



Transcriptome analyses of primitively eusocial wasps reveal novel insights into the evolution of sociality and the origin of alternative phenotypes

Ferreira *et al.*

RESEARCH

Open Access

# Transcriptome analyses of primitively eusocial wasps reveal novel insights into the evolution of sociality and the origin of alternative phenotypes

Pedro G Ferreira<sup>1,6</sup>, Solenn Patalano<sup>2,3</sup>, Ritika Chauhan<sup>4</sup>, Richard Ffrench-Constant<sup>4</sup>, Toni Gabaldón<sup>1</sup>, Roderic Guigó<sup>1</sup> and Seirian Sumner<sup>2,5\*</sup>

## Abstract

**Background:** Understanding how alternative phenotypes arise from the same genome is a major challenge in modern biology. Eusociality in insects requires the evolution of two alternative phenotypes - workers, who sacrifice personal reproduction, and queens, who realize that reproduction. Extensive work on honeybees and ants has revealed the molecular basis of derived queen and worker phenotypes in highly eusocial lineages, but we lack equivalent deep-level analyses of wasps and of primitively eusocial species, the latter of which can reveal how phenotypic decoupling first occurs in the early stages of eusocial evolution.

**Results:** We sequenced 20 Gbp of transcriptomes derived from brains of different behavioral castes of the primitively eusocial tropical paper wasp *Polistes canadensis*. Surprisingly, 75% of the 2,442 genes differentially expressed between phenotypes were novel, having no significant homology with described sequences. Moreover, 90% of these novel genes were significantly upregulated in workers relative to queens. Differential expression of novel genes in the early stages of sociality may be important in facilitating the evolution of worker behavioral complexity in eusocial evolution. We also found surprisingly low correlation in the identity and direction of expression of differentially expressed genes across similar phenotypes in different social lineages, supporting the idea that social evolution in different lineages requires substantial *de novo* rewiring of molecular pathways.

**Conclusions:** These genomic resources for aculeate wasps and first transcriptome-wide insights into the origin of castes bring us closer to a more general understanding of eusocial evolution and how phenotypic diversity arises from the same genome.

## Background

Phenotypic plasticity is a fundamental biological process that allows organisms to adapt to changes in their environment [1,2]. Examples of plastic phenotypes include insect castes [3,4], horn-polyphenic beetles [5] and sex differences [6], where they play a crucial role in shaping the ecology and evolution of species and ecosystems. Understanding how alternative phenotypes arise from the same genome is one of the most challenging questions in modern biology [1,2].

The most impressive examples of phenotypic plasticity are found in the eusocial Hymenoptera (ants, some bees

and some aculeate wasps), where specialized reproductive or non-reproductive phenotypes (castes) play a pivotal role in the ecological success of these insects [7]. Eusocial insect castes usually arise through differential expression of shared genes [8,9], but studies to date have been largely restricted to the highly eusocial bees and ants, where caste phenotypes are evolutionarily derived, being greatly modified from their ancestral state (for example, with morphological adaptations that are established early in development). Thus, some of the molecular processes underlying the origins of castes are likely to differ substantially from those underlying highly eusocial species [10,11]. Understanding how castes first arise is best studied in primitively eusocial species, but little is known about their molecular evolution, particularly in the aculeate wasps [12-15]. Thus, there is a need

\* Correspondence: Seirian.Sumner@Bristol.ac.uk

<sup>2</sup>Institute of Zoology, Zoological Society of London, Regent's Park, NW1 4RY, UK

Full list of author information is available at the end of the article

for comprehensive gene-level studies on primitively eusocial wasps.

In the evolution of eusociality, behavioral and physiological traits of the non-social ancestor become decoupled into complementary queen and worker phenotypes [16], but we understand little about the molecular patterns and processes by which this originates and to what extent molecular ground plans are re-organized. These questions are of general biological importance in understanding the evolution of phenotypic diversity [1,2]. Empirical assessment of how gene transcription is decoupled into alternative phenotypes has been difficult to obtain because it requires unbiased quantification of gene expression (of both known and novel (undescribed) genes) across whole genomes [17]. Microarray analyses of the highly eusocial honeybee (*Apis mellifera*) and fire ant (*Solenopsis invicta*) suggest marked functional decoupling of molecular processes in queens and workers, involving large numbers of genes [18-21]. Little is known about the patterns and processes of transcriptional decoupling for castes in primitively eusocial insects [12,14]. A simple prediction is that patterns of transcriptional and functional decoupling should be less pronounced than in highly eusocial species, because phenotypes are in the first stages of merely behavioral decoupling [9,12].

Processes of phenotypic decoupling at the molecular level may occur via several mechanisms [5,22]. Conserved toolkits of ancestral molecular processes may be redeployed to regulate similar alternative phenotypes across species and lineages [23-26]. In eusocial insects, conserved toolkit genes appear to be differentially expressed in castes across species, including bees and wasps [12,27]. An alternative, untested prediction is that at least some of the molecular processes underlying social phenotypes, in different social lineages and among different levels of social complexity, will differ substantially. This is because extensive molecular and developmental re-wiring or new gene evolution may be required for the loss of phenotypic plasticity and elaboration of phenotype-specific traits [10,28]. There may also be rapid evolution of genes associated with the emergence of alternative phenotypes [6,17,29], and there is evidence for this in the highly eusocial honeybees and fire ants [30,31]. Finally, the importance of novel genes (that is, previously undescribed genes that lack any detectable protein-coding homologues from existing sequence data [32]) in the evolution of phenotypic innovations has recently emerged in animals as varied as yeast [33], hydra, reptiles [34-36], and also honeybees [37]. To date the eusocial insect literature has largely focused on conserved molecular pathways for social behavior [12,26,27,38,39], and the relative roles of the

different mechanisms of phenotypic decoupling are unknown.

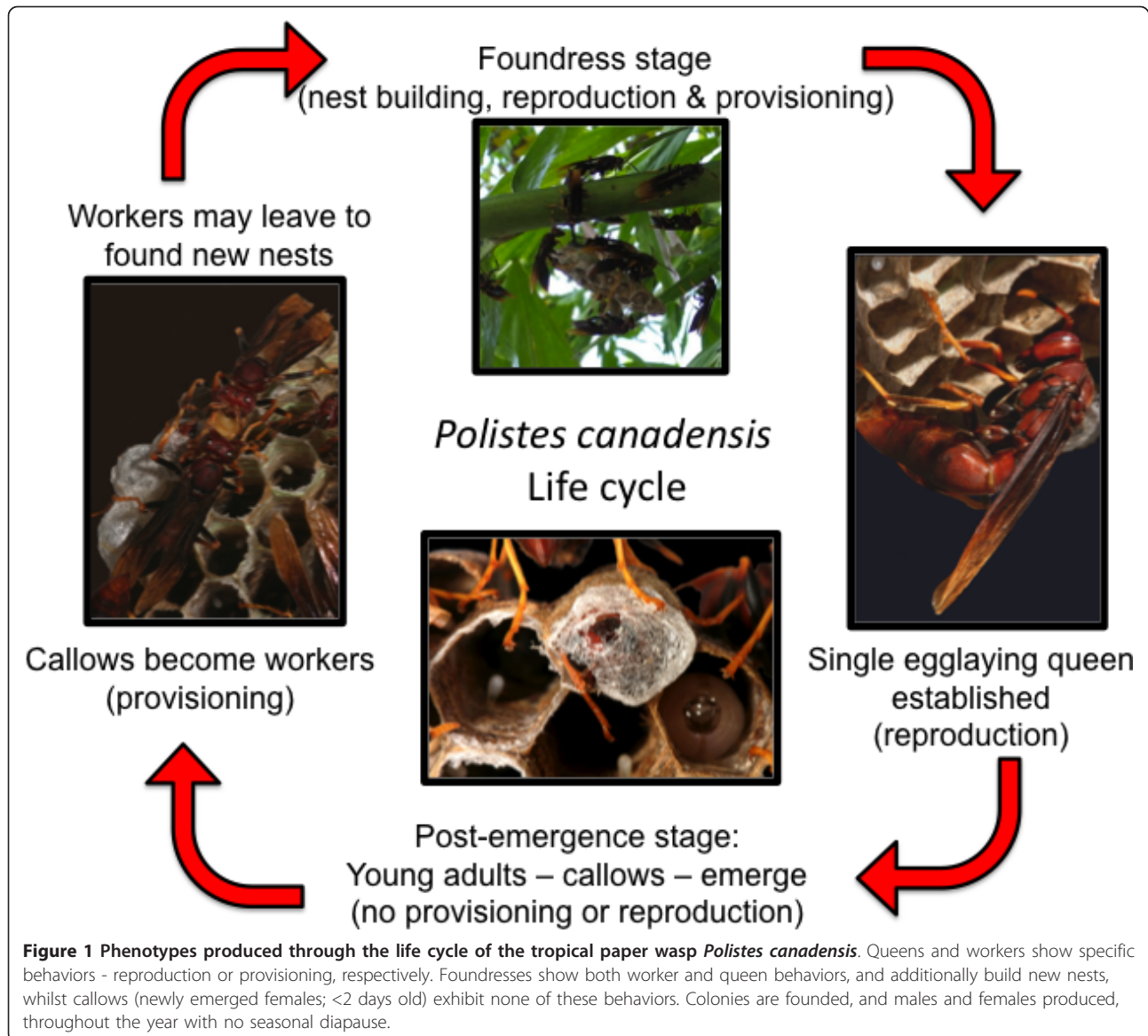
We sequenced over 20 Gbp of genome-wide caste-specific transcripts from the primitively eusocial, tropical paper wasp *Polistes canadensis* using high-throughput sequencing technologies. Castes in *Polistes* differ only in behavior and are determined through differential gene expression during adulthood [12-14], making *Polistes* important models for studying the early stage in the evolution of eusociality and phenotypic plasticity [40]. *Polistes* originated in the tropics, and the secondary adaptations to diapause found in temperate species (for example, [41]) will be absent in *P. canadensis* [42]. Typically a few foundresses build and provision new colonies (Figure 1). One foundress becomes the primary egg-layer (queen). Emerging adults help raise the offspring of their mother queen, but retain the ability to become egg-layers [43-45]. Thus, *Polistes* castes represent an early stage of phenotypic decoupling in the evolution of eusociality [16].

