Comparative Transcriptomics of Convergent Evolution: Different Genes but Conserved Pathways Underlie Caste Phenotypes across Lineages of Eusocial Insects

Ali J. Berens,*[1,2] James H. Hunt,[3,4,5] and Amy L. Toth[1,2,6]

[1]Program in Bioinformatics and Computational Biology, Iowa State University

[2]Department of Ecology, Evolution, and Organismal Biology, Iowa State University

[3]Department of Biological Sciences, North Carolina State University

[4]Department of Entomology, North Carolina State University

[5]W. M. Keck Center for Behavioral Biology, North Carolina State University

[6]Department of Entomology, Iowa State University

***Corresponding author:** E-mail: berens.ali@gmail.com.

Associate editor: Gregory Wray

Abstract

An area of great interest in evolutionary genomics is whether convergently evolved traits are the result of convergent molecular mechanisms. The presence of queen and worker castes in insect societies is a spectacular example of convergent evolution and phenotypic plasticity. Multiple insect lineages have evolved environmentally induced alternative castes. Given multiple origins of eusociality in Hymenoptera (bees, ants, and wasps), it has been proposed that insect castes evolved from common genetic "toolkits" consisting of deeply conserved genes. Here, we combine data from previously published studies on fire ants and honey bees with new data for *Polistes metricus* paper wasps to assess the toolkit idea by presenting the first comparative transcriptome-wide analysis of caste determination among three major hymenopteran social lineages. Overall, we found few shared caste differentially expressed transcripts across the three social lineages. However, there is substantially more overlap at the levels of pathways and biological functions. Thus, there are shared elements but not on the level of specific genes. Instead, the toolkit appears to be relatively "loose," that is, different lineages show convergent molecular evolution involving similar metabolic pathways and molecular functions but not the exact same genes. Additionally, our paper wasp data do not support a complementary hypothesis that "novel" taxonomically restricted genes are related to caste differences.

Key words: **convergent evolution, phenotypic plasticity, social insects, castes, gene expression, transcriptome.**

Article

Introduction

Convergent evolution of similar phenotypes in distant lineages has historically been viewed as the result of divergent, nonhomologous underlying mechanisms (Arendt and Reznick 2008; Scotland 2011). However, some studies of convergent traits such as echolocation in bats and dolphins (Parker et al. 2013) and camera-like eyes in octopi and humans (Ogura et al. 2004) provide evidence that some of the same genes can be recruited multiple times in evolution to regulate the development of convergent traits (Elmer et al. 2010; Scotland 2011; Stern 2013). Other studies point to different genetic mechanisms for convergent traits, but importantly, these cases may still involve genes with overlapping molecular functions that can influence the same biochemical pathways (e.g., pigmentation, reviewed in Kronforst et al. [2012]; wing loss in ants Abouheif and Wray [2002]). These prior studies have focused on fixed phenotypes that characterize a particular species or population. However, many complex traits are plastic, highly responsive to the environment, and the result of the combined expression of many thousands of interacting genes. Are there also convergent molecular mechanisms for plastic traits that result from environmentally responsive gene expression?

Social insects are spectacular examples of both convergent phenotypic evolution and phenotypic plasticity. Across the insects, there have been multiple origins of alternate phenotypes (such as reproductive queen and nonreproductive worker female castes) that are independent of genotype. Environmental factors such as nutritional inequalities during development are associated with caste determination in most social insect species (reviewed in Smith et al. [2008]). Sociality evolved independently in insects multiple times (at least 11 times within the order Hymenoptera [Cameron and Mardulyn 2001; Brady et al. 2006; Hines et al. 2007; Pilgrim et al. 2008; Cardinal et al. 2010; Johnson et al. 2013]), with convergent evolution of nutritionally dependent caste polyphenisms even in taxa that show large differences in their life history and social biology. For example, advanced social insect societies such as honeybees and fire ants are defined by morphologically distinct queen and worker castes that are established early during larval development (Haydak 1970 [*Apis mellifera*—honey bee]; Petralia and Vinson 1979 [*Solenopsis*

invicta—fire ant]). In such species, larval food provisioning affects a developmental switch to produce striking alternative adult forms differing greatly in morphology, physiology, and behavior (Patel et al. 1960; Haydak 1970; Corona 1999; Evans and Wheeler 1999; Hepperle and Hartfelder 2001 [*A. mellifera*—honey bee]; Wheeler 1986; Linksvayer et al. 2006; reviewed in Anderson et al. [2008] [Ants]). In primitively social insect societies, such as paper wasps in the genus *Polistes*, there are no discrete morphological differences between castes, and caste-related behaviors are flexible even in adults (Reeve 1991). Nonetheless, even in *Polistes*, castes are biased during early development, such that poorly nourished first generation females are worker destined, whereas well-nourished second generation females are queen destined (Hunt and Dove 2002; Karsai and Hunt 2002; Hunt and Amdam 2005). Thus, there are common nutritional environmental determinants of caste in all three major eusocial lineages.

Examinations of insect social evolution from an evolutionary developmental (evo-devo) perspective led to the proposal that castes are derived from new arrangements of solitary behavioral and physiological modules (West-Eberhard 1987, 1996; Amdam et al. 2004; Hunt and Amdam 2005) and, along with these, deeply conserved genetic modules (reviewed in Page and Amdam [2007] and Toth and Robinson [2007]). A general hypothesis for convergent evolution of eusociality is that there is a shared "toolkit" of molecular and physiological processes across several independently evolved social insect lineages (Toth and Robinson 2007; Toth et al. 2010). Because queen and worker castes can be produced from the same genome, the genetic toolkit underlying convergent social caste phenotypes depends on the differential expression (DE) of common genes and\or pathways.

In the study of eusociality, a few previous studies on bees and wasps suggested some overlap in gene expression patterns related to social behavior across lineages. Using relatively small sets of a few dozen candidate genes, Toth et al. (2007) and Daugherty et al. (2011) found between 25% and 60% of genes showed similar expression patterns for aspects of division of labor in *P. metricus* and honey bees. Further comparative analysis using microarrays between honey bees and *P. metricus* wasps highlighted commonalities in regulation of foraging/provisioning behavior (Toth et al. 2010) and aggressive behavior (Toth et al. 2014) at the molecular level. However, the extent of overlap across species in these studies was relatively small, and an additional analysis suggested that reproductive behaviors are associated with nonshared gene expression patterns (Toth et al. 2010). One possible explanation for the relatively small overlap is that these studies focused only on individual genes. Instead, general changes in the modulation of consequential pathways and key biological functions may be more important (Toth and Robinson 2007). For example, because queens are nearly always larger and better nourished than workers (Wilson 1971), we would expect important changes in the regulation of metabolic and nutrient signaling pathways (reviewed in Smith et al. [2008]), but the specific genes involved and the direction of their expression may be more evolutionarily labile. Additional research on a genomic scale that includes multiple origins of sociality is needed to advance our understanding of the possible role of shared genes, pathways, and gene networks in social evolution.

In this study, our primary aim was to probe the molecular basis of the convergent evolution of castes by conducting the most comprehensive comparative analysis to date of caste-related global gene expression (table 1). Our study utilizes data from three independent social lineages, represented by a honey bee (*A. mellifera*), a fire ant (*S. invicta*), and a temperate paper wasp (*Polistes metricus*). This analysis was made possible by the release of comparable caste-related developmental transcriptomic resources for these species (Chen et al. 2012 [*A. mellifera*]; Ometto et al. 2011 [*S. invicta*]), including a new *P. metricus* transcriptome (this article). Alternative queen and worker caste phenotypes in these three species are strongly affected by differential nutrition (Haydak 1970; Wheeler 1986; Karsai and Hunt 2002). Thus, we hypothesized there would be evolutionary convergence across the three lineages in the expression of specific genes, pathways, and/or gene networks associated with caste determination, especially those related to nutrition and metabolism. Because of the fact that the lineages are distant (separated by over 100 My) and caste phenotypes are complex and plastic, we predicted there would be less overlap at the level of specific genes, and more overlap on the levels of pathways and biological functions.

Second, we also use the new *P. metricus* transcriptome data to address an emerging complementary hypothesis that posits that "novel" genes are important for the evolution of novel caste phenotypes (table 1; Johnson and Tsutsui 2011; Harpur et al. 2014). Novel genes are defined as previously undescribed genes that have no significant homology with known sequences (Ding et al. 2012). Genes that are novel (or "taxonomically restricted") to Hymenoptera, bees and the genus *Apis*, are more likely to be overexpressed in honey bee workers (Johnson and Tsutsui 2011) and show evidence of positive selection (Harpur et al. 2014), suggesting novel genes may play a role in causing caste differences. Similarly, in the tropical paper wasp *P. canadensis*, there is an overabundance of novel transcripts that show caste DE (Ferreira et al. 2013). However, comparisons of ant genomes suggest an abundance of species-specific novel genes may relate more to derived, species-specific traits possessed by workers, rather than castes differences per se (Simola et al. 2013). Thus, it is important to reassess the hypothesis that novel genes relate to caste differences by examining the association between novel gene expression and caste differences using pairs of more closely related species. In this study, we combined data from the two *Polistes* species to identify a set of well-supported *Polistes*-specific transcripts. If novel genes are relevant to the evolution of sociality in *Polistes*, we predicted *Polistes*-specific transcripts would be more likely to be caste-differentially expressed in both *P. canadensis* and *P. metricus*. The reason for this is that castes evolved only once in the common ancestor of both species, and thus any novel transcripts that contributed to

Table 1. The Two Major Hypotheses Being Tested in This Study Are Outlined, along with the Level at Which the Analysis Was Conducted (KEGG), the Possible Outcomes from Each Analysis, and the Inferred Relationship from each Possible Outcome.

