

Brain transcriptomic analysis in paper wasps identifies genes associated with behaviour across social insect lineages

Amy L. Toth, Kranthi Varala, Michael T. Henshaw, Sandra L. Rodriguez-Zas, Matthew E. Hudson and Gene E. Robinson

Proc. R. Soc. B published online 17 March 2010
doi: 10.1098/rspb.2010.0090

Supplementary data	"Data Supplement" http://rspb.royalsocietypublishing.org/content/suppl/2010/03/13/rspb.2010.0090.DC1.html
References	This article cites 43 articles, 17 of which can be accessed free http://rspb.royalsocietypublishing.org/content/early/2010/03/13/rspb.2010.0090.full.html#ref-list-1
P<P	Published online 17 March 2010 in advance of the print journal.
Subject collections	Articles on similar topics can be found in the following collections molecular biology (220 articles) behaviour (1129 articles) evolution (1536 articles)
Email alerting service	Receive free email alerts when new articles cite this article - sign up in the box at the top right-hand corner of the article or click here

Advance online articles have been peer reviewed and accepted for publication but have not yet appeared in the paper journal (edited, typeset versions may be posted when available prior to final publication). Advance online articles are citable and establish publication priority; they are indexed by PubMed from initial publication. Citations to Advance online articles must include the digital object identifier (DOIs) and date of initial publication.

To subscribe to *Proc. R. Soc. B* go to: <http://rspb.royalsocietypublishing.org/subscriptions>

Brain transcriptomic analysis in paper wasps identifies genes associated with behaviour across social insect lineages

Amy L. Toth^{1,2,7,*}, Kranthi Varala⁴, Michael T. Henshaw⁶, Sandra L. Rodriguez-Zas³, Matthew E. Hudson^{1,4} and Gene E. Robinson^{1,2,5}

¹Institute for Genomic Biology, ²Department of Entomology, ³Department of Animal Biology, ⁴Department of Crop Sciences, and ⁵Neuroscience Program, University of Illinois, Urbana-Champaign, IL, USA

⁶Department of Biology, Grand Valley State University, Allendale, MI, USA

⁷Department of Entomology, Pennsylvania State University, University Park, PA, USA

Comparative sociogenomics has the potential to provide important insights into how social behaviour evolved. We examined brain gene expression profiles of the primitively eusocial wasp *Polistes metricus* and compared the results with a growing base of brain gene expression information for the advanced eusocial honeybee, *Apis mellifera*. We studied four female wasp groups that show variation in foraging/provisioning behaviour and reproductive status, using our newly developed microarray representing approximately 3248 *P. metricus* genes based on sequences generated from high-throughput pyrosequencing. We found differences in the expression of approximately 389 genes across the four groups. Pathways known from *Drosophila melanogaster* to be related to lipid metabolism, heat and stress response, and various forms of solitary behaviour were associated with behavioural differences among wasps. Forty-five per cent of differentially expressed transcripts showed significant associations with foraging/provisioning status, and 14 per cent with reproductive status. By comparing these two gene lists with lists of genes previously shown to be differentially expressed in association with honeybee division of labour, we found a significant overlap of genes associated with foraging/provisioning, but not reproduction, across the two species. These results suggest common molecular roots for foraging division of labour in two independently evolved social insect species and the possibility of more lineage-specific roots of reproductive behaviour. We explore the implications of these findings for the idea that there is a conserved 'genetic toolkit' for division of labour across multiple lineages.

Keywords: paper wasp; honeybee; foraging behaviour; division of labour; gene expression; microarray

1. INTRODUCTION

Studies of the mechanisms underlying social behaviour have the potential to provide important insights into how complex phenotypes can evolve. Recent findings that honeybee social foraging behaviour is regulated by several genes associated with feeding behaviour in *Drosophila melanogaster* (Ben-Shahar *et al.* 2002; Ament *et al.* 2008) suggest that social behaviour can evolve from solitary behaviour *via* changes in the regulation of deeply conserved genes (Toth & Robinson 2007; Smith *et al.* 2008).

Based on such findings, it has been further suggested that there may be a 'genetic toolkit' for eusocial behaviour in the insects, similar to the genetic toolkit that has been elucidated for development, in which a core set of genes is used repeatedly during social evolution to generate novel forms of behaviour (Toth & Robinson 2007). A key prediction of this idea is that some of the same pathways that regulate division of labour in the well-studied honeybee (Smith *et al.* 2008) also regulate division of labour in other, independently evolved social insect lineages. However, to

date, cross-species comparisons of social insects have been limited for two reasons. First, there has been a paucity of advanced genomic resources for most species (reviewed in Smith *et al.* 2008). Second, direct comparisons have been challenging owing to differences across studies in the tissues and developmental stages analysed, as well as the techniques used to measure gene expression (reviewed in Sumner 2006; Smith *et al.* 2008). As a consequence, the extent of overlap across species, and thus the relevance of the 'toolkit' idea to social evolution, is unknown.

There have been several studies with the honeybee (*Apis mellifera*) examining genome-wide expression patterns associated with division of labour, providing an excellent basis to begin cross-species comparisons. This includes work on both the reproductive division of labour between queens and workers (Barchuk *et al.* 2007; Grozinger *et al.* 2007) and division of labour among workers (Whitfield *et al.* 2003, 2006; Cash *et al.* 2005).

Using honeybee gene expression data as a source of candidate genes, Toth *et al.* (2007) performed a comparative genomic analysis by using quantitative real-time PCR (qPCR) to measure brain expression for 32 genes in behaviourally differentiated groups of *Polistes metricus* wasps. Wasps and bees diverged approximately 100–150 Myr ago (Danforth *et al.* 2006), and social paper wasps have an independent origin of sociality from

* Author for correspondence (amytoth@iastate.edu).

† Present address: Department of Ecology, Evolution and Organismal Biology, Iowa State University, Ames, IA, USA.

Electronic supplementary material is available at <http://dx.doi.org/10.1098/rsob.2010.0090> or via <http://rsob.royalsocietypublishing.org>.

honeybees. Paper wasp division of labour is regulated by competition and dominance interactions among reproductively competent adults (Pardi 1948). By contrast, worker–worker division of labour in honeybees involves higher levels of cooperation and pheromone regulation (Winston 1987). Despite these differences in social biology, Toth *et al.* (2007) found a large overlap in patterns of brain gene expression in honeybees and paper wasps, suggesting strong commonalities in the regulation of different systems of division of labour at the molecular level. However, the same study used genes that were carefully selected based on their known associations with honeybee division of labour; thus, the actual extent of overlap between the two species could not be fairly estimated. For a broader set of genes, what is the extent of overlap across species? A more global analysis of *P. metricus* brain gene expression is necessary to further test the hypothesis of a genetic toolkit for division of labour in the social insects. We did this using a newly developed microarray platform, whose characterization is also presented here.