Here, we generate a genome-wide catalogue of aculeate wasp genes expressed in adult females to provide a resource for genomic analyses. We then conduct the first RNA-seq analyses of caste-biased expression in a wasp and primitively eusocial insect to test the above hypotheses on the patterns and processes of molecular decoupling of alternative phenotypes at the early stages in the evolution of eusocial behavior. These datasets allow a first simultaneous assessment of the role of conserved genetic toolkits, novel genes and gene evolution when social behavior first evolves. We also use our data to reassess the phylogenetic relationships between the three aculeate Hymenoptera subfamilies (bees, wasps and ants), which is vital information for interpreting common processes and lineage-specific novelties underlying the evolution of alternative phenotypes. Recently, there has been much debate over the relationships of aculeate bees, wasps and ants, challenging the morphology-based view that Vespoidea (ants plus aculeate wasps) are monophyletic [46,47]. Genome sequences (or similar catalogue of genes) for aculeate wasps will help resolve this controversy and provide a basis for interpreting comparative data on social evolution.

## Results and discussion

### Gene assembly for an aculeate wasp

We first generated a reference assembly of the genes expressed by *P. canadensis* adult female phenotypes (queens, workers, foundresses and callows) (Figure 1; Section 1 in Additional file 1) by sequencing normalized 454 libraries from phenotype-specific pools (37 individuals in total; Section 2 in Additional file 1). Using transcriptome rather than genome sequencing data to provide a gene set is a powerful and accessible approach for initiating



genomic analyses on non-model organisms. Library normalization meant that we were able to capture transcripts at all levels of the expression spectrum. The longer read lengths provided by 454 technology and the high coverage of these datasets increase the likelihood of full transcript assembly. The 454 reads were assembled into a reference transcriptome set of 26,284 isogroups - henceforth referred to as 'genes' for simplicity (Section 2 in Additional file 1; Additional file 2). This gene set included 98% of the core eukaryotic genes with equal coverage for each phenotype (Section 3 in Additional file 1). Additionally, sequences were 100% identical to existing overlapping Sanger sequences for this species ( $n = 36$  genes) [14]. Genes were equally spread across honeybee chromosomes and included 94% of publicly available ESTs for the

temperate paper wasp *Polistes metricus* ( $n = 422$  genes [12]; Section 3 in Additional file 1). Although this assembly is restricted to the genes expressed in adult females, it provides an important step forward in generating a comprehensive genome-wide sequence resource for aculeate wasps.

We characterized in detail the sequence and expression features for two groups of genes. The first group corresponds to the set of genes with detectable homology to protein sequences of any species in the comprehensive GenBank non-redundant (NR) database. We identified homologs for 37.4% of the total *P. canadensis* genes ( $n = 9,839$ ), which were then annotated using Gene Ontology (GO; Table 1; Section 4 in Additional file 1; Additional files 3 and 4): 71% of best hits were with the honeybee,



**Table 1 Number of best BLAST hits between the genes of sequenced genomes for aculeate hymenopterans**

| Species               | <i>P. canadensis</i> | <i>A. mellifera</i> | <i>S. invicta</i> | <i>C. floridanus</i> | <i>H. saltator</i> | <i>N. vitri-pennis</i> |
|-----------------------|----------------------|---------------------|-------------------|----------------------|--------------------|------------------------|
| <i>P. canadensis</i>  | 22,460 (100%)        | 7,040 (31.3%)       | 5,548 (24.7%)     | 6,501 (28.9%)        | 5,880 (26.2%)      | 4,416 (19.7%)          |
| <i>A. mellifera</i>   | 5,764 (52.1%)        | 11,062 (100%)       | 7,093 (71.4%)     | 8,197 (74.1%)        | 6,769 (61.2%)      | 5,366 (48.5%)          |
| <i>S. invicta</i>     | 4,662 (28.2%)        | 6,293 (38.0%)       | 16,522 (100%)     | 8,864 (53.6%)        | 8,341 (50.5%)      | 4,935 (29.9%)          |
| <i>C. floridanus</i>  | 5,422 (31.8%)        | 7,129 (68.6%)       | 11,741 (68.6%)    | 17,064 (100%)        | 12,492 (73.2%)     | 5,552 (32.5%)          |
| <i>H. saltator</i>    | 4,913 (26.5%)        | 6,798 (36.6%)       | 10,591 (57.0%)    | 9,247 (49.8%)        | 18,564 (100%)      | 5,667 (30.5%)          |
| <i>N. vitripennis</i> | 3,861 (20.5%)        | 5,377 (28.6%)       | 6,144 (32.6%)     | 6,600 (35.0%)        | 5,725 (30.4%)      | 18,822 (100%)          |

Absolute number (and percentage of genes) shared between each pair of species are given. Annotations used for each species are: Official Gene Set (OGS)1.2 for the parasitoid wasp *Nasonia vitripennis*, OGS3.3 for the ant *Harpegnathos saltator*, OGS3.3 for the ant *Camponotus floridanus*, OGS2.2.3 for the ant *Solenopsis invicta*, pre-release of OGS2.0 for the honeybee *Apis mellifera*, and the longest ORFs (putative coding genes only - that is, excluding 3,824 potential long non-coding RNAs) for *Polistes canadensis* from our RNAseq datasets. Rows represent the source species, which are compared with the test species in the columns. For instance, of the 11,062 genes found in *A. mellifera*, 5,764 have significant hits with genes found in *P. canadensis*, corresponding to 52.1% of the total 11,062 genes. *P. canadensis* shares 25 to 31% of genes with the other aculeate hymenopterans, but only 19.7% with the parasitoid *N. vitripennis*. Notably, *P. canadensis* does not share more genes with the ants than the bee, in agreement with our maximum-likelihood phylogenetic analysis (Figure 3a).

which reflects the wealth of genomic resources for this species rather than any close phylogenetic relationship (see below). This level of homology is comparable with data for the temperate wasp *P. metricus* where 39% of the approximately 391K sequence fragments had putative orthologs [13]. In comparison, 60% to 70% of genes in the genomes of other social insects (ants and bees) had homology with sequences in other insects [48,49].

The second group of genes ( $n = 16,445$ ; 62%) lacked detectable homology with protein sequences in NR databases, and will be referred to here as putative novel genes as they differ significantly in sequence from any described sequences to date (Section 5 in Additional file 1). These include taxonomically restricted genes, and they may be derived from ancestral coding genes, or arise *de novo* from noncoding ancestors [32]. Overall, transcripts with homology were longer (with-hits group,  $1,718 \pm 1,343$ ; no-hits group,  $736 \pm 768$ ) and had longer ORFs (with-hits group,  $256 \pm 205$ ; no-hits group,  $83 \pm 36$ ) than those without (Figure 2a). By scanning the sequences for known protein domains with HMMscan [50] we detected domain homology in 48% of the sequences with NR hits and only 9.9% in the group without hits (Figure 2b). GC content was also lower in the no-hits group (mean GC with-hits, 0.337; no-hits, 0.306; Figure 2c). Analysis of protein coding potential with PORTRAIT [51] revealed a significantly lower potential for protein coding in the group of isotigs without hits (Figure 2d). We then examined whether the novel isotigs could potentially be long non-coding RNAs (lncRNAs). First, we looked more closely at the distribution of isotigs that are longer than 300 bp and have an ORF shorter than 90 amino acids (96% of the *A. mellifera* transcripts have an ORF greater than 90 amino acids). The vast majority of these isotigs lacked homology ( $n = 12,751$  with no homology;  $n = 1,162$  with homology), suggesting that the putative novel isotigs had shorter ORFs. Next, we imposed a PORTRAIT score lower than 0.5 (at which a sequence is more likely to be non-coding than coding), which yielded more isotigs in

