PROCEEDINGS B

royalsocietypublishing.org/journal/rspb

Research

Cite this article: 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.
http://dx.doi.org/10.1098/rspb.2019.1815

Received: 7 August 2019
Accepted: 29 October 2019

Subject Category:
Behaviour

Subject Areas:
behaviour, evolution, genomics

Keywords:
transcriptomics, toolkit genes, social evolution, phenotypic plasticity, time course gene expression, behavioural genetics

Author for correspondence:
Sandra M. Rehan
e-mail: sandra.rehan@gmail.com

Electronic supplementary material is available online at https://doi.org/10.6084/m9.figshare.c.4734701.

THE ROYAL SOCIETY PUBLISHING

Social modularity: conserved genes and regulatory elements underlie caste-antecedent behavioural states in an incipiently social bee

Wyatt A. Shell[1] and Sandra M. Rehan[1,2]

[1]Department of Biological Sciences, University of New Hampshire, 38 Academic Way, Durham, NH 03824, USA
[2]Department of Biology, York University, 4700 Keele Street, Toronto, Ontario, Canada M3 J 1P3

SMR, 0000-0002-6441-5155

The evolutionary origins of advanced eusociality, one of the most complex forms of phenotypic plasticity in nature, have 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, involving deeply conserved and differentially expressed genes in the expression of early social forms.

1. Background

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 [1]. In this way, observable phenotypic plasticity—the ability of a single genotype to produce multiple phenotypes under the influence of varying environmental conditions [2]—can be understood as a product of the organismal system's essential modularity [1]. This is conspicuously demonstrated by the developmental plasticity of eusocial insects, such as honeybees (*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 [3]. Honeybee 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 [4,5]. 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 [6,7].

Accordingly, extensive genomic and transcriptomic research exploring caste determination and behaviour across highly eusocial Hymenoptera has illuminated an array of important genetic and regulatory elements [8,9]. It is now

widely appreciated that eusociality's ontogenetic and behavioural variety may be rooted in the modular nature of transcription factors and regulatory networks [10–12] and their combined influence on differential gene expression [13]. 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 [1,14]. Concurrently, comparative sociogenomic research has shown strong support for the theory that eusocial behavioural phenotypes are underpinned in part by 'toolkit' genes [15,16], 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 [17,18].

It is generally accepted that advanced eusocial Hymenoptera evolved from solitary ancestors, with lineages having gradually gained traits of social complexity as an increasingly well-defined division of labour was established [19]. 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 [20–22] (reviewed in [19,23]). 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 nest-mates along behavioural and ontogenetic lines (e.g. *Ceratina australensis* [18]; *Megalopta genalis* [20]; *Polistes canadensis* [24]). 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 [1,25].

Recent evidence has also provided support for the social ladder hypothesis, which states that differentially expressed and deeply conserved genes probably play an important role in both the early and later stages of social evolution [19,20,26,27]. 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* [18]). There remains, however, a paucity of transcriptomic datasets which capture gene expression and regulatory patterns underlying ontogeny and behavioural plasticity among incipiently social species—data which are necessary for further empirical examination of these hypotheses within a pan-social comparative framework [9,23].

The incipiently social small carpenter bee *Ceratina calcarata* is an emerging model organism for studies of early insect sociality (e.g. [28,29]). After establishing a nest and provisioning her brood, a *C. calcarata* mother guards and cleans her maturing offspring through adulthood [30,31]. Once her brood has matured, the mother resumes foraging to feed her adult offspring, ensuring their survival during the long winter diapause [32]. 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 [32,33]. In these social nests, the mother and her daughter forage and guard at roughly equivalent rates [32,33]; despite roughly a year's difference in age, both females demonstrate conspicuous and phenotypically similar behaviours 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 behavioural 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 behaviour in mothers and daughters; and (iii) determine whether deeply conserved genes appear to play a role in a species demonstrating emergent social traits.

2. Methods

(a) 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–August). Branches were dissected lengthwise to secure females and to assess nest developmental stage (following [30,31]). In May, founding nest mothers excavate or reoccupy a nesting burrow in preparation for brood rearing [31]. By June, mothers have mated and begin actively brooding, provisioning brood cells with pollen balls and laying a single egg on each. By July, mothers of full brood 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–April [29]). Adult females identified at each of these five life-history stages (figure 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 behaviours, we targeted foraging mothers, guarding mothers, foraging daughters and guarding daughters (further details on sample collection protocol can be found in electronic supplementary material, Methods).

A total of 33 individuals were selected for whole head RNA extraction (electronic supplementary material, table S1). Heads were removed on dry ice and immediately processed using the QIAGEN RNeasy Kit and protocol (cat. no. 73404). RNA sample quality was then confirmed on an Agilent Tape Station 2200 and submitted for library prep and 150 base paired-end Illumina HiSeq 2500 sequencing by Genome Quebec. Read data were then aligned to the *C. calcarata* genome [34] before being used for analysis (data accessible via NCBI SRA PRJNA434715).