2. MATERIAL AND METHODS

(a) Wasps

We collected *P. metricus* from naturally occurring nests at Allerton Park in Monticello, Illinois, as in Toth *et al.* (2009). We studied four adult female groups: (i) *foundresses*, which initiate colonies in the spring. Although *P. metricus* sometimes cooperate to found nests, we only analysed solitary foundresses; (ii) *queens*, which are successful foundresses that have reared a set of adult worker offspring. Queens specialize mainly in egg-laying and cease foraging and provisioning brood; (iii) *workers*, which perform several behaviours that relate to colony growth and maintenance including foraging for food and provisioning brood. We focused on foraging workers by collecting only workers with evidence of wing wear, an indicator of foraging experience (Toth *et al.* 2009); and (iv) *gyenes*, which are future reproductive females that emerge late in the colony cycle. We collected gyenes before they dispersed from their natal nests to overwinter, during which time they generally do not work and are also not actively egg-laying. We collected 12 females from each group and attempted to collect queens, workers and gyenes from the same nests whenever possible. Wasps were paint-marked every 3–4 days, so the approximate dates of emergence of workers and gyenes were known. Complete information on collection dates and ages of wasps is in table S1 of the electronic supplementary material. All wasps were freeze-killed by collecting onto dry ice and stored at -80°C . Brains were dissected from frozen specimens as described in Toth *et al.* (2007).

As colony usurpation is common in *Polistes*, microsatellite analyses (Henshaw *et al.* 2004) were performed to verify that the queens were the mothers of the gyenes and workers from their colonies (as in Toth *et al.* 2009). The results were consistent with mother–daughter relationships for all but one gyene from one colony (table S2, electronic supplementary material).

(b) Microarrays

70 mer oligos were synthesized by Invitrogen (Carlsbad, CA, USA) and custom-printed onto microarrays at the Keck Center for Comparative and Functional Genomics at the

University of Illinois. This platform has been previously used successfully with honeybees (Kocher *et al.* 2008; Alaux *et al.* 2009a,b). Probes were designed by Invitrogen based on expressed sequence tag (EST) sequences derived from 454/Roche pyrosequencing (described in Toth *et al.* 2007). The sequences were assembled into consensus contigs using a slightly modified version of the STACKPACK pipeline (<http://www.eugenetics.com/stackpack.html>). ‘High-quality’ probes were designed based on contigs that: (i) contained over 100 sequences or (ii) had a ‘specificity score’ (supplied by Invitrogen as a metric of specificity) greater than 0.32, and either (iii) had a TBLASTX *e*-value $< 1 \text{ E-4}$ against the honeybee predicted genes (HoneyBee Genome Sequencing Consortium 2006) or (iv) had a BLASTX *e*-value $< 1 \text{ E-4}$ against the GenBank proteins. This resulted in 7207 oligos. Then, we excluded redundant matches to the same bee EST sequence (Whitfield *et al.* 2002) unless it matched a different GenBank protein, and redundant matches to the same GenBank protein unless it matched a different honeybee EST. This resulted in 5102 oligos. We then excluded all sequences with BLASTN *e*-values $< 1 \text{ E-6}$ to ribosomal RNA and all sequences with ‘probe scores’ (supplied by Invitrogen as a metric of hybridization quality and synthesis efficiency) less than 250, resulting in 4948 oligos.

Forty additional oligos were designed based on genes from consensus assembled sequences of interest (Toth *et al.* 2007) or their roles in behaviour and physiology in other insects, but were not included based on the aforementioned criteria. The best available oligonucleotide sequence from each targeted contig was used for oligo design. This gave a final number of 4988 *P. metricus* probes, which we estimated to correspond to at least 3248 different *P. metricus* genes based on TBLASTX matches to different honeybee predicted genes or BLASTX matches to different GenBank proteins (*e*-value $< 1 \text{ E-4}$). We also added 240 control probes (supplied by Invitrogen) based on unrelated sequences from other organisms. This gave a total of 5228 probes on the array, each spotted twice for a total of 10 456 spots.

We used a standard two dye (Cy3 and Cy5) microarray protocol as previously described (Alaux *et al.* 2009a). The Maui hybridization system (BioMicro Products, Salt Lake City, UT, USA) significantly improved hybridization efficiency and was used for all experimental arrays. We used a loop design, consisting of 48 individual wasps (12 per group) paired on 72 arrays with dye swaps. Each wasp brain was analysed on either two or four arrays (figure S1, electronic supplementary material).

To validate the performance of the microarray, we prepared a set of two self–self hybridizations derived from pooled samples of wasps. Each pooled sample was split into two, one half labelled with Cy3, the other with Cy5, and these were hybridized together on the same microarray. These tests allowed us to characterize hybridization efficiency of the array probes without the complication of expression differences.

(c) Statistical analyses

Microarray spots were quality filtered by ‘spotfinding’ as in Alaux *et al.* (2009a). Fluorescence intensities were normalized using a LOWESS transformation (Yang *et al.* 2002). Duplicate spots for each gene were averaged and adjusted for microarray and dye effects. A mixed-effects ANOVA model using PROC MIXED in SAS was used to describe

the expression measurements of each oligo. The model included the fixed effects of dye, foraging/provisioning (P), reproduction (R) and the interaction of the two terms, and the random effects of wasp and array. Each group was coded as follows (as in Toth *et al.* 2007, 2009): foundresses (P+/R+), gynes (P-/R-), queens (P-/R+) and workers (P+/R-). Significance was based on a false discovery rate of $p < 0.05$, using an *F1* statistic. After the analysis was completed, we eliminated oligos with less than 60 per cent of the spots above background intensity across all arrays. We performed an alternative ANOVA analysis using PROC MIXED that included the fixed effects of dye and group (which had four levels: 'foundress', 'gyne', 'queen' and 'worker'), and the random effects of wasp, array and colony. These results are not presented in the main text but are included in table S3 of the electronic supplementary material.

The resulting list of oligos showing significant differences in expression ('wasp differentially expressed, full list') was used for subsequent analyses. To look at overall patterns in the data and understand the relationships among overall gene expression patterns between the four groups, we performed several multivariate analyses. We performed hierarchical clustering analysis (HCA) by oligo and wasp using GENESIS (Sturn *et al.* 2002). We performed a linear discriminant analysis (LDA) using the 'lda' function in R. Because LDA requires there to be more individual samples than variables, we randomly selected 40 oligos from the differentially expressed list for LDA.

In addition, we used *p*-values for each factor in the ANOVA to cut the *Polistes* gene list into two lists, one containing genes differentially expressed in association with foraging/provisioning status ('wasp foraging/provisioning sublist') and the second containing genes differentially expressed in association with reproductive status ('wasp reproduction sublist'). We compared these lists with gene lists from prior honeybee studies. We used lists from two studies comparing brain gene expression in nurse and forager honeybees ('bee foraging lists'; Whitfield *et al.* 2003; Alaux *et al.* 2009b) and one comparing brain gene expression in queens and 'sterile' workers ('bee reproduction list'; Grozinger *et al.* 2007).