	Hypothesis	At the Level of	Possible Outcomes	Relationship
Genetic toolkit	Same genetic mechanisms underpin the independent evolution of insect caste systems	Transcript expression	Common caste-associated gene expression	Convergent evolution of expression of specific genes
			Different caste-associated gene expression	Nonconvergent evolution of expression of specific genes
		KEGG pathways	Gene expression changes in common pathways	Convergent evolution of expression of genetic pathways
			Gene expression changes in different pathways	Nonconvergent evolution of expression of genetic pathways
		GO functions	Gene expression changes for common gene functions	Convergent evolution of expression of gene networks
			Gene expression changes for different gene functions	Nonconvergent evolution of expression of gene networks
Novel gene	Novel (i.e., *Polistes*-specific transcripts) genes important for the evolution of caste system in *Polistes* lineage	*Polistes*-specific transcripts	Common caste differentially expressed novel transcripts	Novel genes important for evolution of caste phenotypes in *Polistes* lineage
			Different caste differentially expressed novel transcripts	No evidence for novel genes in expression of caste phenotypes in *Polistes* lineage

Table 2. *Polistes metricus* Transcriptome Assembly Statistics Assembled De Novo with Trinity and Annotated with the mRNAmarkup Protocol.

Number of transcripts	Longest transcript	N50	N90	CEGMA partial	CEGMA complete
74,516	23,236	1,316	279	98.79%	93.55%

caste evolution in the *Polistes* lineage should show common, caste-related gene expression patterns.

The objectives of this study were thus 2-fold: 1) investigate transcriptome-wide molecular mechanisms associated with convergent social caste phenotypes across three major hymenopteran social lineages (bees, ants, and wasps) and 2) address the contribution of *Polistes*-specific transcripts to caste differences in two species of *Polistes* wasps.

Results

Transcriptome Assembly and Annotation

We sequenced 16 transcriptomic libraries, representing four biological replicates of *P. metricus* fifth-instar larvae from two castes (field-collected queen- and worker destined) and two nutritional manipulation levels (laboratory-reared low and high nourishment groups; to be published separately Berens AJ, Hunt JH, Toth AL in preparation). After de novo transcriptome assembly using Trinity (Grabherr et al. 2011) and filtering out chimeric and contaminant sequences using the mRNAmarkup protocol (Brendel V, unpublished data; code available from http://brendelgroup.org/bioinformatics2go/mRNAmarkup.php, last accessed June 18, 2013) the final *P. metricus* Transcriptome Shotgun Assembly (TSA) consisted of 74,516 transcripts with an N50 of 1,316 base pairs. To address quality and completeness, we assessed the *P. metricus* transcriptome assembly using the Core Eukaryotic Gene Mapping Approach (CEGMA) method (Parra et al. 2007); 93.55% of the Core Eukaryotic Genes (CEGs) mapped completely and 98.79% of the CEGs mapped partially to the TSA, thereby indicating a very complete representation of expressed genes. Summary statistics of the TSA are shown in table 2. In total, 16,662 (22.4%) of the *P. metricus* transcripts are putatively homologous to 13,429 unique sequences in National Center for Biotechnology Information (NCBI)'s nonredundant (NR) database. Ninety-six percent (15,997) of the best hits are from the order Hymenoptera (fig. 1 illustrates taxonomic grouping of the *P. metricus* transcriptomic hits to the NR database). With BLAST2GO (Conesa et al. 2005), we made a putative functional annotation of the *P. metricus* TSA based on the best Basic Local Alignment Search Tool (BLAST) hit to *Drosophila melanogaster*. From 12,225 hits to *Drosophila*, 9,736 *P. metricus* transcripts were functionally annotated (table 4 lists the *P. metricus* transcriptome functional annotation and Gene Ontology [GO] enrichment statistics. Supplementary file, Supplementary Material online—Pmet-tsa-r1.1_Annotation.xlsx, includes *P. metricus* GO annotations).

Polistes metricus *Caste DE*

We identified 736 differentially expressed transcripts (DETs) between *P. metricus* queen and worker-destined larvae using DESeq (Anders and Huber 2010) ($n = 4$ per caste, False Discovery Rate (FDR) ≤ 0.05 [Benjamini and Hochberg 1995]). Surprisingly, of these *P. metricus* caste DETs, 91.7% were upregulated in worker-destined relative to queen-destined larvae (fig. 2 is a heat map of the scaled read counts for the *P. metricus* caste DETs by sample; see supplementary file, Supplementary Material online—Pmet-tsa-r1.1_Caste.xlsx, for a list of *P. metricus* caste DETs). Interestingly, there is one worker-destined sample that clustered with the

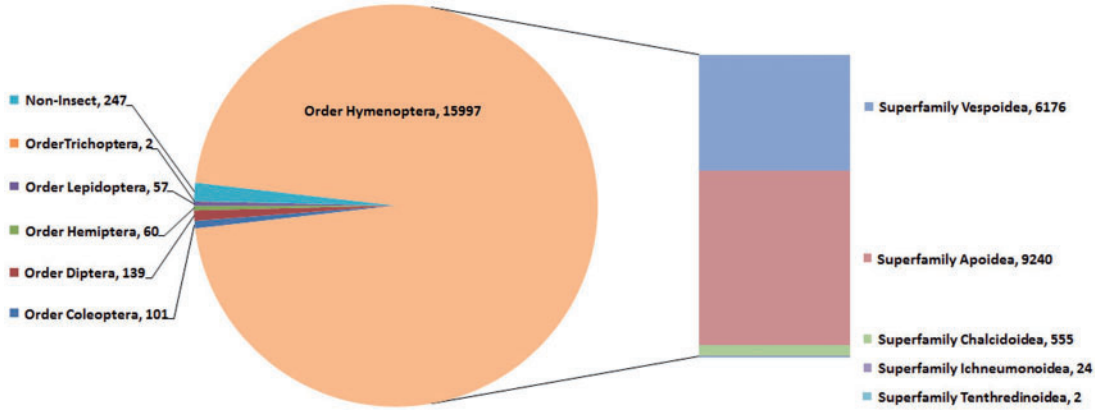

Fig. 1. Taxonomic grouping of the 16,662 *Polistes metricus* transcripts with hits to the NCBI NR database (*E* value ≤ 10e-4). Bar chart indicates hits to superfamilies in the order Hymenoptera. Listed next to each taxonomic class is the number of hits; 59 hits (not shown) did not have taxonomic information.

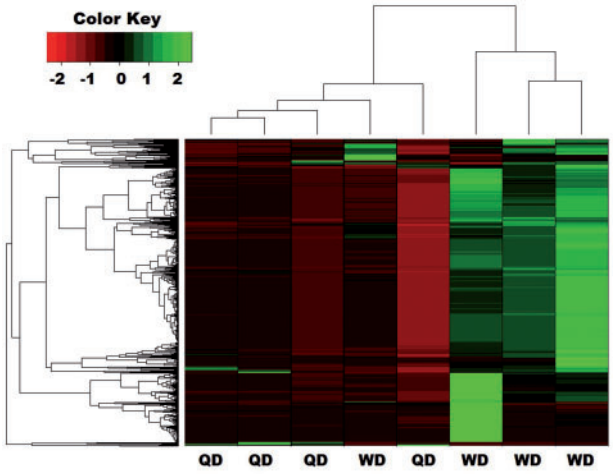

Fig. 2. Heat maps of the scaled sample read counts for the 736 DETs between castes. Groups: QD—queen-destined larvae and WD—worker-destined larvae.

queen-destined samples, which highlights the large biological variation that exists among individuals and is not totally surprising given the fact that some early season females can show queen-like phenotypes and even enter early diapause (reviewed in Hunt [2007]). Strikingly, a previous candidate study (quantitative reverse transcription-polymerase chain reaction [PCR]) identified 16 caste DETs related to lipid metabolism, heat response, and stress response in *P. metricus* (Hunt et al. 2010) and all were downregulated in worker-destined larvae, which is in the opposite direction with respect to the majority of *P. metricus* DETs in this study. A comparison of our RNA-seq data to these published data confirmed the directionality of expression patterns (downregulation in worker-destined larvae, 20% of the significantly differentially expressed between castes for the RNA-seq data) was consistent for these 16 transcripts across the two studies. Supplementary figure S1, Supplementary Material online, is a heat map of the scaled read counts from this study for the 16 caste-related transcripts identified in Hunt et al. (2010). This result suggests a relatively small but important suite of genes including hexamerins, heat shock proteins, and insulin signaling (Hunt et al. 2010) are upregulated in queen-destined larvae, and these results also provide independent validation for our RNA-seq results.

Comparative Caste Transcriptomics of Wasps, Ants, and Bees

We compared the results of our *P. metricus* caste DE analysis to the most recent, comparable data sets from other social insects (Chen et al. [2012] on the honey bee, *A. mellifera* and Ometto et al. [2011] on the fire ant, *S. invicta*). Comparisons were conducted at three levels: The transcript level, at the level of Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways (Kanehisa and Goto 2000; Kanehisa et al. 2014), and GO categories (Gene Ontology Consortium 2000). The transcript level served as a proxy for genes; transcripts representing longest assembled, expressed sequences were used because many genes of interest did not have full gene sequences available for one or more species. KEGG pathways and GO categories served as rough proxies for genetic pathways and gene networks, respectively. KEGG pathways elucidate information about enzymes that are known to interact as part of well-conserved biochemical pathways, and GO categories provide information about potential broad functional changes (i.e., above the level of pathways) associated with differential gene expression.

