.OQ; and Cjap9.YQ + Cjap10.YW).

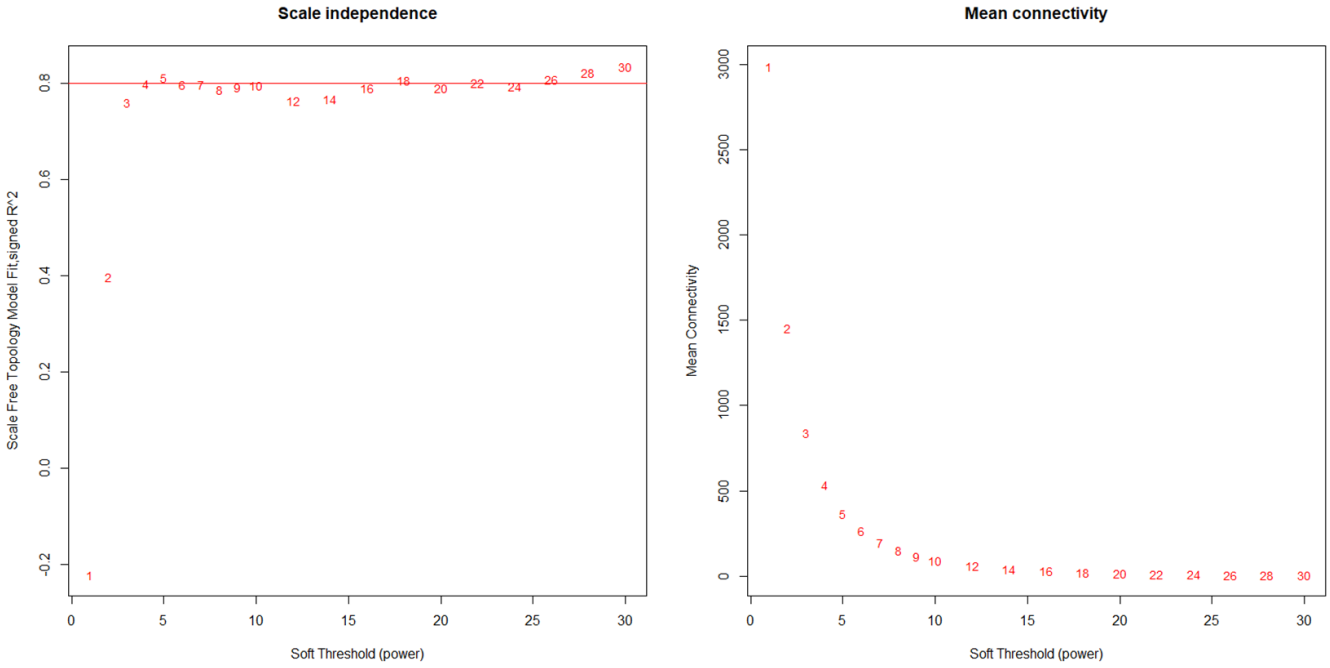

Figure S4.5. Results of automated analysis of network topology testing a range of soft-thresholding powers. The y-axis of the left panel displays the scale-free topology fit index curve as a function of soft-thresholding power (x-axis). The red line indicates an R^2 cut-off value of 0.80. The right panel shows the mean model connectivity (y-axis) as a function of soft-thresholding power (x-axis). As it signifies the inflection point of model fit, we selected a soft-thresholding value of 4 for further analysis.

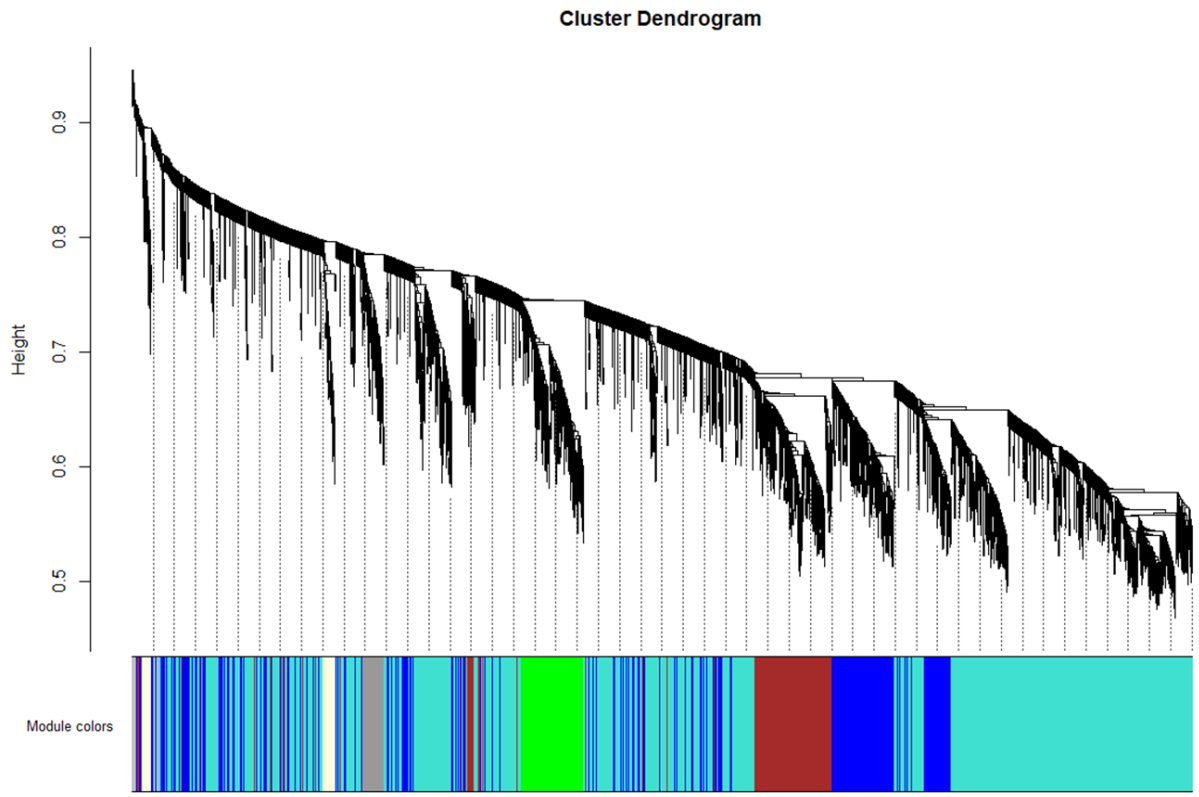

Figure S4.6. Clustering dendrogram of all genes analyzed for class and age, with dissimilarity (height) based on network topological overlap; assigned module colors are indicated below.

Figure S4.7. Module-trait associations across all age and class categories in *C. japonica*. Each row corresponds to a module's eigengene, and each column specifies associated biological traits (i.e. age – young or old; class – queens, workers, or solitary females). Table cells are color-coded by correlation between each module and the listed trait (ranging from -1, green, negative correlation; through +1, red, positive correlation. Each individual cell contains the corresponding summary correlation and p-value for the eigengene of that particular module-trait relationship. All module-trait correlation and significance values can be found in supplementary data **Tables S4.8-S4.12**.

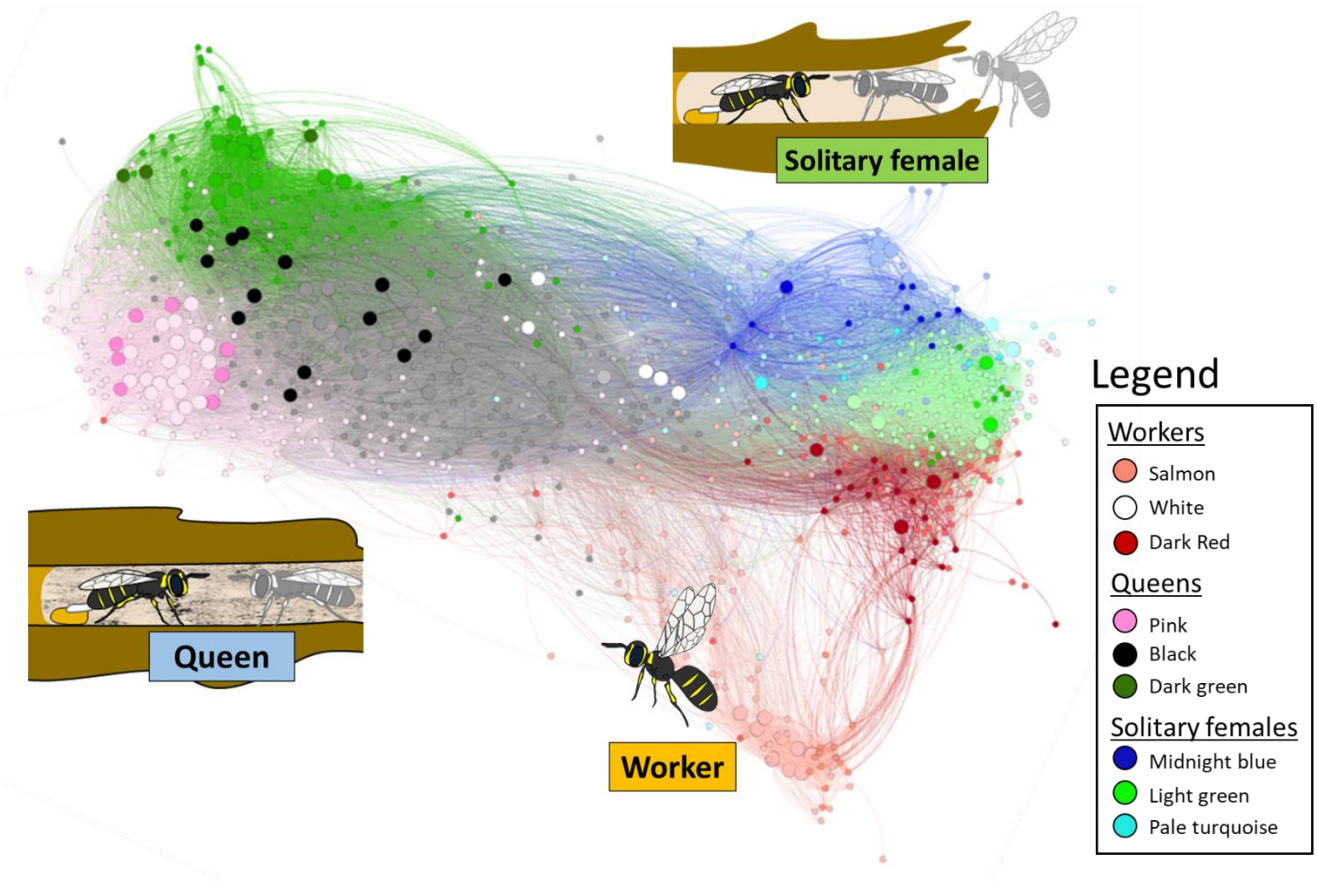

Figure S4.8. Weighted gene co-expression network rendering and colorizing only the top three most significantly and positively correlated gene modules for each class for summary visualization (see Legend; $N_{totalNodes}$ = 903 genes, $N_{totalEdges}$ = 40,174). Queen- and solitary female-associated genes occupy generally distinct regions of an otherwise well-interconnected overall network; although some worker-associated genes are interconnected more tightly with those of solitary females (e.g. dark red), a considerable number cluster tightly on the periphery of the network (salmon), suggesting class-specific expression patterns. Hub genes and network connectivity is discussed in detail in main text; for full WGCNA results across all 37 modules see **Tables S4.8-S4.13**.

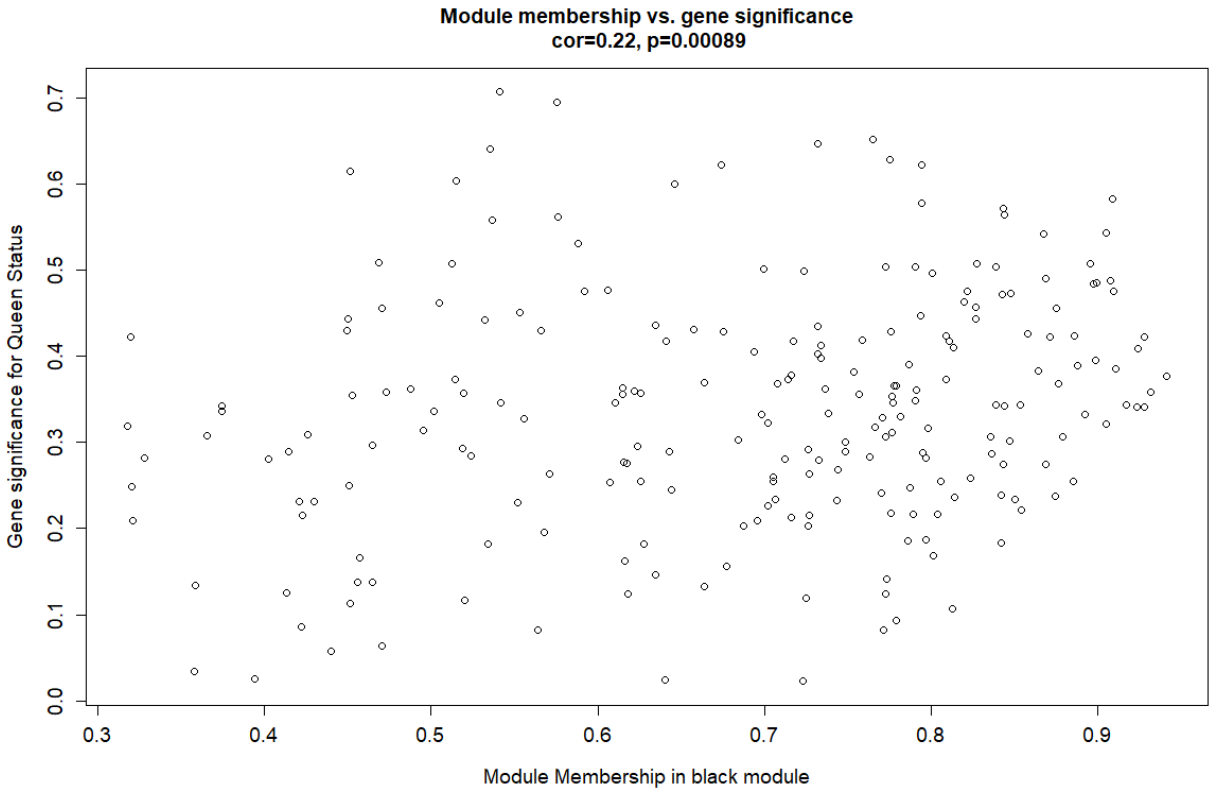

Figure S4.9. A scatterplot of gene significance for association with queens vs. module membership in the black module. The significant positive correlation between association with queen status and membership in the black module suggests expression of genes in this set is closely tied to queen phenotype in *C. japonica*.