We limited the bee–wasp comparisons to wasp oligos for which there was a good TBLASTX hit to a honeybee gene (*e*-value $< 1 \text{ E-}4$) and to genes that were analysed on both bee and wasp array platforms. A gene was counted as overlapping regardless of the direction of expression (whether the expression level was up- or downregulated in one species relative to the other). This approach is justified because several studies suggest that genes (e.g. the *foraging* gene in bees and ants) can retain conserved roles in behaviour across species, but that the direction of gene regulation may be evolutionarily labile (Robinson & Ben-Shahar 2002; Ingram *et al.* 2005). We used two independent statistical tests to examine the overlap between genes that were differentially regulated in wasps to the three aforementioned bee gene lists. First, we used one-tailed Fisher exact tests to compare the number of observed overlaps to a null hypothesis of the number of overlaps expected by random chance. Second, simulation tests were performed by drawing a random set of *x* and *y* genes from each of the two lists being compared, where *x* is the number of genes significant in the wasp list and *y* is the number significant in the bee list. The number of overlapping genes between each set of two samplings represents one randomly drawn overlap. This sampling regime

(without replacement) was repeated 4000 times, creating a distribution of overlaps. The probability of the actual overlap was then calculated based on the simulated distribution.

In addition to the aforementioned comparisons, tests of the probability of overlap (again using Fisher exact tests and simulations) were also performed for five additional wasp gene sublists that reflected group-specific patterns, and the results are reported in the electronic supplementary material.

Possible functional significance was explored with gene ontology (GO) analysis. We tested for enriched terms in the three wasp gene lists: 'wasp differentially expressed, full list', 'wasp foraging/provisioning sublist' and 'wasp reproduction sublist' (using DAVID: Dennis *et al.* 2003; Huang *et al.* 2009). We focused only on genes for which there were good TBLASTX hits to annotated genes in *D. melanogaster* (*e*-value $< 1 \text{ E-}4$). We used a raw *p*-value cut-off of $p < 0.05$ and a minimum of three gene products in a given category, and only considered GO categories at the level of 'biological process'.

3. RESULTS

(a) Validation of microarray performance

In both replicates of the 'self–self' hybridizations, approximately 95 per cent of the spots on the two test arrays hybridized at a 'sum of median intensity in both channels 635/532 nm' (SMI) of greater than 100 (above background intensity, 40–100). We found a wide range of SMI values (mean $\pm$ s.e. = 3569 ± 68). There were control spots on the array consisting of probes with a gradient of incorrect nucleotides (5%, 10%, 15%, 20%, 30%, 40%, 50% mismatch) for 20 randomly selected oligos. As expected, these showed a decrease in average SMI as the per cent mismatch increased (7303 (5%), 1106 (10%), 227 (15%), 310 (20%), 202 (30%), 229 (40%) and 201 (50%)). For each of these 20 oligos, there was a negative control (the same sequence randomly scrambled), which had an average SMI of 149. The average SMI for blank spots on the test arrays was 69 (in the range of background intensity). We also found a significant correlation between the SMI values of duplicate spots ($R = 0.96$, $p < 0.001$). These diagnostics gave us confidence that the array was producing reliable hybridizations.

We also used the complete set of experimental microarrays for further validation. We tested whether the expression patterns agreed with two predictions: (i) oligos with a best match to the same putative wasp gene (e.g. had the same best TBLASTX hit to a honeybee gene) should result in a similar level of hybridization and (ii) the expression ratios for the four groups should be correlated with values from our previous qPCR-based study (Toth *et al.* 2007). For both of these analyses, we used normalized mean expression estimates from the ANOVA described above. To address the first prediction, we compared the average estimate for all pairs of wasp contigs that had a best TBLASTX hit to the same honeybee gene. We found a highly significant correlation across the wasp contig pairs ($R = 0.654$, $p < 0.0001$, $n = 1206$ contigs corresponding to 603 honeybee genes). To address the second prediction, we compared mean fold-difference across six comparisons (foundress–gyne, foundress–queen, foundress–worker, gyne–queen,

Table 1. Some genes associated with both foraging/provisioning in *P. metricus* wasps (this study) and worker foraging behaviour in *A. mellifera* honeybees. (The honeybee reference is shown in parentheses after each *P. metricus* gene. *Polistes metricus* gene names and putative functions are based on significant similarity (TBLASTX *e*-value < 1 E-4) to *D. melanogaster* genes with known functions (Tweedie *et al.* 2009).)

<i>Polistes metricus</i> putative gene	inferred function
<i>Pmforaging</i> (Ben-Shahar <i>et al.</i> 2002)	cGMP-dependent protein kinase, foraging behaviour
<i>PmVitellogenin</i> (Amdam <i>et al.</i> 2004)	egg yolk protein
<i>Pmalpha-2-macroglobulin related</i> (Whitfield <i>et al.</i> 2003)	phagocytosis, engulfment
<i>PmNa channel subunit alpha</i> (Whitfield <i>et al.</i> 2003)	locomotory behaviour, determination of adult lifespan
<i>Pmbellwether</i> (Whitfield <i>et al.</i> 2003; Alaux <i>et al.</i> 2009b)	ATPase activity, growth, spermatid development
<i>PmCG12262-like</i> (Whitfield <i>et al.</i> 2003)	fatty acid beta-oxidation
<i>PmCG4164-like</i> (Whitfield <i>et al.</i> 2003)	response to heat, protein folding
<i>PmCalreticulin</i>	central nervous system development, olfactory behaviour
<i>PmElongation factor 2a</i> (Whitfield <i>et al.</i> 2003)	translation elongation
<i>PmHeat shock factor</i> (Whitfield <i>et al.</i> 2003)	response to heat
<i>PmInositol-3-phosphate synthase</i> (Whitfield <i>et al.</i> 2003)	inositol biosynthetic process
<i>PmInositol 1,4,5,-tris-phosphate receptor 83A</i> (Whitfield <i>et al.</i> 2003)	flight behaviour, mushroom body development
<i>PmOrganic anion transporting polypeptide 33Ea</i> (Whitfield <i>et al.</i> 2003)	organic anion transport
<i>PmPinocchio</i> (Whitfield <i>et al.</i> 2003; Alaux <i>et al.</i> 2009b)	olfactory behaviour, response to chemical stimulus
<i>PmPeroxidase</i> (Whitfield <i>et al.</i> 2003)	phagocytosis, engulfment
<i>Pmslowpoke</i> (Whitfield <i>et al.</i> 2003)	potassium ion transport, circadian behaviour
<i>Pmsynaptogyrin</i> (Whitfield <i>et al.</i> 2003)	synaptic vesicle exocytosis
<i>Pmsynaptotagmin</i> (Whitfield <i>et al.</i> 2003)	neurotransmitter secretion, larval locomotory behaviour
<i>PmThiolester containing protein III</i> (Whitfield <i>et al.</i> 2003)	phagocytosis, engulfment
<i>Pmturtle</i> (Whitfield <i>et al.</i> 2003)	adult locomotory behaviour, flight behaviour
<i>PmUcp4A</i> (Whitfield <i>et al.</i> 2003)	proton transport, mitochondrial transport
<i>PmCholine acetyltransferase</i> (Alaux <i>et al.</i> 2009b)	neuromuscular synaptic transmission
<i>Pmswiss cheese</i> (Alaux <i>et al.</i> 2009b)	brain development
<i>PmEcdysone induced protein 71CD</i> (Alaux <i>et al.</i> 2009b)	sulphur amino acid metabolism
<i>PmAllatostatin</i> (Alaux <i>et al.</i> 2009b)	neuropeptide hormone
<i>Pmnervana 2</i> (Alaux <i>et al.</i> 2009b)	cation transmembrane transporter
<i>PmHelix Loop Helix protein 106</i> (Alaux <i>et al.</i> 2009b)	transcription factor
<i>PmLarval serum protein 2</i> (Alaux <i>et al.</i> 2009b)	nutrient reservoir activity
<i>PmElongation factor 2b</i> (elongation factor 2b) (Alaux <i>et al.</i> 2009b)	translation elongation
<i>PmCytochrome p450 reductase</i> (Alaux <i>et al.</i> 2009b)	oxidation reduction
<i>PmRabphilin</i> (Alaux <i>et al.</i> 2009b)	neurotransmitter secretion
<i>PmCad87A</i> (Alaux <i>et al.</i> 2009b)	homophilic cell adhesion