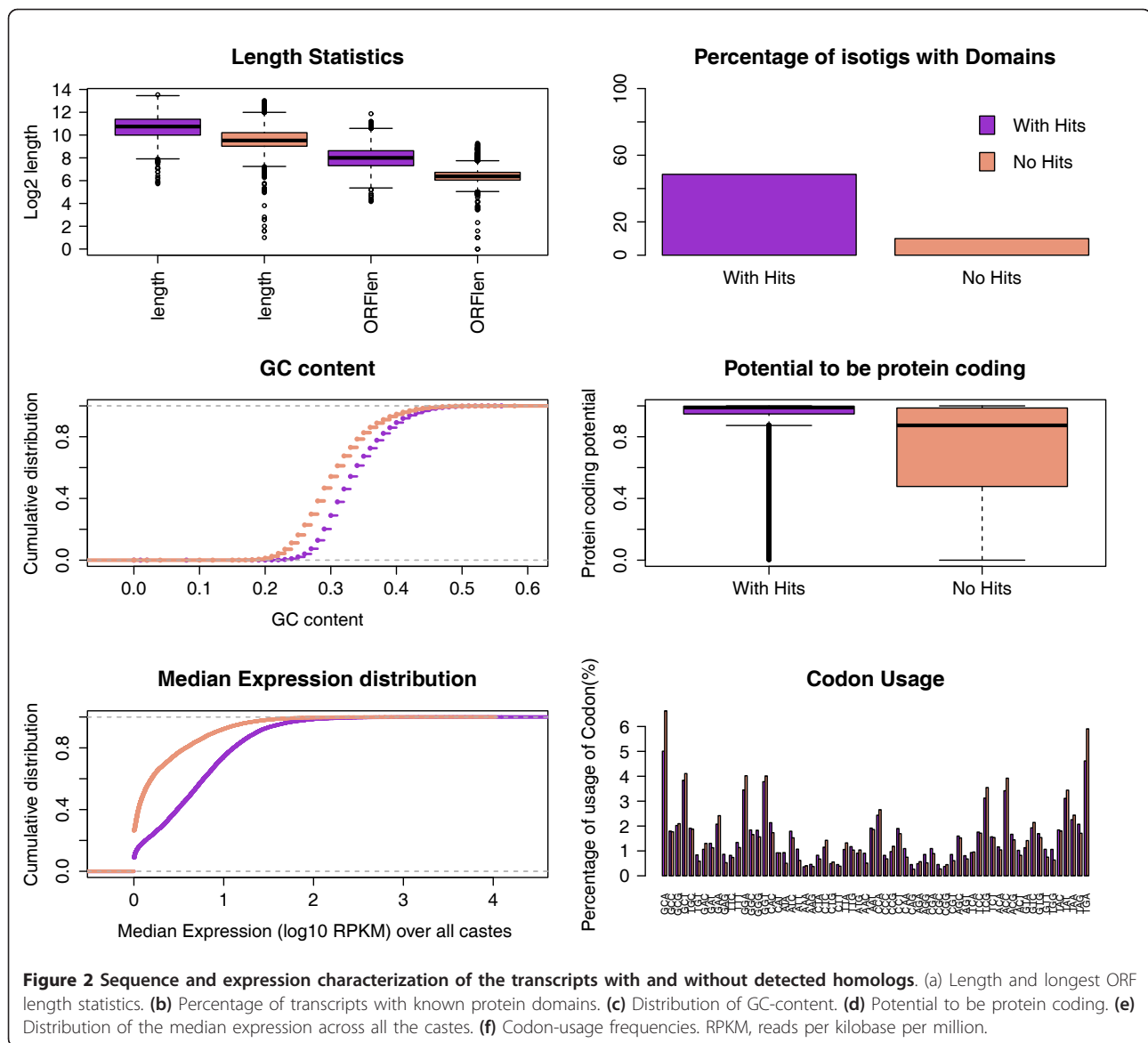
the set lacking homology ( $n = 4,096$ ), than those with ( $n = 314$ ). Finally, we imposed the condition of no homology with known protein domains, which again yielded more isotigs in the set without homology ( $n = 3,824$ ) than with ( $n = 262$ ). These latter two results suggest lower protein coding potential in the putative novel isotigs. The distribution of the median expression across all castes shows that, in general, isotigs with hits have higher expression values than those without (Figure 2e; but see also expression analysis below). The pattern of codon usage was similar in both groups ( $P = 1$ , two sample *t*-test; Figure 2f).

The characteristics of isotigs lacking homology - namely shorter ORFs, lower protein coding potential, lower GC content and lower expression values [52] - suggest that a large proportion of novel isotigs correspond to non-protein coding transcripts. But low sequence conservation across species, lack of genomic information for aculeate wasps and the typical tissue specificity of the non-coding RNAs [52] may also explain the novelty of these genes.

We also verified that the abundance of novel genes was not an artifact resulting from assembly of transcript fragments rather than complete transcript sequences (Section 6 in Additional file 1). Our analyses suggested that less than 10% of genes were affected by this, and thus do not alter significantly our main analyses (below). However, without a genome sequence we cannot exclude the possibility that some genes may represent incomplete assemblies.

#### Paraphyly of Vespoidea

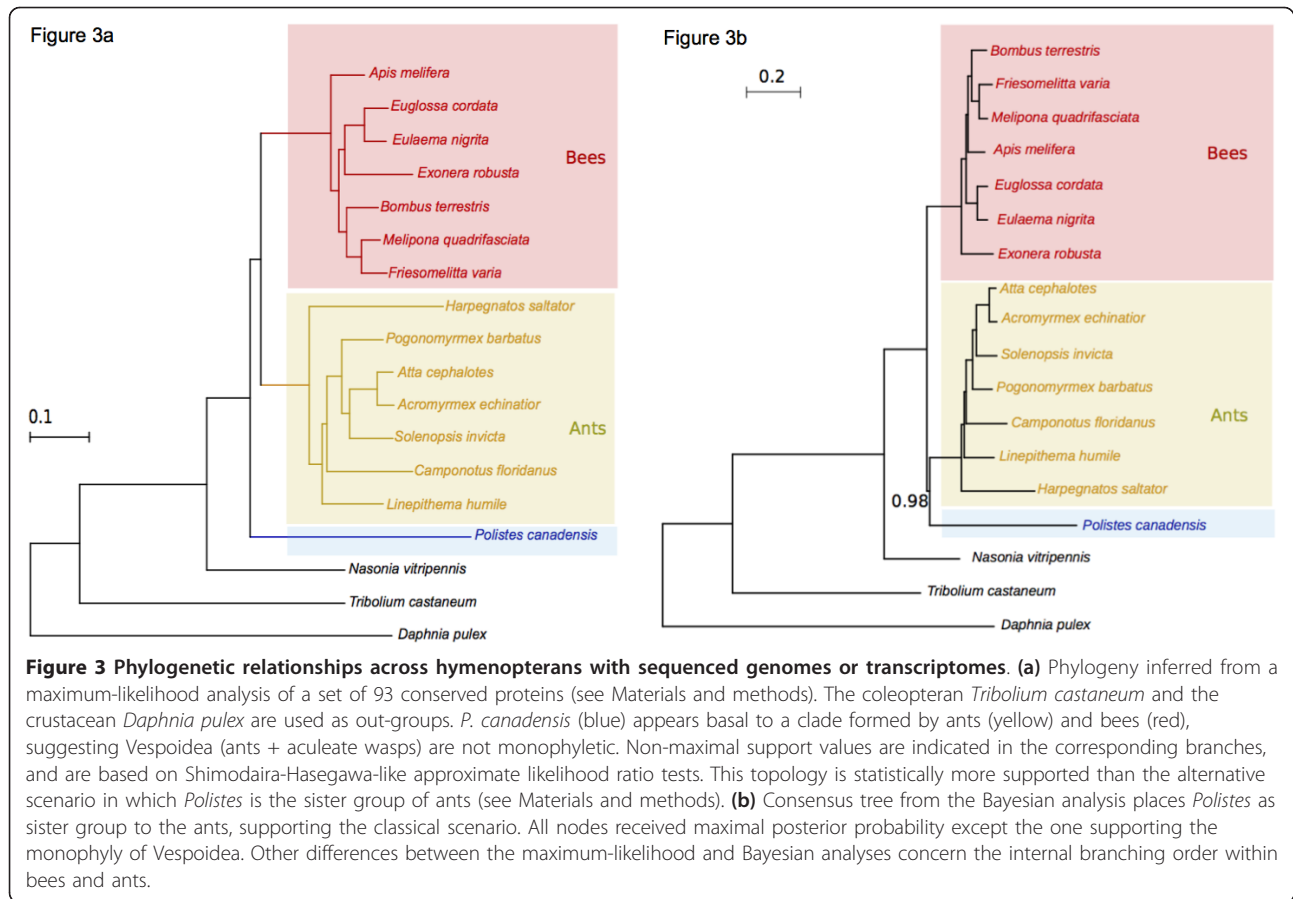
Long-standing analyses based on the fossil record, cladistics and phylogenetic analyses of the aculeate Hymenoptera place ants and aculeate wasps as a monophyletic clade - the Vespoidea - with the bees as a sister group. This has received mixed support from recent molecular analyses [47,53-55], but these studies were limited to a maximum of four genes and sparse data matrices. High-throughput transcriptomics data are a reliable source of



phylogenetic information [56]. Our transcriptome affords a first assessment of the monophyly of Vespoidea using a large and complete data matrix. We analyzed 93 conserved one-to-one orthologs of *P. canadensis* genes, derived from 196 single-copy insect phylogeny markers in the eight fully sequenced hymenopteran genomes [57] and nine bee transcriptomes [38] (Additional file 5). A maximum-likelihood (ML) analysis of these genes across all 32 insect species placed *Polistes* as the basal, sister group to the bees and ants (Figure 3a). This topology was significantly more supported ( $P < 0.05$ , Shimodaira-Hasegawa test) than the earlier proposal that Vespoidea (ants and *Polistes* in our study) are monophyletic. This result is consistent with the levels of gene sharing across aculeates (Table 1), where *P. canadensis* does not share more

genes with ants than bees, as would be expected if wasps and ants were monophyletic.

We further tested the phylogenetic relationships using a Bayesian framework. In contrast, this gave greater support to the classical view of aculeate wasps and ants being monophyletic (Figure 3b; Section 10 in Additional file 1). The reason for the discrepancy in the results of these two methods is likely to be due to the wasp clade being the most poorly represented in our analysis (with only one species), and moreover, it is the clade with the longest branch. Our dataset offers a first chance to examine sub-family relationships across large numbers of genes but, clearly, phylogenomic data on more species of aculeate wasps are required to determine whether the term 'Vespoidea' should be dropped, or reclassified as paraphyletic.



This finding would have important general implications for our understanding of eusociality as it would suggest that bees and ants shared an aculeate wasp-like common ancestor, that ants are wingless wasps [58], and that bees are wasps that lost predacious behaviors. Primitively eusocial wasps may therefore hold some important clues to the early stages of eusociality, if they are the basal descendants of non-aculeate Hymenoptera.

### Transcriptional and functional decoupling at the origin of alternative phenotypes

We next conducted RNA-seq analyses of phenotype-biased expression to provide a first genome-wide look at molecular decoupling of alternative phenotypes at the early stages of eusocial evolution. We quantified genome-wide transcription in individuals ( $n = 2$  to 12 per phenotype) from each of the four adult female phenotypes by sequencing at least 3 Gbp of Illumina short-reads of brain cDNA per phenotype (Additional files 6, 7, 8 and 9). These samples appeared to capture most transcription across the genome since 99.8% of the 26,284 genes in the reference transcriptome were detected in our pooled Illumina dataset (Section 7 in Additional file 1).