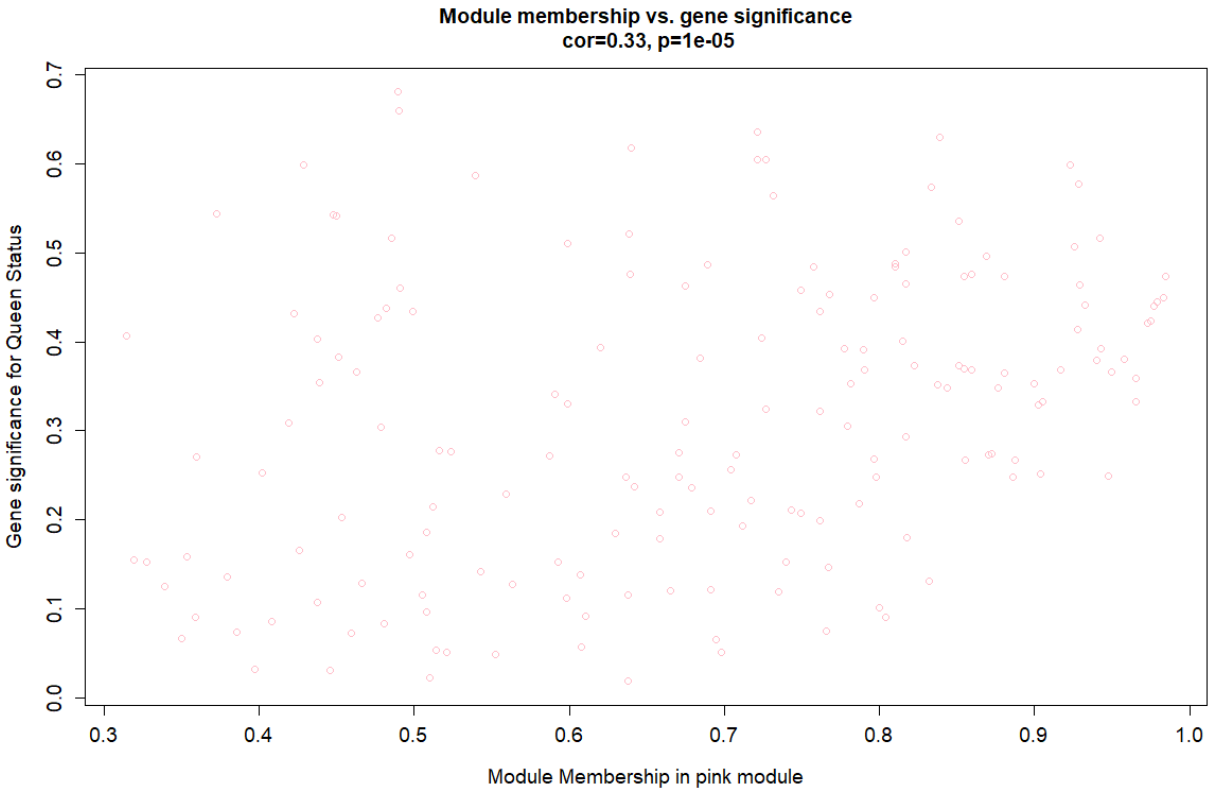

Figure S4.10. A scatterplot of gene significance for association with queens vs. module membership in the pink module. The highly significant positive correlation between association with queen status and membership in the pink module suggests expression of genes in this set is closely tied to queen phenotype in *C. japonica*.

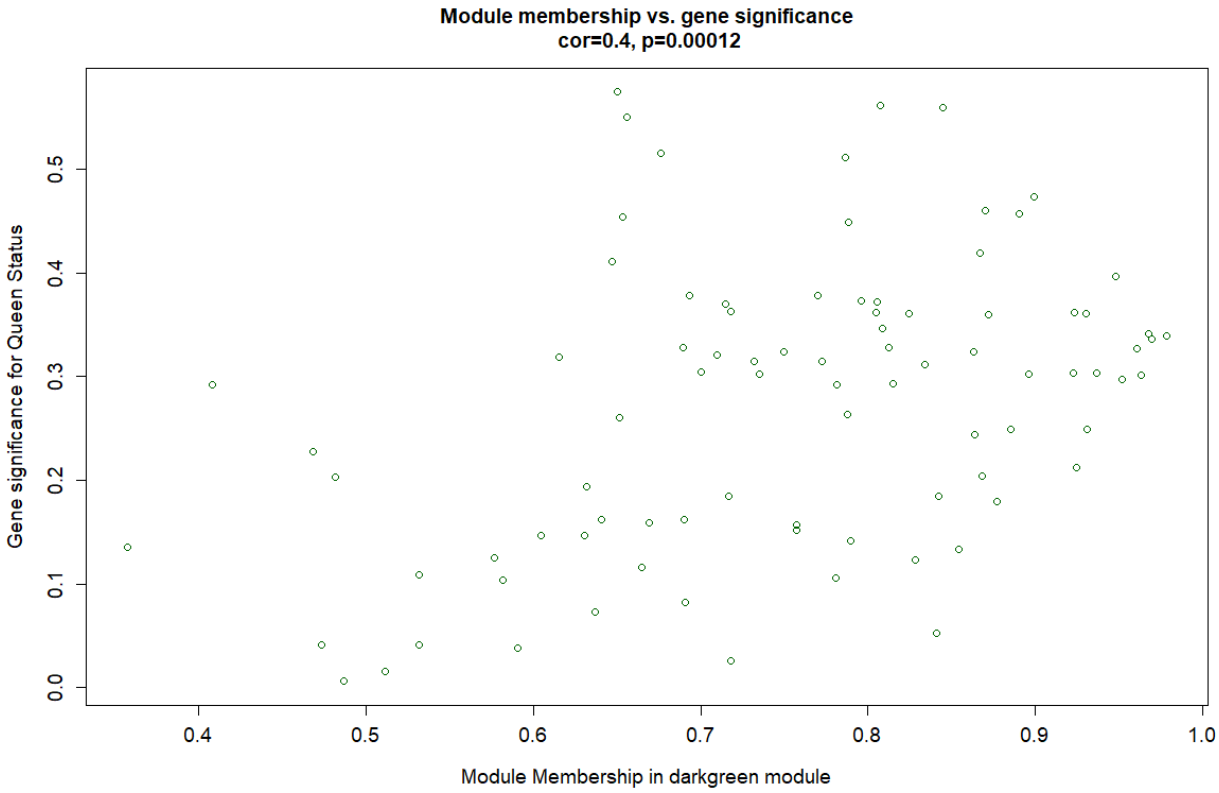

Figure S4.11. A scatterplot of gene significance for association with queens vs. module membership in the dark green module. The significant positive correlation between association with queen status and membership in the dark green module suggests expression of genes in this set is closely tied to queen phenotype in *C. japonica*.

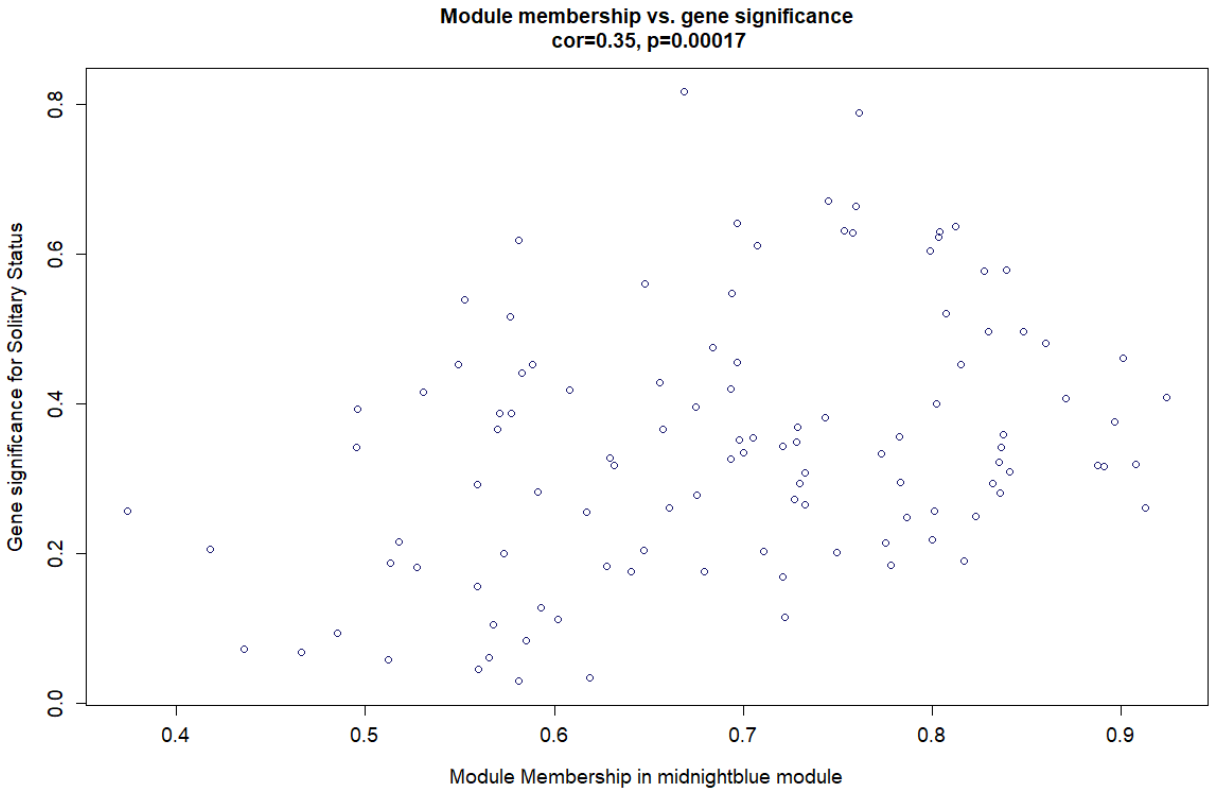

Figure S4.12. A scatterplot of gene significance for association with solitary females vs. module membership in the midnight blue module. The significant positive correlation between association with solitary female status and membership in the midnight blue module suggests expression of genes in this set is closely tied to solitary female phenotype in *C. japonica*.

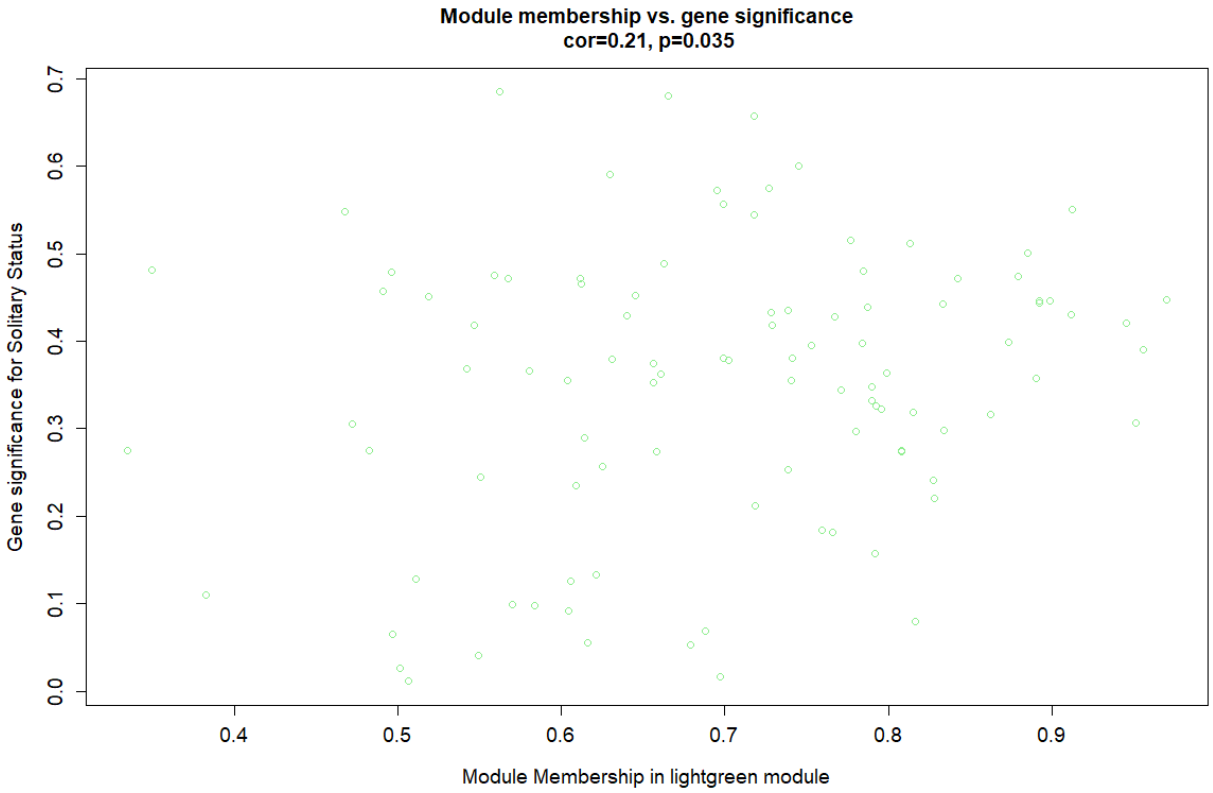

Figure S4.13. A scatterplot of gene significance for association with solitary females vs. module membership in the light green module. The significant positive correlation between association with solitary female status and membership in the light green module suggests expression of genes in this set is closely tied to solitary female phenotype in *C. japonica*.