gyne–worker and queen–worker). For 10 oligos based on the same exact EST contigs from both qPCR and the microarray, we found a significant correlation between the fold-differences ($R = 0.30$, $p = 0.02$, $n = 10$). Although the correlation was not high, we considered it to be confirmatory considering the two sets of analyses were performed on different wasps collected in different years.

(b) Differentially expressed genes

A total of 447 oligos were differentially expressed across the four groups. These 447 oligos corresponded to 389 different genes (based on best TBLASTX hits to honeybee predicted genes or best BLASTX hits to GenBank proteins), which is approximately 12 per cent of the genes represented on the microarray. Of these, expression differences for 45 per cent (201 oligos) were associated with differences in foraging/provisioning status, 14 per cent (63 oligos) with reproduction and 38 per cent (169 oligos) with the interaction between these two terms. A description of the expression patterns of candidate genes from our previous study (Toth *et al.* 2007)

including *Pmforaging* and *PmVitellogenin* can be found in the electronic supplementary material. A complete list of all differentially expressed oligos (and BLAST hits for each oligo) is given in table S3 of the electronic supplementary material, and a subset of the ‘wasp foraging/provisioning sublist’ is shown in table 1.

(c) Overall patterns

HCA by oligo showed several different patterns in the expression data (figure 1). For example, there were clusters of genes that reflected group-specific patterns and both reproduction and foraging/provisioning-related patterns. In some clusters, gynes and foundresses showed similar patterns. Although gynes and foundresses have nothing in common with respect to foraging/provisioning or reproductive states, they both have limited social contact in comparison with queens and workers (i.e. gynes at this stage are not part of a dominance hierarchy, and foundresses are alone).

HCA by group (figure 1) showed a different topology from that of the previously published qPCR study on these same four groups (Toth *et al.* 2007). In the current

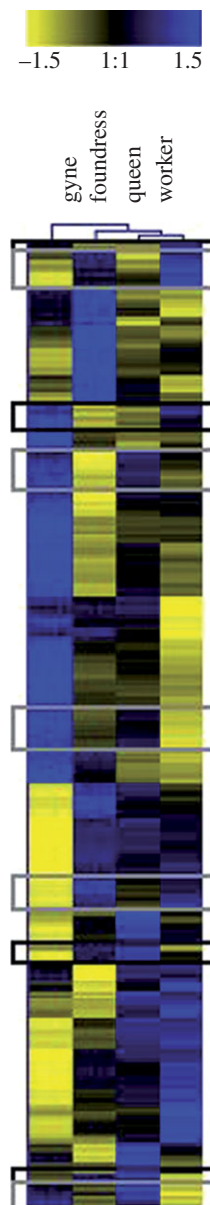


Figure 1. Predominant patterns of brain gene expression as revealed by HCA by oligo and group for the 447 differentially regulated oligos. Mean expression values for each oligo are colour coded: blue represents upregulation, yellow represents downregulation and black represents equal expression. Numerous expression patterns are present, and those that show reproduction-related patterns are highlighted with black boxes and foraging-provisioning-related patterns with grey boxes ($n = 477$ oligos).

study, workers and queens had the most similar expression patterns, in contrast to workers and foundresses having the most similar expression patterns in the earlier study. The previous study (Toth *et al.* 2007) involved only genes known to be associated with honeybee division of labour, whereas this study involved a broader set. In both studies, gynes had the most divergent expression patterns.

LDA resulted in three linear discriminants (LDs), which accounted for 41, 37 and 22 per cent of the variation in the data. A scatterplot of LD1 and LD2 (78% of the variation) lends further support to the result that gynes had the most divergent expression patterns (figure 2). In addition, the LDA further supported the

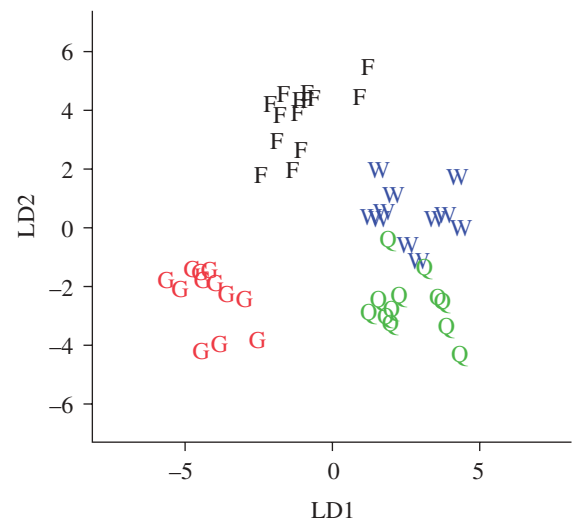


Figure 2. Linear discriminants (LD) 1 and 2 from an LDA on a randomly selected subset of 40 significantly different oligos. F, foundress; G, gyne; Q, queen and W, worker. LD1 versus LD2 accounted for 41% and 37% of the variation in the data, respectively.

finding from HCA that workers and queens have the most similar expression patterns. However, the LD scatterplot shows minimal overlap between all groups, even queens and workers, indicating each group has a distinct expression profile.

(d) *Overlap with honeybee transcriptomic studies*

We examined the extent of overlap with previous studies of honeybee brain gene expression related to foraging and reproduction. The number of wasp genes overlapping for each bee gene list is shown in figure 3*a–c*. Genes on the ‘wasp reproduction sublist’ showed no significant overlap with the ‘bee foraging lists’ (Whitfield *et al.* 2003; Alaux *et al.* 2009*b*; figure 3*a,b*, respectively). By contrast, based on both Fisher exact tests and simulations, we found significant overlap between the ‘wasp foraging/provisioning sublist’ and the ‘bee foraging list’ from Alaux *et al.* (2009*b*), and a marginally significant overlap between the ‘wasp foraging/provisioning sublist’ with the ‘bee foraging list’ from Whitfield *et al.* (2003). Table 1 gives a list of a subset of the genes that overlapped.