DET Analysis

We compared the list of *P. metricus* caste DETs to lists of caste DETs from *A. mellifera* (Chen et al. 2012) and *S. invicta* (Ometto et al. 2011). Overall, there are 15 shared orthologous DETs (of 2,475 orthologous sequences, *E* value ≤ 1e-4) across these three social insect species. In pairwise comparisons of orthologous sequences between *A. mellifera* and *P. metricus* and between *A. mellifera* and *S. invicta*, there is an overrepresentation and underrepresentation of shared caste DETs, respectively (Fisher's exact tests, *P* values ≤ 0.05). Between *P. metricus* and *S. invicta* orthologous transcripts, the observed number of caste DETs does not deviate significantly from the number of expected DETs (Fisher's exact test,

A **Differentially Expressed Transcripts**

	Apis mellifera (honey bee)	*Polistes metricus* (temperate paper wasp)
Polistes metricus (temperate paper wasp)	110 (88) of 14947	
Solenopsis invicta (fire ant)	302 (436) of 2493	12 (19) of 4706

B **Modulated KEGG Pathways**

	Apis mellifera (honey bee)	*Polistes metricus* (temperate paper wasp)
Polistes metricus (temperate paper wasp)	19 (16) of 118	
Solenopsis invicta (fire ant)	19 (17) of 118	8 (3) of 118

C **Enriched GO Terms**

	Apis mellifera (honey bee)	*Polistes metricus* (temperate paper wasp)
Polistes metricus (temperate paper wasp)	22 (11) of 3390	
Solenopsis invicta (fire ant)	30 (5) of 3416	21 (10) of 1045

Fig. 3. Number of observed and expected (in parentheses): (*A*) Common caste DETs out of total number of shared transcripts, (*B*) common modulated KEGG pathways out of total number KEGG pathways, (*C*) common caste GO-enriched terms out of total number GO-enriched terms (FDR ≤ 0.05; two tailed) between pairs of species. The color indicates either a statistically significant overrepresentation (blue) or a statistically significant underrepresentation (yellow) of common DETs, KEGG pathways, or GO-enriched terms (Fisher's exact test, *P* value ≤ 0.05).

see fig. 3*A* for the observed and expected number of shared caste DETs by species pairs). In addition to a relatively small overlap in caste DETs, directionality (upregulation in worker- or queen-destined samples) was not well-conserved across species (fig. 4*A* is a heat map of the caste DETs, colored by directionality, for the three species of social insects; supplementary table S5, Supplementary Material online, summarizes directionality of DETs between pairs of species). In *Polistes* and *S. invicta*, the majority (91.7% and 69%, respectively) of the caste-biased transcripts were upregulated in worker-destined samples, whereas 67.5% of genes were upregulated in queen-destined larvae in *A. mellifera*. Additionally, correlations between expression levels of orthologous DETs across species were weak; $\log_2$ fold changes between *P. metricus* and *S. invicta* showed a very weak negative correlation and between *P. metricus* and *A. mellifera* a very weak positive correlation (Pearson's correlation test, *P* value < 0.05; supplementary table S6 and fig. S2, Supplementary Material online).

KEGG Pathway Analysis

Using the lists of caste DETs, we identified KEGG pathways (Kanehisa and Goto 2000) from each of the three focal species with large numbers of differentially expressed genes; such pathways were designated as "modulated." We define a modulated metabolic pathway as a KEGG pathway where, across all enzymes within the pathway, the sum of the proportion of DETs per enzyme is ≥ 1 (determination of the modulated threshold is described in the Materials and Methods section below). Across all three species, there are eight common modulated KEGG pathways (see fig. 4*B* for a Venn diagram of the number of modulated KEGG pathways between *A. mellifera*, *P. metricus*, and *S. invicta*, see supplementary file, Supplementary Material online—SpeciesComparison.xlsx, for the full list of modified KEGG pathways) including arginine/proline metabolism and glycolysis/gluconeogenesis (see fig. 5 for an abridged glycolysis/gluconeogenesis pathway, and supplementary fig. S5, Supplementary Material online, for an abridged arginine/proline metabolism pathway that highlights caste DETs for *A. mellifera*, *P. metricus*, and *S. invicta*). Fisher's exact tests showed a significant overlap in the number of modulated KEGG pathways between *A. mellifera–P. metricus* and *P. metricus–S. invicta*. There is no deviation from the expected number of modulated KEGG pathways between *A. mellifera* and *S. invicta* (fig. 3*B* contains the observed and expected number of modulated KEGG pathways between pairs of species).

Although our study focuses on gene expression, previous work in bees identified signatures of selection at the sequence level (Woodard et al. 2011). Woodard et al. (2011) found that genes associated with carbohydrate metabolism, specifically within the glycolysis pathway, showed evidence of positive selection in multiple social lineages of bees. Additionally, they suggest that within the *Apis* lineage, there is an abundance of sequence changes within the glycolysis pathway. Our results suggest evolutionary changes in gene regulation may have also occurred within the glycolysis pathway, because we identified a very large number of differentially expressed enzymes within this pathway in *A. mellifera* (see fig. 5 for an abridged glycolysis/gluconeogenesis metabolism pathway). In fact, two-thirds of the glycolysis/gluconeogenesis differentially expressed enzymes are in *A. mellifera* (but note that *A. mellifera* had many more DETs overall). In contrast to the large number of modifications within the *A. mellifera* glycolysis pathway, there are only two enzymes corresponding to *P. metricus* caste DETs—ec:1.2.1.3 (NAD+ dehydrogenase) and ec:1.1.1.2 (NADP+ dehydrogenase). For both *A. mellifera* and *P. metricus*, NAD+ dehydrogenase is upregulated in queen-destined samples, whereas NAD+ dehydrogenase is not differentially expressed in *S. invicta*. NADP+ dehydrogenase

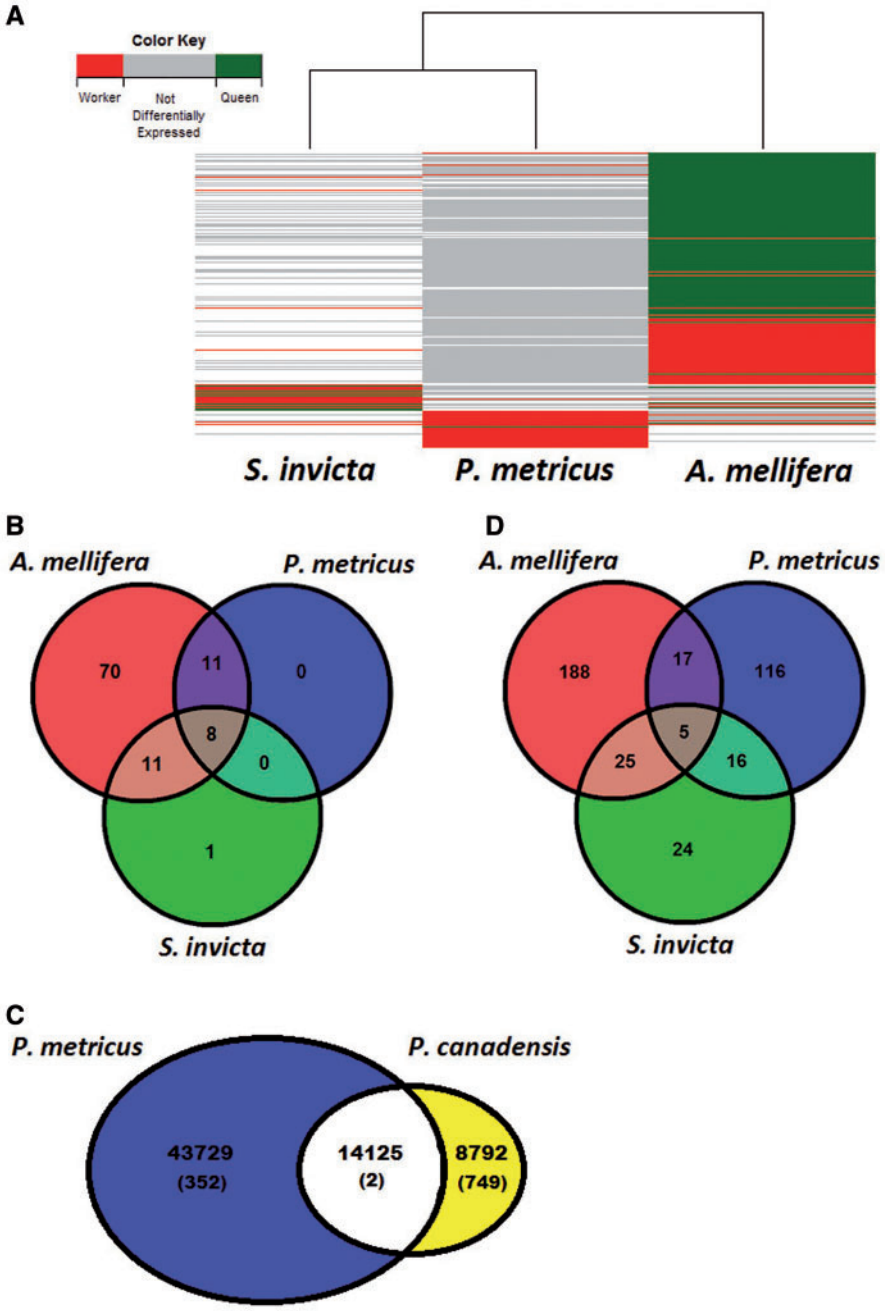

FIG. 4. (*A*) Heat map of the DETs for three species of social insects (*Apis mellifera*, *Polistes metricus*, and *Solenopsis invicta*). For each species, putative homologs of the DETs were determined for all other species by BLAST. White indicates that no putative homolog was found (*E* value > 1e-4). Based on the log fold change, each transcript was classified as upregulated in queen destined (green), upregulated in worker destined (red), or not differentially expressed between castes (Gray). (*B*) Venn diagram of the number of interrupted KEGG pathways (threshold 1.0) for *A. mellifera*, *P. metricus*, and *S. invicta*. (*C*) Venn diagram of the number of GO terms significantly enriched (FDR < 0.05; two-tailed) between differentially expressed caste transcripts and the remaining transcriptome for *A. mellifera*, *P. metricus*, and *S. invicta*. (*D*) Venn diagram of the number of novel transcripts (transcripts without a hit [*E* value > 1e-4] to the NCBI NR database) with the number of caste differentially expressed novel transcripts in parentheses for *P. canadensis* and *P. metricus*.

is upregulated in *P. metricus* worker-destined larvae, upregulated in *S. invicta* queen-destined samples and not present in *A. mellifera* samples. Again, this indicates that directionality is not conserved across taxa. From Woodard et al. (2011), GB15039 (ec:4.2.1.11, hydratase) was classified as having accelerated evolution within highly social lineages compared with primitively social and nonsocial lineages. We found hydratase significantly upregulated in workers in highly social honey bees and fire ants but not differentially expressed in the primitively social paper wasp *P. metricus*.