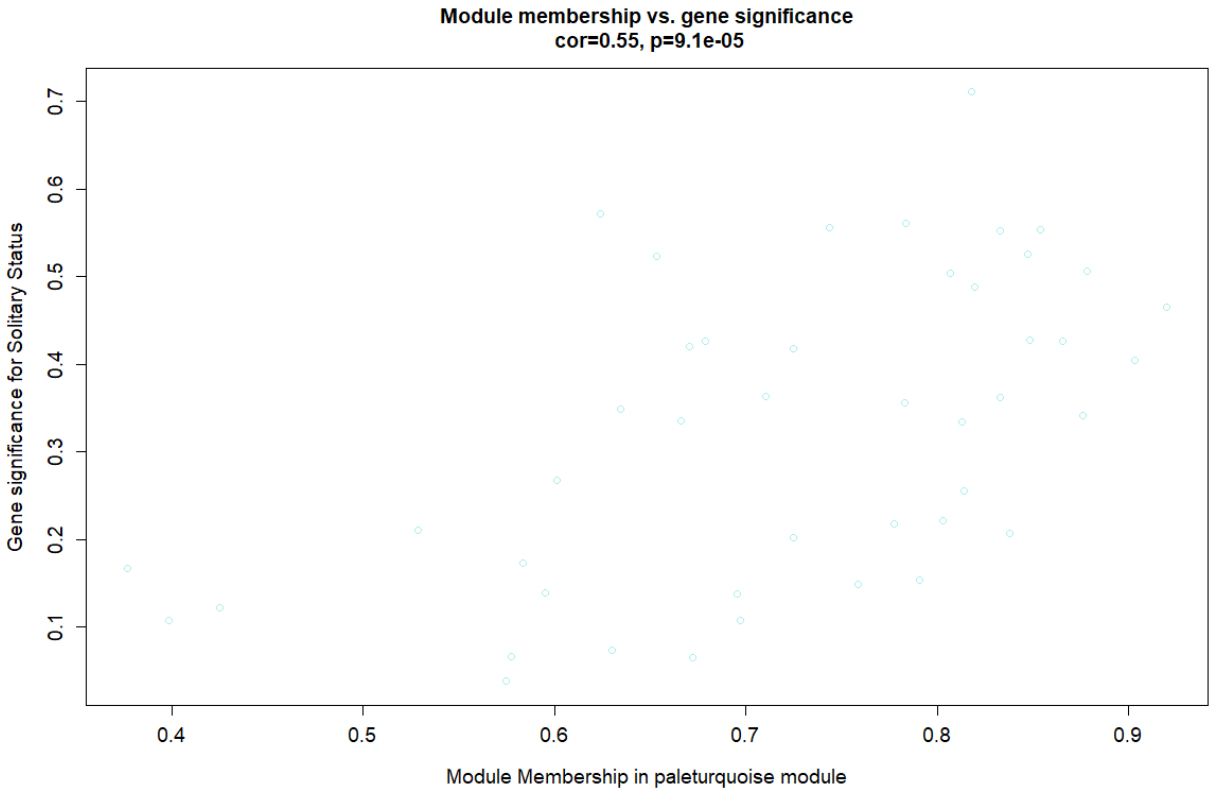

Figure S4.14. A scatterplot of gene significance for association with solitary females vs. module membership in the pale turquoise module. The highly significant positive correlation between association with solitary female status and membership in the pale turquoise module suggests expression of genes in this set is closely tied to solitary female phenotype in *C. japonica*.

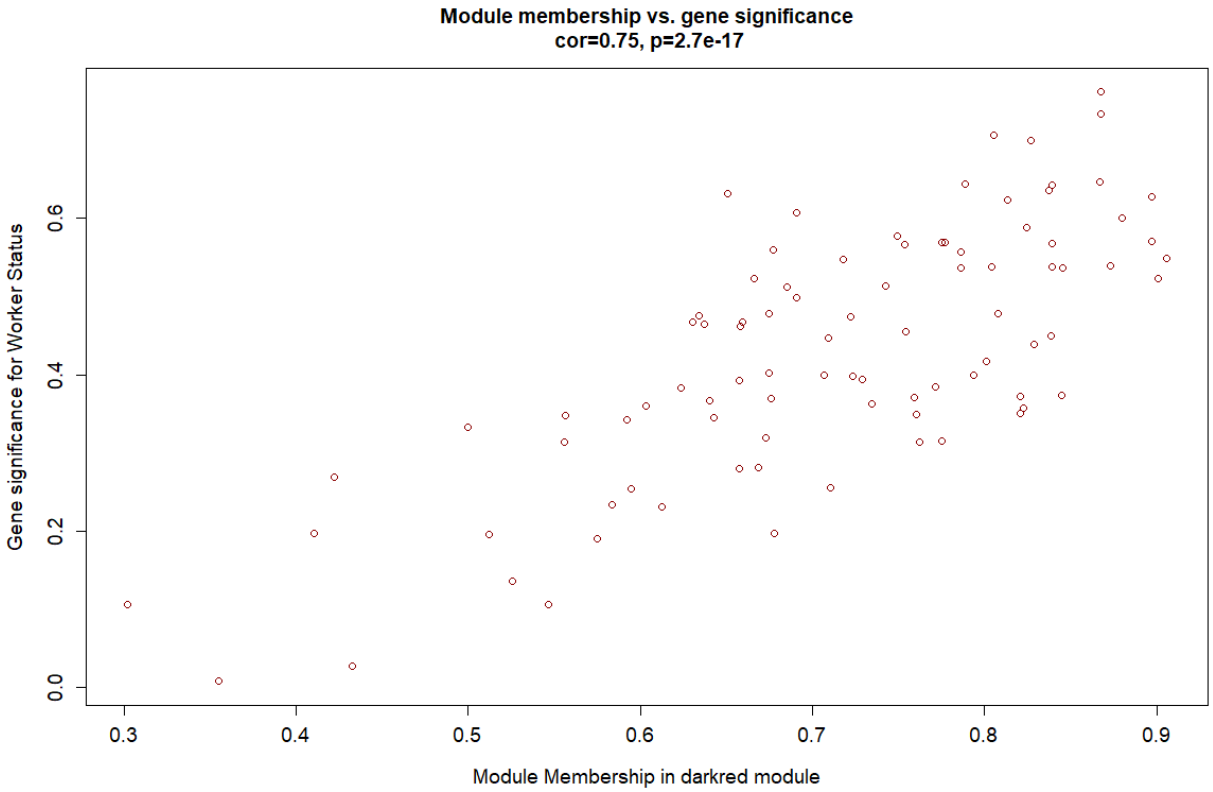

Figure S4.15. A scatterplot of gene significance for association with workers vs. module membership in the dark red module. The highly significant and positive correlation between association with worker status and membership in the dark red module suggests expression of genes in this set is closely tied to worker phenotype in *C. japonica*.

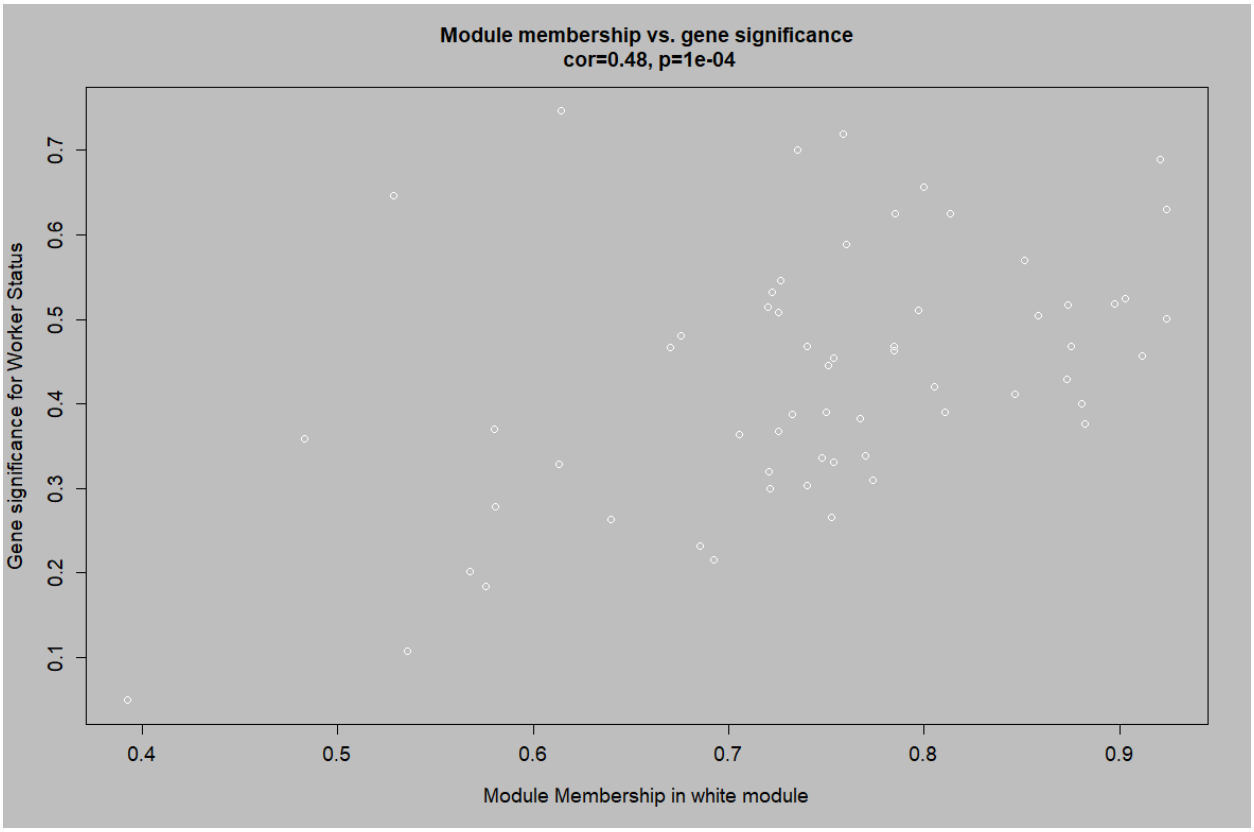

Figure S4.16. A scatterplot of gene significance for association with workers vs. module membership in the white module (panel has been manually darkened to allow visualization of white-colored data points). The highly significant and positive correlation between association with worker status and membership in the white module suggests expression of genes in this set is closely tied to worker phenotype in *C. japonica*.

