JOURNAL OF Evolutionary Biology

doi: 10.1111/jeb.12691

Caste-biases in gene expression are specific to developmental stage in the ant *Formica exsecta*

C. MORANDIN*†, K. DHAYGUDE*, J. PAVIALA*, K. TRONTTI*, C. WHEAT‡ & H. HELANTERÄ*†

*Centre of Excellence in Biological Interactions, Department of Biosciences, Helsinki University, Helsinki, Finland
†Tvärminne Zoological Station, University of Helsinki, Hanko, Finland
‡Department of Zoology, Population Genetics, Stockholm University, Stockholm, Sweden

Keywords:

caste;
gene expression;
genomics;
life stages;
RNA sequencing;
social insects.

Abstract

Understanding how a single genome creates and maintains distinct phenotypes is a central goal in evolutionary biology. Social insects are a striking example of co-opted genetic backgrounds giving rise to dramatically different phenotypes, such as queen and worker castes. A conserved set of molecular pathways, previously envisioned as a set of 'toolkit' genes, has been hypothesized to underlie queen and worker phenotypes in independently evolved social insect lineages. Here, we investigated the toolkit from a developmental point of view, using RNA-Seq to compare caste-biased gene expression patterns across three life stages (pupae, emerging adult and old adult) and two female castes (queens and workers) in the ant *Formica exsecta*. We found that the number of genes with caste-biased expression increases dramatically from pupal to old adult stages. This result suggests that phenotypic differences between queens and workers at the pupal stage may derive from a relatively low number of caste-biased genes, compared to higher number of genes required to maintain caste differences at the adult stage. Gene expression patterns were more similar among castes within developmental stages than within castes despite the extensive phenotypic differences between queens and workers. Caste-biased expression was highly variable among life stages at the level of single genes, but more consistent when gene functions (gene ontology terms) were investigated. Finally, we found that a large part of putative toolkit genes were caste-biased at least in some life stages in *F. exsecta*, and the caste-biases, but not their direction, were more often shared between *F. exsecta* and other ant species than between *F. exsecta* and bees. Our results indicate that gene expression should be examined across several developmental stages to fully reveal the genetic basis of polyphenisms.

Introduction

The ability of an organism to change its phenotype in response to external stimuli has fascinated evolutionary biologists for well over a century (Darwin, 1859; Stearns, 1989; West-Eberhard, 1989). In particular, the mechanisms of phenotypic plasticity, which allow genetically similar individuals to dramatically vary in phenotype (e.g. morphology, physiology, behaviour, life histories), are poorly understood. Here, we focused upon an extreme example of phenotypic plasticity manifest in the reproductive queen and nonreproductive worker castes in social Hymenoptera (social bees, wasps and ants), where phenotypic plasticity is the foundation of their social organization and evolutionary success (Wilson, 1971).

Eusociality has evolved independently in several hymenopteran lineages (Bourke, 2011). Motivated by the evolutionary developmental biology idea that these

Correspondence: Claire Morandin, Centre of Excellence in Biological Interactions, Department of Biosciences, University of Helsinki, P.O. Box 65, FI-00014 Helsinki, Finland. Tel.: +358 2 941 57703; fax: +358 9 191 57694; e-mail: claire.morandin@helsinki.fi

instances of convergent evolution could have a shared genetic basis (Ferrier & Holland, 2001; Carroll *et al.*, 2005; Carroll, 2008), researchers have hypothesized that the independent instances of insect sociality could have a shared genetic basis, a so-called genetic toolkit (Toth & Robinson, 2007; Toth *et al.*, 2010; Fischman *et al.*, 2011; Woodard *et al.*, 2011; Wilkins, 2014). In line with this hypothesis, many of the genes shown to be differentially expressed between castes in the honeybee were also shown to often have caste-biased expression patterns in other lineages of bees and wasps (Toth & Robinson, 2007). These observations suggest that at least some caste-biased genes are shared across lineages in a convergent, toolkit-like manner (Toth *et al.*, 2010; Woodard *et al.*, 2011; Ferreira *et al.*, 2013). These observations have recently been generalized to functional groups of genes, as Berens *et al.* (2014) showed that even if only a few genes are similarly caste-biased across the three main social hymenopteran lineages (bees, wasps and ants), there is more similarity in expression patterns at the level of pathways and biological functions. Thus, the idea of a genetic toolkit has begun to evolve, moving beyond an emphasis on the expression patterns of single genes to that of their biological functions.

Female caste differences in social insects first arise during the larval stages (Noirot, 1991; Miura, 2001, 2004), in response to various extrinsic and social factors (Evans & Wheeler, 1999; Pereboom *et al.*, 2005; Sumner *et al.*, 2006; Barchuk *et al.*, 2007; Hunt *et al.*, 2010). Genes in the toolkit have been identified during development, affecting caste determination [e.g. insulin/insulin-like and target of rapamycin (TOR) pathways (Hunt *et al.*, 2010; Fischman *et al.*, 2011; Berens *et al.*, 2014; Schrader *et al.*, 2015)], as well as later in life, affecting reproduction and behaviour (Toth & Robinson, 2007; Toth *et al.*, 2010). Only a few studies have compared several life stages, finding that caste-biases at the level of individual genes are unstable across life stages, with more genes being caste-biased at the adult stage (Hoffman & Goodisman, 2007; Harrison *et al.*, 2015a). What emerges from these findings is that to understand the genetic toolkit, individuals must be studied across multiple life stages.

Here, to study the genetic toolkit from a life stage perspective, we analysed levels of gene expression during three life stages (pupae, emerging adult and old adult) in queens and workers of the ant *Formica exsecta* with three aims. First, we wanted to quantify the magnitude of gene expression differences among the castes and among life stages within castes. Second, we wanted to test whether the caste-biases in expression of genes are consistent across life stages, and whether the stability is seen at the level of individual genes or at the more general level of pathways and functions. Third, we wanted to further explore the genetic toolkit by comparing the occurrence and directionality of caste-biased gene expression in different life stages of *F. exsecta* to expression patterns observed in other eusocial hymenopteran lineages.

Materials and methods

Model organism *Formica exsecta*

Formica exsecta is a common mound-building Palaearctic ant with nests usually containing 1000–10 000 workers (Helanterä & Sundström, 2007). There is a clear morphological separation between queens and workers. Queens may live up to 20 years, whereas adult workers will overwinter once and usually have lifespans slightly over 1 year (Pamilo, 1991). Queens lay eggs during spring and early summer, and new queens and males will emerge from cocoons in June/July, and new workers later in July. Their diet consists mainly of honeydew and insect prey (Czechowski *et al.*, 2002), which can be replaced with a egg–agar–honey diet under laboratory conditions (Bhatkar & Whitcomb, 1970).

Sample collection

All samples of *F. exsecta* were collected from colonies around the Tvärminne Zoological Station in the Hanko Peninsula, southwestern Finland (GenBank Biosample SAMN02046301–SAMN02046306). Old adult workers and queens were collected in April 2011 at the colony surface, when overwintered queens and workers gather on the colony mound surface to warm up, and this is the only time point when the egg-laying queens can be sampled from the wild; after the surrounding ground warms up, the queens move into the subterranean parts of the colony. Age of overwintered queens could not be controlled; however, they were all found in large mature colonies and were reproductive at the time of sampling with greatly enlarged abdomen due to egg production. Pupae were collected in the field in July 2011 and maintained in the laboratory for a couple of days until they matured to the right age. The pupae samples were a mix of early (white cuticle without eye)-, intermediate (white cuticle with dark eyes)- and late (brown cuticle with dark eyes)-stage pupae. Different developmental stages of pupae were included in the analysis to keep the within-pool variation in age similar in pupae and other samples. Emerging queens were collected in June and emerging workers in July, right after they emerged from the cocoons. After collection, these samples were frozen immediately at −80 °C for RNA extraction. Due to strongly seasonal reproduction in *F. exsecta*, it is impossible to obtain samples at the same time from the different age classes.

Library preparation

Total RNA for library preparation was extracted from whole bodies using a standard TRIzol protocol (TRIsure;

Table 1 Number of individuals pooled for library preparations for each replicate (FIMM and BGI) with the number of source colonies. In total, 12 libraries were prepared from these samples.