GO Enrichment Analysis

For each social insect lineage, we functionally annotated the transcriptomes based on the best BLAST hits to *D. melanogaster* sequences and identified enriched GO categories (FDR ≤ 0.05; supplementary file, Supplementary

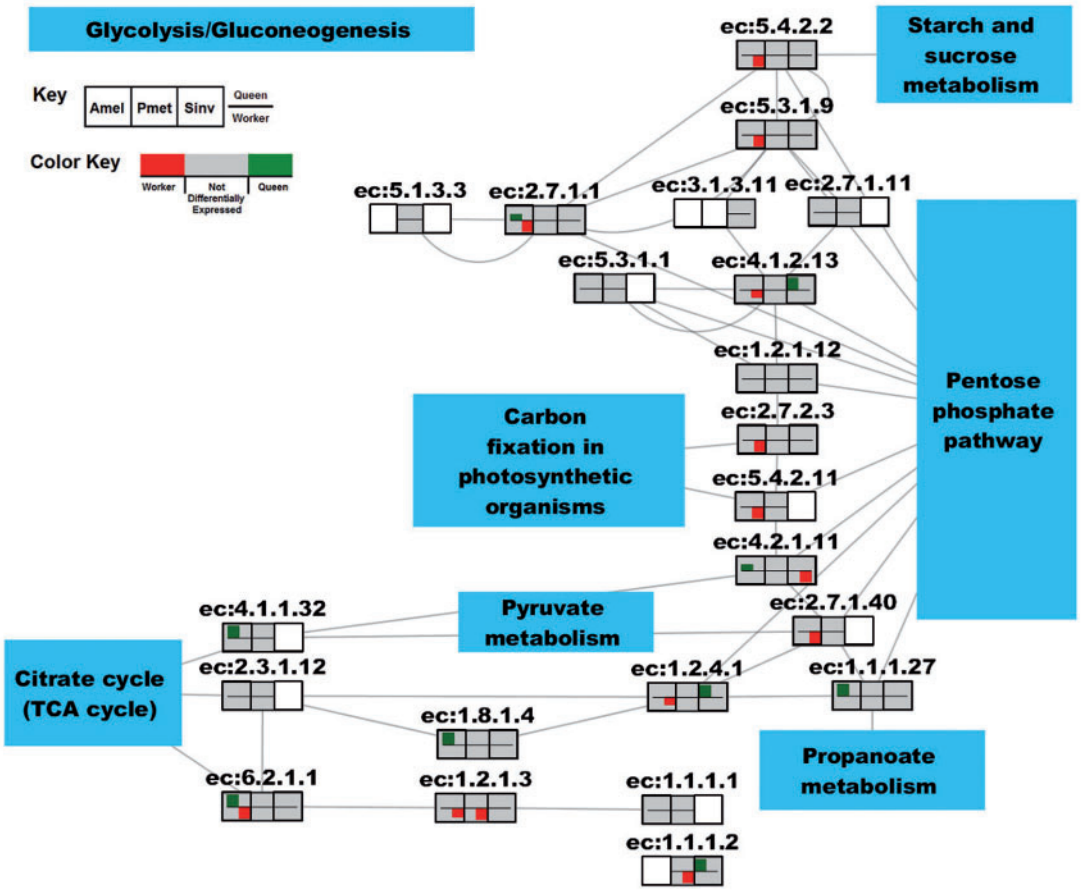

Fig. 5. Abridged glycolysis/gluconeogenesis metabolism pathway (KEGG map:00010), which highlights caste DETs for *Apis mellifera*, *Polistes metricus*, and *Solenopsis invicta*. Based on the log fold change, each transcript was classified as upregulated in queen-destined samples (green, positive direction), upregulated in worker-destined samples (red, negative direction), or not differentially expressed (gray). For some enzymes, there were multiple transcripts that mapped to the enzyme. Thus, the relative proportion of DETs was calculated for each enzyme and displayed as the bar height. White indicates that no putative homolog was found (*E* value $>$ 1e-4).

Material online—SpeciesComparison_Annotations.xlsx, contains all and significantly enriched GO terms for *A. mellifera* and *S. invicta*. Supplementary files, Supplementary Material online—Pmet-tsa-r1.1_Annotation.xlsx and Pmet-tsa-r1.1_Caste.xlsx, include the *P. metricus* functional annotation and significantly enriched GO terms). Pairwise comparisons of caste GO enrichment terms revealed a significant overrepresentation of shared terms for all species pairs (Fisher's exact tests, *P* values ≤ 0.05; see fig. 3C for the pairwise observed and expected number of enriched GO terms). Five significantly enriched caste-biased GO terms (GO:0044424—intracellular part, GO:0005622—intracellular, GO:0009653—anatomical structure morphogenesis, GO:2001141—regulation of RNA biosynthetic process, and GO:0006355—regulation of transcription, DNA dependent) are shared in common between *A. mellifera*, *P. metricus*, and *S. invicta*, which suggests that similar gene networks may be affected across species (fig. 4C is a Venn diagram of the number of overlapping enriched GO categories).

Polistes-*Specific Transcripts*

To investigate the role of taxonomically restricted transcripts in caste differences, we examined the expression patterns of transcripts that were specific to the *Polistes* lineage. First, we reanalyzed the raw data from Ferreira et al. (2013) using the same pipeline (Bowtie2-Express-DESeq) used on our own data to control for methodological differences in the identification of DETs. This resulted in the identification of 1,320 DETs (see supplementary file, Supplementary Material online—SpeciesComparison.xlsx) with our methods compared with 2,543 DETs from Ferreira et al. (2013). With this more comparable set of DETs, we still identified an overabundance of *P. canadensis* novel transcripts represented within the list of caste DETs (Fisher's exact test, odds ratio = 1.30, *P* value = 3.011e-6), as was observed in Ferreira et al. (2013). Next, we identified 57,854 *P. metricus* novel transcripts (without a BLAST hit [*E* value $>$ 1e-4] to the NCBI NR database), of which 354 were differentially expressed across castes (novel DETs). Forty-one (12%) of the *P. metricus* novel DETs are upregulated in queen-destined larvae, and the remaining 313 (88%) *P. metricus* novel DETs are upregulated in worker-destined larvae. These are similar to the proportions for all caste-biased transcripts in *P. metricus* (8% upregulated in queen destined [61 transcripts] and 92% upregulated in worker destined [675 transcripts]). Overall, there is an underrepresentation of caste DETs within novel transcripts in *P. metricus* (only 354 of the 57,854 novel transcripts; Fisher's

exact test, odds ratio = 0.26, $P < 2e\text{-}16$). This is in stark contrast to the reported overrepresentation of caste DETs in *P. canadensis* novel transcripts (Ferreira et al. 2013) (see fig. 4*D* for a Venn diagram of the number of novel transcripts and DETs for *P. canadensis* and *P. metricus*). Comparing *P. metricus* with *P. canadensis* transcripts, we identified 14,125 shared *Polistes* taxonomically restricted transcripts. Very few (only 2 or 0.01%) of these showed DE across castes (in both *P. metricus* and *P. canadensis*), and this overlap was not statistically significant (Fisher's exact test, odds ratio = 1.26, $P = 0.67$; fig. 4*D*).

Discussion

Comparative Transcriptomics of Convergent Evolution

The genetic toolkit hypothesis for the evolution of sociality predicts that there are common caste-related genes and genetic pathways shared across the order Hymenoptera (table 1). In this analysis, we identified a very small number (15) of common caste DETs between representatives of three hymenopteran social lineages (fire ants, honey bees, and paper wasps). Furthermore, directionality of caste-related gene expression is not conserved across species. This suggests that the same genes are not necessarily involved in caste differences across different origins of eusociality.

Instead, our data suggest that there may be similar molecular functional changes leading to convergent caste phenotypes through modifications within common metabolic pathways and gene networks. We identified five common KEGG pathways, including arginine/proline metabolism and glycolysis/gluconeogenesis (fig. 5, supplementary file, Supplementary Material online—SpeciesComparison.xlsx, supplementary fig. S5, Supplementary Material online) that were modulated in all three species. Previously, Woodard et al. (2011) proposed a link between the glycolysis pathway and multiple origins of sociality within bees based on sequence changes across the *Apis* lineage. The results from our cross-species gene expression analysis suggest that the glycolysis pathway may play a role in caste determination across Hymenoptera.