Genes on the ‘wasp reproduction sublist’ showed no significant overlap with the ‘bee reproduction sublist’ (Grozing *et al.* 2007; figure 3*c*), and likewise for the two ‘bee foraging lists’.

(e) *GO functional analysis of gene lists*

GO results for enriched ‘biological process’ categories are summarized in table 2 (and detailed results are presented in table S4 of the electronic supplementary material). For the ‘wasp foraging/provisioning sublist’, we found genes in three main biological processes were overrepresented: lipid metabolism, response to stress and heat, and behaviour (including response to other organisms and locomotory behaviour). For the ‘wasp reproduction sublist’, one main biological process was overrepresented—response to heat and chemical stimuli.

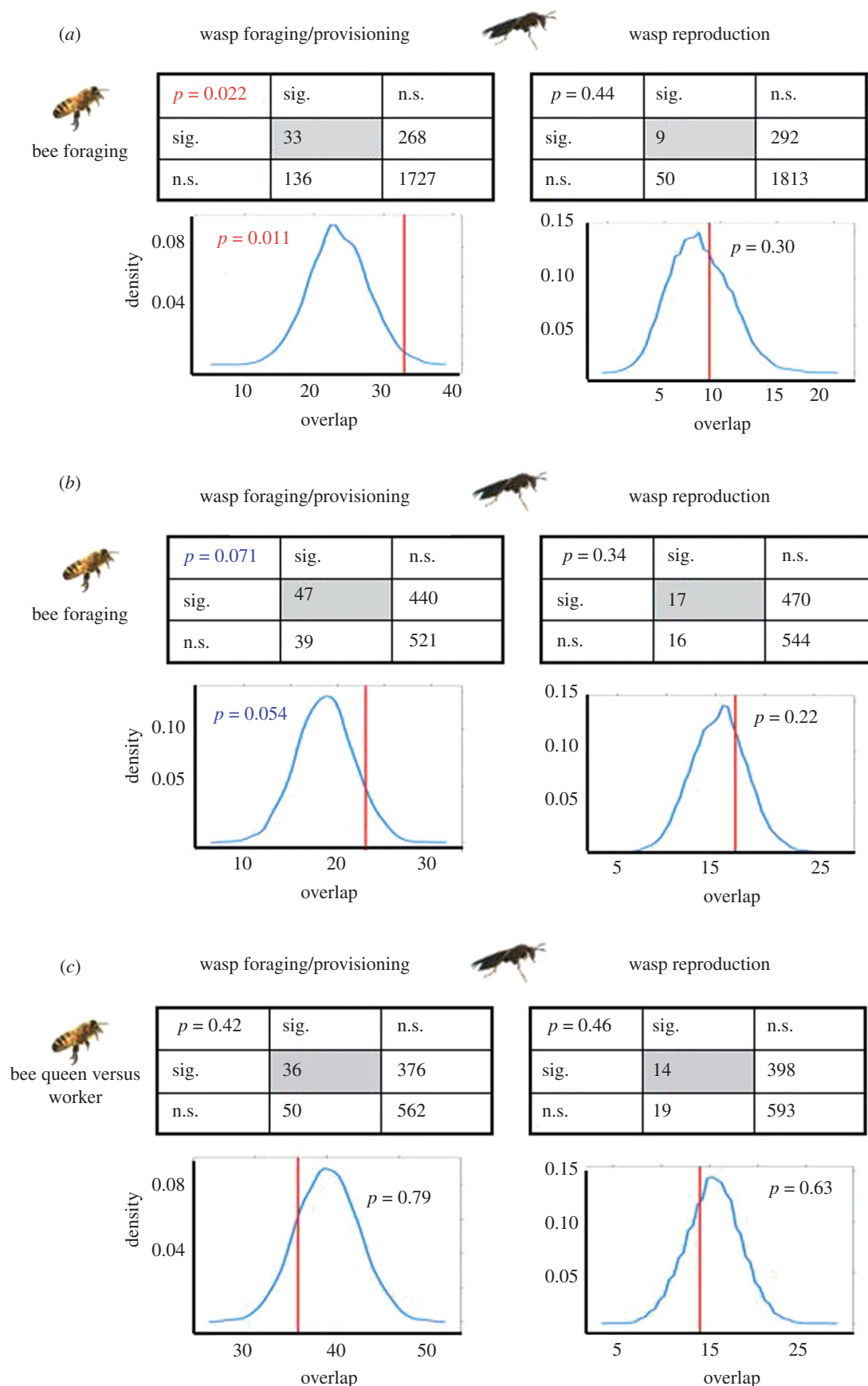


Figure 3. Tests of the extent of overlap between lists of differentially regulated genes from three different microarray studies in honeybees (*A. mellifera*, in *a–c*) and two lists of differentially regulated genes from the present study of *P. metricus* wasps. For each comparison, the upper panel is a table used for a Fisher exact test, which shows the breakdown of the genes found to be significantly different (sig.) or not different (n.s.) in each study; the p -value from the test is shown in the upper left cell of each table. The lower panel shows the simulated distribution of random overlaps for the two studies, the red line indicates the observed overlap, and the associated probability of that observation is given. Comparisons with honeybees were based on studies measuring gene expression in hive workers versus foragers: (a) Alaux *et al.* (2009b) and (b) Whitfield *et al.* (2003), and queens versus workers (c) Grozinger *et al.* (2007).

Table 2. Summary of GO functional analysis of wasp differentially expressed gene lists (subsets with significant similarity (TBLASTX e -value $< 1 \times 10^{-4}$) to *D. melanogaster* annotated genes). (Results are shown for the 'wasp differentially expressed, full list', the 'wasp foraging/provisioning sublist' and the 'wasp reproduction sublist'. 'no. genes' corresponds to the number of genes for which there were *D. melanogaster* genes with significant similarity or the number of genes that were represented in each significant category. Detailed results are given in table S4 of the electronic supplementary material.)

gene list	no. genes	GO category
all significant	374	
	16	heat shock, protein folding
	8	transport of amino and carboxylic acids
	15	multiorganism process, including mating, courtship and behavioural interaction
	15	metabolism of alcohols, including glycerol
	10	regulation of alternative nuclear mRNA splicing
provisioning	216	
	9	mesoderm development
	15	metabolism of lipids and fatty acids
	9	locomotory behaviour
	12	response to biotic stimulus, multiorganism process
	14	stress response, response to heat
	10	regulation of cell morphogenesis
reproduction	93	
	9	response to heat and chemical stimulus

4. DISCUSSION

This paper provides, to our knowledge, the first large-scale transcriptomic analysis of brain gene expression in a social insect species other than the honeybee. We detected 389 genes (approx. 12% of genes on the array) that differed in brain expression across four female groups in a paper wasp society. Several different patterns of gene expression were found, some that represent group-specific patterns and others that correlate with differences in foraging/provisioning behaviour and reproductive state.