(b) Gene expression and cis-regulatory element enrichment analyses

Brain gene expression data were used to perform several analyses of gene expression (electronic supplementary material, table S1). First, a time course analysis of gene expression was performed using maSigPro v. 3.7 [35] in R v. 3.4.3 [36] to assess consistencies in gene expression variation during female ageing. Analyses of differentially expressed genes (DEGs) using DESeq2 [37] were then performed in R to explore foraging and guarding behaviour in mothers and daughters. We also performed a complementary weighted gene co-expression network analysis (WGCNA [38]) to further inspect these data. Gene ontology (GO) term enrichment was then determined for gene lists from maSigPro and DESeq2 analyses using topGO v. 3.7 [39], and significantly enriched

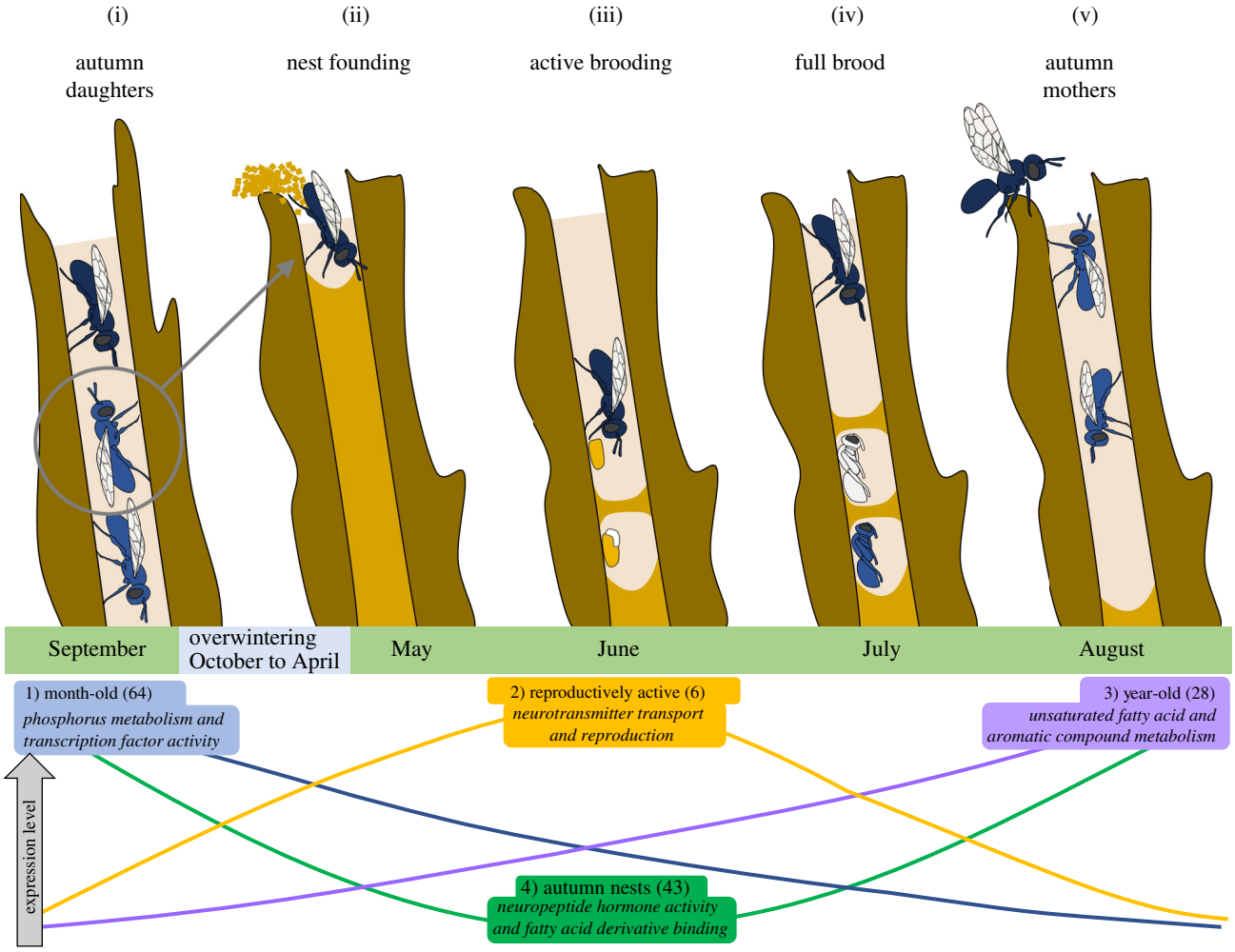

Figure 1. Timeline of *Ceratina calcarata* adulthood illustrating five major waypoints associated with reproductive maturation and senescence: (i) a recently eclosed autumn daughter (circled) occupies her natal nest wherein she may assist her mother (dark blue) in foraging and guarding; (ii) after overwintering, 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 four captures genes upregulated in both AM and daughters. Gene counts (in parentheses) and representative GO terms enriched for each set are provided. For full list of genes and GO terms, see electronic supplementary material, tables S5, S6, S8 and S10. (Online version in colour.)

GO term lists ($p < 0.05$) were further reduced using Revigo to select terms which featured a dispensability rating less than or equal to 0.5 [40]. We identified transcription factor binding site (TFBS) motif enrichment among all DEGs determined for each of our focal conditions using the programs STUBB [41] and CIS-METALYSIS [42], searching against the JASPAR Insect and Vertebrate databases for motif reference [43] (full details on maSigPro, DESeq2, WGCNA and CIS-METALYSIS analyses can be found in electronic supplementary material).