We identified genes that were differentially expressed between phenotypes using NOIseq [59]. Of the 26,284 genes, 2,442 (9.3%) were differentially expressed in any one phenotype (Table 2; Section 8 in Additional file 1; Additional file 10). This suggests that a very small part of the adult female transcriptome is decoupled to produce alternative phenotypes in this species. This is similar to that detected in the temperate wasp *P. metricus*, where 12% ( $n = 389$ ) of genes on a microarray were caste-biased [12]. Because RNA-seq is a new methodology, we also confirmed that the number of differentially expressed genes was not artificially elevated due to gene fragmentation (Section 9 in Additional file 1). In a microarray study on whole body samples of the highly eusocial fire ant *Solenopsis invicta*, the proportion of examined genes that were differentially expressed between adult queens and workers (19% of 14,467 genes) [21] was twice that of *Polistes*. RNA-seq analyses on highly eusocial species like the fire ant are required to explore these patterns of transcriptional decoupling further.

Workers appear to upregulate significantly more caste-biased genes ( $n = 2,224$ ; 91% of caste-biased genes) than other phenotypes (callows, foundresses and queens)

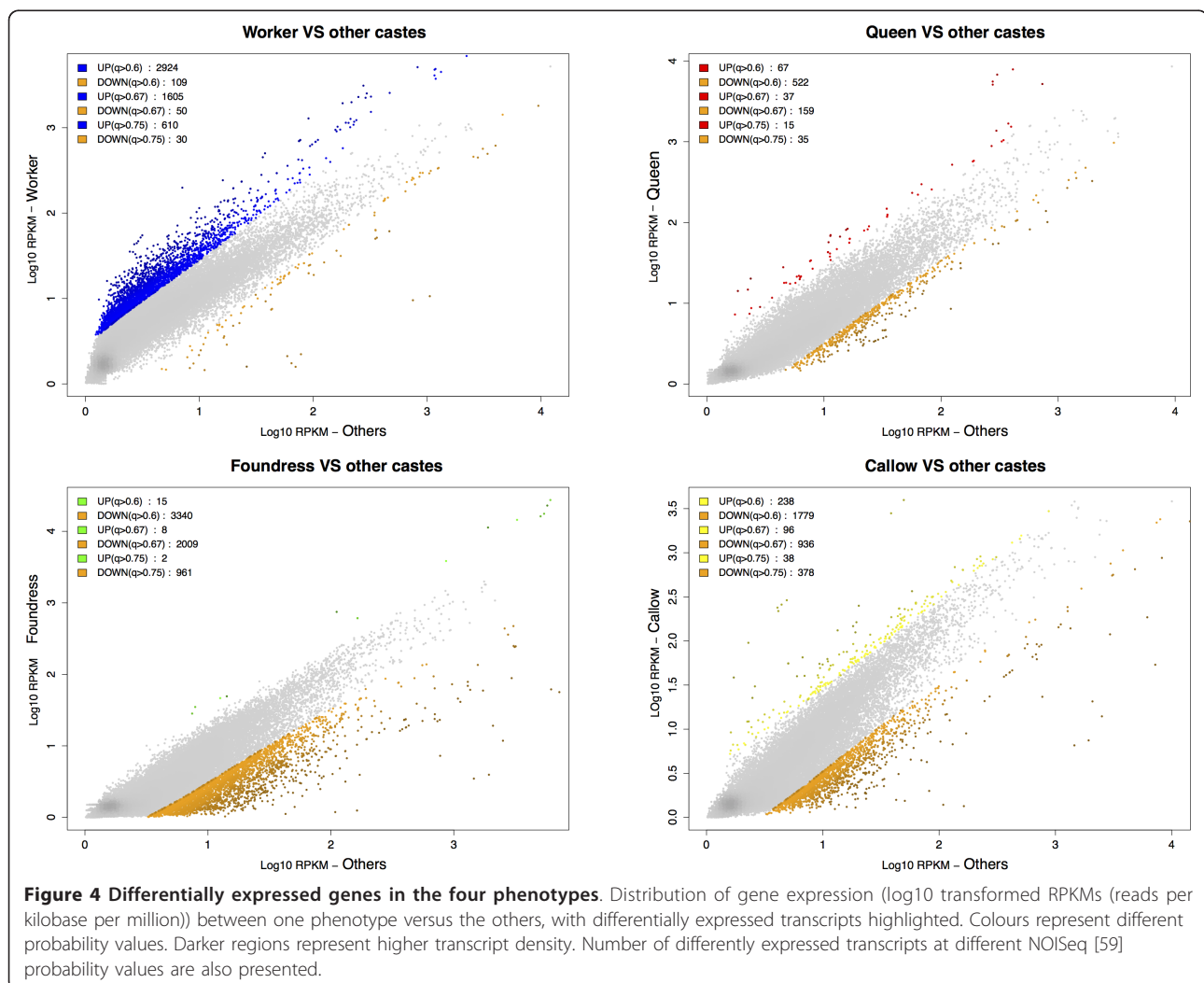
**Table 2 Isotig and isogroup differential expression statistics for the four adult female phenotypes in *P. canadensis***

| Phenotype | Up-regulated (q > 0.6) isotigs (isotigs) | Isogroups with recognized homologs | Isogroups with annotated homologs | Down-regulated (q > 0.6) isotigs (isotigs) | Isogroups with recognized homologs | Isogroups with annotated homologs |
|-----------|------------------------------------------|------------------------------------|-----------------------------------|--------------------------------------------|------------------------------------|-----------------------------------|
| Queen     | 47 (67)                                  | 39 (82.9%)                         | 20 (42.5%)                        | 402 (522)                                  | 63                                 | 38                                |
| Worker    | 2,222 (2,924)                            | 412 (18.5%)                        | 238 (10.7%)                       | 84 (109)                                   | 65                                 | 38                                |
| Foundress | 11 (15)                                  | 8 (72.7%)                          | 4 (36.4%)                         | 2,389 (3,340)                              | 776                                | 507                               |
| Callow    | 162 (238)                                | 142 (87.6%)                        | 112 (69.1%)                       | 1,454 (1,779)                              | 322                                | 194                               |
| Total     | 2,442 (3,244)                            | 596 (24.4%)                        | 370 (15.2%)                       | 4,329 (5,750)                              | 1,311                              | 858                               |

We show the number of isogroups and isotigs found as differentially expressed by NOISeq at a threshold  $q > 0.6$ ; number and percentage of differentially expressed genes with significant homologs recognized in GenBank non-redundant databases; the number and percentage of annotated homologs; and values across all castes are given in the final row of the table. Distribution of up-regulated isogroups, recognized homologs and annotated homologs are highly significantly different between queens and workers (chi-square test; respective chi-square values: 1,566 (1d.f.),  $P < 2.2e-16$ ; 4,795 (1d.f.),  $P = 9.1 \times 10^{-13}$ ; 5,382 (1d.f.),  $P < 2.2e-16$ ).

(Figure 4 and Table 2). This result holds for all different NOISeq [59] selection thresholds (Figure 4) and is independent of the number of biological replicates per phenotype analyzed (Section 8 and Table S2 in Additional file 1). Our original aim was to examine decoupling in queens

and workers. We therefore re-analyzed the data using only these two phenotypes and found an even greater asymmetry with workers up-regulating 93.1% ( $n = 1,797$ ) of 1,909 worker/queen caste-biased genes (queens up-regulate only 112 genes; Table 3 and Figure 5). Upregulation of caste-





**Table 3 Isotig and isogroup differential expression in the comparison of only queen and worker castes**

| Caste  | Up-regulated ( $q > 0.6$ ) isotigs (isotigs) | Isogroups with recognized homologs | Isogroups with annotated homologs | Putative lncRNAs (total identified = 3,824) |
|--------|----------------------------------------------|------------------------------------|-----------------------------------|---------------------------------------------|
| Queen  | 112 (163)                                    | 91 (55.8%)                         | 51 (45.5%)                        | 6 (5.3%)                                    |
| Worker | 1,797 (2,380)                                | 347 (19.3%)                        | 217 (12.1%)                       | 339 (18.86%)                                |

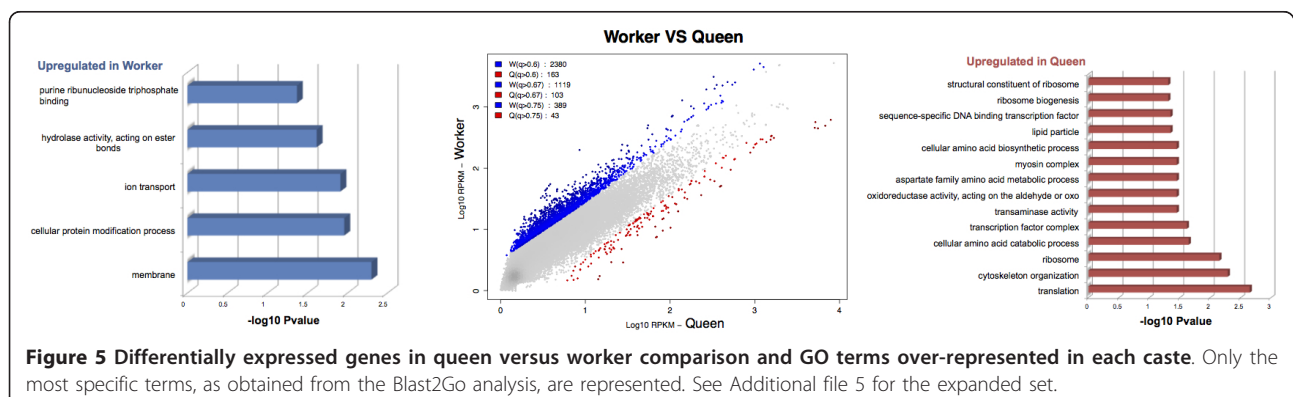
Significant differences between queens and workers in the numbers of isogroups and isotigs found as differentially expressed by NOISeq at a threshold of  $q > 0.6$  (chi-square value, 1,083 (1d.f.);  $P < 2.2e-16$ ); number and percentage of differentially expressed genes with significant homologs recognized in GenBank NR database (chi-square value 102 (1d.f.);  $P < 2.2e-16$ ); the number and percentage of annotated homologs (chi-square value, 70 (1d.f.);  $P < 2.2e-16$ ); number and percentage of putative lncRNAs (chi-square value, 236 (1d.f.);  $P < 2.2e-16$ ) are given.

biased genes in workers, rather than queens, at the early stages of caste evolution may facilitate the expansion of worker trait complexity, typified by workers in the highly eusocial descendants of primitively eusocial species. These results support the emerging picture that much of phenotypic evolution occurs in workers, rather than queens [21,37], and provide the first suggestion that molecular evolution of the worker caste is in place early in social evolution.