Even beyond specific metabolic pathways, our data indicate that common molecular functions are associated with caste differences across species, suggesting there may be shared gene networks related to eusocial evolution in the three lineages. We identified a common enrichment of five GO functional categories related to transcription regulation and morphogenesis across fire ants, honey bees, and paper wasps (figs. 3*B* and 4*B*, supplementary file, Supplementary Material online—SpeciesComparison.xlsx). This multilevel analysis suggests that caste phenotypes are associated with evolutionary lability at the level of genes and similarity on the level of pathways and GO categories (fig. 6). Thus, the genetic toolkit appears to be rather "loose"; social insect caste systems have evolved to utilize metabolic pathways and biological functions rather than the same genes (table 1).

How do these results inform our understanding of convergent evolution? Although debated in the literature (Arendt and Reznick 2008), some authors (e.g., Scotland 2011) suggest that focusing on underlying mechanisms can help distinguish between cases of clear convergent evolution

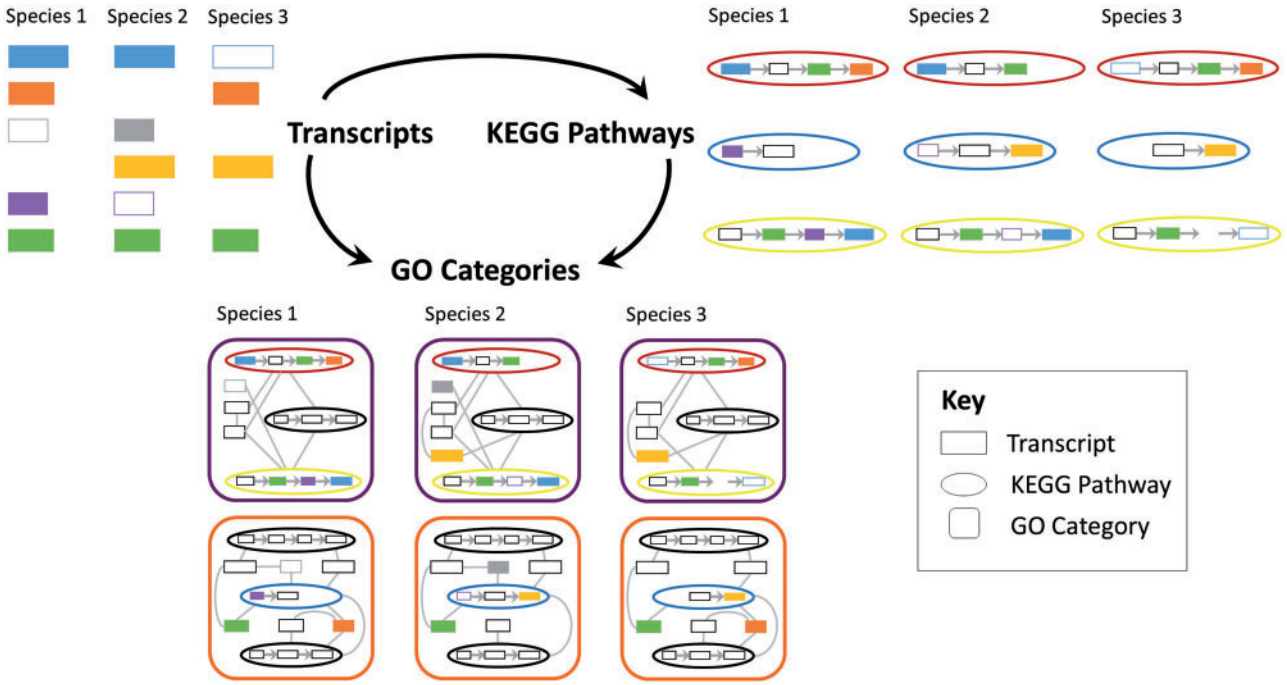

FIG. 6. Conceptual visualization of multilevel effects of transcripts, metabolic pathways, and biological functions. (*a*) Transcripts were identified as being differentially expressed (filled rectangle) or not significantly different (outlined rectangle), for all putative orthologs (row). Species without a putative ortholog were left blank. (*b*) KEGG pathways contain multiple, interacting transcripts. Similar phenotypic effects may occur across the three species due to modulation of KEGG pathways at differing transcripts. Both transcripts and KEGG pathways can affect enrichment of (*c*) GO categories.

Table 3. Summary of Transcriptome-Wide Analyses on Caste Determination in Social Insects: *Apis mellifera, Polistes metricus*, and *Solenopsis invicta*.

Species	Number of DETs between Castes	Life Stage	Tissue	Technology	Reference
A. mellifera (honey bee)	4,363	Fifth-instar larvae	Whole body	RNA-Seq (SOLiD)	Chen et al. (2010)
P. metricus (temperate paper wasp)	736	Fifth-instar larvae	Heads	RNA-Seq (Illumina)	This study
S. invicta (fire ant)	854	Pupae	Whole body	Microarray	Ometto et al. (2011)

Table 4. *Polistes metricus* Transcriptome Functional Annotation Statistics and GO Enrichment Statistics.

Sequences with				GO Terms	
Blast	Mapping	Annotation	Annotation and InterPro Scan	Total	Enriched
12,225 (16.4%)	10,177 (13.7%)	9,532 (12.8%)	9,736 (13.1%)	187	49 (26.2%)

NOTE.—Functional annotation was determined using BLAST2GO based on the best BLAST hit to *Drosophila melanogaster*. In parentheses, the proportion of sequences with BLAST results, mapping results, or annotated compared with the total number of transcripts or the proportion of caste enriched GO terms compared with the total GO terms, respectively.

(same phenotype, different mechanisms, and often in distant taxa) and cases of parallel evolution (same phenotype, same mechanisms, and often in more closely related taxa). Caste-containing societies of ants, bees, and wasps have convergently evolved queen and worker phenotypes from superficially similar yet distinctly different ancestral solitary states, stemming from various forms of maternal care behavior in each ancestral lineage (West-Eberhard 1996, Schwarz et al. 2011). Queen and worker phenotypes are multicomponent states that represent a collection of distinct morphological, behavioral, and physiological traits (Amdam and Page 2010), and they are plastic and highly polygenic (Grozinger et al. 2007). Therefore, our results on gene expression in social insect castes do not fit neatly into prior definitions of convergence or parallelism that focus more on discrete, fixed traits (Scotland 2011). However, our results are illuminating in that they suggest that the convergent evolution of complex traits is to some extent associated with repeatable molecular changes, at least on the level of pathways and biological functions.

It is important to note that the fire ant, honey bee, and paper wasp studies were published independently and not designed to be directly comparable, so there are a number of differences between these studies (table 3) in tissue types (whole body vs. head), life stages (fifth-instar larvae vs. pupae), and technologies (microarray vs. RNA-sequencing—both SOLiD and Illumina platforms). The *S. invicta* microarray study may have identified fewer DETs because arrays are limited to a few thousand candidate genes, and in *A. mellifera*, the large number of DETs compared with the two other species may be due in part to a lack of biological replicates. Despite the technical differences (above), ecological differences (Oster and Wilson 1978; Ross and Keller 1995), and evolutionary distance (Johnson et al. 2013) between these species, it is noteworthy that we were still able to uncover significant overlaps from our meta-analysis of transcriptomic data. In the future, studies that use more directly comparable approaches (life stages, platforms and tissues) and additional species along the solitary to social spectrum could provide an even more complete assessment of the importance of conserved transcripts and pathways in social evolution.

Our RNA-seq results are robust; we were able to replicate expression patterns from previously published real-time quantitative PCR data on 16 genes associated with insulin signaling pathways and storage proteins (Hunt et al. 2010) (supplementary fig. S1, Supplementary Material online). This verification is an excellent test of the data, because all 16 of the previously identified transcripts have higher expression in queen-destined larvae, whereas the majority of DETs from this study show the opposite pattern. Thus, the specific expression of these 16 genes highlights the consistency between our RNA-Seq and previous quantitative PCR results, despite differences in technologies. Furthermore, these results highlight that certain key processes including those related to insulin signaling pathways and storage proteins are upregulated in queen-destined larvae (Hunt et al. 2003, 2007, 2011; Wheeler et al. 2006; Patel et al. 2007), whereas many other pathways and functions are downregulated.

Novel Genes

An emerging hypothesis for social evolution posits that DE of novel (previously undescribed) transcripts is important for the evolution of novel caste phenotypes (Johnson and Tsutsui 2011; Ding et al. 2012; Ferreira et al. 2013; Harpur et al. 2014). As a corollary to this hypothesis, we predicted that species within the same social lineage should share common novel DETs. In our analysis of two paper wasps, *P. canadensis* and *metricus*, this prediction was not supported. Unlike a previously published report in which *P. canadensis* showed a significant bias toward novel transcripts being caste related, novel transcripts were actually significantly underrepresented among the caste-related transcripts of *P. metricus*. In addition, there were very few (only nine) novel caste-related transcripts shared between the two *Polistes* species (fig. 4C). One possible explanation for the difference between *Polistes* species is that the two studies examined different life stages (larval for *P. metricus* compared with adult for *P. canadensis*). As it now stands, there are conflicting reports

on the importance of novel genes in social evolution. Studies in honey bees paved the way for studies of novel genes, with data suggesting these genes are particularly important in worker behavior in this advanced eusocial species (Johnson and Tsutsui 2011; Harpur et al. 2014). On the contrary, a recent report in advanced eusocial fire ants (*S. invicta*) found that colony social organization (queen founding strategy) does not appear to be associated with the expression of novel genes (Manfredini et al. 2013). This collection of results suggests there may be variation in the importance of novel genes for different types of behavior, in specific life stages, or at different levels of sociality. In addition, there are some challenging technical issues associated with defining and detecting novel genes, in particular because genomic resources and gene discovery tools are rapidly advancing and changing (Khalturin et al. 2009).