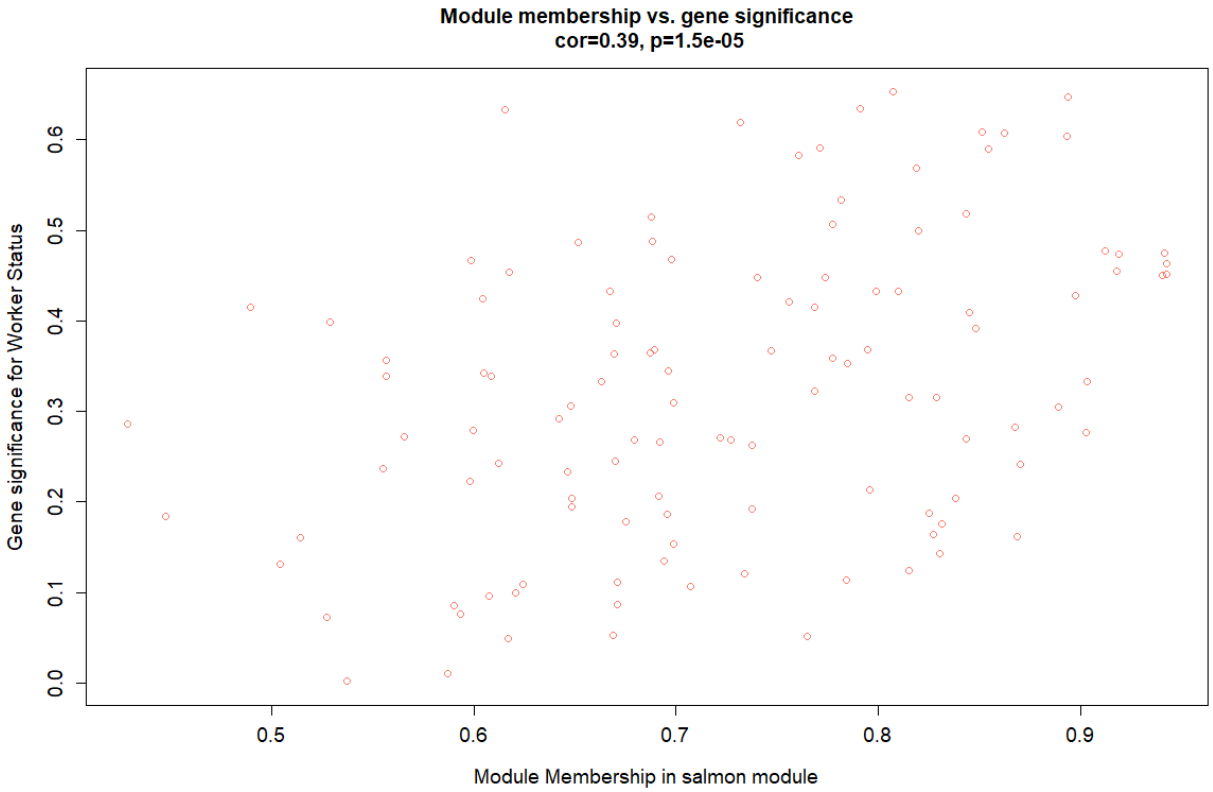

Figure S4.17. A scatterplot of gene significance for association with workers vs. module membership in the salmon module. The highly significant and positive correlation between association with worker status and membership in the salmon module suggests expression of genes in this set is closely tied to worker phenotype in *C. japonica*.

REFERENCES

Abouheif E., Favé M. J., Ibarrarán-Viniegra A. S., Lesoway M. P., Rafiqi A. M., Rajakumar R. 2014 Eco-evo-devo: the time has come. in: Landry, C. R., Aubin-Horth, N. (Eds.) Ecological Genomics. Springer, Netherlands pp. 107-125

Alexa A, Rahnenfuhrer J. 2016 topGO: Enrichment analysis for gene ontology. R package version 2.28.0. *CRAN*

Alexander RD. 1974 The evolution of social behavior. *Annu. Rev. Ecol. Evol. Syst*. **5**, 325-383.

Ament SA, Blatti CA, Alaux C, Wheeler MM, Toth AL, Conte YL, Hunt GJ, Guzmán-Novoa E, DeGrandi-Hoffman G, Uribe-Rubio JL, Amdam GV, Page RE, Rodriguez-Zas SL, Robinson GE, Sinha S. 2012 New meta-analysis tools reveal common transcriptional regulatory basis for multiple determinants of behavior. *PNAS* **109**, E1801-E1810

Amdam G. V., Norberg K., Fondrk M. K., Page R. E. 2004 Reproductive ground plan may mediate colony-level selection effects on individual foraging behavior in honey bees. *Proc. Natl. Acad. Sci. U.S.A.* **101** (21), 11250-11255

Amdam G. V., Csondes A., Fondrk M. K., Page R. E. 2006 Complex social behaviour derived from maternal reproductive traits. *Nature*. **439** (7072), 76-78

Ament S. A., Corona M., Pollock H. S., Robinson G. E. 2008 Insulin signaling is involved in the regulation of worker division of labor in honey bee colonies. *Proc. Natl. Acad. Sci. U.S.A* **105** (11), 4226-4231

Anders S, Huber W. 2010 Differential expression analysis for sequence count data. *Genome Biology* **11**, R106.

Arneson J, Wcislo WT. 2003 Dominant-subordinate relationships in a facultatively social, nocturnal bee, *Megalopta genalis* (Hymenoptera: Halictidae). *J. Kans. Entomol. Soc*. **76** (2), 183-193.

Augusto SC, Garófalo CA. 2004 Nesting biology and social structure of *Euglossa (Euglossa) townsendi* Cockerell (Hymenoptera, Apidae, Euglossini). *Insect. Soc.* **51** (4), 400-409.

Azuma N, Takahashi J, Kidokoro M, Higashi S. 2005 Isolation and characterization of microsatellite loci in the bee *Ceratina flavipes*. *Mol Ecol Notes*. **5**, 433-435.

Baer B, Schmid-Hempel P. 2001 Unexpected consequences of polyandry for parasitism and fitness in the bumblebee, *Bombus terrestris*. *Evolution*. **55** (8), 1639-1943.

Barchuk A., Cristino A. S., Kucharski R., Costa L. F., Simões Z. L. P., Maleszka R. 2007 Molecular determinants of caste differentiation in the highly eusocial honeybee *Apis mellifera*. *BMC Dev. Biol.* **7** (70), 1-19

Bartomeus I, Ascher JS, Gibbs J, Danforth BN, Wagner DL, Hedtke SM, Winfree R. 2013 Historical changes in northeastern US bee pollinators related to shared ecological traits. *P Natl Acad Sci USA.* **110**, 4656-4660.

Bastian M, Heymann S, Mathieu J. 2009 Gephi: an open source software for exploring and manipulating networks. *International AAAI Conference on Weblogs and Social Media*

Bentzur A, Shmueli A, Omesi L, Ryvkin J, Knapp J-M, et al. 2018 *Odorant binding protein 69a* connects social interaction to modulation of social responsiveness in *Drosophila*. *PLoS Genet.* **14**, e1007328

Berens A. J., Hunt J. H., Toth A. L. 2014 Comparative transcriptomics of convergent evolution: different genes but conserved pathways underlie caste phenotypes across lineages of eusocial insects. *Mol. Biol. Evol*. **32** (3), 690-703

Bewick A. J., Vogel K. J., Moore A. J., Schmitz R. J. 2017 The evolution of DNA methylation and its relationship to sociality in insects. *BioRxive* (doi: https://doi.org/10.1101/062455)

Blacher P, Huggins TJ, Bourke AFG. 2017 Evolution of ageing, costs of reproduction and the fecundity-longevity trade-off in eusocial insects. *Proc. R. Soc. B.* **284**, 20170380

Bloch G., Grozinger C. M. 2011 Social molecular pathways and the evolution of bee societies. *Phil. Trans. R. Soc. B.* **366**, 2155-2170

Boomsma J. J. 2009 Lifetime monogamy and the evolution of eusociality. *Phil. Trans. R. Soc. B.* **364**, 3191-3207

Bourke AFG. 2011 The validity and value of inclusive fitness theory. *Proc. R. Soc. B.* **278** (1723), 3313-3320.

Brady S. G., Sipes S., Pearson A., Danforth B. N. 2006 Recent and simultaneous origins of eusociality in halictid bees. *Proc. R. Soc. B.* **273**, 1643-1649

Brand N., Chapusat M. 2012 Born to be bee, fed to be worker? The caste system of a primitively eusocial insect. *Front. Zool*. **9** (35), 1-9

Branstetter M. G., Danforth B. N., Pitts J. P., Faircloth B. C., Ward P. S., Buffington M. L., Gates M. W., Kula R. R., Brady S. G. 2017 Phylogenomic insights into the evolution of stinging wasps and the origins of ants and bees. *Curr. Biol*. **27**, 1019-1025

Brittain C, Williams N, Kremen C, Klein AM. 2012 Synergistic effects of non-*Apis* bees and honey bees for pollination services. *Proc. R. Soc. B.* **280**, 20122767.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, *et al*. 2009 BLAST+: architecture and applications.

Cameron SA, Lozier JD, Strange JP, Koch JB, Cordes N, Solter LF, Griswold TL. 2011 Patterns of widespread decline in North American bumble bees. *P Natl Acad Sci USA*. **108**, 662-667.

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

Cardoso SD, Teles MC, Oliveira RF. 2015 Neurogenomic mechanisms of social plasticity. *J. Exp. Biol*. **218**, 140-149

Carroll S. B. 2005 Endless forms most beautiful. W. W. Norton

Carroll S. B. 2008 Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell*. **134**, 25-36

Cash AC, Whitfield CW, Ismail N, Robinson GE. 2005 Behavior and the limits of genomic plasticity: power and replicability in microarray analysis of honeybee brains. *Genes Brain Behav*. **4**, 267-271

Castoe TA, Poole AW, Gu W, De Koning APJ, Daza JM, Smith EN, Pollock DD. 2010 Rapid identification of thousands of copperhead snake (*Agkistrodon contortrix*) microsatellite loci from modest amounts of 454 shotgun genome sequence. *Mol. Ecol. Resour*. **10**, 341-347.

Černá K., Zemenová M., Macháčková L., Kolínová Z., Straka J. 2013 Neighborhood society: nesting dynamics, usurpations and social behaviour in solitary bees. *PLoS One*. **8** (8), e73806

Chandrasekaran S., Ament S. A., Eddy J. A., Rodriguez-Zas S. L., Schatz B. R., Price N. D., Robinson G. E. 2011 Behavior-specific changes in transcriptional modules lead to distinct and predictable neurogenomic states. *Proc. Natl. Acad. Sci. U.S.A.* **108** (44), 18020-18025

Chen S., Krinsky B. H., Long M. 2013 New genes as drivers of phenotypic evolution. *Nature Rev. Genet*. **14**, 645-660

Chen W, Fang L, Liu L, He Z, Hu HY. 2015 Isolation and characterization of polymorphic microsatellite loci for *Pachycrepoideus vindemmiae* (Rondani) (Hymenoptera: Pteromalidae). *Genet. Mol. Res*. **14**, 1798-1801.

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

Conesa A, Nueda MJ. 2018 maSigPro: significant gene expression profile differences in time course gene expression data. R package version 1.54.0 (DOI: 10.18129/B9.bioc.maSigPro)

Corona M, Hughes KA, Weaver DB, Robinson GE. 2005 Gene expression patterns associated with queen honey bee longevity. *Mechanisms of Ageing and Development* **126**, 1230-1238

Corona M., Libbrecht R., Wheeler D. E. 2016 Molecular mechanisms of phenotypic plasticity in social insects. *Curr. Opin. Insect. Sci*. **13**, 55-60

Craig, R. 1979 Parental manipulation, kin selection, and the evolution of altruism. *Evolution.* **33** (1), 319-334.

Crespi BJ, Ragsdale JE. 2000 A skew model for the evolution of sociality via manipulation: why it is better to be feared than loved. *Proc. R. Soc. B.* **267** (1445), 821-828.

Crespi BJ. 1992 Eusociality in Australian gall thrips. *Nature*. **359** (6397), 724-726.

Crozier RH, Pamilo P. 1996 Evolution of social insect colonies. Oxford University Press, Oxford, UK.

Danforth B. N. 2002 Evolution of sociality in a primitively eusocial lineage of bees. *Proc. Natl. Acad. Sci. USA* **99** (1), 286-290

Danforth B. N., Cardinal S., Praz C., Almedia E. A. B., Michez D. 2013 The impact of molecular data on our understanding of bee phylogeny and evolution. *Annu. Rev. Entomol*. **58**, 57-78

Danforth B. N., Conway L., Shuqing J. 2003 Phylogeny of eusocial *Lasioglossum* reveals multiple losses of eusociality within a primitively eusocial clade of bees (Hymenoptera: Halictidae) *Syst. Biol*. **52** (1), 23-36

Danforth B. N., Eickwort G. C. 1997 The evolution of social behavior in the augochlorine sweat bees (Hymenoptera: Halictidae) based on a phylogenetic analysis of the genera. in: Choe J. C., Crespi B. J. (Eds.) The Evolution of Social Behavior in Insects and Arachnids. Cambridge University Press, Cambridge, UK. pp. 270-292

Dani FR. 2009 Cuticular lipids as semiochemicals in paper wasps and other social insects. *Ann. Zool. Fennici.* **43**, 500-514

Davison P. J., Field J. 2016 Social polymorphism in the sweat bee *Lasioglossum* (*Evylaeus*) *calceatum*. *Insect. Soc*. **63**, 327-338

Dew R. M., Tierney S. M., Schwarz M. P. 2016 Social evolution and casteless societies: needs for new terminology and a new evolutionary focus. *Insect. Soc*. **63**, 5-14

Doganitz KA, Harpur BA, Rodrigues A, Beani L, Toth AL, Zayed A. 2018 Insects with similar social complexity show convergent patterns of adaptive molecular evolution. *Sci. Rep.* **8**:10388

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

Doums C, Moret Y, Benelli E, Schmid-Hempel P. 2002 Senescence of immune defence in *Bombus* workers. *Ecol. Entomol.* **27**, 138-144

Dreier S, Redhead JW, Warren IA, Bourke AFG, Heard MS, Jordan WC, Sumner S, Wang J, Carvell C. 2014 Fine-scale spatial genetic structure of common and declining bumble bees across an agricultural landscape. *Mol Ecol*. **23**, 3384-3395.

Duffy JE. 1996 Eusociality in a coral-reef shrimp. *Nature*. **381** (6582), 512.