Interestingly, foundresses and callows accounted for the majority (89%) of down-regulated genes. Some of these could be attributed directly to a corresponding up-regulation in workers: specifically, 55% and 75%, respectively, of the down-regulated genes in foundresses and callows. But, the remaining foundress and callow down-regulated genes were not upregulated in other castes. This is also reflected in the GO classification (Additional file 5), where none of the 22 GO terms over-represented among the down-regulated genes in foundresses and only 5 of the 23 terms over-represented among the down-regulated genes in callows were up-regulated in workers. Although preliminary, these results raise some important questions: they suggest that callows are born neither worker- nor queen-like, but perhaps instead as an undifferentiated state, as expected in a species where all females retain full reproductive totipotency; down-regulation in callows may also reflect maturation processes that take place in the first days after emergence as an adult. Foundresses are thought to be hopeful reproductives, co-founding a nest in the

hope of becoming the dominant egg-layer (queen); our data suggest that foundresses are quite different from established queens, even though they fit the queen phenotype in all other respects (that is, they are active egglayers, rarely absent from the nest, and are behaviorally dominant). Taken together, these results highlight how the definition of a phenotype in this species is more complex than a simple distinction between queens and workers.

Using the 268 genes differentially expressed between queens and workers that could be annotated (Table 3), we then looked for any signs of functional specialization between these castes. Of the GO terms corresponding to these genes (Additional file 4), 2.4% (53 out of 2,242) were significantly enriched between queens and workers (Additional file 5; 40 terms (1.7%) over-represented in queens, 13 terms (0.57%) over-represented in workers). Although many of our genes did not have any functional information, there are some interesting first insights from these analyses. The putative lack of substantial functional specialization between castes in *Polistes* contrasts with the marked functional specialization in highly eusocial species, such as honeybees [19]. This may reflect selection on females to retain reproductive totipotency in primitive societies, like *P. canadensis*, in order that they can exploit alternative reproductive strategies [43], and minimize the risks of specialization associated with small colony size [60]. General terms overrepresented in workers include cellular protein modification processes, membrane and ion transport,



which may suggest elevated cellular activity in worker brains. Among the over-represented terms in queens, several candidates associated with cell structure and cytoskeleton were found, such as homologs of myosin chains, actin isoforms and troponin, which may be associated with synaptic plasticity and memory [61]. However, because so few of the genes up-regulated in workers could be annotated, and because of the lack of significant levels of enrichment between castes, these inferred functional differences between castes should be taken with caution and require further analyses once some functionality can be assigned to the unknown genes.

### Role of novel genes in the evolution of alternative phenotypes

We explored the hypothesis that novel genes (those lacking homologs in NR databases; see above) may play an important role in the early stages of phenotypic decoupling in *P. canadensis*. The importance of novel genes was recently highlighted in organisms as varied as *Hydra* to reptiles for their contribution to phenotypic diversity, through generating both behavioral and morphological variation [34-36]. Of the 2,442 up-regulated caste-biased genes (Table 2), 24.4% ( $n = 596$  genes) had known homologues, indicating that both novel and described genes are important in *P. canadensis* social behavior. Interestingly, 81.5% of worker-biased genes lacked homologs, compared to only 17.1% of queen-biased genes (Table 2). Foundresses and callows show similar levels of homology to queens. Lack of homology cannot be attributed to fragmentation or other anomalies arising from using transcriptome data (Section 6 in Additional file 1) and they are also valid for the stricter comparison of queens versus workers (Table 3). These results are exactly in line with Johnson and Tsutsui [37] and Barchuk *et al.* [62], who found an over-representation of novel (taxonomically restricted) genes up-regulated in honeybee workers (adults [37] and brood [62]), and suggested that these novel genes may facilitate expansion of worker behavioral complexity. Our data provide the first suggestion that novel gene transcription may be important at the early stages of worker evolution, confirming the potentially broad role of novel genes in eusocial evolution [37] and adding to the emerging general importance of novel genes in the evolution of phenotypic diversity [34-36]. It is important to note that microarray methods (as used in the honeybee studies) are largely biased towards conserved genes [37], as sequence conservation facilitates gene prediction. Thus, RNA-seq methods as used in our study offer new potential to explore the role of putative novel genes in phenotypic evolution.

Our ability to interpret what these novel genes are is limited without genome and proteome analyses, which is beyond the scope of this study. But, we explored the

hypothesis that many of the caste-biased novel genes were in fact non-coding RNAs, as identified from our whole transcriptome analysis (above). Interestingly, we found a significant difference in the proportion of novel transcripts that were putatively lncRNAs, with an over-representation in workers relative to queens (Table 3). lncRNAs are thought to play a role in gene regulation [63], but also may be evidence of *de novo* gene origin [33]. Future work will determine the significance of these processes in worker, rather than queen, evolution.

### Conserved molecular toolkits associated with alternative phenotypes

Social lineages evolve from non-social ancestors [64]. Alternative phenotypes (social insect castes) may therefore evolve through the decoupling of conserved sets of genes that regulated changes in hormone titers underlying the provisioning and reproductive phases of the non-social ancestor [16]. A prevailing hypothesis is that changes in the patterns of expression of these ancestral ground-plan genes could account for caste evolution across social taxa [26,28]. Using a candidate gene approach, sets of such 'toolkit' genes and molecular processes have been shown to have conserved roles in caste regulation across bees and wasps [12,27,38]. We also found some evidence of this in *P. canadensis*: vitellogenin (four isogroups), insulin (one isogroup) and major royal jelly protein (two isogroups) were queen-biased; juvenile hormone (two isogroups) and methyltransferases (two isogroups) were worker-biased. However, other toolkit genes (hexamerin, *Malvolio*, *Amfor*, P450) were present in our dataset, but were not identified as significantly caste-biased (Additional file 4) [12,27,65,66].

As a genome-wide, unbiased representation of genes expressed by alternative phenotypes in wasps, our datasets also offer the opportunity to test to what extent a conserved molecular toolkit in general explains alternative social phenotypes across eusocial taxa [12,27]. Interestingly, we found little correlation in the general identity and direction of expression of genes underlying castes in other eusocial insects, as identified from microarray analyses. For example, only 6.5% of honeybee (*A. mellifera*) caste-biased genes [19] were caste-biased in *P. canadensis* and there was no consistency in the direction of expression. Of 50 cDNAs predictive of honeybee worker behavior [18], only 8 were found in *P. canadensis* and they were not significantly worker biased (Additional file 5). There was no correlation in direction and identity of gene expression in castes of the highly eusocial fire ant *S. invicta* ( $r \sim 0$ ) [21]. Finally, only 22 of the 423 caste-biased genes of the temperate wasp *P. metricus* [12] were also caste-biased in *P. canadensis*, and they showed a low correlation in the direction of phenotype-specific expression between species (Pearson

correlation of 0.25 for worker, 0.14 for queen, 0.24 for foundress and 0.23 between gyne and callow; Additional file 5). This was unexpected since these two species are close relatives, share a recent common ancestor, and have similar social and behavioral traits. Differences in ecology of tropical and temperate *Polistes*, and possible secondary adaptation to temperate living [42], may contribute to these differences. These analyses suggest that the social toolkit represents a small portion of the genome involved in phenotype decoupling of *P. canadensis*, and that there is much lineage and ecology-specific transcription underlying caste differences. This supports the hypothesis that the genes underlying castes in different social lineages will not necessarily be conserved, because extensive molecular and developmental re-wiring may be required in the evolution of caste commitment and eusociality [10,11].

#### Insights into gene evolution in alternative phenotypes

Rapid evolution of genes has been associated with alternative phenotypes in a range of animals [6,17,29]. Caste-biased genes evolve rapidly in the honeybee and fire ant [30,31], possibly because fast-evolving genes are preferentially recruited into caste-biased gene expression [30]. Interestingly, we found no evidence of this in *P. canadensis* (Section 10 in Additional file 1; Additional file 5), suggesting that at the early stages of phenotypic decoupling genes are maintained through purifying selection. This agrees with social evolutionary theory that predicts that females retain the ability to exploit alternative reproductive opportunities (for example, to switch from worker to queen) to maximize individual-level inclusive fitness at the early stages of sociality [43]. Selection may be relaxed after workers commit to a non-reproductive pathway, as found in highly eusocial species [10,28], resulting in gene evolution, new genes and new gene networks [31,67-69]. Analyses of incipient phenotypic decoupling in other animals will help determine whether accelerated gene evolution only occurs after key innovations or fundamental shifts in life-history traits.

We next looked for evidence of processes that may have been lost or gained during phenotypic decoupling in eusocial evolution (Section 10 in Additional file 1; Additional file 5). We identified 431 caste-biased genes from across eusocial insect species that are also conserved in the 19 insect species genomes available. From this conserved set of genes, 12 GO terms were over-represented in highly eusocial insects (see Section 10 in Additional file 1 for species list) relative to *P. canadensis*, and 217 terms were under-represented (Section 10 in Additional file 1; Additional file 5). Under-represented groups (which include regulatory and signaling processes) may be less important in complex eusociality and hence not maintained by stabilizing selection. However, we cannot exclude the possibility that these differences are wasp-specific.