Caste	Stage	FIMM	BGI	Colonies
Queen	Old adult	6	4	10
	Emerging adult	10	8	9
	Pupae	24	18	21
Worker	Old adult	18	30	14
	Emerging adult	10	15	11
	Pupae	24	18	21

Bioline, Luckenwalde, Germany). RNA was pooled after extraction into libraries (Table 1). The worker pools had more individuals extracted than the queen pools due to the worker's smaller size and resulting mRNA yield differences. We expect the larger number of individuals in the worker pool to have a negligible effect, as each pool was composed of con-colonial workers (instead of adding number of colonies) that are highly related [i.e. often full sisters (Sundström *et al.*, 2005)] and from comparatively similar environments (i.e. same colony). Thus, mRNA yield was increased while minimizing any increase in genetic or environmental variation in worker samples compared to queen samples. The final data set consisted of six classes of samples: overwintered workers and queens (old adults), newly emerged workers and virgin queens (emerging adults), and queen and worker pupae. From each class of samples, two sets of paired-end libraries (12 libraries in total) (P-E 91–99) were prepared by BGI (Shenzhen, China) and FIMM (Helsinki, Finland) and sequenced on Illumina HighSeq 2000 platform using the provider's pipeline. For the libraries, the mRNA was selected using two rounds of poly-A selection (Poly-A Purist Kit; Ambion, Austin, TX, USA), fragmented and size-selected to approximately 200–700 bp. Each library was barcoded uniquely, pooled and sequenced in two lanes, mitigating any lane effects between the libraries (same procedure for both BGI and FIMM). For more details about library composition and preparations, see Johansson *et al.* (2013) and GenBank Biosample SAMN02046301–SAMN02046306.

Formica exsecta reference transcriptome assembly

Quality of raw reads was initially assessed with FastQC version 0.10.1 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/), followed by trimming of the low-quality bases from 3′-ends using FastX toolkit v. 0.0.13, and reducing read length to 86 bases at both ends (average ≥ 20 Q for all base positions Phred scale). Following quality filtering, we constructed a *de novo* transcriptome using Trinity software [Release 2012-05-18 (Grabherr *et al.*, 2013; Haas *et al.*, 2013)] from the 322 M trimmed read pairs. We used a range of k-mer settings ($k = 23$–31) to maximize transcript coverage (Surget-Groba & Montoya-Burgos, 2010). The resulting contigs were combined and scaffolded using Vmatch (Abouelhoda *et al.*, 2004). After scaffolding, the transcriptome contigs were cleaned from probable exogenous RNAs (Johansson *et al.*, 2013), and only contigs that showed significant BLAST hits against published ant genomes (Bonasio *et al.*, 2010; Nygaard *et al.*, 2011; Smith *et al.*, 2011a, b; Suen *et al.*, 2011; Wurm *et al.*, 2011) and other insect genome annotations were kept for further analysis [*Apis* (The Honeybee Genome Sequencing Consortium, 2006), *Nasonia* (Werren *et al.*, 2010), *Tribolium* (Richards *et al.*, 2008) and *Drosophila* (Adams *et al.*, 2000)] [BLASTx, *e*-value threshold 10^{-3}, minimum requirement of 70% amino acid identity and at least 100 bases (33 amino acids) alignment length]. The final transcriptome assembly had an average read coverage of > 300 ×. Altogether, 39 262 mRNA transcripts were aligned to the seven currently published ant genomes and 5010 found homology to other insects.

Gene expression and whole-gene count estimation

The trimmed paired-end reads were mapped to *F. exsecta de novo* reference transcriptome using splice-aware aligner Bowtie2 (Langmead & Salzberg, 2012) with default settings. We used the software Samtools (Li *et al.*, 2009) to estimate the total raw counts of reads aligned with the assembled transcriptome. To quantify whole-gene expression, we assigned each *F. exsecta de novo* contig to a database of predicted proteins from the phylogenetically closest sequenced ant genome, *Camponotus floridanus* (version 3.3 retrieved from Hymenoptera database), using BLASTx (*e*-value threshold $10^{-3)}$. This predicted protein set of *C. floridanus* was a nonredundant database, generated by a self-blast of these predicted proteins followed by taking the longest representative of a given cluster of proteins that shared ≥ 90% identity at amino acid level over at least 50 amino acids. The following criteria were used to assign *F. exsecta* contigs to this *C. floridanus* database: bit score ≥ 50, *e*-value ≤ 1×10^{-3}, and over a length of 33 amino acids. This relational table between the *F. exsecta* contigs and their putative *C. floridanus* orthologs was used to generate whole-gene expression values (using custom Python scripts) for each *C. floridanus* ortholog; it is the sum of the mapped RNA-Seq reads to the set of *F. exsecta* contigs sharing the same *C. floridanus* ortholog. Overall, 42 476 contigs were assembled with Trinity for the *F. exsecta* reference transcriptome, and 36 961 of these contigs were matched with a *C. floridanus* ortholog. These 36 961 contigs were found to be expressed at the emerging and pupae stage for both castes. At the old adult stage, 64 (worker) and 119 contigs (queen) were not expressed. We used this approach as it has been shown to produce accurate whole-gene expression

levels that are highly similar to mapping RNA-Seq reads against the predicted genes from a well-characterized genome (Hornett & Wheat, 2012). The final whole-gene expression data thus consisted of count data per *C. floridanus* gene, with minimal template biases (i.e. SNP or alternative splicing effects). Thus, our analysis only covers genes with identified orthologs in *C. floridanus.*

We took this approach for several reasons. First, *de novo* transcriptome assemblies enable functional genomic studies of nonmodel organisms, but also have the potential for mis-assemblies and thus misinterpreting the overall pattern of caste-biased expression (Li *et al.*). *De novo* assemblies contain a plethora of contigs with low abundances, which are difficult to distinguish from 'true' contigs (Li *et al.*, 2014). Second, many assemblies can contain SNP-specific contigs, resulting in mapping that reflects the relatedness of individuals to specific contigs. Third, both of these effects as well as alternative splicing and its overprediction result in an inflation of the number of contigs per unique gene in the organism, significantly complicating downstream analysis of redundant results. The whole-gene quantification approach we used provides a robust approach to these potential biases (Hornett & Wheat, 2012; Finseth & Harrison, 2014). Finally, as our main focus was on conserved toolkit genes, excluding *F. exsecta*-specific genes using the *C. floridanus* gene set is justified.

Gene expression analysis

Differential expression was analysed using the R Bioconductor package EdgeR (Robinson *et al.*, 2010). Reads generated by BGI and FIMM were used as replicates. Similarity between our replicates was assessed by the biological coefficient of variation implemented in EdgeR. The biological coefficient of variation measures gene expression variability across libraries and is defined as 'the coefficient of variation with which the (unknown) true abundance of the gene varies between replicate RNA samples' (McCarthy *et al.*, 2012). We first filtered out transcripts with very low read counts, by removing loci lower than 1 per kilobase exon per million fragments mapped in at least half of the sequenced libraries. This step removed ~15% of the total number of genes from the analysis. Trimmed mean of *M*-values normalization was applied prior to estimating differentially expressed genes (DEGs) to account for compositional difference between libraries, following EdgeR recommendations. Expression differences were considered significant after correction for multiple testing using a false discovery rate (FDR) < 0.05. The number of DEGs between the different castes and life stages was visualized by Venn diagrams using Venny (freely accessible at http://bioinfogp.cnb.csic.es/tools/venny/index.html). A two-tailed Fisher's exact test, as implemented in the exact2x2 R package (Fay, 2010), was used to compare the proportion of genes showing bias towards either caste relative to the overall number of genes. A multidimensional scaling (MDS) plot was drawn to measure the similarity between samples/libraries using EdgeR package.

Functional analysis of expressed genes

Blast2GO (http://www.blast2go.org) was used to infer functional (gene ontology or GO terms) annotation of the entire *C. floridanus* gene set using structural similarity (BLASTx with an *e*-value cut-off $\leq 10^{-3}$ to obtain the best homolog). Additional GO terms were obtained using InterProScan (Quevillon *et al.*, 2005) before being merged with the BLAST results. The GOstats package for R (Beissbarth & Speed, 2004) was used to conduct GO term enrichment analysis on the differentially expressed gene sets, using the set of all genes for which GO terms were available as the universe. Enriched GO terms (FDR < 0.01) were subsequently clustered using Revigo (Supek *et al.*, 2011) to avoid redundancy of GO terms.

The R package GOSemSim (Yu *et al.*, 2010) was used to compute semantic similarity among samples using the set of caste-/life stage-enriched GO terms and based on Biological Processes Ontology. GOSemSim assigns a value between 0 and 1 with higher values indicating greater similarity between groups of GO terms. Subsequently, semantic similarity data were used to cluster samples according to their similarity in terms of biological functionality on the GO term level, using heatmap clustering in GOSemSim.