Durant DR, Berens AJ, Toth AL, Rehan SM. 2016 Transcriptional profiling of overwintering gene expression in the small carpenter bee, *Ceratina calcarata*. *Apidologie*. **47** (4), 572-582.

Dussaubat C, Maisonnasse A, Crauser D, Tehamitchian S, Bonnet M, Cousin M, Kretzschmar A, Brunet J-L, Le Conte Y. 2016 Combined neonicontinoid pesticide and parasite stress alter honey bee queens' physiology and survival. *Scientific Reports* **6**, 31430

Eickwort G. C. 1986 First steps into eusociality: the sweat bee *Dialictus lineatulus*. *Fla. Entomol*. **69**, 742-754

Eickwort G. C., Eickwort J. M., Gordon J., Eickwort M. A. 1996 Solitary behavior in a high-altitude population of the social sweat bee Halictus rubicundus (Hymenoptera: Halictidae). *Behav. Ecol. Sociobiol*. **38** (4), 227-233

Ekblom R, Galindo J. 2011 Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity*. **107**, 1-15.

Emlen D. J., Nijhout H. F. 2000 The development and evolution of exaggerated morphologies in insects. *Annu. Rev. Entomol*. **45**, 661-708

Emms DM, Kelly S. 2015 OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* **16**, 157

Espinosa-Soto C, Wagner A. 2010 Specialization can drive the evolution of modularity. *PLOS Comput. Biol.* **6**, e1000719

Estoup A, Scholl A, Pouvreau A, Solignac M. 1995 Monoandry and polyandry in bumble bees (Hymenoptera; Bombinae) as evidenced by highly variable microsatellites. *Mol. Ecol*. **4** (1), 89-93.

Evans H. E. 1977 Commentary: extrinsic versus intrinsic factors in the evolution of insect sociality. *Bioscience*. **27** (9), 613-617

Evans JD, Aronstein K, Chen YP, Hetru C, Imler J-L, Jiang H, Kanost M, Thompson GJ, Zou Z, Hultmark D. 2006 Immune pathways and defence mechanisms in honey bees *Apis mellifera*. *Insect Molecular Biology* **15**, 645-656

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

Evans J. D., Wheeler D. E. 2001 Gene expression and the evolution of insect polyphenisms. *BioEssays*. **23**, 62-68

Fan J, Francis F, Liu Y, Chen JL, Cheng DF. 2011 An overview of odorant-binding protein functions in insect peripheral olfactory reception. *Genet. Mol. Res.* **10**, 3056-3069

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

Ferreira P. G., Patalano S., Chauhan R., Ffrench-Constant R., Gabaldón T., Guigó R., Sumner S. 2013 Transcriptome analyses of primitively eusocial wasp reveal novel insights into the evolution of sociality and the origin of alternative phenotypes. *Genome Biol*. **14** (2), 1-14

Field J., Paxton R. J., Soro A., Bridge C. 2010 Cryptic plasticity underlies a major evolutionary transition. *Curr. Biol*. **20** (22), 2028-2031

Field J, Paxton R, Soro A, Craze P, Bridge C. 2012 Body size, demography and foraging in a socially plastic sweat bee: a common garden experiment. *Behav. Ecol. Sociobiol*. **66** (5), 743-756.

Fischman B. J., Woodard S. H., Robinson G. E. 2011 Molecular evolutionary analyses of insect societies. *Proc. Natl. Acad. Sci. U.S.A.* **108** (Supplement 2), 10847-10854

Fjerdingstad E. J., Crozier R. H. 2006 The evolution of worker caste diversity in social insects. *Amer. Nat*. **167** (3), 390-400

Forêt S., Kucharski R., Pellegrini M., Feng S., Jacobsen S. E., Robinson G. E., Maleszka R. 2012 DNA methylation dynamics, metabolic fluxes, gene splicing, in alternative phenotypes in honey bees. *Proc. Natl. Acad. Sci. U.S.A.* **109** (13), 4968-4973

Forêt S, Maleszka R. 2006 Function and evolution of a gene family encoding odorant binding-like proteins in a social insect, the honey bee (*Apis mellifera*). *Genome Research* **16**, 1404-1413

Foster KR, Wenseleers T, Ratnieks FLW. 2006 Kin selection is the key to altruism. *Trends Ecol. Evol*. **21** (2), 57-60.

Fu F., Kocher S. D., Nowak M. A. 2015 The risk-return trade-off between solitary and eusocial reproduction. *Ecol. Lett.* **18**, 74-84

Gadagkar R. 1990 Evolution of eusociality: the advantage of assured fitness returns. *Phil. Trans. R. Soc. B.* **329** (1252), 17-25.

Gadagkar R. 1997 The evolution of caste polymorphism in social insects: genetic release followed by diversifying evolution. *J. Genet*. **76** (3), 167-179

Gadagkar R. 2001 The social biology of *Ropalidia marginata*: toward understanding the evolution of eusociality. Harvard University Press.

Gadagkar R. 2010 Sociobiology in turmoil again. *Curr. Sci*. **99** (8), 1036-1041.

Geib JC, Strange JP, Galen C. 2015 Bumble bee nest abundance, foraging distance, and host-plant reproduction: implications for management and conservation. *Ecol Appl*. **25**, 768-778.

Gibbs J., Brady S. G., Kanda K., Danforth B. N. 2012 Phylogeny of halictine bees supports a shared origin of eusociality for *Halictus* and *Lasioglossum* (Apoidea: Anthophila: Halictidae). *Mol. Phylogenet. Evol*. **65**, 926-939

Gilbert S. F. 2012 Ecological developmental biology: environmental signals for normal animal development. *Evol. Dev*. **14** (1), 20-28

Gilbert S. F., Bosch T. C. G., Ledón-Rettig C. 2015 Eco-evo-devo: developmental symbiosis and developmental plasticity as evolutionary agents. *Nature Rev. Genet*. **16**, 611-622

Glastad KM, Arsenault SV, Vertacnik KL, Geib SM, Kay S, *et al*. 2017 Variation in DNA methylation is not consistently reflected by sociality in Hymenoptera. *GBE* **9**, 1687-1698

Glastad K. M., Hunt B. G., Yi S. V., Goodisman M. A. D. 2011 DNA methylation in insects: on the brink of the epigenetic era. *Insect Mol. Biol*. **20** (5), 553-565

Glenn TC, Schable NA. 2005 Isolating microsatellite DNA loci. *Method Enzymol*. **395**, 202-222.

Goldberg A. D., Allis C. D., Bernstein E. 2007 Epigenetics: a landscape takes shape. *Cell*. **128**, 635-638

Gospocic J, Shields EJ, Glastad KM, Lin Y, Penick CA, Yan H, Mikheyev AS, Linksvayer TA, Garcia BA, Berger SL, Liebig J, Reinberg D, Bonasio R. 2017 The neuropeptide corazonin controls social behavior and caste identity in ants. *Cell*. **170**, 748-759

Grover A, Sharma PC. 2016 Development and use of molecular markers: past and present. *Crit Rev Biotechnol*. **32**, 290-302.

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

Hamilton WD. 1964 The genetical evolution of social behavior. *J. Theor. Biol*. **7**, 1-52.

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

Harman D. 1992 Free radical theory of aging. *Mutation Research/DNAging* **275**, 257-266

Harpur B. A., Kent C. F., Molodtsova D., Lebon J. M. D., Alqarni A. S., Owayss A. A., Zayed A. 2014 Population genomics of the honey bee reveals strong signatures of positive selection on worker traits. *Proc. Natl. Acad. Sci. U.S.A.* **111** (7), 2614-2619

Harpur B. A., Zayed A. 2013 Accelerated evolution of innate immunity proteins in social insects: adaptive evolution or relaxed constraint? *Mol. Biol. Evol*. **30** (7), 1665-1674

Hirata M., Higashi S. 2008 Degree-day accumulation controlling allopatric and sympatric variations in the sociality of sweat bees, *Lasioglossum* (*Evylaeus*) *baleicum* (Hymenoptera: Halictidae). *Behav. Ecol. Sociobiol*. **62**, 1239-1247

Holbrook C. T., Clark R. M., Jeanson R., Bertram S. M., Kukuk P. F., Fewell J. H. 2009 Emergence and consequences of division of labor in association of normally solitary sweat bees. *Ethology*. **115**, 301-310

Honeybee Genome Sequencing Consortium 2006 Insights into social insects from the genome of the honeybee *Apis mellifera*. *Nature*. **443**, 931-949

Hsu C-Y, Hsieh Y-S. 2014 Oxidative stress decreases in the trophocytes and fat cells of worker honey bees during aging. *Biogerontology* **15**, 129-137

Hudson ME. 2008 Sequencing breakthroughs for genomic ecology and evolutionary biology. *Mol Ecol Resour*. **8**, 3-17

Hughes CR, Strassmann JE. 1988 Age is more important than size in determining dominance among workers in the primitively eusocial wasp, *Polistes instabilis*. *Behaviour*. **107**, 1-14

Hughes WOH, Oldroyd BP, Beekman M, Ratnieks FLW. 2008 Ancestral monogamy shows kin selection is key to the evolution of eusociality. *Science*. **320** (5880), 1213-1216

Hunt B. G., Glastad K. M., Yi S. V., Goodisman M. A. D. 2013 The function of intragenic DNA methylation: insights from insect epigenomes. *Integr. Comp. Biol*. **53** (2), 319-328

Hunt J. H., Wolschin F., Henshaw M. T., Newman T. C., Toth A., L., Amdam G. V. 2010 Differential gene expression and protein abundance evince otogenetic bias toward castes in a primitively eusocial wasp. *PLoS One*. **5** (5), e10674

Iovinella I, Dani FR, Niccolini A, Sagona S, Michelucci E, Gazzano A, Turillazzi S, Felicioli A, Pelosi P. 2011 Differential expression of odorant-binding proteins in the mandibular glands of the honey bee according to caste and age. *Journal of Proteome Research* **10**, 3439-3449

Jacomy M, Venturini T, Heymann S, Bastian M. 2014 ForceAtlas2, a continuous graph layou0t algorithm for handy network visualization designed for the Gephi software. *PloS One* **9**, e98679

Jandt JM, Tibbets EA, Toth AL. 2014 Polistes paper wasps: a model genus for the study of social dominance hierarchies. *Insect. Soc*. **61** (1), 11-27.

Jarvis JUM. 1981. Eusociality in a mammal: cooperative breeding in naked mole-rat colonies. *Science*. **212** (4494), 571-573.

Jeanson R., Kukuk P. F., Fewell J. H. 2005 Emergence of division of labour in halictine bees: contributions of social interactions and behavioural variance. *Anim. Behav*. **70**, 1183-1193

Jeanson R., Clark R., Holbrook C., Bertram S., Fewell J., Kukuk P. 2008 Division of labour and socially induced changes in response thresholds in associations of solitary halictine bees. *Anim. Behav*. **76** (3), 593-602

Johnson B. R., Linksvayer T. A. 2010 Deconstructing the superorganism: social physiology, groundplans, and sociogenomics. *Q. Rev. Biol.* **85** (1), 57-79

Johnson B. R., Tsutsui N. D. 2011 Taxonomically restricted genes are associated with the evolution of sociality in the honey bee. *BMC Genomics*. **12** (164), 1-10

Johnson MD. 1988 The relationship of provision weight to adult weight and sex ratio in the solitary bee, *Ceratina calcarata*. *Ecol. Entomol*. **13**, 165-170.

Jones B. M., Kingwell C. J., Wcislo W. T., Robinson G. E. 2017 Caste-biased gene expression in a facultatively eusocial bee suggests a role for genetic accommodation in the evolution of eusociality. *Proc. R. Soc. B*. **284** (1846), 1-9

Jones B. M., Wcislo W. T., Robinson G. E. 2015 Developmental transcriptome for a facultatively eusocial bee, *Megalopta genalis*. *G3 Genses Genom. Genet*. **5**, 2127-2135

Kalinowski ST, Wagner AP, Taper ML. 2006 ML-Relate: a computer program for maximum likelihood estimation of relatedness and relationship. *Mol. Ecol. Notes* **6**, 576-579.

Kapheim K. M. 2016 Genomic sources of phenotypic novelty in the evolution of eusociality in insects. *Curr. Opin. Insect Sci*. **13**, 24-32

Kapheim KM, Bernal SP, Smith AR, Nonacs P, Wcislo WT. 2011 Support for maternal manipulation of developmental nutrition in a facultatively eusocial bee, *Megalopta genalis* (Halictidae). *Behav. Ecol. Sociobiol*. **65** (6), 1179-1190

Kapheim KM, Chan T-Y, Smith AR, Wcislo WT, Nonacs P. 2016 Ontogeny of division of labor in a facultatively eusocial sweat bee *Megalopta genalis*. *Insect. Soc*. **63** (1), 185-191

Kapheim KM, Nonacs P, Smith AR, Wayne RK, Wcislo WT. 2015 Kinship, parental manipulation and evolutionary origins of eusociality. *Proc. R. Soc. B.* **282** (1803), 20142886.