Over-represented terms may be important in complex eusocial behavior [30]. These included processes previously highlighted for putative roles in caste determination (for example, oxidation reduction, metabolic processes) [39], suggesting they reflect the level of eusociality rather than differences in ecology or lineage.

#### Conclusions

Our RNAseq analyses of phenotypic decoupling in *P. canadensis* provide some intriguing novel insights into the evolution of alternative phenotypes, and the early stages of caste evolution and sociality. First, they highlight the importance of novel genes in phenotypic diversity, emphasizing the need to expand existing genomic datasets beyond established model organisms to include a wider range of taxonomic groups. Secondly, they add important insights to the current emerging picture that much of the molecular changes that accompany social evolution occur predominantly in the worker rather than queen caste, with an over-representation of novel caste-biased genes in workers. These features may be important in facilitating the evolution of behavioral complexity in the worker caste. We show that these patterns are evident even at the early stages of sociality, where castes retain plasticity. Further, contrary to highly eusocial species, genes involved in caste differentiation do not appear to be subject to relaxed selection at the early stages of sociality. Diversity in the molecular regulation of castes across social lineages is expected if many pleiotropic genes involved with sexual conflict are lost in the early stages of social evolution, when ancestral monogamy is required to generate the conditions for the evolution of worker behavior [10]. Subsequent evolution of queen and worker behaviors, therefore, may take very different molecular pathways in different lineages, requiring new gene networks to evolve independently [28]. Further RNA-seq studies on other species (for example, [70]) and also genome sequencing to explore the potential role of alternative splicing (for example, [71]) will help determine the complementary roles of conserved and novel molecular processes in shaping social and other polyphenisms, bringing us closer to understanding how genomes give rise to phenotypic diversity in general.

#### Materials and methods

##### Source material

*P. canadensis* wasps of known behavioral repertoires were collected from wild populations in Panama, in July 2009 (Punta Galeta, Colon). All wasps were collected directly off their nests with forceps around midday during the active periods (that is, sunny weather) and preserved immediately in RNAlater (Ambion, Invitrogen, Applied Biosystems), and stored at -20°C until analysis (Section 1 in Additional file 1).



### Transcriptome sequencing, assembly and analyses

454 sequencing of pooled samples of 37 wasps across phenotypes (5 to 18 individuals per phenotype; 2.1 million reads, 80% brain, 10% abdomen, and 10% antennae) was used to generate a reference transcriptome (Section 2 in Additional file 1). Newbler v2.3 was used to generate the final assembled gene set (Table S1 in Additional file 1). Transcripts were annotated using GO categories assigned using BLASTx of GenBank NR databases with a conservative e-value threshold of  $10^{-5}$ , and Blast2Go was used to assess enrichment of GO terms among phenotypes (Section 4 in Additional file 1). Illumina sequencing of 14 biological replicates (>377 million reads) across 5 lanes was conducted to quantify differential gene expression, expressed as RPKM (reads per kilobase per million) values (Section 8 in Additional file 1). We trialed a number of methods for identifying differentially expressed genes and settled on a novel non-parametric method (NOISeq [59]) This method infers the noise distribution from the data and performs pairwise comparison of the samples to identify differentially expressed genes. A variety of probability thresholds were tested (Section 8 in Additional file 1). For the GO analysis we used a q-value >0.6 that represents a 50% chance that the gene is differentially expressed rather than not differentially expressed.

### Phylogenetic analyses

Protein sequences were aligned using MUSCLE [72], with default parameters. This is a multiple sequence aligner that includes an iterative alignment refinement phase to overcome known pitfalls of the progressive alignment strategy. Subsequently, poorly aligned regions of the alignment were trimmed with trimAl v1.3. [73] to remove columns with gaps in more than 30% of the sequences. A maximum-likelihood analysis was conducted on the concatenated alignment containing 33,506 sites using PhyML v3.0 [74]. Out of a total of five evolutionary models (LG, JTT, WAG, VT, BLOSUM62), the general replacement model LG (after 'Lee and Gascuel') was found to be the best fitting model using the AIC criterion [75]. In all cases four categories of evolutionary rates were used, estimating the gamma shape parameter and the proportion of invariable sites from the data. Branch supports were obtained using an approximate likelihood ratio test as implemented in PhyML ('Minimum of SH and chi-square' option). The resulting topology was compared to an alternative topology placing *Polistes* as a sister group to the ants. To do so, the phylogeny was re-computed, fixing the monophyly of *Polistes* and ants, but allowing the rest of the topology to be optimized. Support for the two topologies were compared using a Shimodaira-Hasegawa test, as implemented in CONSEL [76]. A Bayesian analysis was conducted as implemented in PhyloBayes [77], using the default CAT model and

running two independent MCMC runs during 300,000 generations, and sampling every 100 generations. Consensus trees were built after removing the first 20% sampled trees and using a majority consensus rule.

### Data access

Raw sequence data are available at the European Read Archive (accession number ERP001342). The Transcriptome Shotgun Assembly project has been deposited at DDBJ/EMBL/GenBank under the accession GAFR0100001-GAFR01045087. The version described in this paper is the first version, GAFR01000000. All data and datasets can also be accessed at [78].

### Additional material

**Additional file 1: Sections 1 to 10 and Tables S1 and S2.** Additional methods and results referred to in the main text can be found here. Table S1: comparison of transcriptome assemblies. Number of transcripts, genes and respective transcript length statistics for the optimal assemblies generated with Version 2.3 and Version 2.5 of GS Newbler Assembler, and Version 3.0.5 of Mira Assembler and Oases (with Illumina paired-end sequence data). Table S2: numbers of differentially expressed ( $q > 0.6$ ) genes and transcripts was robust to different numbers of biological replicates.

**Additional file 2: Read coverage from the 454 transcriptome assembly.** (A) Distribution of the number of 454 reads per transcript (truncated to transcripts with less than 500 reads). (B) Distribution weighted by the transcript length (truncated plot). The majority of transcripts had more than 5.19 reads, on average.

**Additional file 3: Functional groups identified in the *P. canadensis* transcriptome.** Distribution of the number of genes according to different GO terms from the 454 pooled transcriptome assembly. Categories were filtered by a minimum number of sequences: 50 for cellular component; 200 for biological process; and 100 for molecular function.

**Additional file 4: Information on the best BLAST hits in the NR database for each transcript; noncoding RNA potential for genes with and without hits in NR.** Tabs: 'besthits\_with\_NR', information extracted from the best hits for each transcript; 'PortraitScore\_transcriptsWithNOHits', 'PortraitScore\_transcriptsWithHits', score from the portrait program.

**Additional file 5: Information on functional analyses and comparison with existing caste expression data from other species.** Tabs are as follows. 'Toth2010\_Pmetricus\_Pcan', mappings between *P. metricus* and *P. canadensis* transcripts and their respective expression levels and whether they are significantly differentially expressed in *P. canadensis*. 'Whitfield2003\_50cDNAs', mappings between 50 most discriminative cDNAs for worker behaviors in *A. mellifera* from Whitfield et al. [18] and *P. canadensis*, with respective expression values. 'GO-Enrichment Differential Expressed (DE) Genes', list of over/under-represented GO terms for genes significantly up- and down-regulated in each caste comparison. Genes were categorized according to biological processes (P), molecular function (F) and cellular component (C). Most specific terms were retrieved with Blast2GO and correspond to the leaf nodes in the GO tree, this excludes cases where parent and child nodes may be included. Column labels are: Test (number of genes in the test set annotated with this GO term); Ref (number of genes in the ref set annotated with this GO term); notAnnotTest (number of genes in the test set NOT annotated with this GO term); notAnnotRef (number of genes in the reference set NOT annotated with this GO term), where 'test' set is the set of genes up-regulated in queens, and the 'ref' set those up-regulated genes in workers, and vice-versa. 'GO-Enrichment', list of over/under-represented GO terms for the *P. canadensis* genes that



are conserved across the genomes of 6 eusocial insects. 'genes\_with\_Behavior\_GOterm', genes classified with GO term thought to be relevant to behavior, their respective descriptions and direction of differential expression in *P. canadensis*. 'genes\_accelerated\_Pol\_Bee\_Ants', genes found accelerated in *Polistes*, bee and ants, and direction of differential expression in *P. canadensis*. 'genes\_accelerated\_polistes', genes found accelerated only in *Polistes*; and direction of differential expression in *P. canadensis*. 'phylogenetic marker identifiers', a list of the 93 marker genes used in the phylogenetic analyses. These correspond to phylomeDB codes in the Aphid phylome used as the reference markers. Sequences for these identifiers can be obtained from Huerta-Cepas et al. [57].

**Additional file 6: Number of mapped reads per individual.**

**Additional file 7: Basic information about the assembled transcripts.** Tabs: 'transcript\_length', length of each transcript; 'transcript\_to\_gene', correspondence between transcript and gene.

**Additional file 8: Cumulative distribution of gene expression for the four phenotypes.**

**Additional file 9: Expression values in RPKM (reads per kilobase per million) for all the individuals, indicating which caste each isotig is up-regulated in.**

**Additional file 10: NOISeq probability values for each transcript for the comparison of each caste (that is, W (worker), Fo (foundress), Q (queen), or C (callow)) versus the others, and between queens and workers.** Expression values for each caste and comparison, differential expression statistics ('M' and 'D' [12]), probability of differential expression ('prob'); 'ranking', which is a summary statistic of 'M' and 'D'.

#### Abbreviations

bp: base pair; GO: Gene Ontology; lncRNA: long non-coding RNA; NR: non-redundant; ORF: open reading frame; RPKM: reads per kilobase per million.

#### Authors' contributions

SS conceived the idea and designed the research. SP and SS performed the fieldwork, SP performed wet-lab work. RC and Ff-C conducted the MIRA assembly; PGF and RG designed and conducted all other bioinformatics analyses. TG conducted the gene evolution and phylogenetic analyses. SS and PGF wrote the paper, with comments from all authors. All authors read and approved the final manuscript.

#### Competing interests

The authors declare that they have no competing interests.