Conclusion

Our comparative transcriptomic analysis across eusocial lineages and levels of molecular analysis suggests significant overlap in the types of metabolic pathways and gene functions associated with the convergent evolution of castes (especially those related to carbohydrate and amino acid metabolism, morphogenesis, oxidation–reduction, and transcriptional regulation). Individual transcripts show less conserved expression patterns and relationships to caste phenotypes across species. This suggests that convergent social behavioral phenotypes across lineages are the product of convergent evolution of molecular mechanisms at the level of metabolic pathways and gene networks, which can be modulated at many different places within such networks (i.e., transcript expression, fig. 6). Thus, although evolution may indeed be a "molecular tinkerer" (Jacob 1977) because different genes can produce similar phenotypes, selection may be constrained to act within a limited set of pathways and gene networks.

Materials and Methods

We collected four biological replicates of *P. metricus* fifth-instar larvae for each of the following groups: Unmanipulated castes (queen- and worker destined) and two nutritional manipulation levels (low and high). The nutritional manipulation samples are only used for transcriptome assembly and will be published separately (Berens et al., in preparation). For the unmanipulated caste-destined larvae, we collected samples from eight naturally occurring *P. metricus* nests at the Iowa 4-H Center (Madrid, IA). On the basis of the colony life cycle of paper wasps (Hunt et al. 2011), we collected foundress-reared (worker-destined) and worker-reared (queen-destined) larvae late spring and late summer 2010, respectively. Both males and females may be produced during the late summer. We removed males from the experiment based on the patterns and coloration of the eighth and ninth abdominal segments on the thawed bodies (laboratory protocol based in part on Cotoneschi et al. [2007]). Larvae were flash frozen in liquid nitrogen and stored at −80 °C until RNA extraction.

RNA Extraction and RNA Sequencing

We extracted total RNA from individual larval heads using an RNeasy Kit (Qiagen). RNA was quality controlled using spectrophotometry (NanoDrop) and a Bioanalyzer (Agilent). The High-Throughput Sequencing and Genotyping Unit of the W.M. Keck Center (University of Illinois at Urbana-Champaign) prepared mRNA Seq libraries with Illumina's "TruSeq RNAseq Sample Prep kit" and sequenced the libraries on a HiSeq 2000 (Illumina). This library preparation included poly(A) RNA purification, fragmentation using sonification, cDNA synthesis from size selected fragments (270 nucleotides on average) using random primers, and bar coding for each of the 16 samples. In a balanced incomplete block design (Auer and Doerge 2010), four lanes of the sequencer were used to generate over 1 billion 100 base paired-end reads with one subject from each treatment in each lane (supplementary table S1, Supplementary Material online, contains read counts for each sample). Raw sequence data have been deposited to the NCBI's Short Read Archive (BioProject ID: PRJNA242774, accession numbers: SRX511425, SRX511426, SRX511427, SRX511430, SRX511432, SRX511433, SRX511434, SRX511435).

Data Preprocessing

Visualization

We visualized the raw reads from each paired end file using FastQC (Andrews S, unpublished data; code available from http://www.bioinformatics.babraham.ac.uk/projects/fastqc/, last accessed November 28, 2011) and SolexaQA (Cox et al. 2010) to determine data quality and identify potential problems with the data. Visualization of all samples using FastQC (Andrews S, unpublished data; code available from http://www.bioinformatics.babraham.ac.uk/projects/fastqc/, last accessed November 28, 2011) and SolexaQA (Cox et al. 2010) suggested that the overall quality of the data is very high. However, there were bases of lower quality, especially at the end of the reads; this is typical of Illumina data and does not indicate a problem if addressed by quality filtering.

Adapter Removal

As part of the library preparation, the High-Throughput Sequencing and Genotyping Unit of the W. M. Keck Center (University of Illinois at Urbana-Champaign) adhered adapter sequences to each end (for pair-end sequencing) of the cDNA fragments. We removed the extraneous sequence before transcriptome assembly using the fastx_clipper tool from the Fastx Toolkit Version 0.0.13 (Hannon GJ, unpublished data; code available from http://hannonlab.cshl.edu/fastx_toolkit/, last accessed Dec 20, 2011).

Quality Filtering

We filtered reads for quality (threshold ≥20) using the Trim Perl script (Joshi N, unpublished data; code available from http://wiki.bioinformatics.ucdavis.edu/index.php/Trim.pl, last accessed November 11, 2011) with a length threshold of 50 bases. After filtering the data, approximately 2% of the reads (600,000–925,000) were removed from each library.

Transcriptome Assembly and Annotation

We assembled the groomed transcriptomic short reads de novo using Trinity (Grabherr et al. 2011, Version r2012-06-08). The final trinity-produced TSA was annotated with the mRNAmarkup protocol (Brendel V, unpublished data; code available from http://brendelgroup.org/bioinformatics2go/mRNAmarkup.php, last accessed June 18, 2013) mRNAmarkup splits potential chimeric assemblies and discards likely contaminants identified as sequences with strong similarities to UniVec (http://www.ncbi.nlm.nih.gov/VecScreen/VecScreen.html, last accessed June 18, 2013) entries or *Escherichia coli* genomic sequences. We identified potential full-length mRNAs within the TSA using MuSeqBox (Xing and Brendel 2001) as distributed with mRNAmarkup (parameter option −F 5 5 10 10 90 60). The final TSA was annotated with the most significant BLASTp hit using stringent criteria (MuSeqBox option −A 45 10 75). With this final cleaned-up and annotated TSA assembly, we assessed quality and completeness using the CEGMA (Version 2.4.010312) method and identified putative homologs to NCBI NR databases using BLASTx with an *E*-value threshold of 1e-4. This TSA project has been deposited at DDBJ/EMBL/GenBank under the accession GBGV00000000. The version described in this article is the first version, GBGV01000000.

Read Mapping, Abundance Estimation, and DE

We aligned the raw paired-end reads to the TSA using Bowtie2 (Langmead and Salzberg 2012, Version 2.1.0) and quantified the abundances of the transcripts for each library with eXpress (Roberts and Pachter 2013, Version 1.3.1) (supplementary table S2, Supplementary Material online, includes read alignment results to the *P. metricus* transcriptome by sample). The raw read counts have been deposited in NCBI's Gene Expression Omnibus (Edgar et al. 2002) and are accessible through GEO Series accession number GSE61960 (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE61960, last accessed October 2, 2014). We used the R (R Core Team 2013, Version 3.0.1) statistical packages DESeq (Anders and Huber 2010, Version 1.12.0) and EdgeR (Robinson et al. 2010, Version 3.2.3) from the Bioconductor repository (Gentleman et al. 2004) to test for DE (FDR $\leq$ 0.05, Benjamini and Hochberg [1995]) between castes. Dispersion was computed per-condition with DESeq. For EdgeR, a common dispersion was calculated across all samples, then it was adjusted for each condition. DESeq produced more robust results than EdgeR; thus subsequent analyses utilized the results from DESeq (supplementary files, Supplementary Material online—Pmet-tsa-r1.1_Caste.xlsx, contains output from DESeq including fold changes for all *P. metricus* transcripts and a list of *P. metricus* caste-related DETs).

Methods for Comparative Analyses

Choice of Studies for Comparative Analysis

After an extensive review of the literature (including related studies reviewed in Smith et al. [2008]), we focused our comparative analysis on studies that were the most directly comparable to the *P. metricus* data set by the following criteria: 1) the data sets compared late-stage subadult queens and workers (fifth-instar larvae or pupae), 2) the data sets were of sufficiently large scale, that is, they used a genome-wide approach such as microarrays or RNA-seq, and 3) they were the most recent studies on this topic. This led to the choice of Chen et al. (2012), an RNA-seq study on queen and worker honey bee (*A. mellifera*) fifth-instar larvae (as opposed to an earlier study that used microarrays, Barchuk et al. 2007), and Ometto et al. (2011), a microarray study on queen and worker fire ant (*S. invicta*) pupae. Numerous other studies have compared gene expression in queen and worker social insects (reviewed in Smith et al. [2008], see also Weil et al. [2009], Cameron et al. [2013], Feldmeyer et al. [2014]), but because of their limited scale (e.g., few differentially expressed genes were assayed or identified) or focus on completely different life stages and tissues, we deemed them suitable for comparative analysis with our data set.

Putative Orthologs

We defined putative orthologous sequences across social insect species as the best BLAST hit (*E* value $\leq$ 1e-4) between pairs of species. For this analysis, we used the *A. mellifera* Official Gene Set (OGS) v1.1 transcriptomic sequences and the *S. invicta* OGS v2.2.3 transcriptomic sequences (supplementary table S3, Supplementary Material online, holds the number of putative homologs between pairs of species). Note that only the hits to the transcriptomic sequences on the *S. invicta* microarray were considered for this analysis.

Transcript DE Analysis

For the transcript DE analysis, we used previously published lists of DE loci between *A. mellifera* fifth-instar queen- and worker-destined larvae (RNA-Seq, single library per caste, $P < 0.01$, Chen et al. [2012]) and DE oligos between *S. invicta* queen- and worker-destined pupa (microarray, pools from 6 colonies per caste, FDR $\leq$ 0.05, http://www.ncbi.nlm.nih.gov/geo/geo2r/?acc=GSE35217, last accessed January 9, 2013, Ometto et al. 2011) (supplementary table S4, Supplementary Material online, lists the number of putative homologous DETs between pairs of species). Of the 4,787 DE loci between *A. mellifera* fifth-instar queen- and worker-destined larvae, only 4,363 loci corresponded to *A. mellifera* transcriptomic sequences. The 2,415 DE oligos between *S. invicta* castes corresponded to 854 *S. invicta* transcripts. Table 3 provides information about the transcriptome-wide analyses on caste bias for the social insect species: *A. mellifera*, *P. metricus*, and *S. invicta*. We performed Fisher's exact tests between pairs of species to determine whether there was an over- or underrepresentation of DETs compared with the background, that is, all shared DETs between all orthologous transcripts (fig. 3; supplementary table S5, Supplementary Material online, summarizes the directionality of DETs between pairs of species; supplementary file, Supplementary Material online—SpeciesComparison.xlsx, contains a conversion table of DETs and a list of overlapping DETs between all three social insect species).