Comparison with earlier studies

Formica exsecta gene expression patterns were compared to earlier studies to determine the extent to which patterns of caste-biased expression of putative toolkit genes are shared among different lineages of eusocial Hymenoptera. Toolkit genes from previous studies were selected based on the following criteria (Toth *et al.*, 2010; Bloch & Grozinger, 2011; Fischman *et al.*, 2011): (i) genes from previous studies exhibiting caste-biased gene expression across multiple social hymenopteran species or (ii) genes belonging to conserved molecular pathways or functions involved in caste determination [e.g. TOR (Patel *et al.*, 2007; Wheeler *et al.*, 2013)]. The full list includes ten genes belonging to the Yellow protein family (Tian *et al.*, 2004; Drapeau *et al.*, 2006; Gräff *et al.*, 2007; Feldmeyer *et al.*, 2014), one major royal jelly protein (Buttstedt *et al.*, 2013), four hexamerin genes (Martins *et al.*, 2008, 2010), four genes belonging to the Toll pathway (Colgan *et al.*, 2011; Koch *et al.*, 2013), nine genes belonging to the insulin/TOR pathway (Wheeler *et al.*, 2006, 2013; Corona *et al.*, 2007; Grozinger *et al.*, 2007; Feldmeyer *et al.*, 2014) and three juvenile hormone genes (Mackert *et al.*, 2010; Li *et al.*,

2013; Bomtorin *et al.*, 2014). Seven immune-related genes (Colgan *et al.*, 2011; Koch *et al.*, 2013) and three DNA methyltransferase enzymes (Liu *et al.*, 2012) were also investigated, because of their highly conserved function across lineages (Feng *et al.*, 2010). Only studies using life stages comparable to ours that clearly mentioned the caste-biased patterns of the genes of interest were included in the comparisons (Tian *et al.*, 2004; Cunha *et al.*, 2005; Bitondi *et al.*, 2006; Drapeau *et al.*, 2006; Wheeler *et al.*, 2006, 2013; Corona *et al.*, 2007; Gräff *et al.*, 2007; Grozinger *et al.*, 2007; Mackert *et al.*, 2008; Martins *et al.*, 2008; Colgan *et al.*, 2011; Liu *et al.*, 2012; Buttstedt *et al.*, 2013; Koch *et al.*, 2013; Li *et al.*, 2013; Bomtorin *et al.*, 2014; Feldmeyer *et al.*, 2014). The final species set is composed of three bee species (*Apis mellifera, Bombus terrestris* and *Apis cerana*) and four ant species (*Solenopsis invicta, Lasius niger, Temnothorax longispinosus* and *Atta vollenweideri*).

To look for the exact gene match in our data set, protein and CDS sequences of the target genes were obtained from the *A. mellifera* genome (NCBI database), which is considered to be the most complete and best annotated social insect genome available. BLASTp and tBLASTn searches were performed against the official *C. floridanus* gene set (version 3.3, obtained from Ants Genome Portal) to identify the putative orthologs in our database, using the following criteria to assign orthology: identity $\geq 70\%$, *e*-value cut-off $\leq 1 \times 10^{-10}$ and query coverage $\geq 70\%$.

To investigate how many of the potential toolkit genes are shared across all three social lineages, and how many are specific to ants or bees, the presence and direction of caste/life stage bias (e.g. up-regulated in queens or workers) were compared between *F. exsecta* and the two above-mentioned groups. Furthermore, we compared whether the patterns of sharing are dependent on the life stage investigated.

Results

Overview of gene expression patterns

Visual inspection of the MDS plot (Fig. 1) shows three main results. First, the two library replicates were very similar to each other. This is consistent with a low biological coefficient of variation between replicates, which varied between 0.30 (emerging queens) to 0.38 (queen pupae). Although some technical variation may occur among our replicates due to noise from experimental effects, our BCV values were similar to previous studies from laboratory-based RNA-Seq on *Arabidopsis thaliana* and humans (Robinson *et al.*, 2010), indicating that variation due to experimental noise and true biological variation among replicate libraries are relatively small in our analysis. Second, caste differences were rather small in pupae and emerging adults, relative to old adults. Third, caste differences are largest in the old adults.

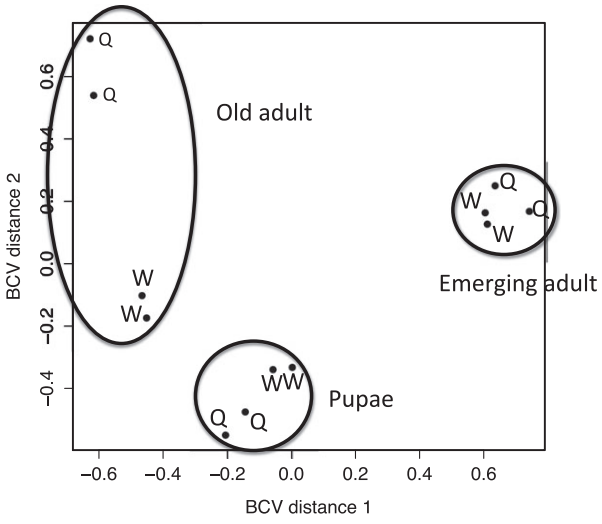

Fig. 1 Multidimensional scaling plot showing transcriptional similarity between the samples. Different castes in the same life stage are more similar to each other than different life stages of the same caste, whereas caste differences are more pronounced in the old individuals.

Caste-biased expression patterns

Numbers of caste-biased genes

The number of caste-biased genes increased significantly with age; the lowest number of differentially expressed genes was found at the pupal stage and the highest at the old adult stage (Fig. 2, all pairwise comparisons $P < 0.05$, Fisher's exact test). For both the emerging and old adult stages, the workers up-regulated twice as many genes (366 and 1384, respectively) as the queens (182 and 654, $P < 0.001$, Fisher's exact test). At the pupal stage, more genes were up-regulated by the queens (92 genes) than workers (11 genes) ($P < 0.001$, Fisher's exact test) (Fig. 2). Caste-biases were in general not stable among the different life stages. Only one gene (annotated as a 'hypothetical protein') was found consistently caste-biased in the same direction (queen biased) in all life stages (Fig. 3). However, many genes were caste-biased in opposite directions in different life stages: 146 genes between old and emerging adults, six genes between pupae and emerging adults, and 15 genes between old and pupal stages. In total, 15 genes were found to be consistently caste-biased in all three life stages, but in varying directions (Fig. 4).

Caste-biased gene enrichment analysis

Caste-biased expression patterns were more consistent at the level of functional classes than single genes. We used GOSemSim to estimate pairwise similarities of GO terms in the overlap between each caste- and life stage-up-regulated gene set. A value of zero would indicate complete lack of overlap, and a value close to

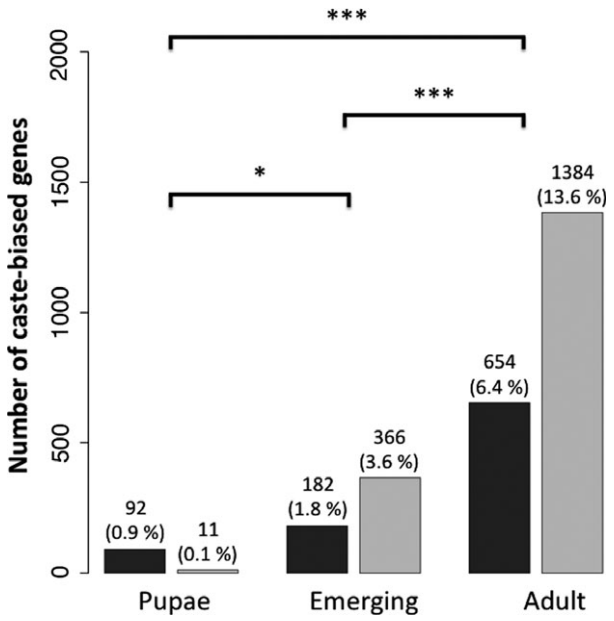

Fig. 2 The number of caste-biased genes increased significantly with age across life stages. The *x*-axis indicates the three life stages (pupae, emerging adults and old adults). The *y*-axis indicates the number of differentially expressed genes found up-regulated by queens (dark grey) and workers (light grey). Numbers of caste-biased genes are given together with their proportion compared to the total number of genes in parentheses, $*P < 0.05$, $**P < 0.01$ and $***P < 0.001$.