The extent of overlap between paper wasps and honeybees for genes related to foraging behaviour was higher than that predicted by chance, suggesting there may be common elements in the gene networks regulating division of labour in the two lineages, despite their independent origins of sociality. We also found numerous differences in gene expression between the two species, pointing to many lineage-specific mechanisms. This is not surprising, considering that bees and paper wasps are separated by over 100 million years of evolution. Overall, our results suggest a combination of conserved and unique mechanisms regulate social behaviour in these phylogenetically distant social insect species.

Results for foraging, but not reproduction, provide support for the idea of a genetic 'toolkit' for division of labour (Toth & Robinson 2007). Several of our analyses

show that genes associated with worker foraging in honeybees are likely to be associated with foraging/provisioning behaviour in wasps. First, we found that 12 per cent of the genes on the microarray were differentially regulated across the four female groups, which contrasts sharply with the results of our previous (qPCR) study, in which 63 per cent of genes were differentially regulated (Toth et al. 2007). This large discrepancy suggests that Toth et al. (2007) indeed examined a select gene set; the genes were chosen based on their known associations with division of labour in honeybees. Second, for this select set, overall expression patterns in wasps were most similar in foundresses and workers, the two foraging/provisioning groups (Toth et al. 2007). By contrast, for all 389 differentially regulated genes from the current study, expression patterns were largely distinct for each group (figure 2). Thus, foraging/provisioning-related expression patterns in wasps were more prevalent for genes that relate to foraging in honeybees. Finally, meta-analysis of microarray data showed a higher than random overlap between genes related to wasp foraging/provisioning and honeybee foraging. These results demonstrate common patterns of gene expression associated with foraging behaviour in wasps and bees.

Genes for lipid metabolism, locomotory behaviour and response to heat stress were overrepresented on the list of genes regulated in association with wasp foraging/provisioning. Although the overlapping wasp and bee gene lists were too small for GO analysis, it is notable that several genes in these same categories are also differentially regulated in other species. Heat shock protein and metabolic genes are differentially expressed in several honeybee brain transcriptomic studies related to foraging (Whitfield et al. 2003, 2006; Ament et al. 2008; Alaux et al. 2009b). Moreover, similar genes have been identified in whole body gene expression analyses of other social insect species. *Polistes canadensis* wasp queens and workers were found to differ in expression of *Vitellogenin*, a heat shock protein, and numerous metabolic enzymes (Sumner et al. 2006). Graff et al. (2007) discovered queen-worker differences in *Vitellogenin* expression in the ant *Lasius niger*. Hoffman & Goodisman (2007) found differences in yellow jackets in *Hexamerin* storage proteins between queen and worker castes across different life stages. Synthesizing this information with the results of the current study, we suggest that storage proteins such as *Vitellogenin* and *Hexamerin*, genes related to stress and heat shock, and metabolic enzymes involved in lipid metabolism may have far-reaching effects on, or may be strongly affected by, caste differences across a wide variety of species and tissue types.

In contrast to the foraging results, we did not find significant overlap in brain gene expression patterns for wasp reproduction and honeybee queen-worker caste differences. These results suggest that different molecular mechanisms regulate reproductive behaviour in the two species. This is consistent with the fact that reproductive physiology differs fundamentally between these two species. In *Polistes*, ovary development and reproductive behaviours are regulated by the insect gonadotropin, juvenile hormone (JH) (Bohm 1972; Roseler et al. 1985; Sledge et al. 2004). In honeybees, JH levels are uncoupled from ovarian development and egg-laying behaviour in queens, a situation that may also characterize advanced eusocial stingless bees (Hartfelder et al. 2001),

but not primitively eusocial bumble bees (Geva *et al.* 2005). Thus, it is possible that the genetic mechanisms regulating egg-laying behaviour in paper wasps may more closely resemble those of solitary ancestors, whereas a different suite of genes regulate adult reproduction in the derived social system of the honeybee. However, like any finding of this nature, it is also possible that this is a 'false negative' result. This can be addressed in the future when more genes are added to the *P. metricus* microarray after additional sequencing, or with the application of more powerful techniques than comparing lists of differentially expressed genes, e.g. inferring gene regulatory networks based on gene expression and transcription factor binding data (Ament *et al.* in press). Comparative studies of global brain gene regulation in additional social and solitary Hymenoptera will also help to resolve this issue.

The discrepancies we noted for foraging- versus reproduction-related gene expression are consistent with other recent findings. For example, a study with *Cryptocercus* termites found that only three out of 10 genes associated with queen-worker differences showed conserved expression patterns across two closely related species (Weil *et al.* 2009). By contrast, other studies of aggression and reward-seeking behaviour (Alaux *et al.* 2009b; Barron *et al.* 2009) suggest that there may be some common elements regulating these forms of behaviour across a wide range of animal taxa, from arthropods to vertebrates. Clearly, additional genomic-level studies on diverse species will be necessary to gain further insight into issues related to how pervasive genetic toolkits are for behaviour.

Our finding that 12 per cent of the genes on the microarray were differentially regulated across the four female *Polistes* groups contrasts with results from honeybee studies, in which different behaviourally defined groups of individuals differ for 25–40% of the genes on the microarray (Whitfield *et al.* 2003; Grozinger *et al.* 2007; Alaux *et al.* 2009b). This observation suggests that division of labour in primitively eusocial insects might be associated with more subtle differences in brain gene expression when compared with advanced eusocial insects. This is consistent with the fact that primitively eusocial individuals retain more behavioural totipotency than in advanced eusocial insects. Perhaps more changes in gene regulation are required for increased caste specialization. Alternatively, this finding might also change when more genes are added to the *Polistes* microarray after additional sequencing. Our results provide an important starting point for exploring whether differences in the number of differentially expressed genes are correlated with differences in the degree of behavioural specialization in social insects.

Some cross-species comparisons of gene expression (Jordan *et al.* 2005) suggest that variation in gene expression for some genes can be tolerated without having an effect on phenotype, in a form of 'neutral variation' in expression level. Our data show differential gene regulation between groups of the same population associated with behavioural differences, as inferred from the significant differences in gene expression. If this difference were owing to high levels of phenotype-neutral variance within the population, we would not expect these changes to be statistically significant between these groups. Since the groups were sampled from the same population; population structure, drift or fixation cannot

account for these differences. In addition, we see a significantly higher than random overlap of genes related to foraging division of labour in two species, indicating a possible functional/mechanistic component for the observed differences. By contrast, the neutral hypothesis would predict the same extent of overlap for both reproduction- and foraging-related lists.

It remains to be seen which differences in gene expression cause differences in behaviour and which differences in gene expression occur as a result of changes in behaviour. If some are causal, it will be important to consider whether shared mechanisms are the result of convergence on the same mechanism or conservation of function across lineages (Toth & Robinson 2007). One possibility is that convergent phenotypes, such as division of labour for foraging, use convergent mechanisms. This would suggest that certain molecular pathways may be more evolutionarily labile and are recruited multiple times during the evolution of sociality to affect social behaviour. There is evidence for this phenomenon from evolutionary studies of development, e.g. homeotic developmental genes have been recruited to control various forms of morphological novelties in different vertebrate lineages (Carroll *et al.* 2005).