(c) Comparative analysis

We used BLASTn [44] to assess gene homology between the *C. calcarata* genome [34] and publicly available genomic and transcriptomic datasets from five bees, three ants, two wasps and the fruit fly using cut-offs of more than 65% shared ID and p-values $< 1.0 \times 10^{-5}$ (electronic supplementary material, table S2). ORTHOFINDER v. 2.3.2 [45] 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. four bees, one ant and one wasp) to further examine orthology. We then compared homologous genes between *C. calcarata* and a total of 32 additional studies collectively examining variations in gene expression by reproductive development, age, caste and/or behavioural state in five bees, five ants, two wasps, the house mouse, stickleback fish and fruit fly (electronic supplementary material, table S3). Where possible, these studies were also used to compare significantly enriched GO terms and TFBS motifs.

3. Results

(a) Read mapping and time course analyses

Paired-end Illumina sequencing generated an average of 31.8 MB of raw sequence data for each of our 33 whole head samples (1.05 GB in total; electronic supplementary material, table S4). Better than 99% of the raw data for each sample met cutoff criteria for alignment, allowing our sequence data to map back to an average of 10 855 genes at 32× 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 1; electronic supplementary material, tables S5 and S6). 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 (electronic supplementary material, table S6) identified multiple metabolic processes associated with energy production (e.g. phosphorus metabolic process). Six genes were upregulated as females reached their 10th 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 (electronic supplementary material, table S6).

The remaining 43 genes were upregulated in both pre- and post-reproductive females, suggesting this set is not so closely regulated by ontogeny (figure 1; electronic supplementary material, table S5). Notable genes in this set included *histone h2a* and *pro-corazonin preproprotein*; and enriched GO terms (electronic supplementary material, table S6) included neuropeptide hormone activity and protein transmembrane transport.

(b) Differential gene expression by behavioural state and age

Analyses of gene expression among behavioural states (i.e. foraging versus guarding versus nesting individuals) captured 2017 distinct significantly DEGs overall (at FDR < 0.05; electronic supplementary material, table S7). Comparing genes uniquely upregulated among phenotypes, 292 genes were upregulated specifically in nesting females, compared to 483 genes associated with foragers and 1104 genes upregulated in guards (figure 2). Forager-associated genes were conserved across bees (*C. australensis*; *A. mellifera*), wasps (*Polistes metricus*) and ants (*Temnothorax longispinosus* and *Solenopsis invicta*; figure 3), and were enriched for gene silencing and immune-related processes (electronic supplementary material, tables S7 and S10). Guard-associated genes were well-conserved among the limited datasets that allowed for comparison, which included *C. australensis* and *A. mellifera* (electronic supplementary material, tables S8 and S9); this set was enriched for processes involved in carbohydrate derivative metabolism, regulation of defence response and neurotransmitter transport (electronic supplementary material, tables S7 and S10).

We identified a total of 5527 significantly DEGs across all behaviour by age comparisons (at FDR < 0.05; electronic supplementary material, table S11). Comparing the effects of age on either behavioural phenotype, 616 genes were found to distinguish foraging mothers ($n_{DEGs} = 198$) from foraging daughters ($n_{DEGs} = 418$, electronic supplementary material, figure S1D and table S11). 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 behaviour in mothers ($n_{DEGs} = 1292$) from daughters ($n_{DEGs} = 1196$; electronic supplementary material, figure S1E). Notable genes upregulated in guarding daughters included *syntaxin 1a*, *glutamate* and many glutamate-associated genes

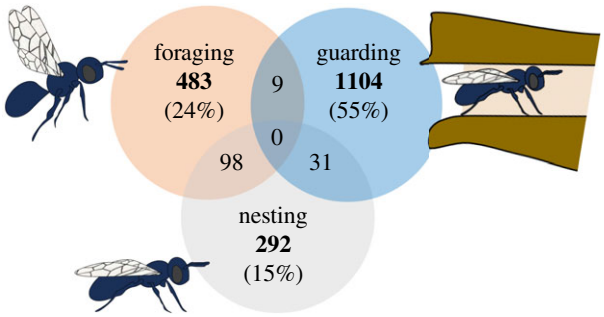

Figure 2. Venn diagram of upregulated genes differentiating guarding and foraging individuals in comparison to non-foraging, non-guarding 'nesting' individuals. Unique gene counts by category are provided in bold, and respective percentages of total genes are indicated in parentheses (6% of DEGs are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

(figure 4). Mother guard enrichment indicated roles for positive regulation of translation and histone deacetylase activity (electronic supplementary material, table S11). Daughter guards were enriched for demethylation, lipid metabolic process and neurotransmitter transport.

A total of 114 genes distinguished guarding from foraging behaviour among mothers, most of which were upregulated in guarding individuals ($n_{DEGs} = 73$; electronic supplementary material, figure S1A and table S11). 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 behaviour in daughters ($n = 1581$, 52%; electronic supplementary material, figure S1B and table S11). 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 signalling pathways (figure 4; electronic supplementary material, table S11), 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* ($n_{DEGs} = 16$; electronic supplementary material, tables S8–S10). GO 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*; electronic supplementary material, table S9).