Kapheim K. M., Pan H., Li C., Salzber S. L., Puiu D., et al. 2015 Genomic signatures of evolutionary transitions from solitary to group living. *Science.* **348** (6239), 1139-1143

Kapheim K. M., Smith A. R., Nonacs P., Wcislo W. T., Wayne R. K. 2013 Foundress polyphenism and the origins of eusociality in a facultatively eusocial sweat be, *Megalopta genalis* (Halictidae). *Behav. Ecol. Sociobiol*. **67**, 331-340

Keller L, Jemielity S. 2005 Social insects as a model to study the molecular basis of ageing. *Experimental Gerontology* **41**, 553-556

Kent CF, Minaei S, Harpur BA, Zayed A. 2012 Recombination is associated with the evolution of genome structure and worker behavior in honey bees. *Proc. Natl. Acad. Sci.* **109**, 18012-18017

Khamis AM, Hamilton AR, Medvedeva YA, Alam T, Alam I, et al. 2015 Insights into the transcriptional architecture of behavioral plasticity in the honeybee *Apis mellifera*. *Scientific Reports* **5**, 11136

Khan A, Fornes O, Stigliani A, Gheorghe M, Castro-Mondragon JA, et al. 2018 JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic Acids Res.* **46**, D260-D266

Kijas JMH, Fowler JCS, Garbett CA, Thomas MR. 1994 Enrichment of microsatellites from the citrus genome using biotinylated oligonucleotide sequences bound to streptavidin-coated magnetic particles. *BioTechniques*. **16**, 656-662.

Kirby KS. 1956. Isolation and characterization of ribosomal ribonucleic acid. *Biochem. J*. **96**, 266-269

Klein AM, Vaissière BE, Cane JH, Steffan-Dewenter I, Cunningham SA, Kremen C, Tscharntke T. 2007 Importance of pollinators in changing landscapes for world crops. *Proc. R. Soc. B.* **274**: 303-313

Koch SI, Groh K, Vogel H, Hannson BS, Kleineidam CJ, Grosse-Wilde E. 2013 Caste-specific expression patterns of immune response and chemosensory related genes in the leaf-cutting ant, Atta vollenweideri. *PLoS One*. **8**, e81518

Kocher S. D., Li C., Yang W., Tan H., Yi S. V., et al. 2013 The draft genome of a socially polymorphic halictid bee, *Lasioglossum albipes*. *Genome Biol*. **14** (12), 1-14

Kocher SD, Mallarino R, Rubin BER, Yu DW, Hoekstra HE, Pierce NE. 2018 The genetic basis of a social polymorphism in halictid bees. *Nat. Comm.* **9**, 4338

Kocher S. D., Paxton R. 2014 Comparative methods offer powerful insights into social evolution in bees. *Apidologie*. **45**, 289-305

Kocher S. D, Pellisier L., Veller C., Purcell J., Nowak M. A. et al. 2014 Transitions in social complexity along elevational gradients reveal a combined impact on season length and development time on social evolution. *Proc. R. Soc. B.* **281**, 1-8

Kohlmeier P, Alleman AR, Libbrecht R, Foitzik S, Feldmeyer B. 2019 Gene expression is stronger associated with behaviour than with age and fertility in ant workers. *Mol. Ecol*. **28**, 658-670

Koressaar T, Remm M. 2007 Enhancements and modifications of primer design program Primer 3. *Bioinformatics*. **23**, 1289-1291

Kremen C, Williams NM, Thorp RW. 2002 Crop pollination from native bees at risk from agricultural intensification. *Proc. Natl. Acad. Sci. USA*. **99**, 16812-16816

Laidlaw HH, Page RE. 1984 Polyandry in honey bees (*Apis mellifera* L.): sperm utilization and intracolony genetic relationships. *Genetics*. **108** (4), 985-997.

Langfelder P, Horvath S. 2008 WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. **9**, 559

Langfelder P, Horvath S. 2014 Tutorial for the WGCNA package for R. https://horvath.genetics.ucla.edu/html/CoexpressionNetwork/Rpackages/WGCNA/Tutorials/

Larsen J, Nehring V, d'Ettorre P, Bos N. 2016 Task specialization influences nestmate recognition ability in ants. *Behav. Ecol. Sociobiol*. **70** (9), 1433-1440

Lawson SP, Ciaccio KN, Rehan SM. 2016 Maternal manipulation of pollen provisions affects worker production in a small carpenter bee. *Behav. Ecol. Sociobiol*. **70** (11), 1891-1900

Leadbeater E, Carruthers JM, Green JP, Rosser NS, Field J. 2011 Nest inheritance is the missing source of direct fitness in a primitively eusocial insect. *Science*. **333** (6044), 874-876

Leonhardt SD, Menzel F, Nehring V, Schmitt T. 2016 Ecology and evolution of communication in social insects. *Cell*. **164**, 1277-1287

Lewis V, Richards MH. 2017 Experimentally induced alloparental care in a solitary carpenter bee. *Anim. Behav*. **123**, 229-238.

Liang ZS, Mattila HR, Rodriguez-Zas SL, Southey BR, Seeley TD, Robinson GE. 2014 Comparative brain transcriptomic analyses of scouting across distinct behavioural and ecological contexts in honeybees. *Proc. R. Soc. B.* **281**, 20141868

Liang Z. S., Nguyen T., Mattila H. R., Rodriguez-Zas S. L., Seeley T. D., Robinson G. E. 2012 Molecular determinants of scouting behavior in honey bees. *Science*. **335**, 1225-1228

Li-Byarlay H, Cleare X. 2020 Current trends in the oxidative stress and ageing of social hymenopterans. *Advances in Insect Physiology* **59**, 43-69

Li-Byarlay H., Li Y., Stroud H., Feng S., Newman T. C. 2013 RNA interference knockdown of DNA methyl-transferase 3 affects gene alternative splicing in the honey bee. *Proc. Natl. Acad. Sci. U.S.A.* **110** (31), 12750-12755

Libbrecht R, Oxley PR, Kronauer DJC. 2018 Clonal raider ant brain transcriptomics identifies candidate molecular mechanisms for reproductive division of labor. *BMC Genomics*. **16**,1

Lin N., Michener C. D. 1972 Evolution of sociality in insects. *Q. Rev. Biol*. **47** (2), 131-159

Linksvayer T. A., Wade M. J. 2005 The evolutionary origin and elaboration of sociality in the aculeate Hymenoptera: maternal effects, sib-social effects, and heterochrony. *Q. Rev. Biol*. **80** (3), 317-336

Lockett GA, Almond EJ, Huggins TJ, Parker JD, Bourke AFG. 2016 Gene expression differences in relation to age and social environment in queen and worker bumble bees. *Experimental Gerontology*. **77**, 52-61

Love MI, Huber W, Anders S. 2014 Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*. **15**, 550

Lucas ER, Keller L. 2019 The co-evolution of longevity and social life. *Functional Ecology* **34**, 76-87

Lyko F., Foret S., Kucharski R., Wolf S., Falckenhayn C., Maleszka R. 2010 The honey bee epigenomes: differential methylation of brain DNA in queens and workers. *PLoS Biol*. **8** (11), e1000506

Ma M, Jia H, Cui X, Zhai N, Wang H, Guo X, Xu B. 2018 Isolation of carboxylesterase (esterase FE4) from *Apis cerana cerana* and its role in oxidative resistance during adverse environmental stress. *Biochimie* **144**, 85-97

Marshall JAR. 2011 Group selection and kin selection: formally equivalent approaches. *Trends Ecol. Evol*. **26** (7), 325-332.

McFrederick QS, Rehan SM. 2016 Characterization of pollen and bacterial community composition in brood provisions of a small carpenter bee. *Mol Ecol*. **25**, 2302-2311

McKenzie S. K., Fetter-Pruneda I., Ruta V., Kronauer D. J. C. 2016 Transcriptomics and neuroanatomy of the clonal raider ant implicate an expanded clade of odorant receptors in chemical communication. *Proc. Natl. Acad. Sci. USA*. **113**, 14091-14096

Michener C. D. 1969 Comparative social behavior of bees. *Ann. Rev. Entomol*. **14**, 299-342

Michener C. D. 1974 The social behaviour of the bees: a comparative study. Harvard University Press, Cambridge, MA

Michener C. D. 2007 *The bees of the world, 2nd edn.* Baltimore, MD: John Hopkins University Press.

Mikát M, Franchino C, Rehan SM. 2017 Sociodemographic variation in foraging behavior and the adaptive significance of worker production in the facultatively social small carpenter bee, Ceratina calcarata. *Behav. Ecol. Sociobiol*. **71**,135

Miyanaga R., Maeta Y., Sakagami S. F. 1999 Geographical variation of sociality and size-linked color patterns in *Lasioglossum (Evylaeus) apristum* (Vachal) in Japan (Hymenoptera, Halictidae) *Insectes Soc*. **46**, 224-232

Molodtsova D., Harpur B. A., Kent C. F., Seevananthan K., Zayed A. 2014 Pleiotropy constrains the evolution of protein but not regulatory sequences in a transcription regulatory network influencing complex social behaviors. *Front. Genet*. **5** (431), 1-7

Morandin C., Dhaygude K., Paviala J., Trontti K., Wheat C., Helanterä H. 2015 Caste-biases in gene expression are specific to developmental stage in the ant *Formica exsecta*. *J. Evol. Biol*. **28**, 1705-1718

Morandin C., Tin M. M. Y., Abril S., Gómez C., Pontieri L., et al. 2016 Comparative transcriptomics reveals the conserved building blocks involved in parallel evolution of diverse phenotypic traits in ants. *Genome Biol*. **17** (43), 1-19

Moreira AS, Horgan FG, Murray TE, Kakouli-Duarte K. 2015 Population genetic structure of *Bombus terrestris* in Europe: Isolation and genetic differentiation of Irish and British populations. *Mol Ecol*. **24**, 3257-3268.

Munshi-South J, Wilkinson GS. 2010 Bats and birds: exceptional longevity despite high metabolic rates. *Ageing Research Reviews* **9**, 12-19

Negroni MA, Foitzik S, Feldmeyer B. 2019 Long-lived *Temnothorax* ant queens switch from investment in immunity to antioxidant production with age. *Scientific Reports* **9**, 7270

Nijhout H. F. 2003 Development and evolution of adaptive polyphenisms. *Evol. Dev*. **5** (1), 9-18

Nowak MA, Tarnita CE, Wilson EO. 2010. The evolution of eusociality. *Nature*. **466** (7310), 1057-1062.

Olejarz J. W., Allen B., Veller C., Gadagkar R., Nowak M. A. 2016 Evolution of worker policing. *J. Theor. Biol*. **339,** 103-116

O'Rourke T, Boeckx C. 2019 Converging roles of glutamate receptors in domestication and prosociality. bioRxiv (DOI: 10.1101/439869)

Ostrander EA, Jong PM, Rine J, Duyk G. 1992 Construction of small-insert genomic DNA libraries highly enriched for microsatellite repeat sequences. *Proc. Natl. Acad. Sci. USA*. **89**, 3419-3423

Palmer KA, Oldroyd BP. 2000 Evolution of multiple mating in the genus *Apis*. *Apidologie*. **31** (2), 235-248

Page Jr. R. E., Amdam G. V. 2007 The making of social insect: development architectures of social design. *BioEssays*. **29** (4), 334-343

Patalano S., Vlasova A., Wyatt C., Ewels P., Camara F., et al. 2015 Molecular signatures of plastic phenotypes in two eusocial insect species with simple societies. *Proc. Natl. Acad. Sci. U.S.A.* **112** (45), 13970-13975

Paxton RJ, Ayasse M, Field J, Soro A. 2002 Complex sociogenetic organization and reproductive skew in a primitively eusocial sweat bee, Lasioglossum malachurum, as revealed by microsatellites. *Mol. Ecol*. **11** (11), 2405-2416.

Paxton RJ, Thorén PA, Tengö J, Estoup A, Pamilo P. 1996 Mating structure and nestmate relatedness in a communal bee, *Andrena jacobi* (Hymenoptera, Andrenidae), using microsatellites. *Mol. Ecol*. **5**, 511-519.

Peakall R, Smouse PE. 2006 GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Mol. Ecol. Notes*. **6**, 288-295.

Peakall R, Smouse PE. 2012 GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research – an update. *Bioinformatics*. **28**, 2537-2539.

Pech MEC, May-Itzá WdeJ, Medina LAM, Quezadoa-Euán JJG. 2008 Sociality in *Euglossa* (*Euglossa*) *viridissima* Friese (Hymenoptera, Apidae, Euglossini). *Insect. Soc.* **55** (4), 428-433.

Perry JC, Rowe L. 2011 Rapid microsatellite development for water striders by next-generation sequencing. *J. Hered*. **102**, 125-129.

Peters R. S., Krogmann L., Mayer C., Donath A., Gunkel S., et al. 2017 Evolutionary history of the Hymenoptera. *Curr. Biol.* **27**, 1-6

Plaisier SB, Taschereau R, Wong JA, Graeber TG. 2010 Rank-rank hypergeometric overlap: identification of statistically significant overlap between gene-expression signatures. *Nucleic Acids Research* **38**, e169

Plateaux-Quénu C. 2008 Subsociality in halictine bees. *Insectes Soc*. **55**, 335-346

Plateaux-Quénu C., Plateaux I., Packer L. 2000 Population-typical behaviours are retained when eusocial and non-eusocial forms of *Evylaeus albipes* (F.) (Hymenoptera, Halictidae) are reared simultaneously in the laboratory. *Insect. Soc*. **47**, 263-270

Powell W, Machray GC, Provan J. 1996 Polymorphism revealed by simple sequence repeats. *Trends Plant Sci*. **1**, 215-222.