KEGG Pathway Analysis

For each species, we used BLAST2GO (Conesa et al. 2005) to annotate enzyme codes and metabolic (KEGG) pathways (Kanehisa and Goto 2000; Kanehisa et al. 2014) for each transcriptomic sequence. We developed a novel method to focus on the most informative DET containing pathways. This was necessary because of the large number of *A. mellifera* DETs (likely including false positives), resulting in all *A. mellifera* KEGG pathways including at least one DET. With our novel method, we aimed to restrict the number of KEGG pathways to the most informative subset of pathways for which there was more evidence (multiple DETs). If an enzyme was represented by several transcripts, but only one was differentially expressed, it was given less weight (as the evidence for a change in the pathway was less well supported). Supplementary table S7, Supplementary Material online, shows that as intended, mainly poorly supported pathways with only a single DET were removed from the analysis.

To account for multiple transcript mappings to each enzyme, we calculated the proportion of DETs per enzyme. We considered the KEGG pathway to be "modulated" by the DETs if the sum of the proportions of DETs overall all enzymes was greater than a threshold. We investigated modulation thresholds ranging from 0.1 to 3 incremented by 0.1 (supplementary fig. S3, Supplementary Material online, is a plot of the number of KEGG pathways above the modulation threshold). All three species had a large decrease in the number of modulated KEGG pathways from thresholds 1 to 1.1. For more robust results, we set the threshold for modulated KEGG pathways to 1 (see supplementary fig. S4, Supplementary Material online, for a Venn diagram of the number of modulated KEGG pathways above threshold 1.1 between all three species). We performed Fisher's exact tests between pairs of species to determine whether there was statistically significant number of common modulated KEGG pathways (Fisher's exact test results summarized in fig. 3; see supplementary file, Supplementary Material online—SpeciesComparison.xlsx, for *A. mellifera*, *P. metricus*, and *S. invicta* KEGG pathway annotations, modulated KEGG pathways by species, and overlapping modulated KEGG pathways).

GO Enrichment Analysis

We used BLAST2GO (Conesa et al. 2005) for functional annotation of the transcriptomic sequences for each species based on best BLAST hits (*E* value $\leq$ 1e-3) to *D. melanogaster* sequences in NCBI Entrez Protein database. For each species, BLAST2GO was used to assess enrichment of GO terms (FDR $\leq$ 0.05; two tailed) between caste-biased transcripts and the remaining transcriptome (supplementary files, Supplementary Material online—Pmet-tsa-r1.1_Annotation.xlsx and Pmet-tsa-r1.1_Caste.xlsx, include *P. metricus* GO annotations and Enriched GO terms; supplementary file, Supplementary Material online—SpeciesComparison.xlsx, contains *A. mellifera* and *S. invicta* GO annotations, *A. mellifera*- and *S. invicta*-enriched GO terms, and list of overlapping enriched GO terms between *A. mellifera*, *P. metricus*, and *S. invicta*). We performed Fisher's exact tests between pairs of species to determine whether there was statistically significant number of shared GO-enriched terms (Fisher's exact test results summarized in fig. 3).

Polistes-*Specific Transcripts*

Polistes metricus novel transcripts were identified as transcripts without a BLAST hit (*E* value $>$ 1e-4) to the NCBI NR database. *Polistes* taxonomically restricted transcripts were defined as the best BLAST hit (*E* value $\leq$ 1e-4) between *P. canadensis* novel Isotig sequences (Ferreira et al. 2013, http://genome.crg.es/~pferreira/pcandata/pcan.htm, last accessed April 19, 2013) and *P. metricus* novel transcripts. We reanalyzed the raw data from Ferreira et al. (2013) using the same pipeline (Bowtie2-Express-DESeq) used for *P. metricus* to identify DETs between *P. canadensis* queen and worker adults. We used a Fisher's exact test to determine whether there was an overrepresentation of DETs within the taxonomically restricted transcripts.

Supplementary Material

Supplementary files, tables S1–S7, and figures S1–S5 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

The authors thank Amy Geffre and Cecile Mercado for help with RNA extractions, and Daniel Standage and Volker Brendel for their assistance with the mRNA markup protocol. They thank members of the Toth laboratory, Christina Grozinger, and Gene Robinson for their insightful comments during the preparation of this manuscript. This work was supported by the National Science Foundation IOS 1146410.

References

Abouheif E, Wray GA. 2002. Evolution of the gene network underlying wing polyphenism in ants. *Science* 297(5579):249–252.

Amdam GV, Norberg K, Fondrk MK, Page RE. 2004. Reproductive ground plan may mediate colony-level selection effects on individual foraging behavior in honey bees. *Proc Natl Acad Sci U S A.* 101(31):11350–11355.

Amdam GV, Page RE. 2010. The developmental genetics and physiology of honeybee societies. *Anim Behav* 79(5):973–980.

Anders S, Huber W. 2010. Differential expression analysis for sequence count data. *Genome Biol.* 11(10):R106.

Anderson KE, Linksvayer TA, Smith CR. 2008. The causes and consequences of genetic caste determination in ants (Hymenoptera: Formicidae). *Myrmecol News* 11:119–132.

Arendt J, Reznick D. 2008. Convergence and parallelism reconsidered: what have we learned about the genetics of adaptation? *Trends Ecol Evol.* 23(1):26–32.

Auer PL, Doerge RW. 2010. Statistical design and analysis of RNA sequencing data. *Genetics* 185(2):405–416.

Barchuk AR, Cristino AS, Kucharski R, Costa LF, Simões ZL, Maleszka R. 2007. Molecular determinants of caste differentiation in the highly eusocial honeybee *Apis mellifera*. *BMC Dev Biol.* 7(1):70.

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B (Methodol)* 57(1):289–300.

Brady SG, Sipes S, Pearson A, Danforth BN. 2006. Recent and simultaneous origins of eusociality in halictid bees. *Proc R Soc Lond Ser B* 273(1594):1643–1649.