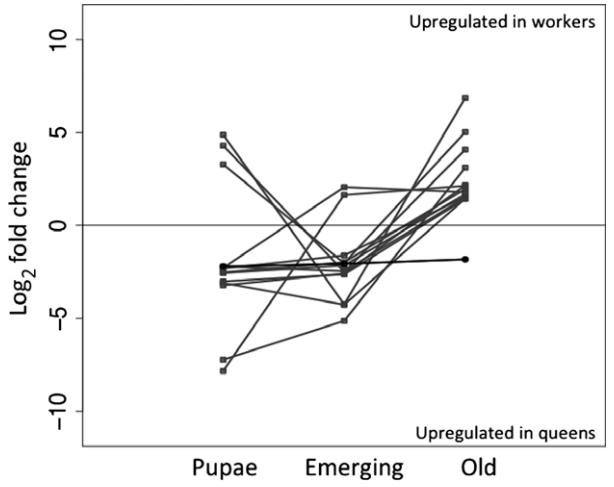

Fig. 3 Caste-bias direction of the 15 genes found to be always differentially expressed among caste varies across developmental stages. Only one gene was found to be consistently up-regulated in the same direction, in the queens, a hypothetical protein.

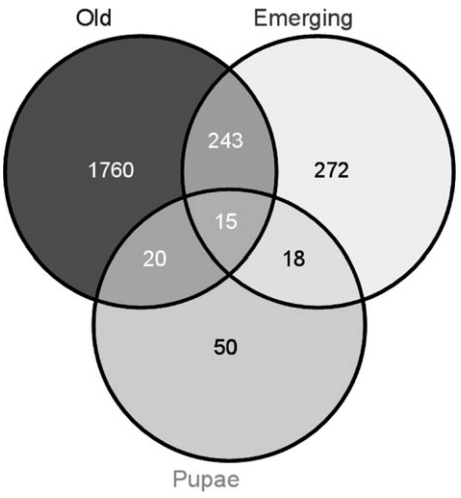

Fig. 4 Venn diagram summarizing the number of caste-biased genes shared among life stages using EdgeR with false discovery rate < 0.05. Only 15 genes were found to be consistently caste-biased over the three developmental stages.

one indicates complete similarity between groups of GO terms. GOSemSim estimates varied between 0.2 and 0.7 across all pairwise comparisons. Based on semantic similarity of enriched GO terms, we show that some similarity of GO terms categories can be found across all pairwise comparisons (Fig. 5a). This suggests that even if the identity of the up-regulated genes varies at the level of individual genes, at least some of the functional classes remain caste-associated through development (Fig. 5a). However, at the level of enriched individual GO terms, no similarities among life stages within caste were observed. No enriched GO terms were shared between the old queens and emerging queens, and a very low number between old/emerging and pupae queen (18 and 6 terms, respectively). Similarly, no enriched GO terms were found similar between emerging workers and worker pupae, and a low number between old workers and emerging workers/worker pupae (93 and 4, respectively).

Among the enriched GO terms per sample class, we found that old queens up-regulated a large amount of genes (35%) linked to the Regulation of Gene Expression and Epigenetic Mechanisms, whereas old workers up-regulated genes involved in Muscle System Processes and Metal Ion Transport. The gene set up-regulated in emerging queens was enriched for Muscle Development and ATP Synthesis Decoupled Proton Transport, whereas the gene set up-regulated in emerging workers was enriched for Single Organism Signaling, Biological Regulation and Central Nervous System Formation. At the pupal stage, Detection of Stimulus was the only GO term enriched for the workers, whereas queen pupae up-regulated genes linked to a variety of developmental processes such as Somatic Muscle Development and Regulation of Growth Rate. Noteworthy, for every caste at each developmental stage, except the worker pupae, genes linked to metabolism were enriched in the

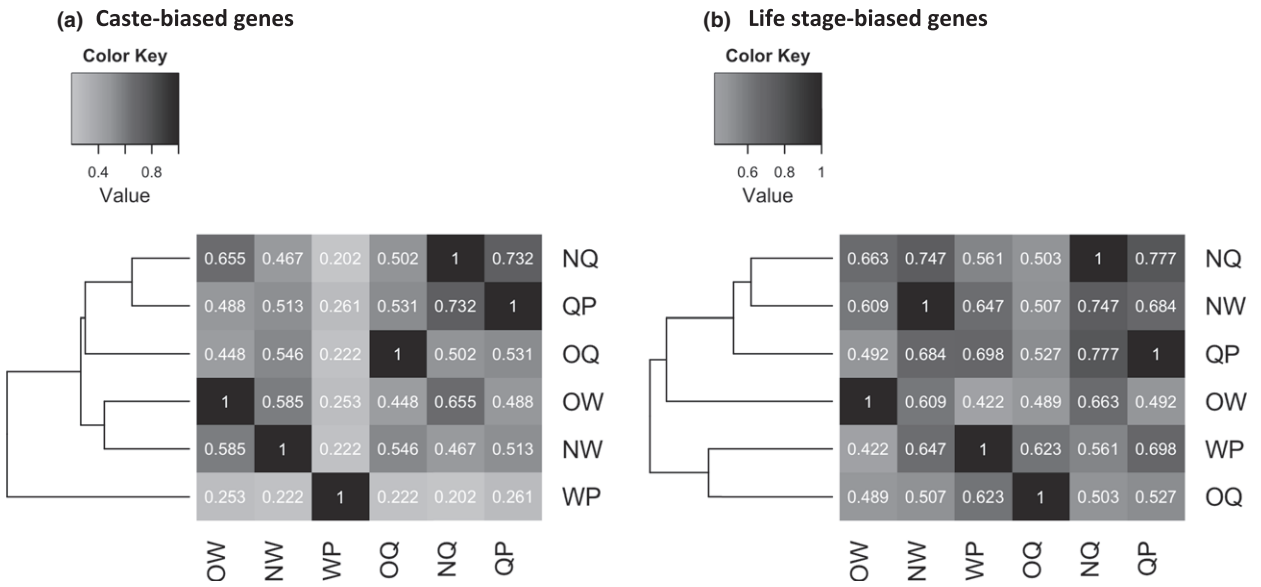

Fig. 5 The heatmap represents semantic similarity values among gene ontology Biological Process terms computed using GOSemSim. Rows and columns represent samples used in this study (OQ – old queen; OW – old worker; NQ – emerging queen; NW – emerging worker; QP – queen pupae; and WP – worker pupae). (a) Caste-biased genes. (b) Life stage-biased genes.

up-regulated gene set, confirming the prediction that metabolic genes are likely to be parts of caste-biased toolkit (Fig. S1). A simplified figure of the GO terms and the complete list of enriched terms can be found in the Fig. S1 and Table S1.

Life stage expression patterns

Number of life stage-biased genes

More DEGs could be found among life stages within each caste than between castes. When doing pairwise comparisons across life stages within each caste, we found that the highest number of DEGs (3669 genes) was between the old adult queens and the emerging queens (Fig. S2). Of the genes differentially expressed between emerging and old adult queens, more were up-regulated in the emerging queens (2025 genes vs. 1644 genes). The pupae up-regulated more genes compared to the old adult queens (1692 genes vs. 1114 genes) from the genes differentially expressed between them. A similar number of genes were up-regulated by the emerging queens and the pupae (617 and 669 genes). A significantly lower number of genes were differentially expressed among life stages within the worker caste than the queen caste (7761 total DEGs within queens, 3398 total DEGs within workers, Fisher's exact test $P < 0.001$). More DEGs (1654 genes) were found between old workers and worker pupae (Fig. S3). Of the genes differentially expressed between emerging and old adult workers, emerging workers up-regulated more genes compared to the old adult workers (588 genes vs. 134). When looking at genes differentially expressed between pupae and old adult worker, pupae up-regulated more genes compared to the old adults (1063 vs. 591 genes). A very similar number of genes could be found up-regulated by both emerging adults and pupae when looking at genes differentially expressed between them (458 and 563 genes, respectively).

Life stage-biased gene enrichment analysis

Clustering based on semantic similarity of enriched GO terms revealed that life stage-biased genes are linked to biologically different functions between the two castes (Fig. 5b). Analyses of individual GO terms showed a similar pattern, as few enriched GO terms were shared between the old queens and old workers when comparing based solely on their GO term ID (1.4% of the old adult-enriched GO terms, 14.8% of the emerging adult-enriched GO terms and 16% of the pupae-enriched GO terms). Enriched GO term list for the old queens included terms related to reproduction and methylation. At the emerging queen stage, the list includes oxidation processes and muscle physiological process, and at the pupal stage, the list comprises terms related to oxidative stress and developmental processes. For the old adult worker, rRNA processing and oxidation–reduction processes, and for the emerging workers, muscle contraction and energy reserve metabolism were found overrepresented. At the pupal stage, the list contains terms related to response to stimulus and amino sugar metabolism. A simplified figure of the GO

terms and the complete list of enriched GO terms can be found in the Table S2 and Fig. S4.