Purcell J. 2011 Geographic patterns in the distribution of social systems in terrestrial arthropods. *Biol. Rev*. **86**, 475-491

Queller DC, Strassmann JE, Hughes CR. 1993 Microsatellites and kinship. *Trends Ecol. Evol*. **8**, 285-288.

Queller DC, Strassmann JE. 1998 Kin selection and social insects. *BioScience*. **48** (3), 165-175.

Quiñones AE, Pen I. 2017 A unified model of Hymenopteran preadaptations that trigger the evolutionary transition to eusociality. *Nat. Commun*. **8**, 1-13

R Core Team 2013 R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. (http://www.R-project.org/)

Rabeling C, Bollazzi M, Bacci Jr M, Beasley RR, Lance SL, Jones KL, Pierce NE. 2014 Development and characterization of twenty-two polymorphic microsatellite markers for the leafcutter ant, *Acromyrmex lundii*, utilizing Illumina sequencing. *Conserv. Genet. Resour*. **6**, 319-322

Ratnieks FLW, Wenseleers T. 2008 Altruism in insect societies and beyond: voluntary or enforced? *Trends Ecol. Evol.* **23** (1), 45-52.

Rau P. 1928 The nesting habits of the little carpenter-bee, Ceratina calcarata. *Ann. Entomol. Soc. Am.* **21**, 380-396.

Raymond M, Rousset F. 1995 GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *J. Hered*. **86**, 248-249.

Rehan S. M., Berens A. J., Toth A. L. 2014a At the brink of eusociality: transcriptomic correlates of worker behaviour in a small carpenter bee. *BMC Evol. Biol*. **14** (260), 1-10

Rehan S. M., Chapman T. W., Craigie A., Richards M. H., Cooper S. J., Schwarz M. P. 2010 Molecular phylogeny of the small carpenter bees (Hymenoptera: Apidae: Ceratinini) indicates early and rapid global dispersal. *Mol. Phylogenet. Evol*. **55**, 1042-1054

Rehan S. M., Glastad K. M., Lawson S. P., Hunt B. G. 2016 The genome and methylome of a subsocial small carpenter bee, *Ceratina calcarata*. *Genome Biol. Evol*. **8** (5), 1401-1410

Rehan SM, Glastad KM, Steffen MA, Fay CR, Hunt BG, Toth AL. 2018 Conserved genes underlie phenotypic plasticity in an incipiently social bee. *Genome Biol. Evol*. **10**, 2749-2758

Rehan S. M., Leys R., Schwarz M. P. 2012 A mid-Cretaceous origin of sociality in Xylocopine bees with only two origins of true worker castes indicates severe barriers to eusociality. *PLoS ONE*. **7** (4), e34690

Rehan S. M., Richards M. H. 2010a Nesting and life cycle of *Ceratina calcarata* in southern Ontario (Hymenoptera: Apidae: Xylocopinae). *Can. Entomol*. **142** (1), 65-74

Rehan S. M., Richards M. H. 2010b The influence of maternal quality on brood sex allocation in the small carpenter bee, *Ceratina calcarata*. *Ethology*. **116** (9), 876-887

Rehan S. M., Richards M. H. 2013 Reproductive aggression and nestmate recognition in a subsocial bee. *Anim. Behav*. **85** (4), 733-741

Rehan S. M., Richards M. H., Adams M., Schwarz M. P. 2014b The costs and benefits of sociality in a facultatively social bee. *Anim. Behav*. **97**, 77-85

Rehan S. M., Richards M. H., Schwarz M. P. 2009 Evidence of social nesting in *Ceratina* of Borneo. *J. Kans. Entomol. Soc*. **82**, 194-209

Rehan S. M., Richards M. H., Schwarz M. P. 2010 Social polymorphism in the Australian small carpenter bee, *Ceratina* (*Neoceratina*) *australensis*. *Insectes Soc*. **57** (4), 403-412

Rehan S. M., Schwarz M. P. 2015 A few steps forward and no steps back: long-distance dispersal patterns in small carpenter bees suggest major barriers to back-dispersal. *J. Biogeogr*. **42** (3), 485-494

Rehan S. M., Schwarz M. P., Richards M. H. 2011 Fitness consequences of ecological constraints and implications for the evolution of sociality in an incipiently social bee. *Biol. J. Linnean Soc*. **103** (1), 57-67

Rehan SM, Sheffield CS. 2011 Morphological and molecular delineation of a new species in the *Ceratina dupla* species-group (Hymenoptera: Apidae: Xylocopinae) of eastern North America. *Zootaxa*. **2873**, 35-50.

Rehan S. M., Tierney S. M., Wcislo W. T. 2015 Evidence for social nesting in Neotropical creatinine bees. *Insect. Soc*. **62** (2), 465-469

Rehan S. M., Toth A. L. 2015 Climbing the social ladder: the molecular evolution of sociality. *Trends Ecol. Evol*. **30** (7), 426-433

Remolina SC, Hafez DM, Robinson GE, Hughes KA. 2007 Senescence in the worker honey bee *Apis mellifera*. *J. Insect. Physiol.* **53**, 1027-1033

Richards M. H. 2000 Evidence for geographic variation in colony social organization in an obligately social sweat bee, *Lasioglossum malachurum* Kirby (Hymenoptera; Halictidae). *Can. J. Zool*. **78**, 1259-1266

Richards MH, French D, Paxton RJ. 2005 It's good to be queen: classically eusocial colony structure and low worker fitness in an obligately social sweat bee. *Mol. Ecol*. **14** (13), 4123-4133

Richards M. H., Packer L. 1996 The socioecology of body size variation in the primitively eusocial sweat bee, *Halictus ligatus* (Hymenoptera: Halictidae). *Oikos*. **77** (1), 68-76

Rittschof C. C., Bukhari S. A., Sloofman L. G., Troy J. M., Caetano-Anollés D. et al. 2014 Neuromolecular responses to social challenge: common mechanisms across mouse, stickleback fish, and honey bee. *Proc. Natl. Acad. Sci. U.S.A.* **111** (50), 17929-17934

Rittschof CC, Robinson GE. 2016 Behavioral genetic toolkits: toward the evolutionary origins of complex phenotypes. *Curr. Top. Dev. Biol*. **119**, 157-204

Robinson G. E. 2002 Sociogenomics takes flight. *Science*. **297** (5579), 204-205

Robinson G. E., Grozinger C. M., Whitfield C. W. 2005 Sociogenomics: social life in molecular terms. *Nat. Rev. Genet*. **6** (4), 257-270

Romiguier J., Lourenco J., Gayral P., Faivre N., Weinert L. A., et al. 2014 Population genomics of eusocial insects: the costs of a vertebrate-like effective population size. *J. Evol. Biol*. **27**, 593-603

Romiguier J., Cameron S. A., Woodard S. H., Fischman B. J., Keller L., Praz C. J. 2015 Phylogenomics controlling for base compositional bias reveals a single origin of eusociality in corbiculate bees. *Mol. Biol. Ecol*. **33** (3), 670-678

Rousset F. 2008 Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. *Mol. Ecol. Resour.* **8**, 103-106

Rubenstein RR, Ågren JA, Carbone L, Elde NC, Hoekstra HE, Kapheim KM, Keller L, Moreau CS, Toth AL, Yeaman S, Hofmann HA. 2019 Coevolution of genome architecture and social behavior. *Trends in Ecology and Evolution* **34**, 844-855

Rubin BER, Jones BM, Hunt BG, Kocher SD. 2019 Rate variation in the evolution of non-coding DNA associated with social evolution in bees. *Philos. Trans. R. Soc. Lond. B.* **374**, 20180247

Sadd M., Barribeau S. M., Bloch G., de Graaf D. C., Dearden P., et al. 2015 The genomes of two key bumblebee species with primitive eusocial organization. *Genome Biol*. **16** (1), 1-31

Sakagami SF, Maeta Y. 1977 Some presumably presocial habits of Japanese *Ceratina* bees, with nots on various social types in Hymenoptera. *Insectes Sociaux* **24**, 319-343

Sakagami S. F., Maeta Y. 1984 Multifemale nests and rudimentary castes in the normally solitary bee *Ceratina japonica* (Hymenoptera: Xylocopinae) *J. Kans. Entomol. Soc*. **57** (4), 639-656

Sakagami S. F., Maeta Y. 1987 Sociality, induced and/or natural, in the basically solitary small carpenter bees (Ceratina). *Anim Soc. Theor. Fac*. 1-16

Sakagami S. F., Maeta Y. 1989 Compatibility and incompatibility of solitary life with eusociality in two normally solitary bees *Ceratina japonica* and *Ceratina Okinawana* (Hymenoptera: Apoidea), with notes on the incipient phase of eusociality. *Jpn. J. Ent*. **57** (2), 417-439

Sakagami S. F., Maeta Y. 1995 Task allocation in artificially induced colonies of a basically solitary bee *Ceratina* (*Ceratinidia*) *okinawana*, with a comparison of sociality between *Ceratina* and *Xylocopa* (Hymenoptera, Anthophoridae, Xylocopinae). Jpn. J. Ent. **63** (1), 115-150

Sakagami SF, Maeta Y, Nagamori S, Saito K. 1993 Diapause and non-delayed eusociality in a univoltine and basically solitary bee *Ceratina japonica* (Hyemnoptera, Anothophoridae). II. Non delayed eusociality induced by juvenile hormone analog treatment. *Japanese Journal of Entomology* **61**, 443-457

Sakagami S. F., Munakata M. 1972 Distribution and bionomics of a transpalaearctic eusocial Halictine bee, *Lasioglossum* (*Evylaeus*) *calceatum*, in northern Japan, with reference to its solitary life cycle at high altitude. *J. Fac. Sci. Hokkaido Univ. Ser. VI. Zool*. **18** (3), 411-439

Saleh NW, Ramírez S. 2019 Sociality emerges from solitary behaviours and reproductive plasticity in the orchid bee Euglossa dilemma. *Proc. R. Soc. B*. **286**, 20190588

Schlichting C. D., Wund M. A. 2014 Phenotypic plasticity and epigenetic marking: an assessment of evidence for genetic accommodation. *Evolution*. **68** (3), 656-672

Schuelke M. 2000 An economic method for the fluorescent labeling of PCR fragments. *Nature Biotechnol*. **18** (2), 233-234.

Schultz T. R. 2000 In search of ant ancestors. *Proc. Natl. Acad. Sci. U.S.A*. **97** (26), 14028-14029

Schwarz MP. 1986 Persistent multi-female nests in an Australian allodapine bee, *Exoneura bicolor* (Hymenoptera, Anthophoridae). *Insectes Sociaux* **33**, 258-277

Schwarz MP. 1987 Intra-colony relatedness and sociality in the allodapine bee *Exoneura bicolor*. *Behav. Ecol. Sociobiol.* **21**, 387-392

Schwarz M. P., Richards M. H., Danforth B. N. 2007 Changing paradigms in insect social evolution: insights from Halictine and Allodapine bees. *Annu. Rev. Entomol*. **52**, 127-150

Schwarz MP, Tierney SM, Rehan SM, Chenoweth LB, Cooper SJB. 2010 The evolution of eusociality in allodapine bees: workers began by waiting. *Biol. Lett*. **7** (2), 277-280.

Seehuus S-C, Norberg K, Gimsa U, Krekling T, Amdam GV. 2006 Reproductive protein protects functionally sterile honey bee workers from oxidative stress. *Proc. Natl. Acad. Sci. USA* **103**, 962-967

Seeley TD. 1982 Adaptive significance of the age polyethism schedule in honeybee colonies. *Behav. Ecol. Sociobiol*. **11** (4), 287-293.

Seeley TD. *Honeybee Ecology: a study of Adaptation in Social Life* (Princeton Univ. Press, Princeton, 1985)

Seger J. 1983 Partial bivoltinism may cause alternating sex-ratio biases that favour eusociality. *Nature*. **301**, 59-62

Shah JS, Renthal R. 2020 Antennal proteome of the *Solenopsis invicta* (Hymenoptera: Formicidae): caste differences in olfactory receptors and chemosensory support proteins. *Journal of Insect Science* **20**, 1-13

Shell W. A., Rehan S. M. 2016a Recent and rapid diversification of the small carpenter bees in eastern North America. *Biol. J. Linnean Soc*. **117**, 633-645

Shell WA, Rehan SM. 2016b Development of multiple polymorphic microsatellite markers for the small carpenter bee, *Ceratina calcarata* (Hymenoptera: Apidae) using genome-wide analysis. *J. Insect Sci*. **16** (1), 1-4.

Shell WA, Rehan SM. 2018 Behavioral and genetic mechanisms of social evolution: insights from incipiently and facultatively social bees. *Apidologie*. **49**, 13-30

Shell WA, Rehan SM. 2019 Social modularity: conserved genes and regulatory elements underlie caste-antecedent behavioural states in an incipiently social bee. *Proc. R. Soc. B.* **286**, 20191815

Simola D. F., Wissler L., Donahue G., Waterhouse R. M., Helmkampf M., et al. 2013 Social insect genomes exhibit dramatic evolution in gene composition and regulation while preserving regulatory features linked to sociality. *Genome Res*. **23**, 1235-1247

Simpson S. J., Sword G. A., Lo N. 2011 Polyphenism in insects. *Curr. Biol*. **21**, 738-749

Sinha S, Liang Y, Siggia E. 2006 Stubb: a program for discovery and analysis of cis-regulatory modules. *Nucleic Acids. Res*. **34**, W555-W559

Smith A. R., Wcislo W. T., O'Donnell S. 2003 Assured fitness returns favor sociality in a mass-provisioning sweat bee, *Megalopta genalis* (Hymenoptera: Halictidae). *Behav. Ecol. Sociobiol*. **54**, 14-21

Smith AR, Wcislo WT, O'Donnell S. 2007 Survival and productivity benefits to social nesting in the sweat bee Megalopta genalis (Hymenoptera: Halictidae). *Behav. Ecol. Sociobiol*. **61** (7), 1111-1120.