Cameron RC, Duncan EJ, Dearden PK. 2013. Biased gene expression in early honeybee larval development. *BMC Genomics* 14(1):903.
Cameron SA, Mardulyn P. 2001. Multiple molecular data sets suggest independent origins of highly eusocial behavior in bees (Hymenoptera: Apinae). *Syst Biol.* 50(2):194–214.
Cardinal S, Straka J, Danforth BN. 2010. Comprehensive phylogeny of apid bees reveals the evolutionary origins and antiquity of cleptoparasitism. *Proc Natl Acad Sci U S A.* 107(37):16207–16211.
Chen X, Hu Y, Zheng H, Cao L, Niu D, Yu D, Sun Y, Hu S, Hu F. 2012. Transcriptome comparison between honey bee queen-and worker-destined larvae. *Insect Biochem Mol Biol.* 42(9):665–673.
Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21(18):3674–3676.
Corona M, Estrada E, Zurita M. 1999. Differential expression of mitochondrial genes between queens and workers during caste determination in the honeybee *Apis mellifera*. *J Exp Biol.* 202(8):929–938.
Cotoneschi C, Scognamiglio F, Scala C, Cervo R, Strassmann JE, Turillazzi S. 2007. Larval sex identification in the paper wasp *Polistes dominulus* (Vespidae, Hymenoptera). *Insectes Soc.* 54(2):132–135.
Cox MP, Peterson DA, Biggs PJ. 2010. SolexaQA: at-a-glance quality assessment of Illumina second-generation sequencing data. *BMC Bioinformatics* 11(1):485.
Daugherty THF, Toth AL, Robinson GE. 2011. Nutrition and division of labor: effects on foraging and brain gene expression in the paper wasp *Polistes metricus*. *Mol Ecol.* 20(24):5337–5347.
Ding Y, Zhou Q, Wang W. 2012. Origins of new genes and evolution of their novel functions. *Annu Rev Ecol Evol Syst.* 43:345–363.
Edgar R, Domrachev M, Lash AE. 2002. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* 30(1):207–210.
Elmer KR, Fan S, Gunter HM, Jones JC, Boekhoff S, Kuraku S, Meyer A. 2010. Rapid evolution and selection inferred from the transcriptomes of sympatric crater lake cichlid fishes. *Mol Ecol.* 19(s1):197–211.
Evans JD, Wheeler DE. 1999. Differential gene expression between developing queens and workers in the honey bee, *Apis mellifera*. *Proc Natl Acad Sci U S A.* 96(10):5575–5580.
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(1):151–161.
Ferreira PG, Patalano S, Chauhan R, Ffrench-Constant R, Gabaldón T, Guigó R, Sumner S. 2013. Transcriptome analyses of primitively eusocial wasps reveal novel insights into the evolution of sociality and the origin of alternative phenotypes. *Genome Biol.* 14(2):R20.
Gene Ontology Consortium. 2000. Gene ontology: tool for the unification of biology. *Nat Genet.* 25(1):25–29.
Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, et al. 2004. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* 5(10):R80.
Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, et al. 2011. Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nat Biotechnol.* 29(7):644–652.
Grozinger CM, Fan Y, Hoover SE, Winston ML. 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(22):4837–4848.
Harpur BA, Kent CF, Molodtsova D, Lebon JMD, Alqarni AS, Owayss AA, Zayed A. 2014. Population genomics of the honey bee reveals strong signatures of positive selection on worker traits. *Proc Natl Acad Sci U S A.* 111(7):2614–2619.
Haydak MH. 1970. Honey bee nutrition. *Annu Rev Entomol.* 15(1):143–156.
Hepperle C, Hartfelder K. 2001. Differentially expressed regulatory genes in honey bee caste development. *Naturwissenschaften* 88(3):113–116.
Hines HM, Hunt JH, O'Connor TK, Gillespie JJ, Cameron SA. 2007. Multigene phylogeny reveals eusociality evolved twice in vespid wasps. *Proc Natl Acad Sci U S A.* 104(9):3295–3299.
Hunt JH. 2007. The evolution of social wasps. New York: Oxford University Press. p. 144–156.
Hunt JH, Amdam GV. 2005. Bivoltinism as an antecedent to eusociality in the paper wasp genus *Polistes*. *Science* 308(5719):264–267.
Hunt JH, Buck NA, Wheeler DE. 2003. Storage proteins in vespid wasps: characterization, developmental pattern, and occurrence in adults. *J Insect Physiol.* 49(8):785–794.
Hunt JH, Dove MA. 2002. Nourishment affects colony demographics in the paper wasp *Polistes metricus*. *Ecol Entomol.* 27(4):467–474.
Hunt JH, Kensinger BJ, Kossuth JA, Henshaw MT, Norberg K, Wolschin F, Amdam GV. 2007. A diapause pathway underlies the gyne phenotype in *Polistes* wasps, revealing an evolutionary route to caste-containing insect societies. *Proc Natl Acad Sci U S A.* 104(35):4020–14025.
Hunt JH, Mutti NS, Havukainen H, Henshaw MT, Amdam GV. 2011. Development of an RNA interference tool, characterization of its target, and an ecological test of caste differentiation in the eusocial wasp *Polistes*. *PLoS One* 6(11):e26641.
Hunt JH, Wolschin F, Henshaw MT, Newman TC, Toth AL, Amdam GV. 2010. Differential gene expression and protein abundance evince ontogenetic bias toward castes in a primitively eusocial wasp. *PLoS One* 5(5):e10674.
Jacob F. 1977. Evolution and tinkering. *Science* 196(4295):1161–1166.
Johnson BR, Borowiec ML, Chiu JC, Lee EK, Atallah J, Ward PS. 2013. Phylogenomics resolves evolutionary relationships among ants, bees, and wasps. *Curr Biol.* 23(20):2058–2062.
Johnson BR, Tsutsui ND. 2011. Taxonomically restricted genes are associated with the evolution of sociality in the honey bee. *BMC Genomics* 12(1):164.
Kanehisa M, Goto S. 2000. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28(1):27–30.
Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M. 2014. Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res.* 42(D1):D199–D205.
Karsai I, Hunt JH. 2002. Food quantity affect traits of offspring in the paper wasp *Polistes metricus* (Hymenoptera: Vespidae). *Environ Entomol.* 31(1):99–106.
Khalturin K, Hemmrich G, Fraune S, Augustin R, Bosch TC. 2009. More than just orphans: are taxonomically-restricted genes important in evolution? *Trends Genet.* 25(9):404–413.
Kronforst MR, Barsh GS, Kopp A, Mallet J, Monteiro A, Mullen SP, Protas M, Rosenblum EB, Schneide CJ, Hoekstra HE. 2012. Unraveling the thread of nature's tapestry: the genetics of diversity and convergence in animal pigmentation. *Pigment Cell Melanoma Res.* 25(4):411–433.
Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods.* 9(4):357–359.
Linksvayer TA, Wade MJ, Gordon DM. 2006. Genetic caste determination in harvester ants: possible origin and maintenance by cytonuclear epistasis. *Ecology* 87(9):2185–2193.
Manfredini F, Riba-Grognuz O, Wurm Y, Keller L, Shoemaker D, Grozinger CM. 2013. Sociogenomics of cooperation and conflict during colony founding in the fire ant *Solenopsis invicta*. *PLoS Genet.* 9(8):e1003633.
Ogura A, Ikeo K, Gojobori T. 2004. Comparative analysis of gene expression for convergent evolution of camera eye between octopus and human. *Genome Res.* 14(8):1555–1561.
Ometto L, Shoemaker D, Ross KG, Keller L. 2011. Evolution of gene expression in fire ants: the effects of developmental stage, caste, and species. *Mol Biol Evol.* 28(4):1381–1392.
Oster GF, Wilson EO. 1978. Caste and ecology in the social insects. Princeton (NJ): Princeton University Press.
Page RE, Amdam GV. 2007. The making of a social insect: developmental architectures of social design. *Bioessays* 29(4):334–343.
Parker J, Tsagkogeorga G, Cotton JA, Liu Y, Provero P, Stupka E, Rossiter SJ. 2013. Genome-wide signatures of convergent evolution in echolocating mammals. *Nature* 502(7470):228–231.

Parra G, Bradnam K, Korf I. 2007. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* 23(9):1061–1067.

Patel A, Fondrk MK, Kaftanoglu O, Emore C, Hunt G, Frederick K, Amdam GV. 2007. The making of a queen: TOR pathway is a key player in diphenic caste development. *PLoS One* 2(6):e509.

Patel NG, Haydak MH, Gochnauer TA. 1960. Electrophoretic components of the proteins in honeybee larval food. *Nature* 186(4725):633–634.

Petralia RS, Vinson SB. 1979. Developmental morphology of larvae and eggs of the imported fire ant, *Solenopsis invicta*. *Ann Entomol Soc Am.* 72(4):472–484.

Pilgrim E., von Dohlen C, Pitts J. 2008. Phylogenetics of the stinging wasps (Hymenoptera: Vespoidea). *Zool Scr.* 37:539–560.

R Core Team. 2013. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna (Austria). [cited 2014 Nov 21]. Available from: http://www.R- project.org.

Reeve HK. 1991. The social biology of wasps. In: Ross KG, Matthews RW, editors. Polistes. Ithaca (NY): Cornell University Press. p. 99–148.

Roberts A, Pachter L. 2013. Streaming fragment assignment for real-time analysis of sequencing experiments. *Nat Methods.* 10(1):71–73.

Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26(1):139–140.

Ross KG, Keller L. 1995. Ecology and evolution of social organization: insights from fire ants and other highly eusocial insects. *Annu Rev Ecol Syst.* 26:631–656.

Schwarz MP, Tierney SM, Rehan SM, Chenoweth LB, Cooper SJ. 2011. The evolution of eusociality in allodapine bees: workers began by waiting. *Biol Lett.* 7(2):277–280.

Scotland RW. 2011. What is parallelism? *Evol Dev.* 13(2):214–227.

Simola DF, Wissler L, Donahue G, Waterhouse RM, Helmkampf M, Roux J, Nygaard S, Glastad K, Hagen DE, Viljakainen L, et al. 2013. Social insect genomes exhibit dramatic evolution in gene composition and regulation while preserving regulatory features linked to sociality. *Genome Res.* 23(8):1235–1247.

Smith CR, Toth AL, Suarez AV, Robinson GE. 2008. Genetic and genomic analyses of the division of labour in insect societies. *Nat Rev Genet.* 9(10):735–748.

Stern DL. 2013. The genetic causes of convergent evolution. *Nat Rev Genet.* 14(11):751–764.

Toth AL, Robinson GE. 2007. Evo-devo and the evolution of social behavior. *Trends Genet.* 23(7):334–341.

Toth AL, Tooker JF, Radhakrishnan S, Minard R, Henshaw MT, Grozinger CM. 2014. Shared genes related to aggression, rather than chemical communication, are associated with reproductive dominance in paper wasps (*Polistes metricus*). *BMC Genomics* 15:75.

Toth AL, Varala K, Henshaw MT, Rodriguez-Zas SL, Hudson ME, Robinson GE. 2010. Brain transcriptomic analysis in paper wasps identifies genes associated with behaviour across social insect lineages. *Proc R Soc Lond Ser B.* 277(1691):2139–2148.

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

Weil T, Korb J, Rehli M. 2009. Comparison of queen-specific gene expression in related lower termite species. *Mol Biol Evol.* 26(8):1841–1850.

West-Eberhard MJ. 1987. Flexible strategy and social evolution. In: Itô Y, Brown JL, Kikkawa J, editors. Animal societies: theories and fact.. Tokyo (Japan): Japan Scientific Society Press. p. 35–51.

West-Eberhard MJ. 1996. Wasp societies as microcosms for the study of development and evolution. In: West-Eberhard MJ, Turillazzi S, editors. Natural history and evolution of paper wasps.. Oxford: Oxford University Press. p. 290–317.

Wheeler DE. 1986. Developmental and physiological determinants of caste in social Hymenoptera: evolutionary implications. *Am Nat.* 128(1):13–34.

Wheeler DE, Buck N, Evans JD. 2006. Expression of insulin pathway genes during the period of caste determination in the honey bee, *Apis mellifera*. *Insect Mol Biol.* 15(5):597–602.

Wilson EO. 1971. The insect societies.. Cambridge (MA): Belknap Press of Harvard University Press.

Woodard SH, Fischman BJ, Venkat A, Hudson ME, Varala K, Cameron SA, Clark AG, Robinson GE. 2011. Genes involved in convergent evolution of eusociality in bees. *Proc Natl Acad Sci U S A.* 108(18):7472–7477.

Xing L, Brendel V. 2001. Multi-query sequence BLAST output examination with MuSeqBox. *Bioinformatics* 17(8):744–745.