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 (electronic supplementary material, tables S12–S16 and figures S5–S10). 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 ($n_{daughters} = 695$; $p = 4.8 \times 10^{-75}$) compared to mothers ($n_{mothers} = 296$; $p = 8.4 \times 10^{-25}$); and the glutamate family of genes was again found to be

	comparative differential gene expression	*Ceratina calcarata*	*Ceratina australensis*	*Apis mellifera*	*Polistes metricus*	*Temnothorax longispinosus*
foraging	peroxidase-like isoform ×4	■	■	■	■	
	histidine decarboxylase-like	■	■	■		
	aggrecan core	■		■	■	
	heat shock cognate protein 70	■		■		■
	ccaat enhancer-binding	■		■		■
guarding	sodium potassium-transporting atpase subunit beta-2	■	■			
	neurocalcin homologue	■	■			
	ribonuclease 3	■		■		
	ecdysone-induced protein 74ef-like isoform ×1	■		■		
	serine threonine-protein phosphatase 2a 56 kda …	■		■		

Figure 3. An illustrative subset of all genes associated with foraging or guarding behaviour 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 electronic supplementary material, table S8). 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. Light grey boxes indicate no applicable comparison. Overall, genes associated with foraging behaviour in *C. calcarata* appear to be deeply conserved across Hymenoptera. Fewer studies examine the transcriptomics of guarding behaviour, but existing bee literature also reveals important conserved candidate genes. (Online version in colour.)

expanded specifically among guarding daughters (see electronic supplementary material, tables S13 and S15).

(c) *Cis*-regulatory enrichment

Significant enrichment of TFBS motifs associated with behavioural states revealed a total of 705 functionally unique motifs (electronic supplementary material, tables S17 and S18). 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 behaviour (figure 5). These neural-associated transcription factors include some which were themselves differentially regulated (electronic supplementary material, table S8). Comparisons of TFBS motifs enriched among foragers in publicly available datasets (e.g. [11,20]; electronic supplementary material, table S18) revealed BarH1 to have a conserved role underlying differential gene expression in *C. australensis* and foraging behaviour 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 behaviours in *A. mellifera*.

A total of 308 TFBS motifs were enriched across upregulated genes among foraging mothers and daughters (electronic supplementary material, table S17 and figure S18). 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 with daughters, foraging mothers were enriched for fewer motifs given the DEG count ($n_{TFBS} = 115$; χ^2-test, $\chi^2 = 2.47$, d.f. = 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*; electronic supplementary material, table S18). 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* (electronic supplementary material, table S18).

A total of 172 TFBS motifs were enriched among mother and daughter guard-associated DEGs (electronic supplementary material, table S17), including those binding TFs involved in methylation, neural function and circadian rhythmicity (electronic supplementary material, table S18 and figure S11). TFBS motifs specific to guarding daughters ($n = 124$) bind TFs involved in neurological development and activity as well as immune response and functionality (electronic supplementary material, table S18). Compared with daughters, guarding mothers featured significantly reduced TFBS motif enrichment given DEG count ($n_{TFBS} = 48$; χ^2-test, $\chi^2 = 37.14$, d.f. = 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 behaviour in *A. mellifera* [11] or social polyphenism in *C. australensis* [38] (electronic supplementary material, table S18).

4. Discussion

Our brain transcriptomic data captures variations in gene expression underlying both a phenological time course and conspicuous behavioural states in *Ceratina calcarata*, an

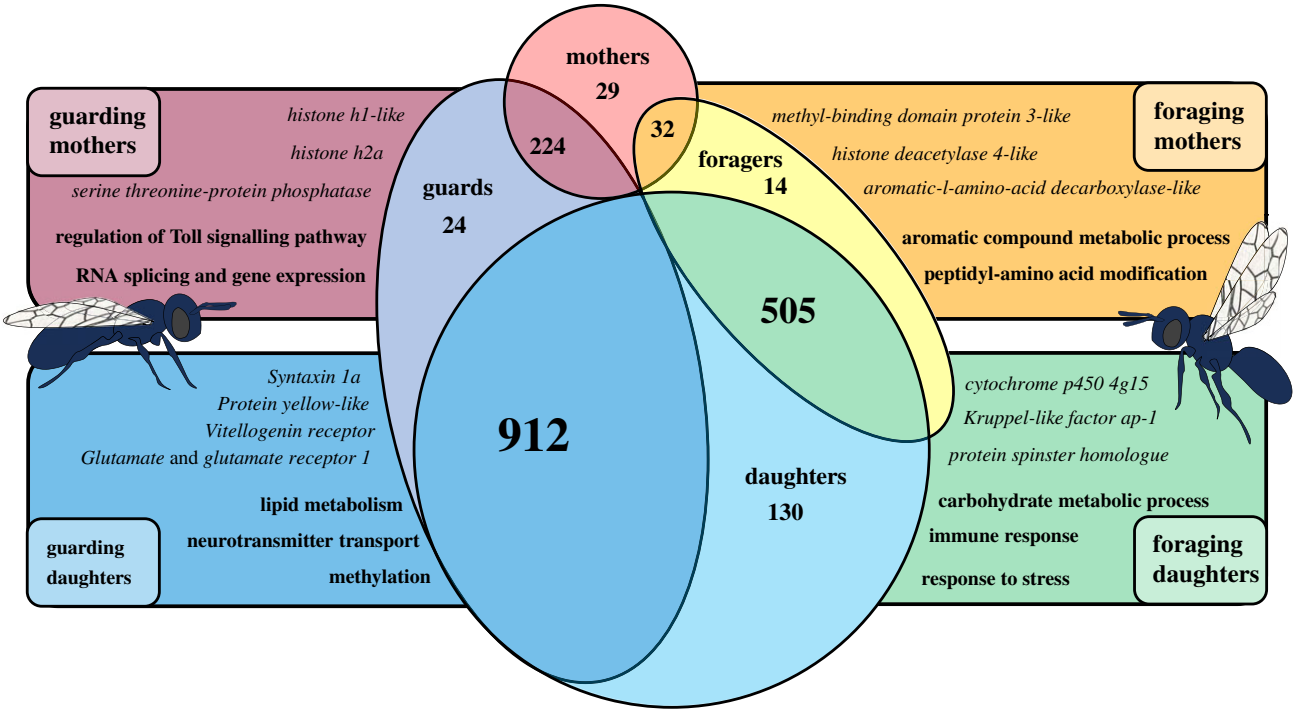

Figure 4. Summary of unique upregulated genes and enriched GO terms identified in each focal age-behavioural state, namely foraging and guarding mothers and daughters. Numbers represent counts of DEGs uniquely upregulated by category (circle sizes and spans of overlap are relative); a selection of notable 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. (Online version in colour.)