Overlap between caste- and life stage-biased genes

Most of the genes classified as caste-biased in the old queens (94%, 613 life stage-biased/654 caste-biased) were found to be also life stage-biased genes. In comparison, only 22% (304/1384) of the old worker-biased genes were found to be also life stage-biased genes (Fisher's exact test $P < 0.001$). For the emerging stage, 95% (174/182) of the queen caste-biased genes and 52% (191/366) of the worker caste-biased genes were found to be life stage-biased genes (Fisher's exact test $P < 0.001$). However, at the pupal stage, 64% (7/11) of worker caste-biased genes were determined to be also life stage-biased genes against only 68% (63/92) for the queen (Fisher's exact test $P = 0.74$).

Toolkit gene expression patterns

All of the toolkit genes we identified from the literature were found expressed at each life stage and caste in the *F. exsecta* transcriptome. Of these, 61% were caste-biased in *F. exsecta*, but in a life stage-specific manner, 37% were caste-biased at old adult and 24% at emerging adult stages, but 0% at the pupal stage. These proportions, apart from the pupal stage, are higher than average proportions of caste-biased genes at these stages (Fisher's exact test, $P < 0.001$), suggesting that the putative toolkit genes are indeed more likely than average genes to be consistently caste-biased when even more species are included in the comparisons. Only six genes were caste-biased in more than one developmental stage. A higher proportion of genes sharing a caste-biased pattern of expression was found between *F. exsecta* and other ants compared to *F. exsecta* and bees ($P = 0.03$, Fisher's test, 73% vs. 34%). However, the caste-biases were in the same direction in only 56% of the genes between *F. exsecta* and other ants, and 35% between *F. exsecta* and bees ($P = 0.08$, Fisher's test). In *F. exsecta*, 38% of the toolkit genes were found to be life stage-biased (49% within the queen caste and 27.5% within the worker caste). The full list of toolkit genes and caste and life stage biases can be seen in the Table S3.

Discussion

Our results reveal that caste-biases in gene expression are life stage dependent. The proportion of caste-biased genes varies among life stages, being the highest among old adults and lowest among pupae. Despite the extensive phenotypic differences, female castes at a given life stage are on average more similar to each other in gene expression profiles than to other life stages within caste. At the level of individual genes, we found widespread life stage specificity in caste-biased expression patterns. Caste-biases were, however, more consistent across life stages when looking at biological functions of gene sets rather than individual genes. Finally, in comparison with other published studies, caste-biased putative toolkit genes were more often shared between *F. exsecta* and other ant species than between *F. exsecta* and bees. However, no obvious links to phylogenetic similarity of species could be found regarding the directionality of the caste-bias.

Pathways and functions, rather than individual genes, are associated with caste phenotypes

We observed that caste-biased gene expression is highly dynamic and variable among life stages at the individual gene level. Life stage specificity has also been observed in caste-biased gene expression in the fire ants (Ometto *et al.*, 2011). In contrast, a more stable pattern of caste-biased expression could be found at the functional annotation level. Indeed, analysis of semantic similarity showed functional similarities among life stages within castes, with castes grouping more strongly than life stages (Fig. 5a); thus, common molecular functions are associated with each caste, rather than specific caste-biased genes. These findings, along with the recent results from Berens *et al.* (2014) and Mikheyev & Linksvayer (2015), suggest that shared functional groups or sets of coexpressed genes, rather than individual genes, are the key to the evolution of caste phenotypes. However, we could not find substantial overlap when comparing castes directly at the level of enriched GO term ID. This contrasting result is likely due to the several hierarchical levels present in the GO term annotations, and thus, more thorough quantitative semantic comparisons are needed to measure similarity among samples (Yu *et al.*, 2010).

The proportion of caste-biased genes increased with age

At the pupal stage, only 1% of the genes were differentially expressed between queens and workers. It is thus possible that smaller differences in expression patterns are required at the pupal stage to create morphological differences, compared to the adult stage where differences in expression patterns may control caste-specific behaviour and physiological processes (Harrison *et al.*, 2015a). It might be the case that both castes use similar mechanisms to develop and grow through metamorphosis. Increasing expression divergence among castes with age, also observed in *Vespula* wasps, bumble bees and fire ants (Hoffman & Goodisman, 2007; Hunt *et al.*, 2013; Harrison *et al.*, 2015a), correlates with the increasing task divergence with age. Queens in mature colonies stay in the nest and lay eggs, whereas workers carry out all other tasks and are in direct contact with

fluctuating environmental conditions. Old queens are clearly separated from the other adult samples. This correlates with the drastic changes in queen physiology and lifestyle after maturation, compared to workers and emerging adult queens that are both mobile and in contact with the external environment, and reproductively inactive. It is also important to note that we studied caste differences only in life stages after the caste fate of the individuals is irreversibly set. Thus, our results about the toolkit genes do not say anything about the causal role of toolkit genes in caste determination *per se*. It remains to be studied whether putative toolkit genes are actually involved in caste determination in multiple taxa, or whether they are sets of genes likely to be repeatedly co-opted into caste-specific use in later life stages.

Evolution of gene plasticity and life stage specificity

The genes that were caste-biased at one or more life stages were also more likely to vary among life stages. This pattern holds especially true for the queens, with for instance 94% of the old queen caste-biased genes found to be also life stage-biased genes. This supports the notion that genes with a regulatory architecture that allows plastic expression, and that are potentially expressed plastically in many contexts (Grishkevich & Yanai, 2013), are also more likely to be co-opted into caste-specific functions (Helanterä & Uller, 2014). For instance, vitellogenin is one of the key genes underlying caste-specific phenotypes in honeybee, but is also expressed in a plastic manner in many contexts such as oxidative stress resistance in queens (Havukainen *et al.*, 2013) and the coordination of workers' nurse–forager transition as the workers age (Guidugli *et al.*, 2005). A large proportion of toolkit genes were expressed in a life stage-specific manner. This suggests that analysing multiple life stages is crucial also for understanding the extent of the shared genetic basis for caste differences. The proportions of genes that fit the toolkit pattern would have been considerably smaller if only one life stage would have been investigated.

Life stage specificity of caste-biased expression patterns has implications on understanding the evolutionary origins and consequences of plastic gene expression. The expression pattern of a gene is predicted to be correlated with its evolutionary rate so that selection should be weaker in the worker-biased genes because workers are sterile and have only indirect fitness (Linksvayer & Wade, 2009; Hall & Goodisman, 2012). Correlations among evolutionary rates and expression patterns have been observed in social insects (Hunt *et al.*, 2011; Harpur *et al.*, 2014), and plastic expression patterns in other systems are often linked to increased evolutionary rates as well (Mank *et al.*, 2007; Mank & Ellegren, 2009; Leichty *et al.*, 2012). However, the instability of expression patterns across life stages suggests that these correlations are more likely to reflect general evolutionary constraints (Hunt *et al.*, 2013), rather than caste-biased expression pattern *per se* (Helanterä & Uller, 2014). As we show that it is rather functional classes than individual genes that correspond to castes across life stages, it remains to be seen whether consistent patterns of evolutionary rates can be seen in such sets of genes.

Studies of gene expression in other polyphenic systems, especially of sex-biased genes (Harrison *et al.*, 2015b), have shown that tissue specificity in expression is a common feature of genes with polymorphic expression. Our decision to use whole-body samples may thus mask some important information for interpreting causes and consequence of an expression pattern. Thus, our study joins many others in suffering from the difficulties of whole-body analysis (e.g. nearly all *Drosophila* studies). Requirements for RNA-Seq experiments also continuously improve, with less quantity of RNA required nowadays, which will further facilitate tissue-specific experiments. Tissue specificity of caste-biased expression in general is certainly a promising direction for future research.

Biases and directionality of toolkit genes

Although there were some expression patterns shared between ants and bees, suggesting the existence of at least some 'true' toolkit genes, not all expression patterns were shared even among ant species. This suggests that caste-biased gene expression is a mixture of toolkit genes and more narrowly taxon-specific patterns. This is not surprising given the fact that bees and ants have diverged ~170 Mya (Ronquist *et al.*, 2012) and expression biases have been shown to vary even among closely related species (Ometto *et al.*, 2011), even in genes of crucial importance, such as vitellogenin (Morandin *et al.* 2014). As expected, the proportion of caste-biased genes shared among *F. exsecta* and other ants, originating from the same eusocial ancestor, was higher than that among *F. exsecta* and bees that represent independent evolutionary origins for eusociality. However, the fact that the higher similarity was not found when directionality of caste-biases was looked at suggests that the precise regulation of toolkit genes may differ even among closely related species. Similar lack of expression directionality pattern exists in highly phylogenetically conserved genes such as the foraging gene (Ingram *et al.*, 2005).