#### Acknowledgements

We thank Centre for Genomic Regulation (CRG) (Heinz Himmelbauer and Francisco C  mara), GenePool and the Babraham Institute (Wolf Reik and Simon Andrews) for support and access to their sequencing facilities. We thank RCUK, CONSOLIDER (CSD2007-0005), the Spanish Ministry of Science (BIO2006-03380, BFU2009-09168), GTEX (NIH Project 1R01MH090941-01), and a Wellcome Trust Senior Investigator Award to Wolf Reik. SS was supported by a UK Research Council (RCUK) fellowship, and PGF by a FCT postdoctoral grant (SFRH/BPD/42003/2007) from the Portuguese Ministry of Science and Technology. We thank the Smithsonian Tropical Research Institute (particularly William Wcislo, Mary Jane West-Eberhard, Jorge Morales and staff at the Galeta Field Station) for hosting the fieldwork, and Stephanie Dreier and Thibault Lengronne for assistance in the field. The fieldwork was conducted under ANAM Permit No. SE/A-33-09.

#### Author details

<sup>1</sup>Center for Genomic Regulation, Universitat Pompeu Fabra (CRG-UPF), Doctor Aiguader, 88, 08003 Barcelona, Catalonia, Spain. <sup>2</sup>Institute of Zoology, Zoological Society of London, Regent's Park, NW1 4RY, UK. <sup>3</sup>The Babraham Institute, Babraham Research Campus, Cambridge, CB22 3AT, UK. <sup>4</sup>Centre for Ecology and Conservation, Biosciences, University of Exeter, Tremough, Penryn, TR10 9EZ, UK. <sup>5</sup>Current address: School of Biological Sciences, University of Bristol, Woodland Road, Bristol, BS8 1UG, UK. <sup>6</sup>Current address:

Department of Genetic Medicine and Development, University of Geneva Medical School, 25 Rue Michel Servet 1, 1211 Geneva, Switzerland.

Received: 17 December 2012 Revised: 31 January 2013

Accepted: 26 February 2013 Published: 26 February 2013

#### References

1. West-Eberhard MJ: *Developmental Plasticity and Evolution* New York: Oxford University Press; 2003.
2. Nijhout HF: **Development and evolution of adaptive polyphenisms.** *Evol Devel* 2003, **5**:9-18.
3. Wilson EO: **The evolution of caste systems in social insects.** *Proc Amer Phil Soc* 1979, **123**:204-211.
4. Holldobler B, Wilson E: *The Ants* Berlin: Springer; 1990.
5. Moczek AP: **Phenotypic plasticity and diversity in insects.** *Phil Trans R Soc Lond B* 2010, **365**:593-603.
6. Ellegren H, Parsch J: **The evolution of sex-biased genes and sex-biased gene expression.** *Nat Rev Genet* 2007, **8**:689-698.
7. Oster GF, Wilson EO: *Caste and Ecology in the Social Insects* Princeton: Princeton University Press; 1978.
8. Smith CR, Toth AL, Suarez AV, Robinson GE: **Genetic and genomic analyses of the division of labour in insect societies.** *Nature Rev Genet* 2008, **9**:735-748.
9. Sumner S: **Determining the molecular basis of sociality in insects: prospects, progress and potential in sociogenomics.** *Ann Zool Fenn* 2006, **43**:423-442.
10. Boomsma JJ: **Lifetime monogamy and the evolution of eusociality.** *Phil Trans R Soc Lond B* 2009, **364**:3191-3207.
11. Linksvayer TA, Wade MJ: **Genes with social effects are expected to harbor more sequence variation within and between species.** *Evolution* 2009, **63**:1685-1696.
12. Toth AL, Varala K, Henshaw MT, Rodriguez-Zas SL, Hudson ME, Robinson GE: **Brain transcriptomic analysis in paper wasps identifies genes associated with behavior across social insect lineages.** *Proc R Soc Lond B* 2010, **277**:2139-2148.
13. Toth AL, Varala K, Newman TC, Miguez FE, Hutchison SK, Willoughby DA, Simons JF, Egholm M, Hunt JH, Hudson ME, Robinson GE: **Wasp gene expression supports an evolutionary link between maternal behavior and eusociality.** *Science* 2007, **318**:441-444.
14. Sumner S, Pereboom JJM, Jordan WC: **Differential gene expression and phenotypic plasticity in behavioral castes of the primitively eusocial wasp, *Polistes canadensis*.** *Proc R Soc Lond B* 2006, **273**:19-26.
15. Hoffman EA, Goodisman MAD: **Gene expression and the evolution of phenotypic diversity in social wasps.** *BMC Biol* 2007, **5**:23.
16. West-Eberhard MJ: **Wasp societies as microcosms for the study of development and evolution.** In *Natural History and Evolution of Paper-Wasps*. Edited by: Turillazzi S, West-Eberhard MJ. Oxford Science Publications; 1996:290-317.
17. Snell-Rood EC, Cash A, Han MV, Kijimoto T, Andrews J, Moczek AP: **Developmental decoupling of alternative phenotypes: insights from the transcriptomes of horn-polyphenic beetles.** *Evolution* 2011, **65**:231-245.
18. Whitfield CW, Cziko AM, Robinson GE: **Gene expression profiles in the brain predict behavior in individual honey bees.** *Science* 2003, **302**:296-299.
19. Grozinger CM, Fan Y, Hoover SE, Winston ML: **Genome-wide analysis reveals differences in brain gene expression patterns associated with caste and reproductive status in honeybees (*Apis mellifera*).** *Mol Ecol* 2007, **16**:4837-4848.
20. Alaux C, Sinha S, Hasadsri L, Hunt GJ, Guzm  n-Novoa E, DeGrandi-Hoffman G, Uribe-Rubio JL, Southey BR, Rodriguez-Zas S, Robinson GE: **Honey bee aggression supports a link between gene regulation and behavioral evolution.** *Proc Natl Acad Sci USA* 2009, **106**:15400-15405.
21. Ometto L, Shoemaker D, Ross KG, Keller L: **Evolution of gene expression in fire ants: the effects of developmental stage, caste, and species.** *Mol Biol Evol* 2011, **28**:1381-1392.
22. Hughes AL: **Evolution of adaptive phenotypic traits without positive Darwinian selection.** *Heredity* 2012, **108**:347-353.
23. Brakefield P, Gates J, Keys D, Kesbeke F, Wijngaarden P, Monteiro A, French V, Carroll S: **Development, plasticity and evolution of butterfly eyespot patterns.** *Nature* 1996, **384**:236-242.