incipiently social small carpenter bee. Here, we reveal the strong and systemic influence of a relatively simple and transient social environment; and examine how differences in behavioural 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 behavioural states in *C. calcarata* include those which may play conserved roles in phenotypic plasticity and social complexity across social taxa.

(a) 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 [46]. By contrast, enrichment among year-old mothers for processes involving organic and aromatic compound metabolism may suggest that mothers increasingly rely on chemical signalling as they begin interacting with their adult brood. Chemical signalling is widely employed among Hymenoptera as a means of nest-mate communication and may reinforce divisions of labour (e.g. *A. mellifera* [47]). Accordingly, chemical signalling may play a role, alongside aggression and social experience, in affecting gene expression and behaviour in *C. calcarata*'s autumn nests [48].

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* [34]. Mothers and daughters are otherwise separated by nearly a year of physiological maturation and experience [30,31] indicating that upregulation in this set of genes is more likely to be associated with the conditions of the autumn nest rather than ontological stage. Considered from an evo-devo approach, continuous variations in phenotype (e.g. behaviour) 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 [1]. 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* [49]), 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 [50]. *Corazonin* contributes to social and behavioural phenotypic plasticity in other species—facilitating shifts between solitary and gregarious phases in two locusts (*Locusta migratoria* and *Schistocerca gregaria* [51]), and reproductive and working states in a ponerine ant (*Harpegnathos saltator* [13])—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 [52]. Further, the molecular dynamics underlying

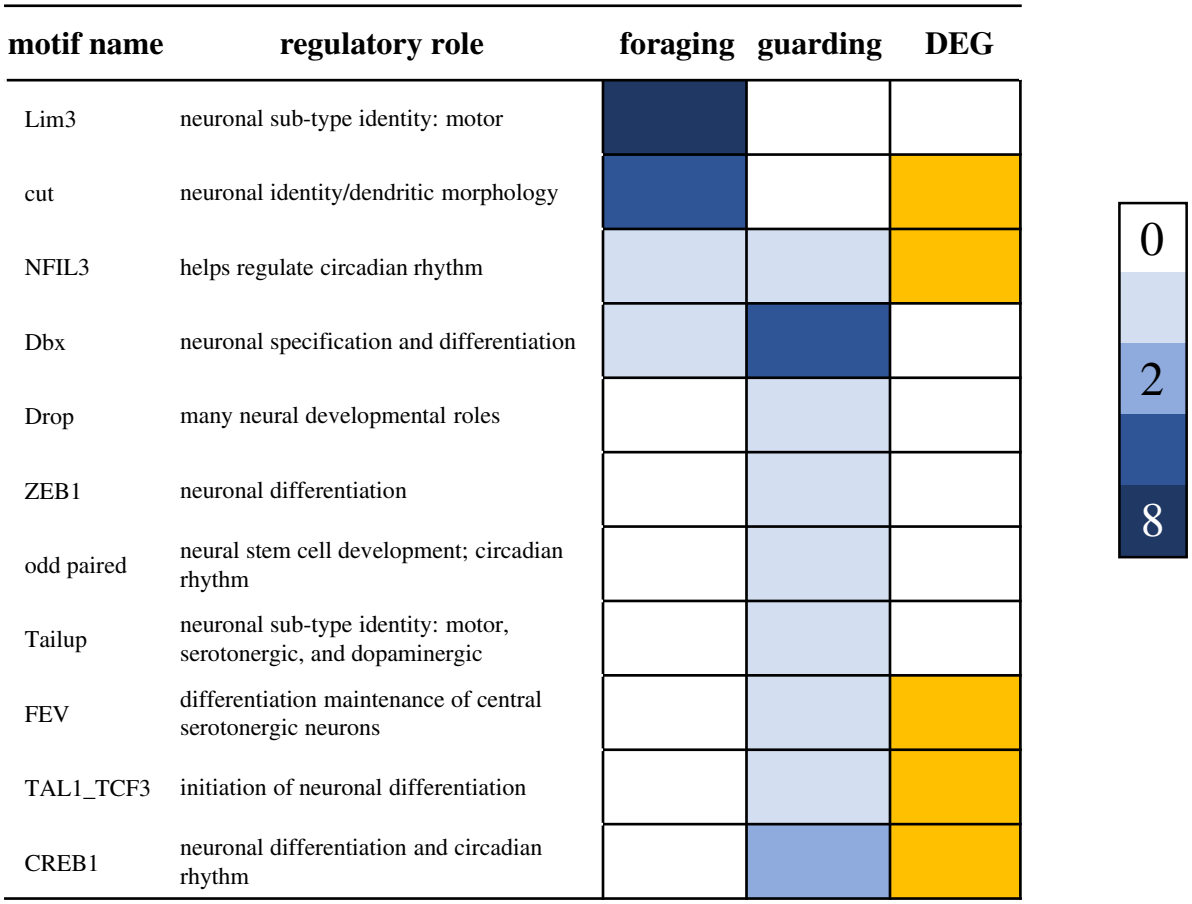

Figure 5. Heat map highlighting TFBS motifs with known neural regulatory roles, significantly enriched in the promoter regions of genes associated with foraging and guarding individuals regardless of age (for full list, see electronic supplementary material, figure S11 and table S18). Motif names are 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 colour (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. (Online version in colour.)