A second possibility is conservation—division of labour in different lineages evolved from a set of behavioural/physiological states shared by their common ancestor, and these behaviours evolved into foraging/non-foraging and reproductive/non-reproductive castes in different lineages. Several of the genes related to foraging behaviour in both bees and wasps (table 1) are also related to flight and locomotor behaviour in *D. melanogaster*, and the *foraging* gene has been implicated in the regulation of foraging in several different social insect species (Ben-Shahar *et al.* 2002; Ingram *et al.* 2005; Tobback *et al.* 2008; Lucas & Sokolowski 2009; present study). West-Eberhard (1996) proposed a similar idea based on ancestral reproductive characters, and additional solitary 'groundplan' hypotheses have recently received some experimental support (Amdam *et al.* 2004; Hunt *et al.* 2007). Further comparative genomic studies across different social and solitary lineages can help to distinguish between these two evolutionary scenarios and thus provide additional insights into the evolutionary processes by which complex behaviour can arise.

We would like to thank K. S. Pruiett, T. D. Daugherty, S. H. Woodard, J. H. Hunt and K. J. Bilof for assistance with fieldwork, T. Nguyen for performing array hybridizations, F. E. Miguez for statistical advice, M. R. Band for advice on microarrays, L. Qi for spotfinding microarrays and J. H. Hunt and members of the C. M. Grozinger laboratory for comments on the manuscript. This work was supported by NSF grant IOS-0641431 (Behavioural Systems Cluster, G.E.R.) and the Illinois Sociogenomics Initiative (G.E.R.). Microarray data have been deposited into ArrayExpress (<http://www.ebi.ac.uk/microarray-as/ae/>), ArrayExpress accession: E-MEXP-2558.

REFERENCES

- Alaux, C., Le Conte, Y., Adams, H. A., Rodriguez-Zas, S., Grozinger, C. M., Sinha, S. & Robinson, G. E. 2009a Regulation of brain gene expression in honey bees by brood pheromone. *Genes Brain Behav.* **8**, 309–319. (doi:10.1111/j.1601-183X.2009.00480.x)

- Alaux, C., Sinha, S., Hasadsri, L., Hunt, G. J., Guzman-Novoa, E., DeGrandi-Hoffman, G., Uribe-Rubio, J. L., Rodriguez-Zas, S. L. & Robinson, G. E. 2009b Honey bee aggression supports a link between gene regulation and behavioral evolution. *Proc. Natl Acad. Sci. USA* **106**, 15 400–15 405. (doi:10.1073/pnas.0907043106)
- 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. USA* **101**, 11 350–11 355. (doi:10.1073/pnas.0403073101)
- Ament, S. A., Corona, M., Pollock, H. S. & Robinson, G. E. 2008 Insulin signaling is involved in the regulation of worker division of labour in honey bee colonies. *Proc. Natl Acad. Sci. USA* **105**, 4226–4231. (doi:10.1073/pnas.0800630105)
- Ament, S. A., Wang, Y. & Robinson, G. E. In press. Nutritional regulation of division of labor in honey bees: toward a systems biology perspective. *Wiley Interdisciplinary Rev.: Syst. Biol. Med.* (doi:10.1002/wsbm.73)
- Barchuk, A. R., Cristino, A. S., Kucharski, R., Costa, L. F., Simoes, Z. L. P. & Maleszka, R. 2007 Molecular determinants of caste differentiation in the highly eusocial honeybee *Apis mellifera*. *BMC Dev. Biol.* **7**, 70. (doi:10.1186/1471-213X-7-70)
- Barron, A. B., Maleszka, R., Helliwell, P. G. & Robinson, G. E. 2009 Effects of cocaine on honey bee dance behaviour. *J. Exp. Biol.* **212**, 163–168. (doi:10.1242/jeb.025361)
- Ben-Shahar, Y., Robichon, A., Sokolowski, M. B. & Robinson, G. E. 2002 Influence of gene action across different time scales on behavior. *Science* **296**, 741–744. (doi:10.1126/science.1069911)
- Bohm, M. K. 1972 Effects of environment and juvenile hormone on ovaries of the wasp, *Polistes metricus*. *J. Insect Physiol.* **18**, 1875–1883. (doi:10.1016/0022-1910(72)90158-8)
- Carroll, S. B., Grenier, J. K. & Weatherbee, S. D. 2005 *From DNA to diversity: molecular genetics and the evolution of animal design*, 2nd edn. Malden, MA: Blackwell Publishing.
- Cash, A. C., Whitfield, C. W., Ismail, N. & Robinson, G. E. 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)
- Danforth, B. N., Brady, S. G., Sipes, S. D. & Pearson, A. 2006 Single-copy nuclear genes recover Cretaceous-age divergences in bees. *Syst. Biol.* **53**, 309–326. (doi:10.1080/10635150490423737)
- Dennis Jr, G., Sherman, B. T., Hosack, D. A., Yang, J., Gao, W., Lane, H. C. & Lempicki, R. A. 2003 DAVID: database for annotation, visualization, and integrated discovery. *Genome Biol.* **4**, 3. (doi:10.1186/gb-2003-4-5-p3)
- Geva, S., Hartfelder, K. & Bloch, G. 2005 Reproductive division of labour, dominance, and ecdysteroid levels in hemolymph and ovary of the bumble bee *Bombus terrestris*. *J. Insect Physiol.* **51**, 811–823. (doi:10.1016/j.jinsphys.2005.03.009)
- Graff, J., Jemielity, S., Parker, J. D., Parker, K. M. & Keller, L. 2007 Differential gene expression between adult queens and workers in the ant *Lasius niger*. *Mol. Ecol.* **16**, 675–683. (doi:10.1111/j.1365-294X.2007.03162.x)
- Grozinger, C. M., Fan, Y., Hoover, S. E. R. & Winston, M. L. 2007 Genome-wide analysis reveals differences in brain gene expression patterns associated with caste and reproductive status in honey bees (*Apis mellifera*). *Mol. Ecol.* **16**, 4837–4848. (doi:10.1111/j.1365-294X.2007.03545.x)
- Hartfelder, K., Bitondi, M. M. G., Santana, W. C. & Simões, Z. L. P. 2001 Ecdysteroid titer and reproduction in queens and workers of the honey bee and of a stingless bee: loss of ecdysteroid function at increasing levels of sociality? *Insect Biochem. Mol. Biol.* **32**, 211–216. (doi:10.1016/S0965-1748(01)00100-X)
- Henshaw, M. T., Robson, S. K. A. & Crozier, R. H. 2004 Queen number, queen cycling and queen loss: the evolution of complex multiple queen societies in the social wasp genus *Ropalidia*. *Behav. Ecol. Sociobiol.* **55**, 469–476. (doi:10.1007/s00265-003-0725-x)
- Hoffman, E. A. & Goodisman, M. A. D. 2007 Gene expression and the evolution of phenotypic diversity in social wasps. *BMC Biol.* **5**, 23. (doi:10.1186/1741-7007-5-23)
- HoneyBee Genome Sequencing Consortium 2006 Insights into social insects from the genome of the honey bee *Apis mellifera*. *Nature* **443**, 931–948. (doi:10.1038/nature05260)
- Huang, D. W., Sherman, B. T. & Lempicki, R. A. 2009 Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57. (doi:10.1038/nprot.2008.211)
- Hunt, J. H. & Amdam, G. V. 2005 Bivoltinism as an antecedent to eusociality in the paper wasp genus *Polistes*. *Science* **308**, 264–267. (doi:10.1126/science.1109724)
- Hunt, J. H. et al. 2007 A diapause pathway underlies the gyne phenotype in *Polistes* wasps, revealing an evolutionary route to caste-containing insect societies. *Proc. Natl Acad. Sci. USA* **104**, 14 020–14 025.
- Ingram, K. K., Oefner, P. & Gordon, D. M. 2005 Task-specific expression of the foraging gene in harvester ants. *Mol. Ecol.* **14**, 813–818. (doi:10.1111/j.1365-294X.2005.02450.x)
- Jordan, I. K., Mariño-Ramírez, L. & Koonin, E. V. 2005 Evolutionary significance of gene expression divergence. *Gene* **345**, 119–126. (doi:10.1016/j.gene.2004.11.034)
- Kocher, S. D., Richard, F., Tarpy, D. R. & Grozinger, C. M. 2008 Genomic analysis of post-mating changes in the honey bee queen *Apis mellifera*. *BMC Genomics* **9**, 232. (doi:10.1186/1471-2164-9-232)
- Lucas, C. & Sokolowski, M. B. 2009 Molecular basis for changes in behavioural state in ant social behaviours. *Proc. Natl Acad. Sci. USA* **106**, 6351–6356. (doi:10.1073/pnas.0809463106)
- Pardi, L. 1948 Dominance order in *Polistes* wasps. *Physiol. Zool.* **21**, 1–13.
- Robinson, G. E. & Ben-Shahar, Y. 2002 Social behavior and comparative genomics: new genes or new gene regulation? *Genes Brain Behav.* **1**, 197–203. (doi:10.1034/j.1601-183X.2002.10401.x)
- Roseler, P. F., Roseler, I. & Strambi, A. 1985 Role of ovaries and ecdysteroids in dominance hierarchy establishment among foundresses of the primitively social wasp *Polistes gallicus*. *Behav. Ecol. Sociobiol.* **18**, 9–13.
- Sledge, M. F., Trinca, I., Massolo, A., Boscaro, R. & Turillazzi, S. 2004 Variation in cuticular hydrocarbon signatures, hormonal correlates and establishment of reproductive dominance in a Polistine wasp. *J. Insect Physiol.* **50**, 73–83. (doi:10.1016/j.jinsphys.2003.10.001)
- 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. *Nat. Rev. Genet.* **9**, 735–748. (doi:10.1038/nrg2429)
- Sturn, A., Quackenbush, J. & Trajanoski, Z. 2002 GENESIS: cluster analysis of microarray data. *Bioinformatics* **18**, 207–208. (doi:10.1093/bioinformatics/18.1.207)
- Sumner, S. 2006 Determining the molecular basis of sociality in insects: progress, prospects and potential in sociogenomics. *Ann. Zool. Fenn.* **43**, 423–442.