We investigated the potential toolkit genes, chosen based on earlier observations, as independent genes. However, as highlighted by recent studies, the toolkits can also be described more broadly at the level of gene ontologies, patterns of gene coexpression, or other methods that do not only look at patterns in single genes (cf. Berens *et al.*, 2014; Mikheyev & Linksvayer,

2015). Extending transcriptome-wide studies to phylogenetically extensive comparisons that include several life stages will offer a more complete understanding of what gene networks constitute the conserved toolkit shared across lineages, and which are the more recent taxon-specific elaborations of sociality.

Taxonomically restricted genes were omitted from this study

In contrast to recent studies (Ferreira *et al.*, 2013; Feldmeyer *et al.*, 2014), most of our contigs presented either GO terms or BLAST functional trait annotations in both queens and workers. This is because we relied only on filtered genes that have annotation evidence of being true insect genes (e.g. significant BLAST hits with genes from previously sequenced insect published genomes), rather than potential sequencing artefact contigs (Li *et al.*, 2014). Although our stringent methods increase our confidence in the results of genewise expression patterns, this decision also means that any putative novel or taxonomically restricted genes because of the split between *Camponotus* and *Formica* are missing from our analyses. Given gene birth rates in insects, there could be up to a few thousands of these genes (Wissler *et al.*, 2013). Whereas novel genes are by definition not parts of the toolkit and thus not directly relevant to our aims, they may nevertheless play important roles in species-specific regulation of social behaviour (Mikheyev & Linksvayer, 2015) and evolutionary novelties in social organisms (Jasper *et al.*, 2014). In the future, it would be interesting to see whether taxonomically restricted genes associated with evolutionary novelty have stable caste-biased expression patterns among closely related species, and among life stages.

Potential role of DNA methylation in the maintenance of female differences

One interesting finding is the significant differential expression of one protein involved in DNA methylation, DNA methyltransferase enzyme DNMT3, up-regulated by the workers at the old adult stage. DNMT3 is responsible for establishing new methylation patterns within the genome (Hata *et al.*, 2002; Kato *et al.*, 2007; Glastad *et al.*, 2011). In honeybee, silencing the expression of DNMT3 led to the emergence of queen larvae instead of worker larvae, suggesting that DNA methylation plays an important role in the development of female castes (Kucharski *et al.*, 2008). This discovery, along with the fact that most of the caste-biased genes are found at the old adult stage, underlines the potential role of methylation in the regulation and/or maintenance of caste gene expression in social insects even after the actual caste determination process.

Conclusions

Whether caste-biased gene expression patterns are seen as highly variable among life stages within a species, or among phylogenetically closely related species, depends on the level of investigation. Expression patterns among life stages are unstable at the level of single genes, but more coherent when looked at the level of functional similarities based on GO terms. In phylogenetic comparisons, which genes end up as being defined as toolkit genes, that is genes that have a similar expression pattern across lineages, depends on how many different life stages are investigated and whether the directionality of expression biases is considered. Comprehensive understanding of causes and consequences of caste-biased gene expression needs analysing gene expression in multiple contexts, paying attention to the phylogenetic variation in gene expression, considering both conserved and novel genes, and investigating the genes in question in the light of their functional role. Similar lines of investigation should provide information in understanding of any polyphenic developmental system, not just castes.

Acknowledgments

This work was supported by the Academy of Finland (grant numbers 140990, 135970 and 273029 to H.H., 134561 to K.T., 131155 to C.W. and 252411 to the Centre of Excellence in Biological Interactions), Helsinki University, and the Knut and Alice Wallenberg Foundation 2012.0058 to C.W. The authors would like to thank Jonna Kulmuni, Luke Holman and anonymous referees for helpful comments on earlier versions of the manuscript.

Availability of supporting data

The raw reads of the transcriptome are available on GenBank (http://www.ncbi.nlm.nih.gov/genbank/) under bioproject ID PRJNA213662, sample accession numbers: SAMN02297446, SAMN02297447, SAMN02297448, SAMN02297449, SAMN02297450, SAMN0229745, SAMN02297452, SAMN03799239, SAMN03799240, SAMN03799241, SAMN03799242, SAMN03799243, SAMN03799244 and SAMN03799245.

References

Abouelhoda, M.I., Kurtz, S. & Ohlebusch, E. 2004. Replacing suffix trees with enhanced suffix arrays. *J. Discrete Algorithms* **2**: 53–86.

Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G. *et al.* 2000. The genome sequence of *Drosophila melanogaster*. *Science* **287**: 2185–2195.

Barchuk, A.R., Cristino, A.S., Kucharski, R., Costa, L.F., Simões, Z.L.P. & Maleszka, R. 2007. Molecular determinants

of caste differentiation in the highly eusocial honeybee *Apis mellifera*. *BMC Dev. Biol.* **7**: 70.

Beissbarth, T. & Speed, T.P. 2004. GOstat: find statistically overrepresented Gene Ontologies within a group of genes. *Bioinformatics* **20**: 1464–1465.

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

Bhatkar, A. & Whitcomb, W.H. 1970. Artificial diet for rearing various species of ants. *Florida Entomol.* **53**: 229–232.

Bitondi, M.M.G., Nascimento, A.M., Cunha, A.D., Guidugli, K.R., Nunes, F.M.F. & Simões, Z.L.P. 2006. Characterization and expression of the *Hex*110 gene encoding a glutamine-rich hexamerin in the honey bee, *Apis mellifera*. *Arch. Insect Biochem. Physiol.* **72**: 57–72.

Bloch, G. & Grozinger, C.M. 2011. Social molecular pathways and the evolution of bee societies. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **366**: 2155–2170.

Bomtorin, A.D., Mackert, A., Rosa, G.C.C., Moda, L.M., Martins, J.R., Bitondi, M.M.G. *et al.* 2014. Juvenile hormone biosynthesis gene expression in the corpora allata of honey bee (*Apis mellifera*) female castes. *PLoS One* **9**: e86923.

Bonasio, R., Zhang, G., Ye, C., Mutti, N.S., Fang, X., Qin, N. *et al.* 2010. Genomic comparison of the ants *Camponotus floridanus* and *Harpegnathos saltator*. *Science* **329**: 1068–1071.

Bourke, A.F. 2011. *Principles of Social Evolution*. Oxford University Press, Oxford.

Buttstedt, A., Moritz, R.F.A. & Erler, S. 2013. Origin and function of the major royal jelly proteins of the honeybee (*Apis mellifera*) as members of the yellow gene family. *Biol. Rev. Camb. Philos. Soc.* **89**: 255–269.

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

Carroll, S.B., Grenier, J.K. & Weatherbee, S.D. 2005. *From DNA to Diversity: Molecular Genetics and the Evolution of Animal Design*. Blackwell Science, Malden.

Colgan, T.J., Carolan, J.C., Bridgett, S.J., Sumner, S., Blaxter, M.L. & Brown, M.J. 2011. Polyphenism in social insects: insights from a transcriptome-wide analysis of gene expression in the life stages of the key pollinator, *Bombus terrestris*. *BMC Genom.* **12**: 623.

Corona, M., Velarde, R.A., Remolina, S., Moran-lauter, A., Wang, Y., Hughes, K.A. *et al.* 2007. Vitellogenin, juvenile hormone, signaling and queen honey bee longevity. *Proc. Natl. Acad. Sci. USA* **104**: 7128–7133.

Cunha, A.D., Nascimento, A.M., Guidugli, K.R., Simões, Z.L.P. & Bitondi, M.M.G. 2005. Molecular cloning and expression of a hexamerin cDNA from the honey bee, *Apis mellifera*. *J. Insect Physiol.* **51**: 1135–1147.

Czechowski, W., Radchenko, A. & Czechowska, W.. 2002. The ants (Hymenoptera, Formicidae) of Poland. Museum and Institute of Zoology PAS.

Darwin, C. 1859. *On the Origin of the Species by Natural Selection*. J. Murray, London.

Drapeau, M.D., Albert, S., Kucharski, R., Prusko, C. & Maleszka, R. 2006. Evolution of the Yellow/Major Royal Jelly Protein family and the emergence of social behavior in honey bees. *Genome Res.* **16**: 1385–1394.

Evans, J.D. & Wheeler, D.E. 1999. Differential gene expression between developing queens and workers in the honey bee, *Apis mellifera*. *Proc. Natl. Acad. Sci. USA* **96**: 5575–5580.