C. calcarata's shifts in nest social environment suggest additional support for the role of certain conserved genes in the operation of social plasticity [15,19].

(b) Extensive differential regulation underlies behavioural states

Differential gene expression analysis revealed twofold and fourfold expansions in gene upregulation among foraging and guarding individuals compared with 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 behavioural states in advanced eusocial insects (e.g. ants [53] and honeybees [5]). For example, in the acorn ant (*T. longispinosus*), four times more variation in gene expression was detected among behavioural states than in comparisons involving age or fertility status [53]. Although some studies suggest little variation in gene expression underlying tasks performed by age-matched individuals (e.g. guarding versus undertaking in *A. mellifera*), this may also be attributable to the strongly age-associated polyethism of these species [54].

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* [18], and among eusocial bees [10], wasps [55] and ants [53,56] (electronic supplementary material, tables S8 and S9; figure 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 DEGs across social taxa [18,19,22].

Notably, 517 of the DEGs associated with foraging or guarding behaviour in this study collectively correspond to 58% of all genes previously identified with signatures of positive selection in the *C. calcarata* genome ($n_{\text{total}} = 877$ genes [34]). Theory suggests that, as insects gain traits of social complexity, genetic release from pleiotropic constraint should lead to elevated rates of positive selection among genes associated with potentially caste-antecedent behavioural phenotypes [19,25,34,57]. As modular subunits (e.g. regulatory networks underlying behavioural 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 [1]. The elevated rates of molecular evolution in genes specifically upregulated within narrowly defined and socially mediated behavioural 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 [27]; and recent comparative studies

among both facultative and obligate eusocial species have consistently detected elevated rates of positive selection specifically among genes associated with the worker caste (e.g. *M. genalis* [20]; *A. mellifera* [58]) or working behaviour [57]. In much the same way, it appears to be primarily the genes associated with socially mediated behavioural states that are experiencing genetic release in *C. calcarata* [25,34].

(c) Distinct neural regulatory pathways underlie guarding and foraging behaviours

Neural-associated TFBS motifs enriched in the promoter regions of genes upregulated in foraging or guarding individuals were largely specific to behaviour 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 [48]. The behavioural regulatory patterns in *C. calcarata* thus resemble those detected in the brain transcriptional regulatory network of *A. mellifera*, in which behavioural states were found to be underpinned by role-specific TF modules [11]. As such, the foraging and guarding behavioural phenotypes may represent distinct neurogenomic states, each subject to its own suites of state-specific selective pressures [1,11,52]. Over time, these distinctions could conceivably drive a molecular wedge between foraging and guarding behavioural phenotypes in a *C. calcarata*-like lineage putatively antecedent to more canalized caste roles [1].

(d) Age effects on behavioural 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* [21]). 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* [18]; *A. mellifera* [59]; leaf-cutting ants, *Atta vollenweideri* [60]), and in the adoption and/or expression of social traits among vertebrates (including mice [61], dogs and humans [62]). 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 [15,16].

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, behaviour-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 [1]. While expression of either behavioural state among mothers may involve comparatively constrained regulatory pathways, potentially pleiotropically tied to the ontogenetic aspects of reproductive maturation and activity [25], these same states are evidently more dynamically and expansively regulated among daughters, incorporating many more TFs and genes. Consequently, while foraging and guarding behaviours may yet appear phenotypically similar between mothers and daughters, the associated underlying regulatory pathways and gene expression patterns may be undergoing substantial differentiation and elaboration; a process strongly suggestive of a lineage actively experiencing an appreciable augmentation in social complexity.

5. Conclusion

Ceratina calcarata's social autumn nest stage provides a prime natural experiment to disentangle the effects of age from behavioural state on gene expression and regulation in a species of incipient sociality. Our results suggest that the behavioural plasticity of *C. calcarata* may be underpinned by a dynamic and modular regulatory network, involving conserved and DEGs. More broadly, we find evidence that even a transiently incipiently social nest environment may have major influence on patterns of gene regulation and expression. Overall, our results lend important empirical support to the functional role of multiple mechanisms theorized to contribute to the evolution of social complexity (from broadest to narrowest: modularity [1]; the social ladder [19] and toolkit genes [15]). 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 [23].

Data accessibility. Data are accessible via NCBI SRA PRJNA434715.

Authors' contributions. W.A.S. conducted laboratory work and data analyses; S.M.R. conceived and funded the study; W.A.S. and S.M.R. wrote the manuscript.

Competing interest. We declare we have no competing interests.

Funding. This work was supported by funding from the National Science Foundation IOS-1456296 to S.M.R. and NSF Graduate Research Fellowship 1450271 to W.A.S.

Acknowledgements. We thank Cullen Franchino, Sean Lombard and Michael Mikát for assistance with field observation and collection of specimens and Genome Quebec for RNA library preparation and Illumina sequencing. We thank Charles Blatti for helpful advice on the use of the cis-Metalysis pipeline, as well as members of the Rehan lab for constructive feedback on this manuscript.