- Sumner, S., Pereboom, J. J. M. & Jordan, W. C. 2006 Differential gene expression and phenotypic plasticity in behavioural castes of the primitively eusocial wasp, *Polistes canadensis*. *Proc. R. Soc. B* **273**, 19–26. (doi:10.1098/rspb.2005.3291)
- Tobback, J., Heylen, K., Gobin, B., Wenseleers, T., Billen, J., Arckens, L. & Huybrechts, R. 2008 Cloning and expression of PKG, a candidate foraging regulating gene in *Vespula vulgaris*. *Anim. Biol.* **58**, 341–351. (doi:10.1163/157075608X383665)
- Toth, A. L. & Robinson, G. E. 2007 Evo-devo and the evolution of social behavior. *Trends Genet.* **23**, 334–341. (doi:10.1016/j.tig.2007.05.001)
- Toth, A. L. *et al.* 2007 Wasp gene expression supports an evolutionary link between maternal behaviour and eusociality. *Science* **318**, 441–444. (doi:10.1126/science.1146647)
- Toth, A. L., Bilof, K. B. J., Henshaw, M. T., Hunt, J. H. & Robinson, G. E. 2009 Lipid stores, ovary development, and brain gene expression in *Polistes metricus* females. *Insectes Sociaux* **56**, 77–84. (doi:10.1007/s00040-008-1041-2)
- Tweedie, S. *et al.* 2009 FLYBASE: enhancing *Drosophila* gene ontology annotations. *Nucleic Acids Res.* **37**, D555–D559. (doi:10.1093/nar/gkn788)
- Weil, T., Korb, J. & Rehli, M. 2009 Comparison of queen-specific gene expression in related lower termite species. *Mol. Biol. Evol.* **26**, 1841–1850. (doi:10.1093/molbev/msp095)
- West-Eberhard, M. J. 1996 Wasp societies as microcosms for the study of development and evolution. In *Natural history and evolution of paper wasps* (eds S. Turillazzi & M. J. West-Eberhard), pp. 290–317. New York, NY: Oxford University Press.
- Whitfield, C. W., Band, M. R., Bonaldo, M. F., Kumar, C. G., Liu, L., Pardini, J. R., Robertson, H. M., Soares, M. B. & Robinson, G. E. 2002 Annotated expressed sequence tags and cDNA microarrays for studies of brain and behaviour in the honey bee. *Genome Res.* **12**, 555–566. (doi:10.1101/gr.5302)
- Whitfield, C. W., Cziko, A. M. & Robinson, G. E. 2003 Gene expression profiles in the brain predict behaviour in individual honey bees. *Science* **302**, 296–299. (doi:10.1126/science.1086807)
- Whitfield, C. W., Ben-Shahar, Y., Brillet, C., Leoncini, I., Crauser, D., Leconte, Y., Rodriguez-Zas, S. & Robinson, G. E. 2006 Genomic dissection of behavioural maturation in the honey bee. *Proc. Natl Acad. Sci. USA* **103**, 16 068–16 075. (doi:10.1073/pnas.0606909103)
- Winston, M. L. 1987 *The biology of the honey bee*. Cambridge, MA: Harvard University Press.
- Yang, Y. H., Dudoit, S., Luu, P., Lin, D. M., Peng, V., Ngai, J. & Speed, T. P. 2002 Normalization for cDNA microarray data: a robust composite method addressing single and multiple systemic variation. *Nucleic Acids Res.* **30**, e15. (doi:10.1093/nar/30.4.e15)