Fay, M.P. 2010. Confidence intervals that match Fisher's exact or Blaker's exact tests. *Biostatistics* **11**: 373–374.

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.

Feng, S., Cokus, S.J., Zhang, X., Chen, P.-Y., Bostick, M., Goll, M.G. *et al.* 2010. Conservation and divergence of methylation patterning in plants and animals. *Proc. Natl. Acad. Sci. USA* **107**: 8689–8694.

Ferreira, P.G., Patalano, S., Chauhan, R., Ffrench-Constant, R., Gabaldón, T., Guigó, R. *et al.* 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.

Ferrier, D.E. & Holland, P.W. 2001. Ancient origin of the Hox gene cluster. *Nat. Rev. Genet.* **2**: 33–38.

Finseth, F.R. & Harrison, R.G. 2014. A Comparison of next generation sequencing technologies for transcriptome assembly and utility for RNA-Seq in a non-model bird. *PLoS One* **9**: e108550.

Fischman, B.J., Woodard, S.H. & Robinson, G.E. 2011. Molecular evolutionary analyses of insect societies. *Proc. Natl. Acad. Sci. USA* **108**: 10847–10854.

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

Grabherr, M., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I. *et al.* 2013. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nat. Biotechnol.* **29**: 644–652.

Gräff, 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.

Grishkevich, V. & Yanai, I. 2013. The genomic determinants of genotype x environment interactions in gene expression. *Trends Genet.* **29**: 479–487.

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.

Guidugli, K.R., Nascimento, A.M., Amdam, G.V., Barchuk, A.R., Omholt, S., Simões, Z.L.P. *et al.* 2005. Vitellogenin regulates hormonal dynamics in the worker caste of a eusocial insect. *FEBS Lett.* **579**: 4961–4965.

Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J. *et al.* 2013. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* **8**: 1494–1512.

Hall, D.W. & Goodisman, M.A.D. 2012. The effects of kin selection on rates of molecular evolution in social insects. *Evolution* **66**: 2080–2209.

Harpur, B.A., Kent, C.F., Molodtsova, D., Lebon, J.M.D., Alqarni, A.S., Owayss, A.A. *et al.* 2014. Population genomics of the honey bee reveals strong signatures of positive selection on worker traits. *Proc. Natl. Acad. Sci. USA* **111**: 2614–2619.

Harrison, M.C., Hammond, R.L. & Mallon, E.B. 2015a. Reproductive workers show queenlike gene expression in an

intermediately eusocial insect, the buff-tailed bumble bee *Bombus terrestris*. *Mol. Ecol.* **24**: 3043–3063.

Harrison, P.W., Wright, A.E., Zimmer, F., Dean, R., Montgomery, S.H., Pointer, M.A. *et al.* 2015b. Sexual selection drives evolution and rapid turnover of male gene expression. *Proc. Natl. Acad. Sci. USA* **112**: 4393–4398.

Hata, K., Okano, M., Lei, H. & Li, E. 2002. Dnmt3L cooperates with the Dnmt3 family of *de novo* DNA methyltransferases to establish maternal imprints in mice. *Development* **129**: 1983–1993.

Havukainen, H., Münch, D., Baumann, A., Zhong, S., Halskau, Ø., Krogsgaard, M. *et al.* 2013. Vitellogenin recognizes cell damage through membrane binding and shields living cells from reactive oxygen species. *J. Biol. Chem.* **288**: 28369–28381.

Helanterä, H. & Sundström, L. 2007. Worker Reproduction in *Formica* ants. *Am. Nat.* **170**: E14–E25.

Helanterä, H. & Uller, T. 2014. Neutral and adaptive explanations for an association between caste-biased gene expression and rate of sequence evolution. *Front. Genet.* **28**: 1381–1392.

Hoffman, E.A. & Goodisman, M.A.D. 2007. Gene expression and the evolution of phenotypic diversity in social wasps. *BMC Biol.* **5**: 23.

Hornett, E.A. & Wheat, C.W. 2012. Quantitative RNA-Seq analysis in non-model species: assessing transcriptome assemblies as a scaffold and the utility of evolutionary divergent genomic reference species. *BMC Genom.* **13**: 361.

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

Hunt, B.G., Ometto, L., Wurm, Y., Shoemaker, D., Yi, S., Keller, L. *et al.* 2011. Relaxed selection is a precursor to the evolution of phenotypic plasticity. *Proc. Natl. Acad. Sci. USA* **108**: 15936–15941.

Hunt, B.G., Ometto, L., Keller, L. & Goodisman, M.A.D. 2013. Evolution at two levels in fire ants: the relationship between patterns of gene expression and protein sequence evolution. *Mol. Biol. Evol.* **30**: 263–271.

Ingram, K.K., Oefner, P. & Gordon, D.M. 2005. Task-specific expression of the foraging gene in harvester ants. *Mol. Ecol.* **14**: 813–818.

Jasper, W.C., Linksvayer, T.A., Atallah, J., Friedman, D., Chiu, J.C. & Johnson, B.R. 2014. Large-scale coding sequence change underlies the evolution of postdevelopmental novelty in honey bees. *Mol. Biol. Evol.* **32**: 334–346.

Johansson, H., Dhaygude, K., Lindström, S., Helanterä, H., Sundström, L. & Trontti, K. 2013. A metatranscriptomic approach to the identification of microbiota associated with the ant *Formica exsecta*. *PLoS One* **8**: e79777.

Kato, Y., Kaneda, M., Hata, K., Kumaki, K., Hisano, M., Kohara, Y. *et al.* 2007. Role of the Dnmt3 family in *de novo* methylation of imprinted and repetitive sequences during male germ cell development in the mouse. *Hum. Mol. Genet.* **16**: 2272–2280.

Koch, S.I., Groh, K., Vogel, H., Hansson, B.S., Hannson, B.S., Kleineidam, C.J. *et al.* 2013. Caste-specific expression patterns of immune response and chemosensory related genes in the leaf-cutting ant, *Atta vollenweideri*. *PLoS One* **8**: e81518.

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

Langmead, B. & Salzberg, S.L. 2012. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**: 357–359.

Leichty, A.R., Pfennig, D.W., Jones, C.D. & Pfennig, K.S. 2012. Relaxed genetic constraint is ancestral to the evolution of phenotypic plasticity. *Integr. Comp. Biol.* **52**: 16–30.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N. *et al.* 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**: 2078–2079.

Li, W., Huang, Z.Y., Liu, F., Li, Z., Yan, L., Zhang, S. *et al.* 2013. Molecular cloning and characterization of juvenile hormone acid methyltransferase in the honey bee, *Apis mellifera*, and its differential expression during caste differentiation. *PLoS ONE* **8**: e68544.

Li, B., Fillmore, N., Bai, Y., Collins, M., Thomson, J.A., Stewart, R. *et al.* 2014. Evaluation of *de novo* transcriptome assemblies from RNA-Seq data. *Genome Biol.* **15**: 553.

Linksvayer, T.A. & Wade, M.J. 2009. Genes with social effects are expected to harbor more sequence variation within and between species. *Evolution* **63**: 1685–1696.

Liu, T.-T., Liu, J.-F., Wang, W.-X., Wang, H., Wang, Z.-L., Zeng, Z.-J. *et al.* 2012. Cloning and expression profiling of the DNA methyltransferase Dnmt3 gene in the Chinese honeybee, *Apis cerana cerana* (Hymenoptera: Apidae). *Acta Entomol. Sin.* **3**: 284–290.

Mackert, A., do Nascimento, A.M., Bitondi, M.M.G., Hartfelder, K. & Simões, Z.L.P. 2008. Identification of a juvenile hormone esterase-like gene in the honey bee, *Apis mellifera* L.–expression analysis and functional assays. *Comp. Biochem. Physiol. B. Biochem. Mol. Biol.* **150**: 33–44.

Mackert, A., Hartfelder, K., Bitondi, M.M.G. & Simões, Z.L.P. 2010. The juvenile hormone (JH) epoxide hydrolase gene in the honey bee *Apis mellifera* genome encodes a protein which has negligible participation in JH degradation. *J. Insect Physiol.* **56**: 1139–1146.

Mank, J.E. & Ellegren, H. 2009. Are sex-biased genes more dispensable? *Biol. Lett.* **5**: 409–412.

Mank, J.E., Hultin-Rosenberg, L., Axelsson, E. & Ellegren, H. 2007. Rapid evolution of female-biased, but not male-biased, genes expressed in the avian brain. *Mol. Biol. Evol.* **24**: 2698–2706.