References

1. West-Eberhard MJ. 2003 *Developmental plasticity and evolution*. Oxford, UK: Oxford University Press.
2. Simpson SJ, Sword GA, Lo N. 2011 Polyphenism in insects. *Curr. Biol.* **21**, 738–749. (doi:10.1016/j.cub.2011.06.006:)
3. 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. (doi:10.1073/pnas.96.10.5575)

4. Seeley TD. 1985 *Honeybee ecology: a study of adaptation in social life*. Princeton, NJ: Princeton University Press.
5. Whitfield CW, Cziko A-M, Robinson GE. 2003 Gene expression profiles in the brain predict behavior in individual honey bees. *Science* **302**, 296–299. (doi:10.1126/science.1086807)
6. Toth AL, Robinson GE. 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. (doi:10.1101/sqb.2009.74.026)
7. Abouheif E, Favé MJ, Ibarrarán-Viniegra AS, Lesoway MP, Rafiqi AM, Rajakumar R. 2014 Eco-evo-devo: the time has come. In *Ecological genomics* (eds CR Landry, N Audin-Horth), pp. 107–125. Dordrecht, The Netherlands: Springer.
8. Corona M, Libbrecht R, Wheeler DE. 2016 Molecular mechanisms of phenotypic plasticity in social insects. *Curr. Opin. Insect Sci.* **13**, 55–60. (doi:10.1016/j.cois.2015.12.003)
9. Weitekamp CA, Libbrecht R, Keller L. 2017 Genetics and evolution of social behavior in insects. *Annu. Rev. Genet.* **51**, 219–239. (doi:10.1146/annurev-genet-120116-024515)
10. Khamis AM *et al.* 2015 Insights into the transcriptional architecture of behavioral plasticity in the honey bee *Apis mellifera*. *Sci. Rep.* **5**, 11136. (doi:10.1038/srep11136)
11. Chandrasekaran S, Ament SA, Eddy JA, Rodriguez-Zas SL, Schatz BR, Price ND, Robinson GE. 2011 Behavior-specific changes in transcriptional modules lead to distinct and predictable neurogenomic states. *Proc. Natl Acad. Sci. USA* **108**, 18 020–18 025. (doi:10.1073/pnas.1114093108)
12. Molodtsova D, Harpur BA, Kent CF, 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. (doi:10.3389/fgene.2014.00431)
13. Gospocic J *et al.* 2017 The neuropeptide corazonin controls social behavior and caste identity in ants. *Cell* **170**, 748–759. (doi:10.1016/j.cell.2017.07.014)
14. Espinosa-Soto C, Wagner A. 2010 Specialization can drive the evolution of modularity. *PLoS Comput. Biol.* **6**, e1000719. (doi:10.1371/journal.pcbi.1000719)
15. Toth AL, Robinson GE. 2007 Evo-devo and the evolution of social behavior. *Trends Genet.* **23**, 334–341. (doi:10.1016/j.tig.2007.05.001)
16. Rittschof CC, Robinson GE. 2016 Behavioral genetic toolkits: toward the evolutionary origins of complex phenotypes. *Curr. Top. Dev. Biol.* **119**, 157–204. (doi:10.1016/bs.ctdb.2016.04.001)
17. Morandin C *et al.* 2016 Comparative transcriptomics reveals the conserved building blocks involved in parallel evolution of diverse phenotypic traits in ants. *Genome Biol.* **17**, 1. (doi:10.1186/s13059-016-0902-7)
18. 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. (doi:10.1093/gbe/evy212)
19. Rehan SM, Toth AL. 2015 Climbing the social ladder: the molecular evolution of sociality. *Trends Ecol. Evol.* **30**, 426–433. (doi:10.1016/j.tree.2015.05.004)
20. Jones BM, Kingwell CJ, Wcislo WT, Robinson GE. 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**, 20162228. (doi:10.1098/rspb.2016.2228)
21. 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. (doi:10.1038/s41467-018-06824-8)
22. 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. (doi:10.1098/rspb.2019.0588)
23. Shell WA, Rehan SM. 2018 Behavioral and genetic mechanisms of social evolution: insights from incipiently and facultatively social bees. *Apidologie* **49**, 13–30. (doi:10.1007/s13592-017-0527-1)
24. Ferreira PG, Patalano S, Chauhan R, Ffrench-Constant R, Gabaldón T, Guigó R, Sumner S. 2013 Transcriptome analyses of primitively eusocial wasps reveal novel insights into the evolution of sociality and the origin of alternative phenotypes. *Genome Biol.* **14**, R20. (DOI:10.1186/gb-2013-14-2-r20)
25. Gadagkar R. 1997 The evolution of caste polymorphism in social insects: genetic release followed by diversifying evolution. *J. Genet.* **76**, 167–179. (doi:10.1007/BF02932215)
26. Toth AL, Rehan SM. 2017 Molecular evolution of insect sociality: an eco-evo-devo perspective. *Annu. Rev. Entomol.* **62**, 419–442. (doi:10.1146/annurev-ento-031616-035601)
27. 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. (doi:10.1038/s41598-018-28489-5)
28. Glastad KM *et al.* 2017 Variation in DNA methylation is not consistently reflected by sociality in Hymenoptera. *GBE* **9**, 1687–1698. (doi:10.1093/gbe/evx128)
29. Rubin BER, Jones BM, Hunt BG, Kocher SD. 2019 Rate variation in the evolution of non-coding DNA associated with social evolution in bees. *Phil. Trans. R. Soc. B* **374**, 20180247. (doi:10.1098/rstb.2018.0247)
30. Rehan SM, Richards MH. 2010 Nesting and life cycle of *Ceratina calcarata* in souther Ontario (Hymenoptera: Apidae: Xylocopinae). *Can. Entomol.* **142**, 65–74. (doi:10.4039/n09-056)
31. Rehan SM, Richards MH. 2010 The influence of maternal quality on brood sex allocation in the small carpenter bee, *Ceratina calcarata*. *Ethology* **116**, 876–887. (doi:10.1111/j.1439-0310.2010.01804.x)
32. 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. (doi:10.1007/s00265-017-2365-6)
33. Rehan SM, Berens AJ, Toth A. 2014 At the brink of eusociality: transcriptomic correlates of worker behaviour in a small carpenter bee. *BMC Evol. Biol.* **14**, 260. (doi:10.1186/s12862-014-0260-6)
34. Rehan SM, Glastad KM, Lawson SP, Hunt BG. 2016 The genome and methylome of a subsocial small carpenter bee, *Ceratina calcarata*. *Genome Biol. Evol.* **8**, 1401–1410. (doi:10.1093/gbe/evw079)
35. 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)
36. R Core Team. 2013 *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. See http://www.R-project.org/
37. Love MI, Huber W, Anders S. 2014 Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550. (doi:10.1186/s13059-014-0550-8)
38. Langfelder P, Horvath S. 2008 WGCNA: an R package for weighted correlation network analysis. *BMC Bioinf.* **9**, 559. (doi:10.1186/1471-2105-9-559)
39. Alexa A, Rahnenfuhrer J. 2016 topGO: Enrichment analysis for gene ontology. R package version 2.28.0.
40. Supek F, Bošnjak M, Škunca N, Šmuc T. 2011 REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS ONE* **6**, e21800. (doi:10.1371/journal.pone.0021800)
41. Sinha S, Liang Y, Siggia E. 2006 Stubb: a program for discovery and analysis of *cis*-regulatory modules. *Nucleic Acids. Res.* **34**, W555–W559. (doi:10.1093/nar/gkl224)
42. Ament SA *et al.* 2012 New meta-analysis tools reveal common transcriptional regulatory basis for multiple determinants of behavior. *Proc. Natl Acad. Sci. USA* **109**, E1801–E1810. (doi:10.1073/pnas.1205283109)
43. Khan A *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. (doi:10.1093/nar/gkx1126)
44. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009 BLAST+: architecture and applications. *BMC Bioinf.* 10, 421. (doi:10.1186/1471-2105-10-421)
45. Emms DM, Kelly S. 2015 OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* **16**, 157. (doi:10.1186/s13059-015-0721-2)
46. Ament SA, Corona M, Pollock HS, Robinson GE. 2008 Insulin signaling is involved in the regulation of worker division of labor in honey bee colonies. *Proc. Natl Acad. Sci. USA* **105**, 4226–4231. (doi:10.1073/pnas.0800630105)
47. Leonhardt SD, Menzel F, Nehring V, Schmitt T. 2016 Ecology and evolution of communication in social