24. Abouheif E, Wray GA: **Evolution of the gene network underlying wing polyphenism in ants.** *Science* 2002, **297**:249-252.
25. Moczek AP, Nagy LM: **Diverse developmental mechanisms contribute to different levels of diversity in horned beetles.** *Evol Dev* 2005, **7**:175-185.
26. Toth AL, Robinson GE: **Evo-devo and the evolution of social behavior.** *Trends Genet* 2007, **23**:334-341.
27. Amdam GV, Norberg K, Fondrk MK, Page RE: **Reproductive ground plan may mediate colony-level selection effects on individual foraging behavior in honey bees.** *Proc Natl Acad Sci USA* 2004, **101**:11350-11355.
28. Johnson BR, Linksvayer TA: **Deconstructing the superorganism: social physiology, groundplans, and sociogenomics.** *Quart Rev Biol* 2010, **85**:57-79.
29. Suzuki Y, Nijhout HF: **Evolution of a polyphenism by genetic accommodation.** *Science* 2006, **311**:650-652.
30. Hunt BG, Ometto L, Wurm Y, Shoemaker DD, Soojin VY, Keller L, Goodisman MAD: **Relaxed selection is a precursor to the evolution of phenotypic plasticity.** *Proc Natl Acad Sci USA* 2011, **108**:15936e15941.
31. Hunt BG, Wyder S, Elango N, Werren JH, Zdobnov EM, Yi SV, Goodisman MAD: **Sociality is linked to rates of protein evolution in a highly social insect.** *Mol Biol Evol* 2010, **27**:497-500.
32. Ding Y, Zhou Q, Wang W: **Origins of new genes and evolution of their novel functions.** *Ann Rev Ecol Syst* 2012, **43**:345-363.
33. Carvunis AR, Rolland T, Wapinski I, Calderwood MA, Yildirim MA, Hidalgo A, Barbette J, Santhanam B, Brar GA, Simonis N, Charlotiaux B, Weissman JS, Regev A, Thierry-mieg N, Cusick ME, Vidal M: **Proto-genes and de novo gene birth.** *Nature* 2012, **487**:370-374.
34. Dai H, Chen Y, Chen S, Mao Q, Kennedy D, Landback P, Eyre-Walker A, Du W, Long M: **The evolution of courtship behaviors through the origination of a new gene in Drosophila.** *Proc Natl Acad Sci USA* 2008, **105**:7478-7483.
35. Khalturin K, Anton-Erxleben F, Sassmann S, Wittlieb Jr, Hemmrich G, Bosch TCG: **A novel gene family controls species-specific morphological traits in hydra.** *PLoS Biol* 2008, **6**:e278.
36. Fry BG, Roelants K, Winter K, Hodgson WC, Griesman L, Kwok HF, Scanlon D, Karas J, Shaw C, Wong L, Norman JA: **Novel venom proteins produced by differential domain-expression strategies in beaded lizards and gila monsters (genus Heloderma).** *Mol Biol Evol* 2010, **27**:395-407.
37. Johnson BR, Tsutsui ND: **Taxonomically restricted genes are associated with the evolution of sociality in the honey bee.** *BMC Genomics* 2011, **12**:164.
38. Woodard SH, Fischman BJ, Venkat A, Hudson ME, Varala K, Cameron SA, Clark AG, Robinson GE: **Genes involved in convergent evolution of eusociality in bees.** *Proc Natl Acad Sci USA* 2011, **108**:7472-7477.
39. Fischman BJ, Woodard SH, Robinson GE: **Molecular evolutionary analyses of insect societies.** *Proc Natl Acad Sci USA* 2011, **108**:10847-10854.
40. Turillazzi S, West-Eberhard MJ: *Natural History and Evolution of Paper Wasps* Oxford: Oxford University Press; 1996.
41. Hunt JH, Amdam GV: **Bivoltinism as an antecedent to eusociality in the paper wasp genus Polistes.** *Science* 2005, **308**:264-267.
42. Carpenter JM: **Phylogeny and biogeography of Polistes.** In *Natural History and Evolution of Paper Wasps*. Edited by: Turillazzi S, West-Eberhard MJ. Oxford: Oxford University Press; 1996:18-57.
43. Leadbeater E, Carruthers JM, Green JP, Rosser NS, Field JP: **Nest inheritance is the missing source of direct fitness in a primitively eusocial insect.** *Science* 2011, **333**:874-876.
44. Hughes CR, Beck MO, Strassmann JE: **Queen succession in the social wasp, Polistes annularis.** *Ethology* 1987, **76**:124-132.
45. Sumner S, Kelstrup H, Fanelli D: **Reproductive constraints, direct fitness and indirect fitness explain helping behavior in the primitively eusocial wasp, Polistes canadensis.** *Proc R Soc Lond B* 2010, **277**:1721-1728.
46. Savard J, Tautz D, Richards S, Weinstock GM, Gibbs RA, Werren JH, Tettelin H, Lercher MJ: **Phylogenomic analysis reveals bees and wasps (Hymenoptera) at the base of the radiation of Holometabolous insects.** *Genome Res* 2006, **16**:1334-1338.
47. Heraty J, Ronquist F, Carpenter JM, Hawks D, Schulmeister S, Dowling AP, Murray D, Munro J, Wheeler WC, Schiff N, Sharkey M: **Evolution of the hymenopteran megaradiation.** *Mol Phylo Evol* 2011, **60**:73-88.
48. Smith CR, Smith CD, Robertson HM, Helmkampf M, Zimin A, Yandell M, Holt C, Hu H, Abouheif E, Benton R, Cash E, Croset V, Currie CR, Elhaik E, Elsik CG, Fave M-J, Fernandes V, Gibson JD, Graur D, Gronenberg W, Grubbs KJ, Hagen DE, Viniegra ASI, Johnson BR, Johnson RM, Khila A, Kim JW, Mathis KA, Munoz-Torres MC, Murphy MC, et al: **Draft genome of the red harvester ant Pogonomyrmex barbatus.** *Proc Natl Acad Sci USA* 2011, **108**:5667-5672.
49. Smith CD, Zimin A, Holt C, Abouheif E, Benton R, Cash E, Croset V, Currie CR, Elhaik E, Elsik CG, Fave M-J, Fernandes V, Gadau Jr, Gibson JD, Graur D, Grubbs KJ, Hagen DE, Helmkampf M, Holley J-A, Hu H, Viniegra ASI, Johnson BR, Johnson RM, Khila A, Kim JW, Laird J, Mathis KA, Moeller JA, Munoz-Torres MC, Murphy MC, et al: **Draft genome of the globally widespread and invasive Argentine ant (Linepithema humile).** *Proc Natl Acad Sci USA* 2011, **108**:5673-5678.
50. Eddy SR: **Accelerated profile HMM searches.** *PLoS Comput Biol* 2011, **7**: e1002195.
51. Arrial RT, Togawa RC, Brigido MdeM: **Screening non-coding RNAs in transcriptomes from neglected species using PORTRAIT: case study of the pathogenic fungus Paracoccidioides brasiliensis.** *BMC Bioinform* 2009, **10**:239.
52. Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, Guernec G, Martin D, Merkel A, Knowles DG, Lagarde J, Veeravalli L, Ruan X, Ruan Y, Lassmann T, Carninci P, Brown JB, Lipovich L, Gonzalez JM, Thomas M, Davis CA, Shiekhattar R, Gingeras TR, Hubbard TJ, Notredame C, Harrow J, Guigó R, et al: **The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression.** *Genome Res* 2012, **22**:1775-1789.
53. Davis R, Baldauf S, Mayhew P: **The origins of species richness in the Hymenoptera: insights from a family-level supertree.** *BMC Evol Biol* 2011, **10**:109.
54. Pilgrim EK, von Dohlen CD, Pitts JP: **Molecular phylogenetics of Vespoidea indicate paraphyly of the superfamily and novel relationships of its component families and subfamilies.** *Zool Script* 2008, **37**:539-560.
55. Peters R, Meyer B, Krogmann L, Borner J, Meusemann K, Schütte K, Niehuis O, Misof B: **The taming of an impossible child: a standardized all-in approach to the phylogeny of Hymenoptera using public database sequences.** *BMC Biol* 2011, **9**:55.
56. Dunn CW, Hejnol A, Matus DQ, Pang K, Browne WE, Smith SA, Seaver E, Rouse GW, Obst M, Edgecombe GD, Sorensen MV, Haddock SHD, Schmidt-Rhaesa A, Okusu A, Kristensen RM, Wheeler WC, Martindale MQ, Giribet G, et al: **Broad phylogenomic sampling improves resolution of the animal tree of life.** *Nature* 2008, **452**:745-749.
57. Huerta-Cepas J, Marcet-Houben M, Pignatelli M, Moya A, Gabaldon T: **The pea aphid phylome: a complete catalogue of evolutionary histories and arthropod orthology and paralogy relationships for Acyrthosiphon pisum genes.** *Insect Mol Biol* 2010, **19**(Suppl 2):13-21.
58. Wilson EO, Holldobler B: **The rise of the ants: A phylogenetic and ecological explanation.** *Proc Natl Acad Sci USA* 2005, **102**:7411-7414.
59. Tarazona S, García-Alcalde F, Dopazo J, Ferrer A, Conesa A: **Differential expression in RNA-seq: A matter of depth.** *Genome Res* 2012, **21**:2213-2223.
60. Bourke AFG: *Principles of Social Evolution* Oxford: Oxford University Press; 2011.
61. Rex CS, Gavin CF, Rubio MD, Kramar EA, Chen LY, Jia Y, Haganir RL, Muzyczka N, Gall CM, Miller CA, Lynch G, Rumbaugh G: **Myosin IIb regulates actin dynamics during synaptic plasticity and memory formation.** *Neuron* 2010, **67**:603-617.
62. Barchuk AR, Cristino AS, Kucharski R, Costa LF, Simoes ZLP, Maleszka R: **Molecular determinants of caste differentiation in the highly eusocial honeybee Apis mellifera.** *BMC Dev Biol* 2007, **7**:70.
63. Guttman M, Amit I, Garber M, French C, Lin MF, Feldser D, Huarte M, Zuk O, Carey BW, Cassady JP, Cabili MN, Jaenisch R, Mikkelsen TS, Jacks T, Hacohen N, Bernstein BE, Kellis M, Regev A, Rinn JL, Lander ES: **Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals.** *Nature* 2009, **458**:223-227.
64. Hughes WOH, Oldroyd BP, Beekman M, Ratnieks FLW: **Ancestral monogamy shows kin selection is key to the evolution of eusociality.** *Science* 2008, **320**:1213-1216.
65. Ament SA, Corona M, Pollock HS, Robinson GE: **Insulin signaling is involved in the regulation of worker division of labor in honey bee colonies.** *Proc Natl Acad Sci USA* 2008, **105**:4226-4231.
66. Giray T, Giovanetti M, West-Eberhard MJ: **Juvenile hormone, reproduction, and worker behavior in the neotropical social wasp Polistes canadensis.** *Proc Natl Acad Sci USA* 2005, **102**:3330-3335.

67. Gadagkar R: **The evolution of caste polymorphism in social insects: genetic release followed by diversifying evolution.** *J Genet* 1997, **76**:167-179.
68. Robinson GE, Ben-Shahar Y: **Social behavior and comparative genomics: new genes or new gene regulation?** *Genes Brain Behav* 2002, **1**:197-203.
69. Linksvayer TA, Wade MJ: **The evolutionary origin and elaboration of sociality in the aculeate hymenoptera: Maternal effects, sib-social effects, and heterochrony.** *Quart Rev Biol* 2005, **80**:317-336.
70. Colgan TC, Carolan JC, Bridgett SJ, Sumner S, Blaxter ML, Brown MJF: **Polyphenism in social insects: Insights from a transcriptome-wide analysis of gene expression in the life stages of the key pollinator, *Bombus terrestris*.** *BMC Genomics* 2011, **12**:623.
71. Lyko F, Foret S, Kucharski R, Wolf S, Falckenhayn C, Maleszka R: **The honey bee epigenomes: differential methylation of brain DNA in queens and workers.** *PLoS Biol* 2010, **8**:e1000506.
72. Edgar R: **MUSCLE: multiple sequence alignment with high accuracy and high throughput.** *Nucleic Acids Res* 2004, **32**:1792-1797.
73. Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T: **trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses.** *Bioinformatics* 2009, **25**:1972-1973.
74. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O: **New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0.** *Syst Biol* 2010, **59**:307-321.
75. Le SQ, Gascuel O: **An improved general amino acid replacement matrix.** *Molec Biol Evol* 2008, **25**:1307-1320.
76. Shimodaira H, Hasegawa M: **CONSEL: for assessing the confidence of phylogenetic tree selection.** *Bioinformatics* 2001, **17**:1246-1247.
77. Lartillot N, Philippe H: **Bayesian phylogenetic software based on mixture models.** *Molec Biol Evol* 2004, **21**:1095-1109.
78. **Data and Datasets.** [<http://genome.crg.es/~pferreira/pcandata/pcan.htm>].

doi:10.1186/gb-2013-14-2-r20

**Cite this article as:** Ferreira et al.: Transcriptome analyses of primitively eusocial wasps reveal novel insights into the evolution of sociality and the origin of alternative phenotypes. *Genome Biology* 2013 **14**:R20.

**Submit your next manuscript to BioMed Central  
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