Martins, J.R., Nunes, F.M.F., Simões, Z.L.P. & Bitondi, M.M.G. 2008. A honey bee storage protein gene, hex 70a, expressed in developing gonads and nutritionally regulated in adult fat body. *J. Insect Physiol.* **54**: 867–877.

Martins, J.R., Nunes, F.M.F., Cristino, A.S., Simões, Z.L.P. & Bitondi, M.M.G. 2010. The four hexamerin genes in the honey bee: structure, molecular evolution and function deduced from expression patterns in queens, workers and drones. *BMC Mol. Biol.* **11**: 23.

McCarthy, D.J., Chen, Y. & Smyth, G.K. 2012. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res.* **40**: 4288–4297.

Mikheyev, A.S. & Linksvayer, T. 2015. Genes associated with ant social behavior show distinct transcriptional and evolutionary patterns. *eLife* **4**: e04775.

Miura, T. 2001. Morphogenesis and gene expression in the soldier-caste differentiation of termites. *Insectes Soc.* **48**: 216–223.

Miura, T. 2004. Proximate mechanisms and evolution of caste polyphenism in social insects : from sociality to genes. *Ecol. Res.* **19**: 141–148.

Morandin, C., Havukainen, H., Kulmuni, J., Dhaygude, K., Trontti, K. & Helanterä, H. 2014. Not only for egg yolk-functional and evolutionary insights from expression, selection, and structural analyses of *Formica* ant vitellogenins. *Mol. Biol. Evol.* **31**: 2181–2189.

Noirot, C. 1991. Caste differentiation in *Isoptera*: basic features, role of pheromones. *Ethol. Ecol. Evol.* **3**: 2–7.

Nygaard, S., Zhang, G., Schiøtt, M., Li, C., Wurm, Y., Hu, H. *et al.* 2011. The genome of the leaf-cutting ant *Acromyrmex echinatior* suggests key adaptations to advanced social life and fungus farming. *Genome Res.* **21**: 1339–1348.

Ometto, L., Shoemaker, D., Ross, K.G. & Keller, L. 2011. Evolution of gene expression in fire ants: the effects of developmental stage, caste, and species. *Mol. Biol. Evol.* **28**: 1381–1392.

Pamilo, P. 1991. Life span of queens in the ant *Formica exsecta*. *Insectes Soc.* **38**: 111–119.

Patel, A., Fondrk, M.K., Kaftanoglu, O., Emore, C., Hunt, G., Frederick, K. *et al.* 2007. The making of a queen: TOR pathway is a key player in diphenic caste development. *PLoS ONE* **2**: e509.

Pereboom, J.J.M., Jordan, W.C., Sumner, S., Hammond, R.L. & Bourke, A.F.G. 2005. Differential gene expression in queen-worker caste determination in bumble-bees. *Proc. Biol. Sci.* **272**: 1145–1152.

Quevillon, E., Silventoinen, V., Pillai, S., Harte, N., Mulder, N., Apweiler, R. *et al.* 2005. InterProScan: protein domains identifier. *Nucleic Acids Res.* **33**: W116–W120.

Richards, S., Gibbs, R.A., Weinstock, G.M., Brown, S.J., Denell, R., Beeman, R.W. *et al.* 2008. The genome of the model beetle and pest *Tribolium castaneum*. *Nature* **452**: 949–955.

Robinson, M.D., McCarthy, D.J. & Smyth, G.K. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**: 139–140.

Ronquist, F., Klopfstein, S., Vilhelmsen, L., Schulmeister, S., Murray, D.L. & Rasnitsyn, A.P. 2012. A total-evidence approach to dating with fossils, applied to the early radiation of the hymenoptera. *Syst. Biol.* **61**: 973–999.

Schrader, L., Simola, D.F., Heinze, J. & Oettler, J. 2015. Sphingolipids, transcription factors, and conserved toolkit genes: developmental plasticity in the ant Cardiocondyla obscurior. *Mol. Biol. Evol.* **32**: 1474–1486.

Smith, C.D., Zimin, A., Holt, C., Abouheif, E., Benton, R., Cash, E. *et al.* 2011a. Draft genome of the globally widespread and invasive Argentine ant (*Linepithema humile*). *Proc. Natl. Acad. Sci. USA* **108**: 5673–5678.

Smith, C.R., Smith, C.D., Robertson, H.M., Helmkampf, M., Zimin, A., Yandell, M. *et al.* 2011b. Draft genome of the red harvester ant *Pogonomyrmex barbatus*. *Proc. Natl. Acad. Sci. USA* **108**: 5667–5672.

Stearns, S.C. 1989. The evolutionary significance of phenotypic plasticity. *Bioscience* **39**: 436–445.

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

Sumner, S., 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. London. Ser. B, Biol. Sci.* **273**: 19–26.

Sundström, L., Seppä, P. & Pamilo, P. 2005. Genetic population structure and dispersal patterns in *Formica* ants — a review. *Ann. Zool. Fennici* **42**: 163–177.

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

Surget-Groba, Y. & Montoya-Burgos, J.I. 2010. Optimization of *de novo* transcriptome assembly from next-generation sequencing data. *Genome Res.* **20**: 1432–1440.

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

Tian, H., Vinson, S.B. & Coates, C.J. 2004. Differential gene expression between alate and dealate queens in the red imported fire ant, Solenopsis invicta (Hymenoptera : Formicidae). *Insect Biochem. Mol. Biol.* **34**: 937–949.

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

Toth, A.L., Varala, K., Henshaw, M.T., Rodriguez-Zas, S.L., Hudson, M.E. & Robinson, G.E. 2010. Brain transcriptomic analysis in paper wasps identifies genes associated with behaviour across social insect lineages. *Proc. Biol. Sci.* **277**: 2139–2148.

Werren, J.H., Richards, S., Desjardins, C.A., Niehuis, O., Gadau, J., Colbourne, J.K. *et al.* 2010. Functional and evolutionary insights from the genomes of three parasitoid Nasonia species. *Science* **327**: 343–348.

West-Eberhard, M.J. 1989. Phenotypic Plasticity and the Origins of Diversity. *Annu. Rev. Ecol. Syst.* **20**: 249–278.

Wheeler, D.E., Buck, N. & Evans, J.D. 2006. Expression of insulin pathway genes during the period of caste determination in the honey bee, *Apis mellifera*. *Insect Mol. Biol.* **15**: 597–602.

Wheeler, D.E., Buck, N.A. & Evans, J.D. 2013. Expression of insulin/insulin-like signalling and TOR pathway genes in honey bee caste determination. *Insect Mol. Biol.* **23**: 113–121.

Wilkins, A.S. 2014. *"the genetic tool-kit": The Life-History of an Important Metaphor Advances in Evolutionary Developmental Biology*. John Wiley & Sons, Hoboken, NJ.

Wilson, E.O. 1971. *The Insect Societies*. Harvard University Press, Cambridge, MA.

Wissler, L., Gadau, J., Simola, D.F., Helmkampf, M. & Bornberg-Bauer, E. 2013. Mechanisms and dynamics of orphan gene emergence in insect genomes. *Genome Biol. Evol.* **5**: 439–455.

Woodard, S.H., Fischman, B.J., Venkat, A., Hudson, M.E., Varala, K., Cameron, S.A. *et al.* 2011. Genes involved in convergent evolution of eusociality in bees. *Proc. Natl. Acad. Sci. USA* **108**: 7472–7477.

Wurm, Y., Wang, J., Riba-Grognuz, O., Corona, M., Nygaard, S., Hunt, B.G. *et al.* 2011. The genome of the fire ant *Solenopsis invicta*. *Proc. Natl. Acad. Sci. USA* **108**: 5679–5684.

Yu, G., Li, F., Qin, Y., Bo, X., Wu, Y. & Wang, S. 2010. GOSemSim: an R package for measuring semantic similarity among GO terms and gene products. *Bioinformatics* **26**: 976–978.

Supporting information

Additional Supporting Information may be found in the online version of this article:

Figure S1 Functional enrichment analysis of caste–biased genes.

Figure S2 Venn diagram summarizing the number of life stage biased genes for among queen life stages, using EdgeR with FRD < 0.05.

Figure S3 Venn diagram summarizing the number of simple life stage biased genes for among worker life stages, using EdgeR with FRD < 0.05.

Figure S4 Functional enrichment analysis of life stage–biased genes.

Table S1 List of enriched Goterms for caste-biased genes.

Table S2 List of enriched Goterms for life stage-biased genes.

Table S3 List of genes with caste-specific differential expression patterns between castes and life stages of several social insect species, and *Formica exsecta* caste and life stage-biased direction.

Received 13 March 2015; revised 30 June 2015; accepted 30 June 2015