insects. *Cell* **164**, 1277–1287. (doi:10.1016/j.cell.2016.01.035)

48. Withee JR, Rehan SM. 2017 Social experience, aggression and brain gene expression in a subsocial bee. *Integr. Comp. Biol.* **57**, 640–648. (doi:10.1093/icb/icx005)
49. 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. (doi:10.1186/s12915-018-0558-8)
50. Veenstra JA. 1991 Presence of Corazonin in three insect species, and isolation and identification of [His^7] Corazonin from *Schistocerca Americana*. *Peptides* **12**, 1285–1289. (doi:10.1016/0196-9781(91)90208-7)
51. 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. (doi:10.1016/j.jinsphys.2015.06.009)
52. Cardoso SD, Teles MC, Oliveira RF. 2015 Neurogenomic mechanisms of social plasticity. *J. Exp. Biol.* **218**, 140–149. (doi:10.1242/jeb.106997)
53. 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. (doi:10.1111/mec.14971)
54. 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. (doi:10.1111/j.1601-183X.2005.00131.x)
55. Toth AL, Varala K, Henshaw MT, Rodriguez-Zas SL, Hudson ME, Robinson GE. 2010 Brain transcriptomic analysis in paper wasps identifies genes associated with behaviour across social insect lineages. *Proc. R. Soc. B* **277**, 2139–2148. (doi:10.1098/rspb.2010.0090)
56. 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. (doi:10.1111/mec.12490)
57. 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. USA* **109**, 18 012–18 017. (doi:10.1073/pnas.1208094109)
58. Harpur BA, Kent CF, Molodtsova D, Lebon JMD, Alqarni AS, Owayss AA, Zayed A. 2014 Population genomics of the honey bee reveals strong signatures of positive selection on worker traits. *Proc. Natl Acad. Sci. USA* **111**, 2614–2619. (doi:10.1073/pnas.1315506111)
59. 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. (doi:10.1098/rspb.2014.1868)
60. 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. (doi:10.1371/journal.pone.0081518)
61. Xiao L, Priest MF, Nasebeny J, Lu T, Kozorovitskiy Y. 2017 Biased oxytocinergic modulation of midbrain dopamine systems. *Neuron* **95**, 368–384. (doi:10.1016/j.neuron.2017.06.003)
62. O'Rourke T, Boeckx C. 2019 Converging roles of glutamate receptors in domestication and prosociality. *bioRxiv*. (doi:10.1101/439869)