Smith C. R., Toth A. L., Suarez A. V., Robinson G. E. 2008 Genetic and genomic analyses of the division of labour in insect societies. *Nature Rev. Genet*. **9**, 735-748

Soro A., Field J., Bridge C., Cardinal S. C., Paxton R. J. 2010 Genetic differentiation across the social transition in a socially polymorphic sweat bee, *Halictus rubicundus*. *Mol. Ecol*. **19**, 3351-3363

Soro A, Paxton RJ. 2009 Characterization of 14 polymorphic microsatellite loci for the facultatively eusocial sweat bee *Halictus rubicundus* (Hymenoptera, Halictidae) and their variability in related species. *Mol. Ecol. Resour*. **9**, 150-152.

Soucy S. L. 2002 Nesting biology and socially polymorphic behavior of the sweat bee *Halictus rubicundus* (Hymenoptera: Halictidae). *Ann. Entomol. Soc. Am*. **95** (1), 57-65

Soucy S. L., Danforth B. N. 2002 Phylogeography of the socially polymorphic sweat bee *Halictus rubicundus* (Hymenoptera: Halictidae). *Evolution*. **56** (2), 330-341

Southey B. R., Zhu P., Carr-Markell M. K., Liang Z. S., Zayed A., et al. 2016 Characterization of genomic variants associated with scout and recruit behavioral castes in honey bees using whole-genome sequencing. *PLoS One*. **11** (1), e0146430

Standage D. S., Berens A. J., Glastad K. M., Severin A. J. 2016 Genome, transcriptome and methylome sequencing of a primitively eusocial wasp reveal a greatly reduced DNA methylation system in a social insect. *Mol. Ecol*. **25**, 1769-1784

Steffen MA, Rehan SM. 2020 Genetic signatures of dominance hierarchies reveal conserved brain gene expression underlying aggression in a facultatively social bee. *Genes, Brain and Behavior* **19**, e12597

Strassmann J. 2001 The rarity of multiple mating by females in the social Hymenoptera. *Insect. Soc*. **48** (1), 1-13.

Strassman J. E., Queller D. C. 1989 Ecological determinants of social evolution, in: Breed, M. D., Page, R. E. (Eds.) The genetics of social evolution. Westview Press, Boulder, Colorado pp. 81-101

Sugahara R, Saeki S, Jouraku A, Shiotsuki T, Tanaka S. 2015 Knockdown of the corazonin gene reveals its critical role in the control of gregarious characteristics in the desert locust. *J. Insect Physiol*. **79**, 80-87

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

Sumner S., Kelstrup H., Fanelli D. 2010 Reproductive constraints, direct fitness and indirect fitness benefits explain helping behaviour in the primitively eusocial wasp, *Polistes canadensis*. *Proc. R. Soc. B*. **277**, 1721-1728

Sunnucks P. 2000 Efficient genetic markers for population biology. *Trends Ecol Evol*. **15**, 199-203.

Supek F, Bošnjak M, Škunca N, Šmuc T. 2011 REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One.* **6**, e21800

Suzuki Y., Nijhout H. F. 2006 Evolution of a polyphenism by genetic accommodation. *Science*. **311**, 650-652

Szathmáry E, Smith JM. 1995 The major evolutionary transitions. Nature. **374** (6519), 227-232.

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

Tavares MG, Pietrani NT, Durvale MdeC, Resende HC, Campos LAdeO. 2013 Genetic divergence between *Melipona quadrifasciata* Lepeletier (Hymenoptera, Apidae) populations. *Genet Mol Biol*. **36**: 111-117.

Thiel, T, Michaelek W, Varshney RK, Graner A. 2003 Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theor Appl Genet*. **106**, 411-422.

Toth A. L., Varala K., Newman T. C., Miguez F. E., Hutchison S. K., et al. 2007 Wasp gene expression supports an evolutionary link between maternal behavior and eusociality. *Science*. **318** (5849), 441-444

Toth A. L., Rehan S. M. 2017 Molecular evolution of insect sociality: an eco-evo-devo perspective. *Annu. Rev. Entomol*. **62**, 419-442

Toth A. L., Robinson G. E. 2007 Evo-devo and the evolution of social behavior. *Trends Genet*. **23** (7), 334-341

Toth A. L., Robinson G. E. 2009 Evo-devo and the evolution of social behavior: brain gene expression analyses in social insects. *Cold Spring Harb. Symp. Quant. Biol.* **74**, 419-426

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

Trivers RL, Hare H. 1976 Haplodiploidy and the evolution of the social insects. *Science*. **191** (4224), 249-263.

Trivers RL. 1971 The evolution of reciprocal altruism. *Q. Rev. Biol*. **46** (1), 35-57.

West-Eberhard MJ. 1975 The evolution of social behavior by kin selection. *Q. Rev. Biol*. **50** (1), 1-33.

True J. R., Carroll S. B. 2002 Gene co-option in physiological and morphological evolution. *Annu. Rev. Cell. Dev. Biol*. **18**, 53-80

Udayakumar A, Shivalingaswamy TM. 2019 Nest architecture and life cycle of small carpenter bee, *Ceratina binghami* Cockerell (Xylocopinae: Apidae: Hymenoptera). *Sociobiology* **66**, 61-65

Ulrich Y, Perrin N, Chapuisat M. 2009 Flexible social organization and high incidence of drifting in the sweat bee, *Halictus scabiosae*. *Mol Ecol*. **18**, 1791-1800.

Ungvari Z, Ridgway I, Philipp EE, Campbell CM, McQuary P, Chow T, Coelho M, ES D, Gelino S, Holmbeck MA, Kim I. 2011 Extreme longevity is associated with increased resistance to oxidative stress in *Arctica islandica*, the longest-lived non-colonial animal. *Journals of Gerontology Series A: Biomedical Sciences and Medical Sciences.* **66**, 741-750

Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG. 2012 Primer3 – new capabilities and interfaces. *Nucleic Acids Res*. **40**, e115.

Vannette RL, Mohamed A, Johnson BR. 2015 Forager bees (*Apis mellifera*) highly express immune and detoxification genes in tissues associated with nectar processing. *Scientific Reports* **5**, 16224

Veenstra JA.1991 Presence of Corazonin in three insect species, and isolation and identification of [His7] Corazonin from Schistocerca Americana. *Peptides*. **12**, 1285-1289

Vickruck JL. 2015 Development of sixteen novel microsatellite markers for the eastern carpenter bee, *Xylocopa virginica* (Hymenoptera: Apidae), through paired-end Illumina sequencing. *Conserv Genet Resour*. **7**, 427-429.

Waddington C. H. 1961 Genetic assimilation. *Adv. Genet*. **10**, 257-293

Wade M. J. 2001 Maternal effect genes and the evolution of sociality in haplo-diploid organisms. *Evolution*. **55** (3), 453-458

Wcislo W. T., Arneson L., Roesch K., Gonzalez V., Smith A., Fernández H. 2004 The evolution of nocturnal behaviour in sweat bees, *Megalopta genalis* and *M. ecuadoria* (Hymenoptera: Halictidae): an escape from competitors and enemies? *Biol. J. Linnean Soc*. **83**, 377-387

Wcislo W. T., Cane J. H. 1996 Floral resource utilization by solitary bees (Hymenoptera: Apoidea) and exploitation of their stored foods by natural enemies. *Annu. Rev. Entomol*. **41**, 257-286

Wcislo W. T., Danforth B. N. 1997 Secondarily solitary: the evolutionary loss of social behavior. *Trends Ecol. Evol*. **12** (12), 468-474

Wcislo WT, Gonzalez VH. 2006 Social and ecological contexts of trophallaxis in facultatively social sweat bees, *Megalopta genalis* and *M. ecuadoria* (Hymenoptera, Halictidae). *Insect. Soc.* **53**, 220-225

Weiner S. A., Toth A. L. 2012 Epigenetics in social insects: a new direction for understanding the evolution of castes. *Genet. Res. Int*. **2012**, 1-11

Weitekamp CA, Libbrecht R, Keller L. 2017 Genetics and evolution of social behavior in insects. *Ann. Rev. Genet*. **51**, 219-239

West-Eberhard M. J. 1987 Flexible strategy and social evolution. *Animal Societies: Theories and Facts*. 35-51

West-Eberhard M. J. 1989 Phenotypic plasticity and the origins of diversity. *Annu. Re. Ecol. Syst*. **20**, 249-278

West-Eberhard M. J. 1996 Wasp societies as microcosms for the study of development and evolution, in: Turillazzi, S. & West-Eberhard, M. J. (Eds.) Natural history and evolution of paper-wasps. Oxford University Press, Oxofrd, pp. 291-317

West-Eberhard M. J. 2003 Developmental Plasticity and Evolution. Oxford University of Press, Oxford

Whitfield CW, Cziko A-M, Robinson GE. 2003 Gene expression profiles in the brain predict behavior in individual honey bees. *Science*. **302**, 296-299

Whitman D. W., Agrawal A. A. 2009 What is phenotypic plasticity and why is it important? In: Whitman, D. W., Ananthakrishnan, T. N. (Eds.) Phenotypic plasticity of insects: mechanisms and consequences pp. 1-63

Wild G., Koykka C. 2014 Inclusive-fitness logic of cooperative breeding with benefits of natal philopatry. *Phil. Trans. R. Soc. B.* **369**, 1-9

Williams JB, Roberts SP, Elekonich MM. 2008 Age and natural metabolically-intensiv behavior affect oxidative stress and antioxidant mechanisms. *Experimental Gerontology*. **43**, 538-549

Wilson E. O. 1971 The Insect Societies. Belknap Press of Harvard University Press, Cambridge, Massachusetts

Wilson EO. 2005 Kin selection as the key to altruism: its rise and fall. *Soc. Res*. **72** (1), 159-166.

Wilson E. O., Hölldobler B. 2005 Eusociality: origin and consequences. *Proc. Natl. Acad. Sci. USA*. **102**, 13367-13371

Wilson-Rich N, Dres ST, Starks PT. 2008 The ontogeny of immunity: development of innate immune strength in the honey bee (*Apis mellifera*). *Journal of Insect Physiology* **54**, 1392-1399

Withee JR, Rehan SM. 2016 Cumulative effects of body size and social experience on aggressive behaviour in a subsocial bee. *Behaviour*. **153** (12), 1365-1385.

Withee JR, Rehan SM. 2017 Social experience, aggression and brain gene expression in a subsocial bee. *Integr. Comp. Biol*. **57**, 640-648

Wittwer B, Hefetz A, Simon T, Murphy LEK, Elgar MA, Piece NE, Kocher SD. 2017 Solitary bees reduce investment in communication compared with their social relatives. *Proc. Natl. Acad. Sci. USA.* **114**, 6569-6574

Woodard S. H., Bloch G. M., Band M. R., Robinson G. E. 2014 Molecular heterochrony and the evolution of sociality in bumblebees (*Bombus terrestris*). *Proc. R. Soc. B.* **281**, 1-8

Woodard S. H., Fischman B. J., Venkat A., Hudson M. E., Varala K., et al. 2011 Genes involved in convergent evolution of sociality in bees. *Proc. Natl. Acad. Sci. U.S.A.* **108** (18), 7472-7477

Wyman L. M., Richards M. H. 2003 Colony social organization of *Lasioglossum malachurum* Kirby (Hymenoptera, Halictidae) in southern Greece. *Insect. Soc*. **50**, 201-211

Xiao L, Priest MF, Nasebeny J, Lu T, Kozorovitskiy Y. 2017 Biased oxytocinergic modulation of midbrain dopamine systems. *Neuron*. **95**, 368-384

Yagi N, Hasegawa E. 2012 A halictid bee with sympatric solitary and eusocial nests offers evidence for Hamilton's rule. *Nat. Comm*. **3**, 939.

Yan H., Bonasio R., Simola D. F., Liebig J., Berger S. L., Reinberg D. 2015 DNA methylation in social insects: how epigenetics can control behavior and longevity. *Annu. Rev. Entomol*. **60**, 435-452

Yanega D. 1989 Caste determination and differential diapause within the first brood of *Halictus rubicundus* in New York (Hymenoptera: Halictidae). *Behav Ecol Sociobiol*. **24**, 97-107

Yanega D. 1990 Philopatry and nest founding in a primitively social bee, *Halictus rubicundus*. *Behav. Ecol. Sociobiol*. **27**, 37-42

Yanega D. 1993 Environmental influences on male production and social structure in *Halictus rubicundus* (Hymenoptera: Halictidae). *Insect. Soc*. **40**, 169-180

Zane L, Bargelloni L, Patarnello T. 2002 Strategies for microsatellite isolation: a review. *Mol Ecol*. **11**, 1-16.

Zhang B, Horvath S. 2005 A general framework for weighted gene co-expression network analysis. *Statistical Application in Genetics and Molecular Biology*. **4**, 1-43